



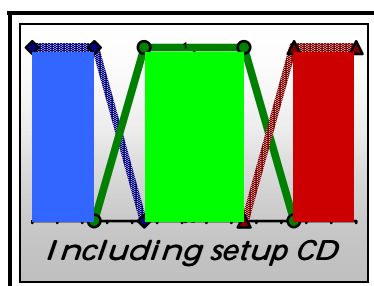
ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ
ΤΕΧΝΟΛΟΓΙΚΗΣ ΠΡΟΒΛΕΨΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΘΕΜΑ: ΜΙΑ ΜΕΘΟΔΟΣ ΚΑΙ ΕΦΑΡΜΟΓΗ ΟΜΑΔΟΠΟΙΗΣΗΣ
ΔΕΔΟΜΕΝΩΝ ΜΕ ΒΑΣΗ ΤΗΝ ΑΝΑΛΥΣΗ ΣΕ ΚΥΡΙΕΣ
ΣΥΝΙΣΤΩΣΕΣ ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΜΟΥ ΤΩΝ ΟΜΑΔΩΝ ΜΕ
ΧΡΗΣΗ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ

ΚΩΣΤΑΣ ΣΑΚΚΟΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:
ΧΡΗΣΤΟΣ ΣΚΙΑΔΑΣ



XANIA
ΑΠΡΙΛΙΟΣ 2000

ΠΡΟΛΟΓΟΣ

Η εργασία αυτή πραγματοποιήθηκε στο Εργαστήριο Ανάλυσης Δεδομένων και Τεχνολογικής Πρόβλεψης του Πολυτεχνείου Κρήτης.

Θα ήθελα να ευχαριστήσω όσους βοήθησαν στην ολοκλήρωσή της. Συγκεκριμένα ευχαριστώ τον καθηγητή του Πολυτεχνείου Κρήτης κ. Σκιαδά Χρήστο για την πολύπλευρη υποστήριξή του καθ' όλη τη διάρκεια εκπόνησης της εργασίας. Επίσης θα ήθελα να ευχαριστήσω τον καθηγητή Δημοτικάλη Ιωάννη για την πολύτιμη βοήθειά του και την υπομονή του κατά τη διάρκεια ανάπτυξης του λογισμικού που συνοδεύει τη μεθοδολογία που προτείνεται στην παρούσα διπλωματική εργασία.

Τέλος θα ήθελα να ευχαριστήσω όσους μου συμπαραστάθηκαν ηθικά και ιδιαίτερα τους γονείς μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ.....	1
ΚΕΦΑΛΑΙΟ 2: ΑΝΑΛΥΣΗ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ	3
2.1 ΕΙΣΑΓΩΓΗ	3
2.2 ΑΝΑΛΥΣΗ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ.....	4
2.3 ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ	5
2.3.1 Κεντροποιημένος Πίνακας.....	5
2.3.2 Ισοσταθμισμένος Πίνακας	6
2.3.3 Πίνακας Συσχετίσεων των Χαρακτηριστικών	6
2.3.4 Πίνακας Ιδιοτιμών	7
2.3.5 Προβολές των Ατόμων στους Νέους Άξονες.....	7
2.3.6 Προβολές των Χαρακτηριστικών στους Νέους Άξονες.....	8
2.3.7 Γράφημα Ανάλυσης σε Κύριες Συνιστώσες.....	8
2.4 ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΧΡΗΣΗΣ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ.....	13
ΚΕΦΑΛΑΙΟ 3: Ο ΑΛΓΟΡΙΘΜΟΣ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ	14
3.1 ΕΙΣΑΓΩΓΗ	14
3.2 ΟΜΑΔΟΠΟΙΗΣΗ ΣΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΑΚΣ.....	15
3.2.1 Υπολογισμός Γεωμετρικών Αποστάσεων	17
3.2.2 Κριτήρια Τερματισμού της Ομαδοποίησης	18
3.3 Η ΑΝΑΠΤΥΞΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ	20
3.3.1 Χρήση Τροποποιημένων Προβολών.	21
3.3.2 Ο Κοντινότερος Γείτονας	30
3.4 Ο ΑΛΓΟΡΙΘΜΟΣ	32
3.4.1 Υπολογισμός Αποστάσεων και Ελάχιστης Απόστασης	33
3.4.2 Προ-Ομαδοποίηση.....	35
3.4.3 Ομαδοποίηση	37
3.4.4 Έλεγχος Αριθμού Ομαδοποιημένων Ατόμων.....	49
3.4.5 Υπολογισμός Πραγματικού Αριθμού Ομάδων.....	50
3.4.6 Έλεγχος Τερματισμού με βάση τον Αριθμό των Ομάδων.....	51

3.4.7 Συναρτήσεις	52
3.5 ΜΙΑ ΑΛΛΗ ΕΚΔΟΧΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ	56
3.6 ΣΥΝΟΨΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ	57
ΚΕΦΑΛΑΙΟ 4: ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ ΟΜΑΔΩΝ ΜΕ ΧΡΗΣΗ ΤΗΣ ΘΕΩΡΙΑΣ ΤΩΝ ΑΣΑΦΩΝ ΣΥΝΟΛΩΝ	58
4.1 ΕΙΣΑΓΩΓΗ	58
4.2 ΑΣΑΦΗ ΣΥΝΟΛΑ	59
4.3 ΛΕΚΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ	61
4.4 ΣΥΝΑΡΤΗΣΕΙΣ ΣΥΜΜΕΤΟΧΗΣ	63
4.4.1 Κατασκευή των Συναρτήσεων Συμμετοχής	64
4.5 ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ ΑΤΟΜΩΝ	66
4.6 ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ ΟΜΑΔΩΝ	68
4.7 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ	69
4.7.1 Ελαστικότητα Συναρτήσεων Συμμετοχής στο χαρακτηρισμό ως “Μεσαίο” ...	69
4.7.2 Συμμετρία των Συναρτήσεων Συμμετοχής	70
4.7.3 Περιορισμοί των Ορίων των Συναρτήσεων Συμμετοχής	70
4.8 ΠΡΟΤΕΙΝΟΜΕΝΟΙ ΚΑΝΟΝΕΣ (Standard Rules)	71
ΚΕΦΑΛΑΙΟ 5: Η ΕΦΑΡΜΟΓΗ ΟΜΑΔΟΠΟΙΗΣΗΣ ΜΕ ΑΣΑΦΗ ΛΟΓΙΚΗ (Fuzzy Data Grouping)	75
5.1 Η ΓΛΩΣΣΑ ΑΝΑΠΤΥΞΗΣ	75
5.2 ΑΠΑΙΤΗΣΕΙΣ	76
5.3 ΧΡΗΣΗ ΑΡΧΕΙΩΝ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ	77
5.3.1 Ο Αρχικός Πίνακας Δεδομένων (Individuals_Variables_Values)	77
5.3.2 Ο Πίνακας των Συναρτήσεων Συμμετοχής (Principal_Axes_MMF)	78
5.4 Η ΑΡΧΙΚΗ ΦΟΡΜΑ	80
5.4.1 Ο Πίνακας Δεδομένων	81
5.4.2 Το Δέντρο των Ορίων των Συναρτήσεων Συμμετοχής	81
5.4.3 Τα Μενού της Αρχικής Φόρμας	82
5.4.4 Η Εργαλειοθήκη	93
5.4.5 Η Γραμμή Κατάστασης	94
5.5 Η ΦΟΡΜΑ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ	95

5.5.1 Το Δέντρο των Ορίων των Συναρτήσεων Συμμετοχής	95
5.5.2 Το Γράφημα των Συναρτήσεων Συμμετοχής	96
5.5.3 Το Γράφημα της Κατανομής των Ατόμων	96
5.5.4 Η Εργαλειοθήκη.....	96
5.6 Η ΦΟΡΜΑ ΤΟΥ ΓΡΑΦΗΜΑΤΟΣ ΤΗΣ ΑΚΣ ΚΑΙ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ.....	97
5.6.1 Τα Μενού	98
5.6.2 Οι Εργαλειοθήκες	103
5.6.3 Pop Up Μενού και Γραμμή Κατάστασης (Status Bar).....	108
5.7 Η ΦΟΡΜΑ Options	111
5.8 Η ΦΟΡΜΑ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ.....	115
5.8.1 Λεπτομέρειες της Ομαδοποίησης (Grouping Details).....	115
5.8.2 Χαρακτηριστικά των Κυρίων Αξόνων (Attributes of Principal Axis)	117
5.8.3 Το Δέντρο των Ομάδων	118
5.8.4 Το Ραβδόγραμμα και οι Πίνακες	118
5.8.5 Η Γραμμή Κατάστασης (Status Bar)	120
ΚΕΦΑΛΑΙΟ 6: ΜΙΑ ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΕΘΟΔΟΥ ΜΕ ΤΟ ΠΡΟΓΡΑΜΜΑ.....	121
6.1 ΑΡΧΙΚΑ ΔΕΔΟΜΕΝΑ	121
6.2 ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ.....	123
6.2.1 Κεντροποιημένος Πίνακας Δεδομένων.	123
6.2.2 Ισοσταθμισμένος Πίνακας Δεδομένων	125
6.2.3 Πίνακας Συσχετίσεων των Χαρακτηριστικών	126
6.2.4 Πίνακας Ιδιοτιμών	127
6.2.5 Προβολές των Χαρακτηριστικών στους Νέους Άξονες.....	128
6.2.6 Προβολές των Ατόμων στους Νέους Άξονες.	129
6.3 ΟΡΙΣΜΟΣ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ	131
6.3.1 Ελαστικός Συμμετρικός κανόνας.....	131
6.3.2 Normal Συμμετρικός κανόνας	133
6.3.3 Αυστηρός Συμμετρικός κανόνας	134
6.3.4 Ελαστικός μη Συμμετρικός κανόνας	136
6.3.5 Normal μη Συμμετρικός κανόνας	137
6.3.6 Αυστηρός μη Συμμετρικός κανόνας.....	138

6.3.7 Τιμές Συμμετοχής των Ατόμων	140
6.4 ΟΜΑΔΟΠΟΙΗΣΗ.....	142
6.4.1 Ομαδοποίηση με χρήση του αλγορίθμου Nearest Neighbour (NN).....	142
6.4.2 Ομαδοποίηση με χρήση του αλγορίθμου Min Distance (MD).....	150
6.5 ΑΝΑΛΥΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΟΜΑΔΟΠΟΙΗΣΗΣ	155
ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΟΝΤΙΚΕΣ ΠΡΟΟΠΤΙΚΕΣ	158
7.1 Η ΜΕΘΟΔΟΣ	159
7.1.1 Ο Αλγόριθμος της Ομαδοποίησης	159
7.1.2 Λεκτικός Χαρακτηρισμός των Ομάδων	160
7.2 Η ΕΦΑΡΜΟΓΗ.....	161
7.3 ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΟΠΤΙΚΕΣ ΚΑΙ ΠΡΟΤΑΣΕΙΣ	161

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Η αναγνώριση των ιδιοτήτων που είναι δυνατό να μοιράζονται πολλά διαφορετικά αντικείμενα αποτελεί μία από τις βασικές ικανότητες του ανθρώπινου νου. Ως αποτέλεσμα, η ομαδοποίηση δεδομένων αποτελεί βασικό πεδίο έρευνας και ενασχόλησης στους περισσότερους τομείς της επιστήμης.

Ιδιαίτερα στα προβλήματα λήψης αποφάσεων, ερχόμαστε συχνά αντιμέτωποι με πίνακες δεδομένων οι οποίοι περιέχουν ένα αριθμό χαρακτηριστικών που περιγράφουν ορισμένα άτομα. Η διαχείριση τέτοιων δεδομένων μπορεί να γίνει με πολυκριτήριες μεθόδους και τεχνικές Ανάλυσης Δεδομένων. Η ομαδοποίηση των δεδομένων σ' αυτές τις περιπτώσεις μπορεί να αποδειχθεί χρήσιμη και αποτελεσματική δίνοντας απαντήσεις σε ποικίλα ερωτήματα.

Απαραίτητη ωστόσο προϋπόθεση αποτελεί η ύπαρξη των κατάλληλων εργαλείων που θα επιτρέπουν την γρήγορη επεξεργασία των δεδομένων κυρίως όταν ο όγκος τους είναι μεγάλος. Η πρόοδος της τεχνολογίας των Η/Υ επιτρέπει σήμερα την ανάπτυξη λογισμικού για την αντιμετώπιση τέτοιων καταστάσεων.

Στην παρούσα διπλωματική εργασία παρουσιάζεται μία μεθοδολογία και εφαρμογή χρήσης Ασαφούς Λογικής για το χαρακτηρισμό ομάδων που προκύπτουν από τη γεωμετρική ομαδοποίηση των ατόμων σε ένα γράφημα Ανάλυσης σε Κύριες Συνιστώσες (ΑΚΣ).

Τα βασικά βήματα που ακολουθήθηκαν είναι:

- Ανάπτυξη αλγορίθμου γεωμετρικής ομαδοποίησης ικανού να διαχειρίζεται ενιαία τα ποικίλα αρχικά δεδομένα της ΑΚΣ.
- Λεκτικός χαρακτηρισμός των ομάδων που προκύπτουν με τη χρήση Ασαφούς λογικής.
- Ανάπτυξη λογισμικού για την εφαρμογή της μεθοδολογίας που προτείνεται.
- Εξαγωγή συμπερασμάτων για την αποτελεσματικότητα τόσο της προτεινόμενης μεθοδολογίας όσο και του λογισμικού με την εφαρμογή σε πραγματικά δεδομένα.

Αναλυτικότερα η παρούσα εργασία αποτελείται από επτά κεφάλαια:

Στο 2ο κεφάλαιο παρουσιάζεται συνοπτικά η μέθοδος της Ανάλυσης σε Κύριες Συνιστώσες με έμφαση στο γράφημά της.

Στο 3ο κεφάλαιο παρουσιάζεται αναλυτικά τόσο η διαδικασία που ακολουθήθηκε για την ανάπτυξη του αλγορίθμου της γεωμετρικής ομαδοποίησης όσο και ο ίδιος ο αλγόριθμος.

Στο 4ο κεφάλαιο δίνεται το θεωρητικό υπόβαθρο και η μεθοδολογία που ακολουθήθηκε για τη λεκτική περιγραφή των ομάδων με χρήση της θεωρίας των Ασαφών Συνόλων.

Στο 5ο κεφάλαιο παρουσιάζεται η εφαρμογή (software) που αναπτύχθηκε προκειμένου να εφαρμοστεί η παραπάνω προτεινόμενη μεθοδολογία. Ουσιαστικά αποτελεί αναλυτικό εγχειρίδιο του Fuzzy Data Grouping Application.

Στο 6ο κεφάλαιο γίνεται η εφαρμογή της μεθοδολογίας σε πραγματικά δεδομένα με τη χρήση του λογισμικού που αναπτύχθηκε. Στόχος είναι η παρουσίαση των δυνατοτήτων και η εξαγωγή συμπερασμάτων για την αποτελεσματικότητα της εφαρμογής Fuzzy Data Grouping.

Τέλος στο 7ο κεφάλαιο αναφέρονται τα συμπεράσματα που προκύπτουν από την εργασία και οι μελλοντικές προοπτικές.

ΚΕΦΑΛΑΙΟ 2: ΑΝΑΛΥΣΗ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ

2.1 ΕΙΣΑΓΩΓΗ

Η Ανάλυση σε Κύριες Συνιστώσες είναι πιθανόν η παλαιότερη και γνωστότερη από τις τεχνικές της πολυκριτήριας ανάλυσης. Προτάθηκε από τον Pearson (1901) και αναπτύχθηκε από τον Hotelling (1933). Όπως πολλές από τις πολυκριτήριες μεθόδους, δεν χρησιμοποιήθηκε έως ότου η εξέλιξη της τεχνολογίας επιτρέψει την ανάπτυξη λογισμικού για την εύκολη εφαρμογή του αλγορίθμου της. Σήμερα συναντάται πλέον στα περισσότερα στατιστικά πακέτα.

Η βασική ιδέα της Ανάλυσης σε Κύριες Συνιστώσες είναι ο περιορισμός της διάστασης ενός συνόλου δεδομένων στο οποίο ένας μεγάλος αριθμός αλληλοσυσχετιζόμενων μεταβλητών καθιστά δύσκολη τη μελέτη των δεδομένων και τη λήψη αποφάσεων. Αυτός ο περιορισμός επιτυγχάνεται με μετασχηματισμό σε νέες μεταβλητές, τους Κύριους Άξονες, οι οποίοι είναι ασυσχέτιστοι και είναι διατεταγμένοι έτσι ώστε οι λίγοι πρώτοι

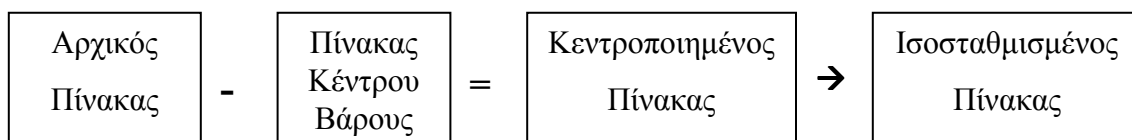
να διατηρούν την περισσότερη δυνατή πληροφορία που αναπαρίσταται στα αρχικά δεδομένα, δηλαδή στο σύνολο των αρχικών μεταβλητών.

Ο υπολογισμός των Κυρίων Αξόνων ουσιαστικά περιορίζεται στην επίλυση ενός προβλήματος ιδιοτιμών-ιδιοδιανυσμάτων ενός συμμετρικού πίνακα. Έτσι ο καθορισμός και υπολογισμός των Κυρίων Αξόνων είναι ακριβής. Αυτή η φαινομενικά απλή τεχνική έχει μία μεγάλη ποικιλία εφαρμογών, τόση όση και η ποικιλία των πηγών των δεδομένων.

2.2 ΑΝΑΛΥΣΗ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ

Στο μάνατζμεντ και στα προβλήματα λήψης αποφάσεων συχνά έχουμε να αντιμετωπίσουμε ένα πίνακα δεδομένων με ένα αριθμό χαρακτηριστικών που περιγράφουν κάποια άτομα. Έστω λοιπόν ότι έχουμε ένα διδιάστατο πίνακα A , τάξεως $N \times K$ όπου N είναι ο αριθμός των γραμμών (ατόμων) και K ο αριθμός των στηλών (χαρακτηριστικά) του πίνακα. Με τη μέθοδο της Ανάλυσης σε Κύριες Συνιστώσες επιτυγχάνεται η παράλληλη μεταφορά των αξόνων ώστε να συμπέσουν με το κέντρο βάρους G του νέφους των σημείων ενώ με κατάλληλη στροφή των αξόνων μεγιστοποιείται ή ελαχιστοποιείται η διακύμανση σε κάθε άξονα. Δηλαδή μεγιστοποιείται ή ελαχιστοποιείται το άθροισμα των τετραγώνων των προβολών των διανυσμάτων που αντιστοιχούν στα διάφορα σημεία πάνω στους νέους άξονες. Η στροφή των αξόνων γίνεται με την εύρεση των ιδιοτιμών και των ιδιοδιανυσμάτων τα οποία δίνουν τις διευθύνσεις των τελικών αξόνων. Οι τελικοί άξονες περνούν από το κέντρο βάρους (G) του νέφους των σημείων.

Συνοπτικά η εργασία που αφορά την Ανάλυση σε Κύριες Συνιστώσες παρουσιάζεται ως εξής:



Πίνακας	Ιδιοτιμές	Ποσοστό	Αθροιστικό Ποσοστό
---------	-----------	---------	--------------------

2.3 ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ

Τα αποτελέσματα της εφαρμογής της μεθόδου ΑΚΣ είναι τα εξής:

2.3.1 Κεντροποιημένος Πίνακας

Αν X_{ij} ($i=1,2,\dots,N$, $j=1,2,\dots,K$) είναι τα στοιχεία του αρχικού πίνακα δεδομένων τότε τα στοιχεία του κεντροποιημένου πίνακα δεδομένων, που έχει τις ίδιες διαστάσεις, είναι αυτά της Σχέσης 2.1.

$$X_{ij} - \bar{X}_j \text{ όπου } \bar{X}_j = \frac{\sum_{i=1}^N X_{ij}}{N} \quad \text{Σχέση 2.1}$$

Για τον Κεντροποιημένο Πίνακα Δεδομένων που χρησιμοποιείται για την εύρεση του Ισοσταθμισμένου Πίνακα Δεδομένων στο επόμενο βήμα της ΑΚΣ ισχύει η Σχέση 2.2.

$$\sum_{i=1}^N (X_{ij} - \bar{X}_j) = 0 \quad \text{Σχέση 2.2}$$

2.3.2 Ισοσταθμισμένος Πίνακας

Τα στοιχεία του Ισοσταθμισμένου Πίνακα Δεδομένων (κεντροποιημένος και κανονικοποιημένος), που είναι διαστάσεων $N \times K$, δίνονται από τη Σχέση 2.3.

$$y_{ij} = \frac{X_{ij} - \bar{X}_j}{\sqrt{\sum_i (X_{ij} - \bar{X}_j)^2}} \quad \text{Σχέση 2.3}$$

Ο πίνακας αυτός χρησιμοποιείται στο επόμενο βήμα για την εύρεση του Πίνακα Συσχετίσεων των Χαρακτηριστικών.

2.3.3 Πίνακας Συσχετίσεων των Χαρακτηριστικών

Ο Πίνακας Συσχετίσεων των Χαρακτηριστικών προκύπτει με πολλαπλασιασμό του ανάστροφου του Ισοσταθμισμένου Πίνακα επί τον Ισοσταθμισμένο Πίνακα. Είναι ένας πίνακας τετραγωνικός ($K \times K$) και τα στοιχεία της κυρίας διαγωνίου του ισούνται με τη μονάδα. Από τον πίνακα αυτό μπορούμε να διαπιστώσουμε τη συσχέτιση των χαρακτηριστικών. Ο βαθμός συσχέτισης είναι μεταξύ μείον ένα και ένα ($\in [-1, 1]$). Τιμές συσχέτισης κοντά στο μηδέν δείχνουν χαρακτηριστικά ασυσχέτιστα μεταξύ τους ενώ τιμές κοντά στο μείον ένα και το ένα δείχνουν υψηλά αρνητική και υψηλά θετική συσχέτιση αντίστοιχα.

2.3.4 Πίνακας Ιδιοτιμών

Ο Πίνακας Ιδιοτιμών είναι ένα διάνυσμα με K συνιστώσες. Το άθροισμα των ιδιοτιμών είναι ίσο με K ($\lambda_1 + \lambda_2 + \dots + \lambda_K = K$) και κάθε ιδιοτιμή αντιστοιχεί σε ένα νέο άξονα. Όσο μεγαλύτερη είναι η ιδιοτιμή που αντιστοιχεί σε ένα νέο άξονα, τόσο μεγαλύτερο είναι το ποσοστό της πληροφορίας του αρχικού πίνακα το οποίο αντιπροσωπεύει. Για τις ιδιοτιμές των κυρίων αξόνων ισχύει ότι $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_K$. Δηλαδή ο πρώτος κύριος άξονας αναπαριστά τη μέγιστη δυνατή πληροφορία, ο δεύτερος τη μέγιστη δυνατή πληροφορία από την υπολειπόμενη πληροφορία κ.τ.λ. Έτσι εξετάζοντας τους λίγους πρώτους άξονες έχουμε την ελάχιστη δυνατή χαμένη πληροφορία. Επίσης είναι γνωστό ότι ισχύει η Σχέση 2.4.

$$\lambda_j = \sum_{i=1}^N P_{ij}^2 = \sigma_j^2 \quad \text{Σχέση 2.4}$$

όπου P_{ij} είναι η προβολή του ατόμου i στον j άξονα και σ_j^2 είναι η διασπορά των N ατόμων στον j κύριο άξονα.

2.3.5 Προβολές των Ατόμων στους Νέους Άξονες

Πρόκειται για ένα πίνακα $N \times K$ που τα στοιχεία του αντιστοιχούν στις προβολές P_{ij} των ατόμων στους νέους άξονες. Λόγω κεντροποίησης (προηγούμενο βήμα) ισχύει ότι το άθροισμα των προβολών των ατόμων για κάθε νέο άξονα είναι μηδέν (Σχέση 2.5).

$$\sum_{i=1}^N P_{ij} = 0 \quad \forall j \quad \text{Σχέση 2.5}$$

Οι προβολές των ατόμων στους νέους άξονες θα χρησιμοποιηθούν στο επόμενο Κεφάλαιο για τη γεωμετρική ομαδοποίηση των ατόμων.

2.3.6 Προβολές των Χαρακτηριστικών στους Νέους Άξονες

Πρόκειται για ένα πίνακα $K \times K$ που τα στοιχεία του αντιστοιχούν στις προβολές Z_{ij} των χαρακτηριστικών στους νέους άξονες. Για τον πίνακα αυτό ισχύει η Σχέση 2.6.

$$\lambda_j = \sum_{i=1}^K Z_{ij}^2 \quad \text{Σχέση 2.6}$$

2.3.7 Γράφημα Ανάλυσης σε Κύριες Συνιστώσες

Ο σχεδιασμός του γραφήματος της ΑΚΣ μπορεί να γίνει από τη στιγμή που υπολογιστούν οι πίνακες προβολών των ατόμων και των χαρακτηριστικών στους νέους άξονες. Από το γράφημα μπορούμε να βγάλουμε συμπεράσματα τόσο για τα άτομα όσο και για τα χαρακτηριστικά. Ωστόσο η ποιοτική ανάλυση του γραφήματος δεν είναι απλή διαδικασία. Εξαρτάται σε μεγάλο βαθμό από το είδος των δεδομένων και σε περίπτωση λήψης αποφάσεων από την εμπειρία του αποφασίζοντα.

Παρακάτω θα προσπαθήσουμε να “διαβάσουμε” ένα γράφημα ΑΚΣ ώστε να δούμε τι είδους πληροφορίες μπορεί να μας δώσει τόσο για τα άτομα όσο και για τα χαρακτηριστικά. Άλλωστε η ανάλυση του γραφήματος της ΑΚΣ αποτελεί και ένα από τους κύριους στόχους της παρούσας εργασίας.

Έστω ο αρχικός πίνακας δεδομένων που φαίνεται στον Πίνακα 2.1.

Ατ.\Χαρ.	Χαρακτηριστικό 1	Χαρακτηριστικό 2	Χαρακτηριστικό 3	Χαρακτηριστικό 4
Άτομο Α	5	44	8	24
Άτομο Β	5	55	12	8
Άτομο C	5	51	17	19
Άτομο D	2	38	12	28
Άτομο E	6	37	11	26
Άτομο F	4	49	9	22
Άτομο G	7	56	14	27
Άτομο H	2	50	12	31

Πίνακας 2.1: Αρχικός Πίνακας Δεδομένων

Ο Πίνακας προβολών των ατόμων και των χαρακτηριστικών στους νέους άξονες φαίνονται στους Πίνακες 2.2 και 2.3 αντίστοιχα. Στους Πίνακες 2.4 και 2.5 φαίνονται οι Ιδιοτιμές των νέων αξόνων και οι συσχετίσεις των χαρακτηριστικών.

ΑΤ.\ΑΞ.	Άξονας 1	Άξονας 2	Άξονας 3	Άξονας 4
Άτομο Α	-0.303	-0.457	0.077	0.116
Άτομο Β	0.701	-0.338	-0.428	-0.134
Άτομο C	0.558	0.427	0.032	-0.275
Άτομο D	-0.665	0.269	-0.149	-0.249
Άτομο E	-0.311	-0.228	0.471	-0.268
Άτομο F	-0.132	-0.278	-0.199	0.191
Άτομο G	0.523	0.173	0.446	0.322
Άτομο H	-0.37	0.432	-0.25	0.296

Πίνακας 2.2: Πίνακας προβολών των ατόμων στους νέους άξονες

ΧΑΡ.\ΑΞ.	Άξονας 1	Άξονας 2	Άξονας 3	Άξονας 4
Χαρακτηριστικό 1	0.62	-0.378	0.686	0.013
Χαρακτηριστικό 2	0.815	0.138	-0.247	0.504
Χαρακτηριστικό 3	0.569	0.754	0.137	-0.295
Χαρακτηριστικό 4	-0.699	0.44	0.432	0.359

Πίνακας 2.3: Πίνακας Προβολών των χαρακτηριστικών στους νέους άξονες

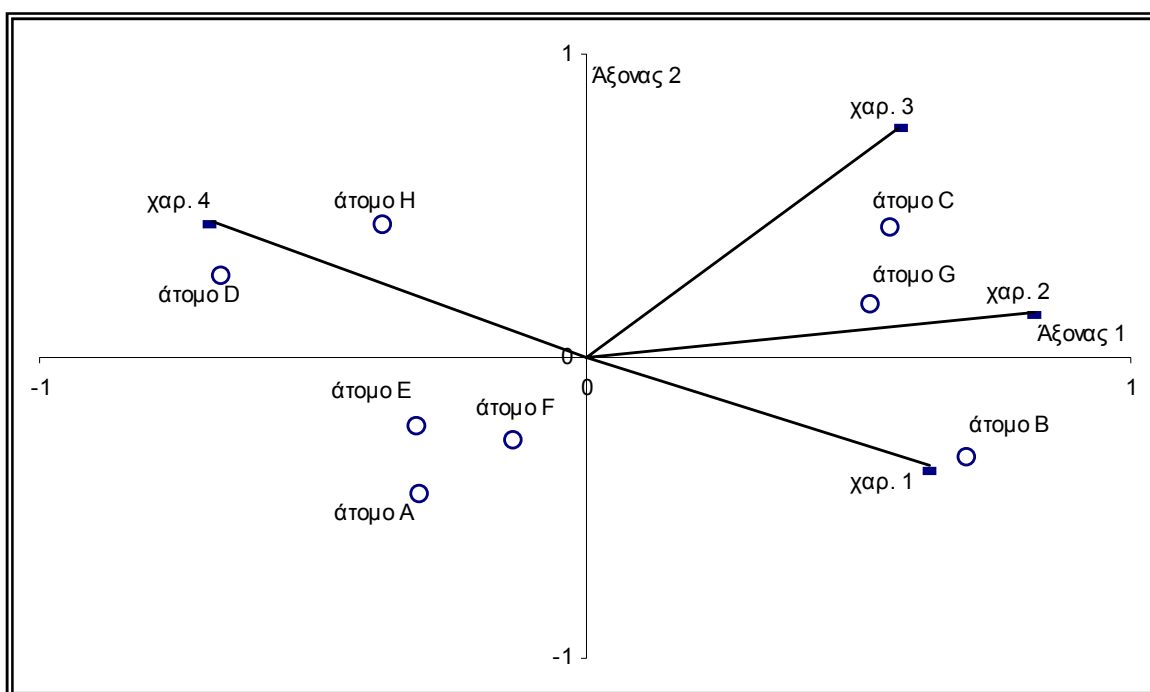
ΙΔΙΟΤΙΜΗ	Ιδιοτιμή	Ποσοστό	Αθροισμα %
Ιδιωτ. 1	1.865	46.63	46.63
Ιδιωτ. 2	0.925	23.13	69.77
Ιδιωτ. 3	0.738	18.45	88.22
Ιδιωτ. 4	0.47	11.77	100

Πίνακας 2.4: Πίνακας Ιδιοτιμών Νέων Αξόνων

ΧΑΡ.\ΧΑΡ.	Χαρ1	Χαρ2	Χαρ3	Χαρ4
Χαρ1	1	0.291	0.158	-0.299
Χαρ2	0.291	1	0.386	-0.435
Χαρ3	0.158	0.386	1	-0.113
Χαρ4	-0.299	-0.435	-0.113	1

Πίνακας 2.5: Πίνακας Συσχετίσεων των Χαρακτηριστικών

Από τους Πίνακες 2.2 και 2.3 προκύπτει το Σχήμα 2.1 που είναι το γράφημα των δύο πρώτων κύριων αξόνων της Ανάλυσης σε Κύριες Συνιστώσες.



Σχήμα 2.1: Γράφημα ΑΚΣ στους δύο πρώτους κύριους άξονες

Από τον πίνακα των Ιδιοτιμών προκύπτει ότι οι άξονες 1 και 2 απεικονίζουν το 69.77% της ολικής αρχικής πληροφορίας, δηλαδή του αρχικού πίνακα δεδομένων.

Από το γράφημα φαίνεται για τα χαρακτηριστικά ότι:

Το χαρακτηριστικό 1 προβάλλεται περισσότερο στον άξονα 1 αλλά σε σημαντικό ποσοστό και στον άξονα 2. Αυτό επαληθεύεται και από τον Πίνακα 2.3. Ωστόσο από τον πίνακα φαίνεται ότι το χαρακτηριστικό 1 προβάλλεται και στον άξονα 3. Η προβολή στον άξονα 1 είναι θετική ενώ αυτή στον άξονα 2 είναι αρνητική κάτι που σημαίνει ότι άτομα με μεγάλες θετικές προβολές στον άξονα 1 και μεγάλες αρνητικές προβολές στον άξονα 2 (με μικρότερη πιθανότητα λόγω του ότι ο άξονας 2 αντιπροσωπεύει μόνο το 23.13% της αρχικής πληροφορίας) αναμένεται να έχουν μεγάλες τιμές στο χαρακτηριστικό αυτό. Πράγματι από το γράφημα φαίνεται ότι τα άτομα B, C και G έχουν μεγάλη προβολή στον άξονα 1 επομένως και μεγάλη τιμή στο χαρακτηριστικό 1. Αντίθετα τα άτομα D και H που έχουν τόσο αρνητικές προβολές στον άξονα 1 όσο και θετικές προβολές στον άξονα 2 αναμένεται να έχουν τις μικρότερες τιμές στο χαρακτηριστικό αυτό κάτι που επαληθεύεται από τον αρχικό πίνακα δεδομένων.

Το χαρακτηριστικό 2 προβάλλεται περισσότερο στον άξονα 1. Επομένως άτομα με μεγάλες προβολές στον άξονα αυτό, αναμένεται να έχουν μεγάλες τιμές και στο χαρακτηριστικό αυτό. Μάλιστα δεδομένου ότι η προβολή του χαρακτηριστικού αυτού είναι μεγαλύτερη από την προβολή του χαρακτηριστικού 1 στον άξονα 1, μπορούμε να εξάγουμε συμπεράσματα με μεγαλύτερη ασφάλεια. Έτσι τα άτομα B, C και G αναμένεται να έχουν τις μεγαλύτερες τιμές στο χαρακτηριστικό αυτό.

Το χαρακτηριστικό 3 προβάλλεται περισσότερο στον άξονα 2 και λιγότερο στον άξονα 1. Οι προβολές είναι και οι δύο θετικές κάτι που σημαίνει ότι τα άτομα με μεγάλες θετικές προβολές στους δύο πρώτους κύριους άξονες αναμένεται να έχουν και μεγάλες τιμές στο χαρακτηριστικό 3. Πράγματι τα άτομα C και G έχουν τις μεγαλύτερες τιμές στο χαρακτηριστικό αυτό. Αντίθετα τα άτομα A, E και F έχοντας αρνητικές προβολές και

στους δύο πρώτους κύριους άξονες, έχουν και τις μικρότερες τιμές στο χαρακτηριστικό 3.

Το χαρακτηριστικό 4 προβάλλεται περισσότερο στον άξονα 1 και λιγότερο στον άξονα 2. Η προβολή του στον άξονα 1 είναι αρνητική ενώ στον άξονα 2 θετική. Αυτό σημαίνει ότι άτομα με μεγάλες αρνητικές προβολές στον άξονα 1 και μεγάλες θετικές τιμές στον άξονα 2 αντίστοιχα αναμένεται να έχουν και μεγάλες τιμές στο χαρακτηριστικό αυτό. Τα άτομα που ικανοποιούν τα παραπάνω είναι το D και το H και όπως φαίνεται και από τον αρχικό πίνακα δεδομένων έχουν τις μεγαλύτερες τιμές στο χαρακτηριστικό 4.

Για τις συσχετίσεις των χαρακτηριστικών από το γράφημα μπορούμε να δούμε ότι τα χαρακτηριστικά 1 και 2 είναι αρνητικά συσχετισμένα με το χαρακτηριστικό 4. Αυτό σημαίνει ότι άτομα με μεγάλες τιμές στα χαρακτηριστικά 1 και 2 αναμένεται να έχουν μικρές τιμές στο χαρακτηριστικό 4 και αντίστροφα. Πράγματι τα άτομα D και H έχουν μεγάλες τιμές στο χαρακτηριστικό 4 και αντίστοιχα μικρές στο χαρακτηριστικό 1.

Επίσης το χαρακτηριστικό 3 είναι μάλλον ασυσχέτιστο με τα χαρακτηριστικά 1 και 4. Πράγματι ενώ τα άτομα B και C έχουν την ίδια τιμή στο χαρακτηριστικό 1, το άτομο C έχει την μεγαλύτερη τιμή στο χαρακτηριστικό 3 ενώ το άτομο A μεσαία να όχι μικρή τιμή σ' αυτό το χαρακτηριστικό. Επίσης ενώ το άτομο H έχει τη μεγαλύτερη τιμή στο χαρακτηριστικό 4 και μία μεσαία τιμή στο χαρακτηριστικό 3, το άτομο C έχει τη μεγαλύτερη τιμή στο χαρακτηριστικό 3 και τη δεύτερη μικρότερη τιμή στο χαρακτηριστικό 4.

Σημαντικό είναι τέλος να τονίσουμε ότι λόγω κανονικοποίησης (κανονικότητας των τιμών των προβολών), ένα άτομο που θα βρισκόταν ακριβώς στο κέντρο των αξόνων, θα είχε μέση τιμή για κάθε χαρακτηριστικό. Το πιο κοντινό στο κέντρο των αξόνων άτομο δείχνει να είναι το F. Πράγματι το άτομο αυτό έχει μεσαία τιμή στα χαρακτηριστικά 1, 2 και 4 με μόνη εξαίρεση το χαρακτηριστικό 3 στο οποίο έχει μικρή τιμή, όχι πάντως την μικρότερη.

2.4 ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΧΡΗΣΗΣ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ

Στο παράδειγμα που μόλις αναλύσαμε, μερικά από τα συμπεράσματα θα μπορούσε να έχουν εξαχθεί και με απ' ευθείας παρατήρηση του αρχικού πίνακα δεδομένων. Τι γίνεται όμως σε περίπτωση που ο αριθμός των ατόμων και πολύ περισσότερο των χαρακτηριστικών είναι τέτοιος που να μην επιτρέπει μία τέτοια παρατήρηση; Το βασικό πλεονέκτημα της Ανάλυσης σε Κύριες Συνιστώσες είναι ο περιορισμός του αριθμού των σημαντικών αξόνων στους λίγους πρώτους Κύριους Άξονες. Έτσι η παρατήρηση του γραφήματος των δύο πρώτων αξόνων αναπαριστά τη μέγιστη δυνατή πληροφορία στο επίπεδο και μας επιτρέπει την εξαγωγή συμπερασμάτων. Ακολουθώντας, τόσο ο περιορισμός των διαστάσεων όσο και η κανονικότητα των τιμών των προβολών στους κύριους άξονες είναι εξαιρετικά χρήσιμα στην περίπτωση γεωμετρικής ομαδοποίησης των ατόμων αλλά και στη λεκτική περιγραφή των ομάδων που θα προκύψουν κάτι με το οποίο θα ασχοληθούμε στα επόμενα κεφάλαια.

ΚΕΦΑΛΑΙΟ 3: Ο ΑΛΓΟΡΙΘΜΟΣ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ

3.1 ΕΙΣΑΓΩΓΗ

Μία από τις πιο βασικές ικανότητες των ζώντων οργανισμών είναι η ομαδοποίηση όμοιων αντικειμένων για τη δημιουργία ταξινόμησης. Ο προϊστορικός άνθρωπος για παράδειγμα θα πρέπει να ήταν ικανός να αναγνωρίσει ότι πολλά ξεχωριστά αντικείμενα μοιράζονταν τις ίδιες ιδιότητες όπως το να είναι φαγώσιμα ή δηλητηριώδη ή άγρια κ.τ.λ. Η ιδέα της ταξινόμησης ομοίων πραγμάτων σε κατηγορίες είναι καθαρά αρχέγονη έννοια η οποία μάλιστα ήταν απαραίτητη για την ανάπτυξη της γλώσσας η οποία περιέχει λέξεις οι οποίες μας βοηθάνε να αναγνωρίζουμε και να συζητάμε τα διάφορα είδη γεγονότων και αντικειμένων με τους ανθρώπων τους οποίους καθημερινά συναλλασσόμαστε. Για παράδειγμα κάθε ουσιαστικό σε μία γλώσσα είναι μία ετικέτα η οποία χρησιμοποιείται για να περιγράψει μία κατηγορία πραγμάτων τα οποία έχουν εξέχοντα κοινά χαρακτηριστικά. Έτσι στην κατηγορία των ζώων ανήκουν οι γάτες, οι σκύλοι, τα άλογα κ.τ.λ., ενώ το κάθε ένα είδος από αυτά αποτελεί μία νέα κατηγορία (ομάδα).

Όντας λοιπόν θεμελιώδης ανθρώπινη δραστηριότητα, η ομαδοποίηση είναι επίσης βασική και στα περισσότερους τομείς και αντικείμενα της επιστήμης.

Οι λόγοι που μπορούν να μας οδηγήσουν στην ομαδοποίηση δεδομένων είναι αρκετοί. Είναι γεγονός ότι η ομαδοποίηση μπορεί να αποδειχθεί αποτελεσματική μέθοδος για την οργάνωση και διαχείριση μεγάλου όγκου δεδομένων έτσι ώστε η ανάκτηση των πληροφοριών να είναι πιο εύκολη. Επίσης περιγράφοντας τις ομοιότητες και τις διαφορές μεταξύ αντικειμένων και ερευνώντας τις ιδιότητες των ομάδων στις οποίες ανήκουν μπορεί να εξασφαλιστεί μία περίληψη των δεδομένων. Τέλος σε περιπτώσεις λήψης αποφάσεων, όπως για παράδειγμα η επιλογή ενός μηχανήματος για μία επιχείρηση ανάμεσα σε δεκάδες, η ομαδοποίηση των δεδομένων μπορεί να βοηθήσει ώστε η απόφαση να είναι γρηγορότερη και ασφαλέστερη.

3.2 ΟΜΑΔΟΠΟΙΗΣΗ ΣΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΑΚΣ

Εμείς θα επιχειρήσουμε την ομαδοποίηση των δεδομένων ενός αρχικού πίνακα σαν αυτού της Ανάλυσης σε Κύριες Συνιστώσες. Βλέποντας τα N σημεία στον K -διάστατο Ευκλείδειο χώρο (ένας άξονας για κάθε χαρακτηριστικό), η γεωμετρική ομαδοποίηση μπορεί να αποτελεί μία λύση. Το πρόβλημα σε αυτή την περίπτωση είναι πρακτικό ακόμα και για λίγα χαρακτηριστικά. Η αναζήτηση ομάδων μπορεί να γίνει περίπλοκη και χρονοβόρα διαδικασία. Η Ανάλυση σε Κύριες Συνιστώσες έχει το βασικό πλεονέκτημα του περιορισμού του αριθμού των σημαντικών αξόνων στους λίγους πρώτους Κύριους Άξονες. Περιορίζοντας λοιπόν τη διάσταση του προβλήματος μπορούμε να αναπαραστήσουμε τα άτομα σε λιγότερων διαστάσεων Ευκλείδειο χώρο, περιορίζοντας έτσι και το χρόνο αλλά και την περιπλοκότητα της ομαδοποίησης (των απαραίτητων υπολογισμών).

Υπάρχουν αρκετοί τρόποι να προσεγγίσει κανείς το πρόβλημα της γεωμετρικής ομαδοποίησης. Ένας θα ήταν να θεωρήσουμε αρχικά το κάθε άτομο ως μία ομάδα και με κατάλληλες διαδικασίες να ενώνουμε τις ομάδες. Μία άλλη ιδέα θα ήταν να θεωρήσουμε

ότι αρχικά όλα τα άτομα ανήκουν στην ίδια ομάδα και με κατάλληλες διαδικασίες να διασπάσουμε αυτή την ομάδα σε μικρότερες. Μία τρίτη ιδέα θα ήταν να θεωρήσουμε ότι κανένα άτομο δεν ανήκει σε ομάδα και να αρχίσουμε να ενώνουμε άτομα δημιουργώντας ομάδες.

Σε κάθε περίπτωση όμως απαραίτητο θεωρείται να υπάρχουν κριτήρια τέλους του αλγορίθμου ώστε να μπορούμε να τον κατευθύνουμε προς την ομαδοποίηση που θέλουμε. Τέτοια κριτήρια μπορεί να είναι η “μέγιστη απόσταση μεταξύ ατόμων της ίδιας ομάδας”, ο “αριθμός των ομάδων”, η “διασπορά των ατόμων μίας ομάδας από το κέντρο της” κ.τ.λ.

Η προσέγγιση που χρησιμοποιούμε είναι η εξής: Αρχικά κανένα άτομο δεν ανήκει σε ομάδα και ο αριθμός των ομάδων είναι μηδέν. Ο αλγόριθμος της ομαδοποίησης έχει τη δυνατότητα να δημιουργεί νέες ομάδες όπου τοποθετεί τα γεωμετρικά πλησιέστερα ή γειτονικά άτομα. Προφανώς στο βήμα αυτό ομάδα που θα δημιουργηθεί θα αποτελείται από τα δύο πιο κοντινά άτομα. Αυτή τη στιγμή τα ομαδοποιημένα άτομα θα είναι δύο και τα μη ομαδοποιημένα $N-2$. Στη συνέχεια ο αλγόριθμος ενώνει μη ομαδοποιημένα άτομα μεταξύ τους ή μη ομαδοποιημένα άτομα με ήδη υπάρχουσες ομάδες ή αν χρειάζεται ομάδες με ομάδες χρησιμοποιώντας τη λογική ότι κάθε άτομο πρέπει να ανήκει στην ίδια ομάδα με το πλησιέστερο σε αυτό γεωμετρικά. Επίσης αναπτύσσονται διάφορες τεχνικές αποφυγής ιδιαίτερων περιπτώσεων, όπου η μορφή των δεδομένων μπορεί να οδηγήσει σε μη “σωστές” ομαδοποιήσεις.

3.2.1 Υπολογισμός Γεωμετρικών Αποστάσεων

Κατά τη διαδικασία της ομαδοποίησης, απαιτείται ο υπολογισμός της γεωμετρικής απόστασης μεταξύ των σημείων. Υπάρχουν γι' αυτό τρεις βασικές μέθοδοι:

α) Η μέθοδος MIN όπου:

Ένα σημείο εντάσσεται σε μία ομάδα, αν είναι κοντινότερο σε ΕΝΑ αντικείμενο της ομάδας αυτής, απ' ό,τι σε ένα σημείο μίας άλλης ομάδας.

β) Η μέθοδος MAX όπου:

Ένα σημείο εντάσσεται σε μία ομάδα, αν είναι κοντινότερο σε ΟΛΑ τα σημεία της ομάδας αυτής, απ' ό,τι σε όλα τα σημεία μίας άλλης ομάδας.

γ) Η μέθοδος MOYENNE όπου:

Ένα σημείο εντάσσεται σε μία ομάδα, αν είναι κοντινότερο κατά μέσο όρο στα σημεία της ομάδας αυτής, απ' ό,τι στα σημεία μίας άλλης ομάδας.

Κάθε μία από τις παραπάνω μεθόδους έχει διαφορετική απόκριση (αποτέλεσμα ομαδοποίησης) στα διάφορα και ποικίλα δεδομένα που μπορεί να εφαρμοστεί η ΑΚΣ. Για παράδειγμα αν τα δεδομένα σχηματίζουν “αλυσίδες” ή είναι συμπαγή ή υπάρχουν σαφείς οπτικοί διαχωρισμοί. Έτσι καθίσταται αναγκαία η δημιουργία ενός ευέλικτου αλγορίθμου που θα αντιμετωπίζει τις ποικίλες μορφές δεδομένων στο σύνολό τους κάτι που θα είναι προς όφελος του χρήστη ο οποίος δεν είναι βέβαιο ότι μπορεί να επιλέξει κάθε φορά την κατάλληλη μέθοδο. Η μέθοδος που χρησιμοποιούμε μοιάζει με τη μέθοδο MIN είναι ωστόσο αρκετά πιο περίπλοκη και έχει ως στόχο την ενιαία αντιμετώπιση κάθε μορφής δεδομένων.

3.2.2 Κριτήρια Τερματισμού της Ομαδοποίησης

Τα κριτήρια τερματισμού του αλγορίθμου της ομαδοποίησης είναι δύο. Δεδομένου ότι δεν υπάρχει αντικειμενικότητα στον αριθμό και η συνεκτικότητα των ομάδων που θα δημιουργηθούν είναι δυνατή η ρύθμιση των παραμέτρων των κριτηρίων από τον χρήστη με κατάλληλες εντολές.

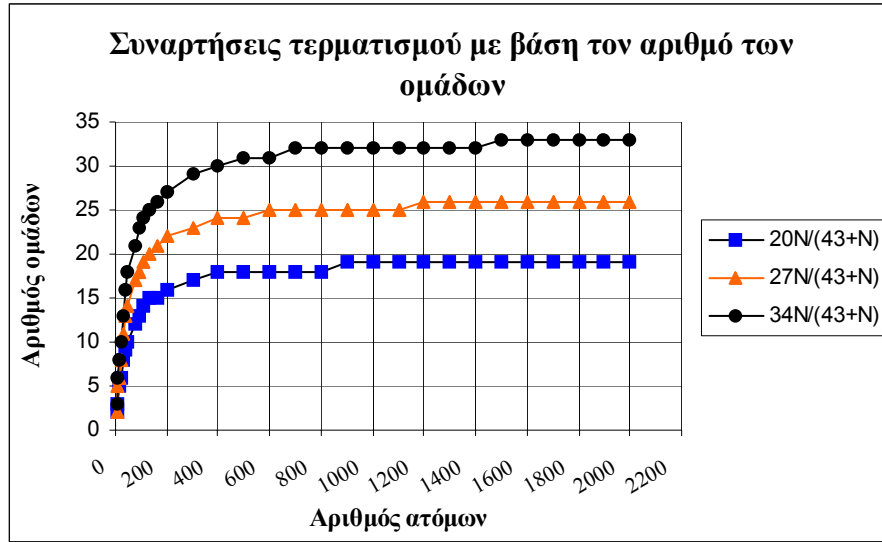
Αριθμός των ομάδων

Υπάρχουν τρεις επιλογές τερματισμού του αλγορίθμου ανάλογα με τις εντολές του χρήστη. Οι εντολές αυτές αντιστοιχούν σε μικρό, μεσαίο και μεγάλο αριθμό ομάδων αντίστοιχα. Οι αριθμοί των ομάδων δίνονται από αυθαίρετα επιλεγμένες συναρτήσεις με μόνο οδηγό την εμπειρία και είναι αυτές που φαίνονται στον Πίνακα 3.1.

Εντολή	Συνάρτηση
Λίγες ομάδες (Few)	$\frac{20N}{43 + N}$
Κανονικές ομάδες (Normal)	$\frac{27N}{43 + N}$
Πολλές ομάδες (Several)	$\frac{34N}{43 + N}$

Πίνακας 3.1: Συναρτήσεις τερματισμού με βάση τον αριθμό των ομάδων (N = αριθμός ατόμων)

Στο Σχήμα 3.1 φαίνονται οι γραφικές παραστάσεις των συναρτήσεων του Πίνακα 3.1.



Σχήμα 3.1: Συναρτήσεις τερματισμού του αλγορίθμου της ομαδοποίησης

Όπως φαίνεται και στο γράφημα η μαύρη γραμμή δίνει τον μεγαλύτερο αριθμό ομάδων τείνοντας στις 34 (για $N \rightarrow \infty$) ενώ η κόκκινη και η μπλε γραμμή τείνουν για άπειρο πλήθος ατόμων σε αριθμούς ομάδων 27 και 20 αντίστοιχα.

Μέγιστη απόσταση μεταξύ ατόμων της ίδιας ομάδας

Η μέγιστη απόσταση μεταξύ ατόμων της ίδιας ομάδας δίνεται από τη Σχέση 3.1.

$$\max \{D_{a,b}\} \text{ για κάθε } a, b \text{ που ανήκουν στην ίδια ομάδα}$$

Σχέση 3.1

όπου $D_{a,b}$ είναι η απόσταση του σημείου a από το σημείο b και δίνεται από την Σχέση 3.2.

Στην εφαρμογή Fuzzy Data Grouping η μέγιστη απόσταση μεταξύ ατόμων της ίδιας ομάδας είναι εκφρασμένη ως ποσοστό του μήκους του κύριου άξονα που φαίνεται στον άξονα X του γραφήματος της ΑΚΣ.

3.3 Η ΑΝΑΠΤΥΞΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ

Η ομαδοποίηση των δεδομένων βασίζεται στις αποστάσεις των σημείων (ατόμων) μεταξύ τους. Οι αποστάσεις αυτές μπορούν να προσδιοριστούν δεδομένης της ύπαρξης του πίνακα προβολών των ατόμων στους νέους άξονες $P_{i,j}$ ($i=1,2,\dots,N$ άτομα, $j=1,2,\dots,K$ άξονες) του οποίου ο προσδιορισμός είναι βήμα της μεθόδου Ανάλυση σε Κύριες Συνιστώσες. Ωστόσο για λόγους που θα εξηγήσουμε παρακάτω οι αποστάσεις μεταξύ των σημείων θα προσδιοριστούν από ένα αντίγραφο του $P_{i,j}$ τον $MP_{i,j}$.

Στον K -διάστατο χώρο (K νέοι άξονες) η απόσταση $D_{a,b}$ μεταξύ δύο σημείων “a” και “b” δίνεται από τη Σχέση 3.2.

$$D_{a,b} = \sqrt{\sum_{j=1}^K (MP_{a,j} - MP_{b,j})^2} \quad \text{Σχέση 3.2}$$

Ο πίνακας $D_{a,b}$ είναι συμμετρικός και τριγωνικός και τα στοιχεία της κύριας διαγωνίου είναι μηδενικά αφού αντιστοιχούν στις αποστάσεις των σημείων με τον εαυτό τους. Για παράδειγμα το στοιχείο $D_{4,5}$ αντιστοιχεί στην απόσταση του σημείου 4 από το σημείο 5.

Έτσι βρίσκοντας την μικρότερη μη μηδενική τιμή του πίνακα $D_{a,b}$ (η οποία αντιστοιχεί στα κοντινότερα σημεία) και ανατρέχοντας στον αριθμό γραμμής και τον αριθμό στήλης του πεδίου του πίνακα της τιμής αυτής βρίσκουμε τα δύο σημεία I και J με τη μικρότερη απόσταση μεταξύ τους. Ο αλγόριθμος προχωράει στη δημιουργία μίας ομάδας και την τοποθέτηση των δύο ατόμων I και J σ’ αυτήν.

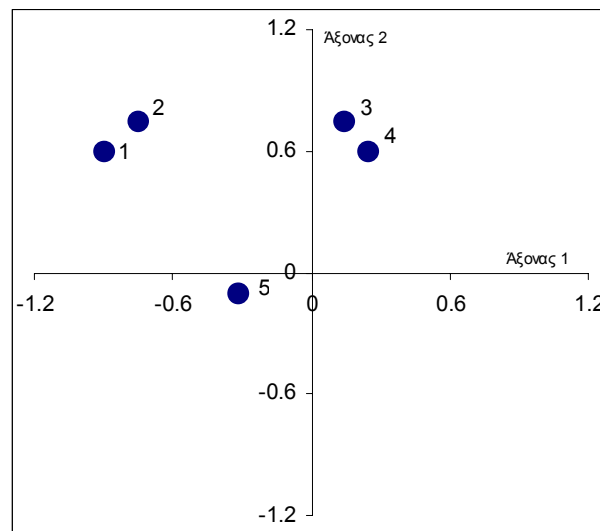
Σε κάθε επανάληψη της διαδικασίας της ομαδοποίησης ομαδοποιούνται κανένα, ένα ή δύο μη ομαδοποιημένα άτομα, ανάλογα με την κατάσταση των I και J (ομαδοποιημένο, μη ομαδοποιημένο) εκείνη τη στιγμή. Μία λογική συνέχεια θα ήταν να ανατρέξουμε ξανά στον πίνακα $D_{a,b}$, να βρούμε την επόμενη μικρότερη τιμή και να συνεχίσουμε ανάλογα.

3.3.1 Χρήση Τροποποιημένων Προβολών

Ας δούμε όμως τι θα προέκυπτε σε μία τέτοια περίπτωση. Έστω τα πέντε άτομα με προβολές στους νέους άξονες $P_{i,j}$ που δίνονται στον Πίνακα 3.2. Θα επιχειρήσουμε ομαδοποίηση στις δύο διαστάσεις (άξονας 1 και άξονας 2) ώστε να έχουμε εποπτική εικόνα. Το γράφημα στους άξονες 1 και 2 φαίνεται στο Σχήμα 3.2 και ο προκύπτων από τον Πίνακα 3.2 και τη Σχέση 3.2 πίνακας αποστάσεων $D_{a,b}$ θα είναι αυτός που φαίνεται στον Πίνακα 3.3.

Σημείο /προβολή	$P_{i,1}$	$P_{i,2}$
1	-0.9	0.6
2	-0.75	0.75
3	0.14	0.75
4	0.25	0.6
5	-0.32	-0.1

Πίνακας 3.2: Οι προβολές (συντεταγμένες) των πέντε σημείων



Σχήμα 3.2: Επίπεδο γράφημα πέντε σημείων

	1	2	3	4	5
1	0				
2	0.212	0			
3	1.051	0.89	0		
4	1.150	1.011	0.186	0	
5	0.909	0.953	0.966	0.903	0

Πίνακας 3.3: Οι αποστάσεις των πέντε σημείων μεταξύ τους στο επίπεδο

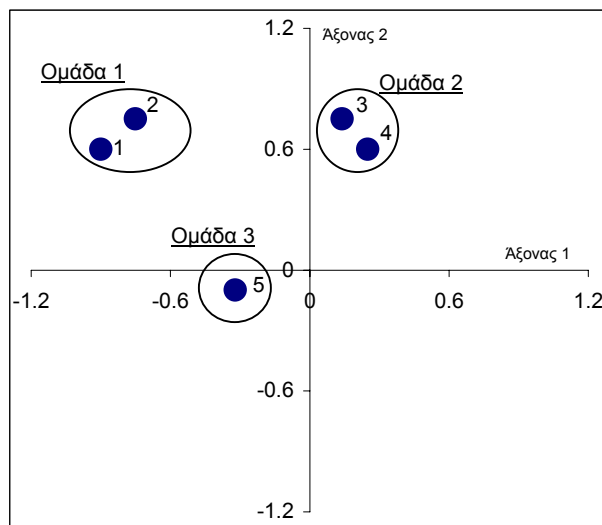
Παρακάτω θα συγκρίνουμε τα αποτελέσματα που προκύπτουν με τη μέθοδο MIN και με τη μέθοδο του Κοντινότερου Γείτονα (Nearest Neighbour) την οποία θα χρησιμοποιήσουμε εμείς. Από τη σύγκριση και τη μελέτη των αποτελεσμάτων θα μπορέσουμε να διακρίνουμε τα βασικά χαρακτηριστικά του αλγόριθμου της ομαδοποίησης που χρησιμοποιούμε στην εφαρμογή Fuzzy Data Grouping.

Μέθοδος MIN:

Η μικρότερη τιμή του Πίνακα 3.3 (0.186) αντιστοιχεί στην απόσταση των σημείων 3 και 4. Επομένως δημιουργείται μία ομάδα όπου τοποθετούνται αυτά τα δύο σημεία. Η αμέσως μικρότερη τιμή (0.212) αντιστοιχεί στα άτομα 1 και 2. Επομένως δημιουργείται μία δεύτερη ομάδα που τα περιέχει. Εάν οι ομάδες στόχος είναι τρεις, ο αλγόριθμος δημιουργεί μία τρίτη ομάδα και τοποθετεί το άτομο 5 σ' αυτήν. Έτσι οι τρεις ομάδες είναι αυτές που φαίνονται στον Πίνακα 3.4 και στο Σχήμα 3.3.

Ομάδα	Σημεία
1	1,2
2	3,4
3	5

Πίνακας 3.4: Ομαδοποίηση των πέντε σημείων σε τρεις ομάδες με τη Μέθοδο MIN

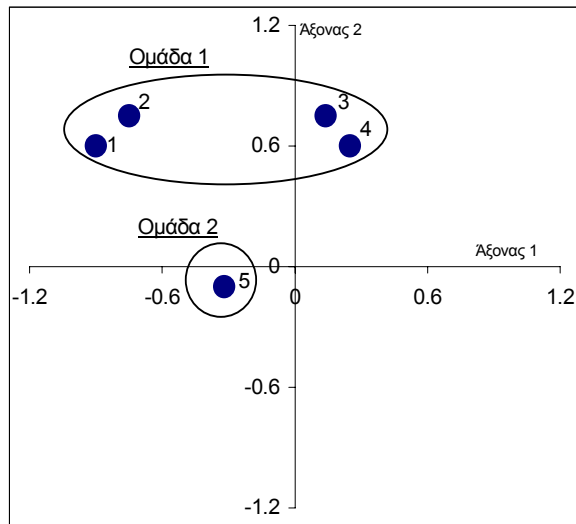


Σχήμα 3.3: Γράφημα της ομαδοποίησης των πέντε σημείων σε τρεις ομάδες

Εάν όμως οι ομάδες στόχος είναι δύο τότε πρέπει να ανατρέξουμε στον πίνακα των αποστάσεων (Πίνακας 3.3) για την επόμενη μικρότερη τιμή. Αυτή (0.89) αντιστοιχεί στα άτομα 2 και 3. Δεδομένου όμως ότι τα άτομα αυτά είναι ήδη ομαδοποιημένα, ο αλγόριθμος προχωράει σε ένωση των ομάδων 1 και 2. Έτσι οι δύο ομάδες είναι αυτές που φαίνονται στον Πίνακα 3.5 και στο Σχήμα 3.4.

Ομάδα	Σημεία
1	1,2,3,4
2	5

Πίνακας 3.5: Ομαδοποίηση των πέντε σημείων σε δύο ομάδες με τη Μέθοδο MIN



Σχήμα 3.4: Γράφημα της ομαδοποίησης των πέντε σημείων σε δύο ομάδες

Μέθοδος Κοντινότερου Γείτονα (Nearest Neighbour):

Ένας εναλλακτικός τρόπος αντιμετώπισης είναι ο εξής: Αφού βρούμε την πρώτη ελάχιστη μη μηδενική τιμή του πίνακα $D_{a,b}$ και τοποθετήσουμε τα άτομα στα οποία αντιστοιχεί στην ίδια ομάδα, μπορούμε να βρούμε το κέντρο βάρους της ομάδας που δημιουργήθηκε και να αντικαταστήσουμε την τιμή της προβολής των ατόμων της ομάδας με αυτή την τιμή. Οι συντεταγμένες του κέντρου βάρους μίας ομάδας όταν εκτελούμε ομαδοποίηση στον d -διάστατο χώρο είναι (m_1, m_2, \dots, m_d) και υπολογίζονται από τη Σχέση 3.3. Αυτό βέβαια προϋποθέτει τον εκ νέου υπολογισμό του πίνακα $D_{a,b}$ και την εύρεση όχι της αμέσως μικρότερης απόστασης αλλά της πρώτης μικρότερης απόστασης του νέου πίνακα αποστάσεων $P_{i,j}$. Προκειμένου όμως να μην χάσουμε τον πίνακα $P_{i,j}$ θα χρησιμοποιήσουμε ένα αντίγραφο του πίνακα αυτού, τον $MP_{i,j}$.

$$m_d = \sum_{i=1}^{N_{Group}} P_{i,d} \quad \text{Σχέση 3.3}$$

όπου είναι:

d ο αριθμός των νέων αξόνων στους οποίους εφαρμόζουμε την ομαδοποίηση,

m_d η μέση τιμή των προβολών των σημείων της ομάδας στον άξονα d ,

$P_{i,d}$ η προβολή του σημείου i στον άξονα d ,

N_{Group} ο αριθμός των σημείων που ανήκουν στην ομάδα.

Έτσι τα βήματα της ομαδοποίησης είναι:

Βήμα 1:

- Η μικρότερη απόσταση στον πίνακα $D_{a,b}$ (Πίνακας 3.6) είναι 0.186 και αντιστοιχεί στην απόσταση των ατόμων 3 και 4.

	1	2	3	4	5
1	0				
2	0.212	0			
3	1.051	0.89	0		
4	1.150	1.011	0.186	0	
5	0.909	0.953	0.966	0.903	0

Πίνακας 3.6: Αποστάσεις μεταξύ των πέντε σημείων στο Βήμα 1

- Δημιουργία της ομάδας 1 που περιέχει τα άτομα 3 και 4.
- Εύρεση του κέντρου βάρους της ομάδας 1 και αντικατάσταση των προβολών των ατόμων στον πίνακα $MP_{i,j}$ (Πίνακας 3.7).

	Άξονας 1	Άξονας 2
1	-0.9	0.6
2	-0.75	0.75
3	0.195	0.675
4	0.195	0.675
5	-0.32	-0.1

Πίνακας 3.7: Τροποποιημένες προβολές σημείων μετά το Βήμα 1

Βήμα 2:

- Υπολογισμός του πίνακα $D_{a,b}$ (Πίνακας 3.8) από τον $MP_{i,j}$.

	1	2	3	4	5
1	0				
2	0.212	0			
3	1.098	0.948	0		
4	1.098	0.948	0	0	
5	0.909	0.953	0.931	0.931	0

Πίνακας 3.8: Αποστάσεις μεταξύ των πέντε σημείων στο Βήμα 2

- Η μικρότερη απόσταση στον $D_{a,b}$ είναι 0.212 και αντιστοιχεί στην απόσταση των ατόμων 1 και 2.
- Δημιουργία της ομάδας 2 που περιέχει τα άτομα 1 και 2.
- Εύρεση του κέντρου βάρους της ομάδας 2 και αντικατάσταση των προβολών των ατόμων στον πίνακα $MP_{i,j}$ (Πίνακας 3.9).

	Άξονας 1	Άξονας 2
1	-0.825	0.675
2	-0.825	0.675
3	0.195	0.675
4	0.195	0.675
5	-0.32	-0.1

Πίνακας 3.9: Τροποποιημένες προβολές σημείων μετά το Βήμα 2

Βήμα 3:

Υπολογισμός του πίνακα $D_{a,b}$ (Πίνακας 3.10) από τον $MP_{i,j}$.

	1	2	3	4	5
1	0				
2	0	0			
3	1.020	1.020	0		
4	1.020	1.020	0	0	
5	0.925	0.925	0.931	0.931	0

Πίνακας 3.10: Αποστάσεις μεταξύ των πέντε σημείων στο Βήμα 3

- Η μικρότερη απόσταση στον πίνακα $D_{a,b}$ είναι 0.925 και αντιστοιχεί στην απόσταση των ατόμων 1 και 5.
- Προσχώρηση του ατόμου 5 στην ομάδα 1 που τώρα περιέχει τα άτομα 1, 2, και 5.
- Εύρεση του κέντρου βάρους της ομάδας 1 και αντικατάσταση των προβολών των ατόμων στον πίνακα $MP_{i,j}$ (Πίνακας 3.11).

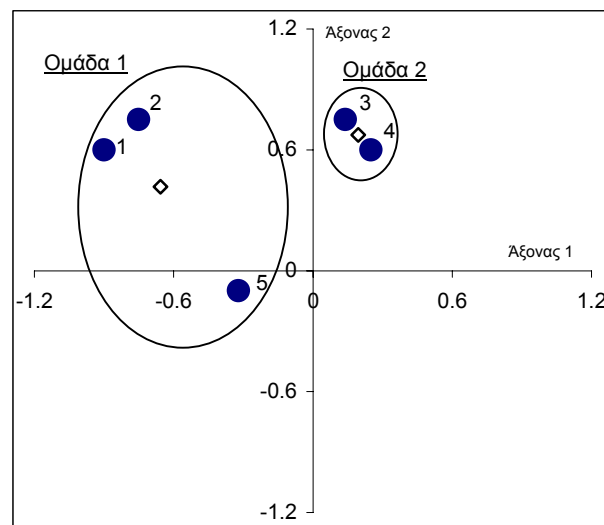
	Άξονας 1	Άξονας 2
1	-0.657	0.417
2	-0.657	0.417
3	0.195	0.675
4	0.195	0.675
5	-0.657	0.417

Πίνακας 3.11: Τροποποιημένες προβολές σημείων μετά το Βήμα 3

Το αποτέλεσμα της ομαδοποίησης με τη μέθοδο αυτή φαίνεται στον Πίνακα 3.12 και στο Σχήμα 3.5.

Ομάδα	Σημεία
1	1,2,5
2	3,4

Πίνακας 3.12: Ομαδοποίηση των πέντε σημείων σε δύο ομάδες με τη Μέθοδο Nearest Neighbour



Σχήμα 3.5: Γράφημα της ομαδοποίησης των πέντε σημείων σε δύο ομάδες

Η παραπάνω περίπτωση κάθε άλλο παρά σπάνια μπορεί να θεωρηθεί και το ερώτημα που τίθεται είναι ποια από τις δύο μέθοδος είναι περισσότερο κατάλληλη και γιατί. Θα συγκρίνουμε τις μεθόδους από την ικανότητά τους να ομαδοποιούν τα συγκεκριμένα δεδομένα σε δύο ομάδες.

Ως κριτήριο θα χρησιμοποιήσουμε τη “μέση απόσταση των σημείων από το κέντρο της ομάδας τους”. Ως κέντρο μίας ομάδας θεωρούμε τη μέση τιμή των προβολών των σημείων της ομάδας για κάθε κύριο άξονα. Οι συντεταγμένες του κέντρου μίας ομάδας συμβολίζονται με $(CG_1, CG_2, CG_3, \dots, CG_K)$. Η ποσότητα CG_j για κάθε άξονα υπολογίζεται από τη Σχέση 3.4, ενώ η μέση απόσταση από τη Σχέση 3.5. Είναι προφανές ότι στόχος είναι η ελαχιστοποίηση του κριτηρίου στο σύνολο των ομάδων.

$$CG_j = \frac{\sum_{i=1}^{N_{Group}} P_{ij}}{N_{Group}} \quad \text{Σχέση 3.4}$$

$$\text{Μέση απόσταση σημείων από το κέντρο της ομάδας τους} = \frac{\sum_{i=1}^{N_{Group}} A}{N_{Group}} \quad \text{Σχέση 3.5}$$

όπου A είναι η απόσταση του σημείου i από το κέντρο της ομάδας του και υπολογίζεται από τη σχέση:

$$A = \sqrt{\sum_{j=1}^K (P_{ij} - CG_j)^2} \quad \text{Σχέση 3.6}$$

Για το παράδειγμα που αναλύσαμε έχουμε για τις δύο μεθόδους τους Πίνακες 3.13 και 3.14

Μέθοδος MIN		
	Σημεία	Μέση απόσταση σημείων από το κέντρο της ομάδας τους
Ομάδα 1	1,2,3,4	0.5156
Ομάδα 2	5	0

Πίνακας 3.13: Ομαδοποίηση σε δύο ομάδες με τη μέθοδο MIN και μέση απόσταση σημείων από το κέντρο της ομάδας τους

Μέθοδος Nearest Neighbour		
	Σημεία	Μέση απόσταση σημείων από το κέντρο της ομάδας τους
Ομάδα 1	1,2,5	0.4225
Ομάδα 2	3,4	0.0930

Πίνακας 3.14: Ομαδοποίηση σε δύο ομάδες με τη μέθοδο Nearest Neighbour και μέση απόσταση σημείων από το κέντρο της ομάδας τους

Διαπιστώνουμε ότι με τη δεύτερη μέθοδο η μεγαλύτερη μέση απόσταση είναι μικρότερη απ’ ότι με την πρώτη μέθοδο. Επομένως κρίνουμε τη δεύτερη μέθοδο πιο αποτελεσματική από την πρώτη δεδομένου ότι η μέση απόσταση των σημείων από το κέντρο της ομάδας τους είναι ένα κριτήριο “συνοχής” ή “ομοιογένειας” των ομάδων.

Δεδομένου επίσης ότι συνήθως στον πρώτο κύριο άξονα της ΑΚΣ έχουμε περισσότερα κοινά χαρακτηριστικά για τα άτομα, προκύπτει ένας ακόμα σημαντικός λόγος να θεωρήσουμε ότι η ομαδοποίηση του Σχήματος 3.5 είναι η περισσότερο επιθυμητή. Εξάλλου η χρήση του πίνακα MP() βοηθάει στην αποφυγή ομαδοποίησης σημείων που σχηματίζουν αλυσίδες. Αυτό είναι κάτι που δεν μας εξυπηρετεί κατά τη διαδικασία χαρακτηρισμού των ομάδων στα παρακάτω κεφάλαια.

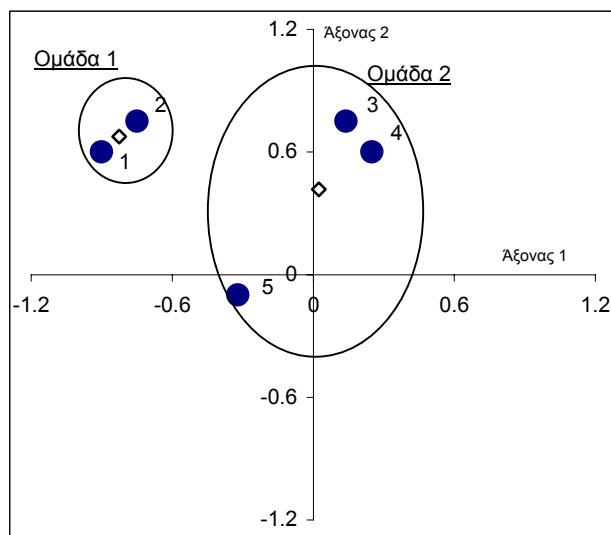
Ένα άλλο ενδιαφέρον σημείο είναι η δυνατότητα του αλγορίθμου να δημιουργεί ομάδες που αποτελούνται από ένα μόνο σημείο (άτομο). Αυτή η δυνατότητα είναι εξαιρετικά χρήσιμη σε περιπτώσεις ύπαρξης πολύ απομακρυσμένων σημείων (πολύ μεγάλη συγκριτικά με τα άλλα σημεία προβολή στον άξονα X είτε στον άξονα Y). Κι’ αυτό γιατί σε αντίθετη περίπτωση θα έπρεπε το απομακρυσμένο σημείο υποχρεωτικά να ανήκει

στην ίδια ομάδα με κάποιο σημείο ίσως και κοντά στο κέντρο των αξόνων. Ωστόσο σε περιπτώσεις σαν την παραπάνω κανένα σημείο δεν είναι τόσο απομακρυσμένο ώστε κάτι τέτοιο να επιβάλλεται. Επομένως μπορούμε να υποστηρίξουμε ότι με τη Μέθοδο Nearest Neighbour όχι μόνο ελαττώσαμε τη μέγιστη μέση απόσταση αλλά πετύχαμε και μία πιο λογική κατανομή των σημείων στις ομάδες.

Το παραπάνω παράδειγμα και είναι κλασικό και αποτελεί πολλές φορές υποπρόβλημα της γεωμετρικής ομαδοποίησης. Άλλωστε ο πειραματισμός τόσο σε δεδομένα προερχόμενα από γεννήτρια τυχαίων αριθμών όσο και σε πραγματικά, μας έδειξαν ότι η δεύτερη μέθοδος δίνει καλύτερα αποτελέσματα.

3.3.2 Ο Κοντινότερος Γείτονας

Ένα αναμενόμενο πρόβλημα της γεωμετρικής ομαδοποίησης με τη Μέθοδο του Κοντινού Γείτονα σε αντίθεση με τη Μέθοδο MIN είναι το ενδεχόμενο κάποια σημεία να ανήκουν σε διαφορετική ομάδα από το γεωμετρικά πλησιέστερό τους. Αυτό μπορεί να συμβεί δεδομένου ότι ο πίνακας $MP_{i,j}$ δεν περιέχει τις πραγματικές τιμές των προβολών των σημείων στους νέους άξονες. Επομένως ούτε ο προκύπτων από αυτόν $D_{a,b}$ περιέχει τις πραγματικές τιμές των αποστάσεων μεταξύ των σημείων. Στο παραπάνω παράδειγμα από παρατήρηση του πίνακα $D_{a,b}$ στο πρώτο βήμα, που περιέχει τις πραγματικές τιμές των αποστάσεων (αφού ο $MP_{i,j}$ είναι αντίγραφο του $P_{i,j}$), προκύπτει ότι το πλησιέστερο του σημείου 5 είναι το σημείο 4. Παρ' όλα αυτά το σημείο 5 προσχωρεί στην ομάδα των σημείων 1 και 2. Προκειμένου να αντιμετωπίσουμε αυτό το πρόβλημα είμαστε υποχρεωμένοι να βρίσκουμε το πλησιέστερο των σημείων που αντιστοιχούν στην μικρότερη τιμή του $D_{a,b}$ χρησιμοποιώντας τις πραγματικές προβολές των σημείων στους νέους άξονες. Αυτού δεδομένου, η σωστή ομαδοποίηση και η μέση απόσταση των σημείων από το κέντρο της ομάδας τους για το παραπάνω παράδειγμα φαίνεται στο Σχήμα 3.6 και στον Πίνακα 3.15 αντίστοιχα.



Σχήμα 3.6: Τελική ομαδοποίηση των πέντε σημείων σε δύο ομάδες με τη μέθοδο Nearest Neighbour

Μέθοδος Nearest Neighbour		
	Σημεία	Μέση απόσταση σημείων από το κέντρο της ομάδας τους
Ομάδα 1	1,2	0.4217
Ομάδα 2	3,4,5	0.1061

Πίνακας 3.15: Ομαδοποίηση σε δύο ομάδες με τη μέθοδο Nearest Neighbour και μέση απόσταση σημείων από το κέντρο της ομάδας τους χρησιμοποιώντας το κριτήριο του Κοντινού Γείτονα

Διαπιστώνουμε ότι με την ομαδοποίηση του σημείου 5 στην ομάδα που ανήκει το γεωμετρικά πλησιέστερο σε αυτό σημείο, η μέγιστη μέση απόσταση ελαττώθηκε κι άλλο κάτι που συμφωνεί με την πρόθεσή μας για ελαχιστοποίηση αυτού του κριτηρίου.

Η περαιτέρω περιγραφή του αλγορίθμου της ομαδοποίησης γίνεται εξαιρετικά περίπλοκη. Κι αυτό γιατί πολλά είναι τα ενδεχόμενα που μπορεί να προκύψουν μετά την εύρεση των πλησιέστερων στα σημεία που αντιστοιχούν στην μικρότερη τιμή του $D_{a,b}$. Γι' αυτό το λόγο θα προχωρήσουμε στην περιγραφή του αλγορίθμου με τη βοήθεια διαγραμμάτων ροής που θα μας διευκολύνουν τόσο στην ακριβή και λεπτομερή διατύπωση όσο και από οργανωτικής πλευράς.

Στα διαγράμματα ροής χρησιμοποιείται απ' ευθείας ο κώδικας του αλγορίθμου της ομαδοποίησης της εφαρμογής Fuzzy Data Grouping.

3.4 Ο ΑΛΓΟΡΙΘΜΟΣ

Ο αλγόριθμος (υπορουτίνα) της Ομαδοποίησης καλείται με τις δύο εξής παραμέτρους:

1. OmadesStoxos: Πόσες ομάδες θέλουμε να έχουμε μετά τον τερματισμό του αλγορίθμου.
2. TrexOmades: Πόσες ομάδες έχουμε σε κάθε βήμα. Εάν δεν έχουμε ομαδοποιημένα άτομα, τότε η παράμετρος αυτή έχει την τιμή μηδέν.

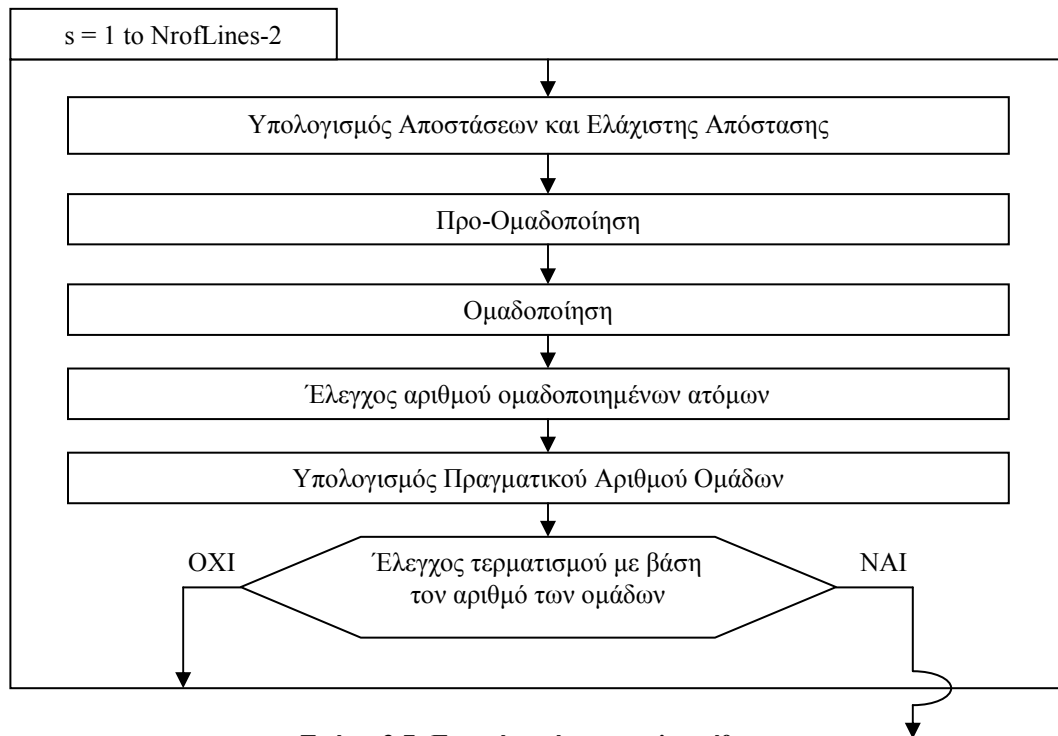
Για παράδειγμα εάν θέλουμε να δημιουργήσουμε 8 ομάδες τότε η υπορουτίνα της ομαδοποίησης καλείται ως εξής:

Omadopoihsh(8,0)

Εάν στη συνέχεια θέλουμε να περιορίσουμε τις ομάδες στις 6 τότε η υπορουτίνα της ομαδοποίησης καλείται ως εξής:

Omadopoihsh(6,8)

Στο Σχήμα 3.7 φαίνεται η επαναληπτική διαδικασία που εκτελεί την ομαδοποίηση των δεδομένων. Κάθε φορά που εκτελείται η διαδικασία αυτή, ομαδοποιούνται κανένα, ένα ή δύο σημεία που δεν έχουν ομαδοποιηθεί μέχρι αυτό το βήμα. Γι' αυτό το λόγο η διαδικασία της Ομαδοποίησης εκτελείται το πολύ $NrofLines-2$ φορές, όπου $NrofLines$ είναι ο αριθμός των σημείων. Η διαδικασία αυτή αποτελείται από έξι βασικά στάδια.



Σχήμα 3.7: Γενική εικόνα του αλγορίθμου

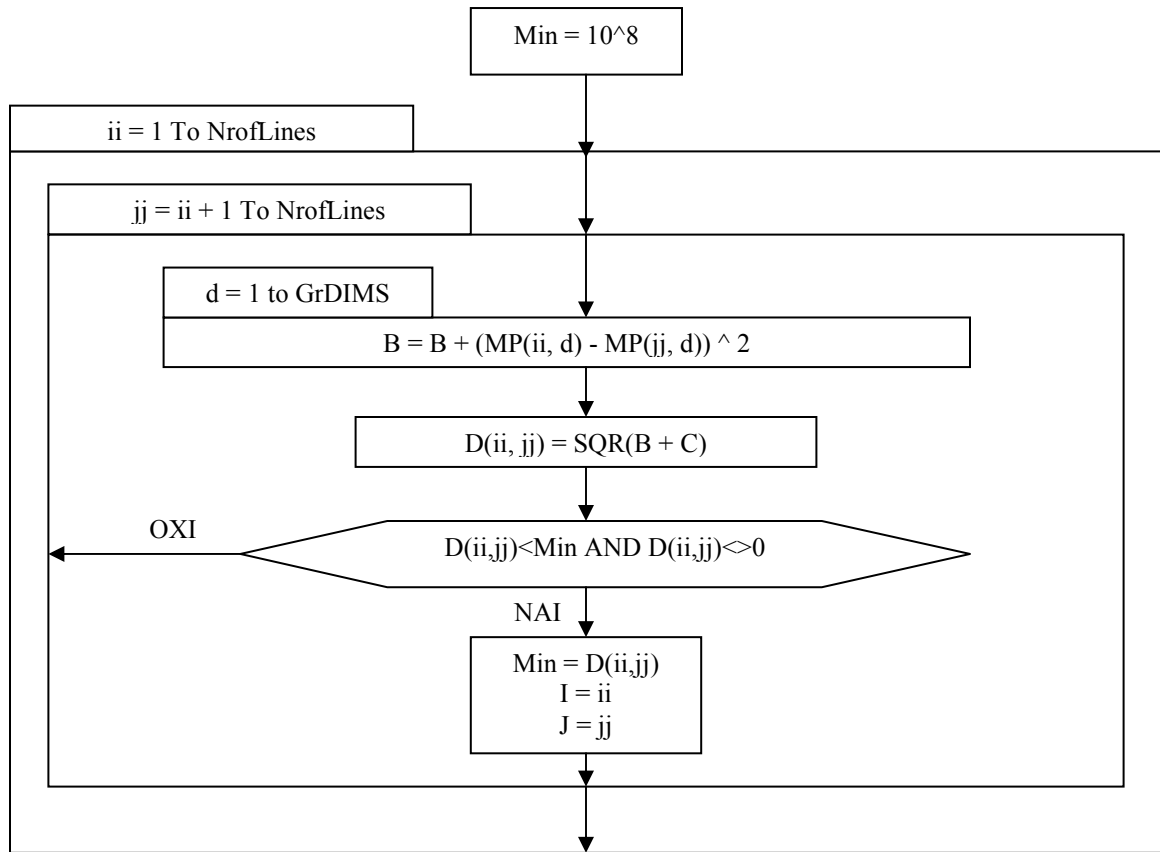
3.4.1 Υπολογισμός Αποστάσεων και Ελάχιστης Απόστασης

Ο παρακάτω αλγόριθμος (Σχήμα 3.8) υπολογίζει τις αποστάσεις μεταξύ όλων των σημείων λαμβάνοντας για κάθε σημείο ως θέση στο σύστημα των αξόνων το κέντρο βάρους όλων των σημείων που ανήκουν στην ίδια ομάδα με αυτό. Οι αποστάσεις αποθηκεύονται στο τριγωνικό τμήμα του πίνακα $D()$.

Για την εύρεση της ελάχιστης μη μηδενικής τιμής του πίνακα $D()$ ανατρέχουμε σε όλες τις τιμές του τριγωνικού μέρους του πίνακα εκτός αυτών της κύριας διαγωνίου. Οι μηδενικές τιμές αντιστοιχούν σε σημεία που ανήκουν στην ίδια ομάδα. Εδώ πρέπει να επαναλάβουμε ότι οι αποστάσεις του πίνακα $D()$ δεν αντιστοιχούν στις πραγματικές γεωμετρικές αποστάσεις μεταξύ των σημείων, με εξαίρεση την πρώτη επανάληψη της διαδικασίας. Κι αυτό γιατί ο πίνακας προκύπτει όχι από τον $P_{i,j}$ δηλαδή τον πίνακα των προβολών των σημείων στους νέους άξονες αλλά από τον $MP_{i,j}$ δηλαδή τον πίνακα των

τροποποιημένων προβολών των σημείων που αντιστοιχούν στο κέντρο βάρους των σημείων που αποτελούν μία ομάδα.

Τελικά η ελάχιστη απόσταση (για την ακρίβεια η τρέχουσα ελάχιστη απόσταση) και τα σημεία που αντιστοιχούν σε αυτή, αποθηκεύονται στις μεταβλητές Min , I και J αντίστοιχα. Η τιμή της μεταβλητής $GrDIMS$ αντιστοιχεί στον αριθμό των νέων αξόνων στους οποίους γίνεται η ομαδοποίηση.



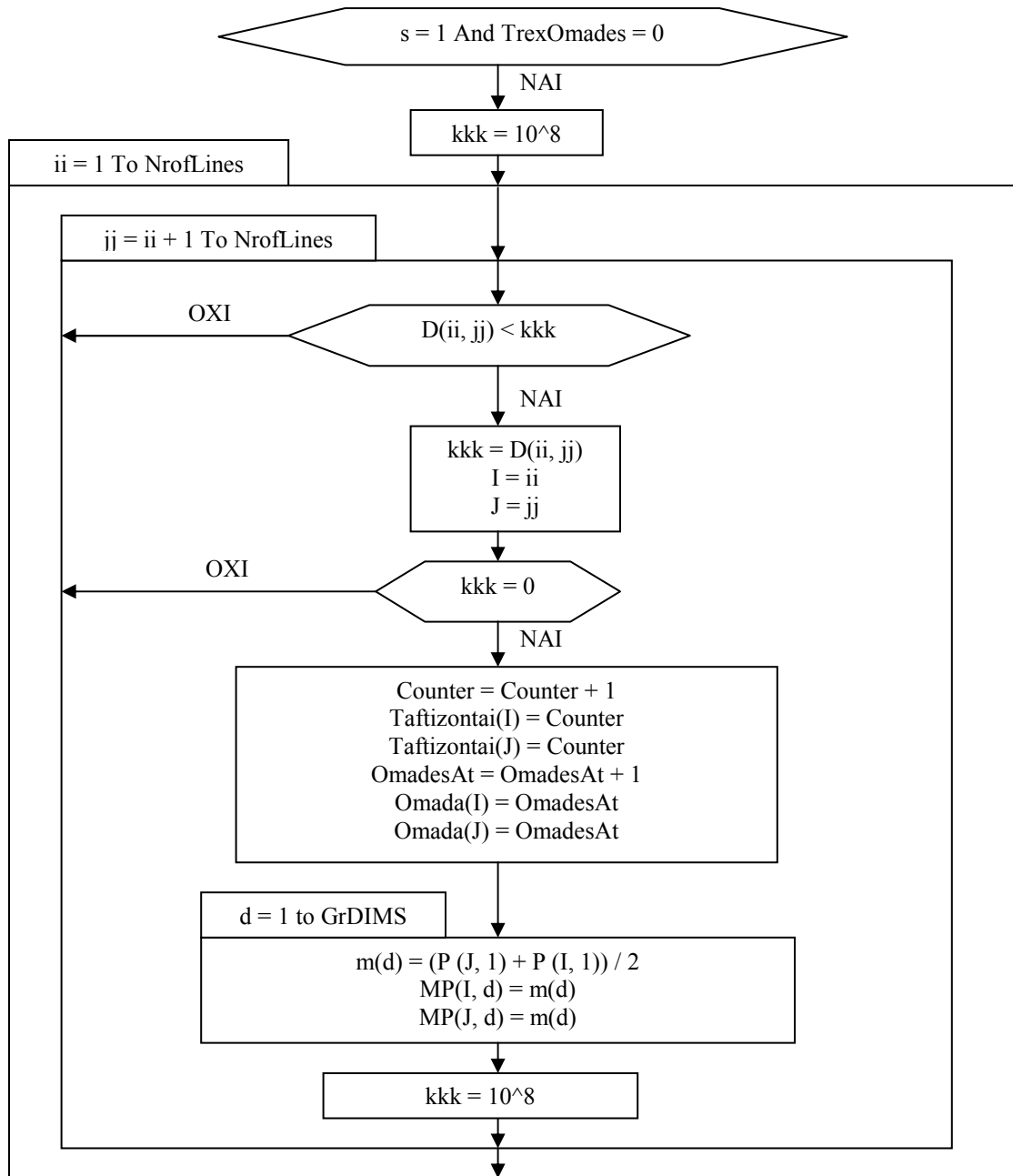
Σχήμα 3.8: Αλγόριθμος υπολογισμού αποστάσεων και ελάχιστης απόστασης

3.4.2 Προ-Ομαδοποίηση

Ο πίνακας $D()$ για τον οποίο μιλήσαμε παραπάνω περιέχει μηδενικά στοιχεία στην κύρια διαγώνιο και μη μηδενικά στοιχεία στο τριγωνικό του μέρος που αντιστοιχούν στις αποστάσεις μεταξύ των σημείων. Επομένως η διαδικασία της εύρεση της μικρότερης απόστασης πρέπει να αποκλείει τα μηδενικά στοιχεία αφού αυτά αντιστοιχούν σε άτομα που ήδη βρίσκονται στην ίδια ομάδα.

Μία ακραία πλην όμως υπαρκτή περίπτωση είναι δύο ή περισσότερα σημεία να έχουν τις ίδιες προβολές σε όλους τους άξονες ή έστω στους άξονες που εφαρμόζουμε την ομαδοποίηση. Τότε η γεωμετρική τους απόσταση θα είναι μηδέν. Σ' αυτή την περίπτωση τα σημεία πρέπει να τοποθετηθούν στην ίδια ομάδα πριν ουσιαστικά αρχίσει η διαδικασία της ομαδοποίησης. Ο αλγόριθμος του σχήματος 3.9 εκτελείται μόνο εφ' όσον βρισκόμαστε στην πρώτη επανάληψη της διαδικασίας και δεδομένου ότι η διαδικασία έχει κληθεί να ομαδοποιήσει τα σημεία εκ του μηδενός.

Ο πίνακας $Omada()$ είναι ένα διάνυσμα στο οποίο αποθηκεύεται για κάθε σημείο ο αριθμός της ομάδας στην οποία ανήκει. Για παράδειγμα αν $Omada(3)=2$, αυτό σημαίνει ότι το σημείο 3 ανήκει στην ομάδα 2.



Σχήμα 3.9: Αλγόριθμος Προ-Ομαδοποίησης των δεδομένων

3.4.3 Ομαδοποίηση

Το παρακάτω κομμάτι του αλγορίθμου (Σχήμα 3.10) αποτελεί την καρδιά όλου του αλγορίθμου της ομαδοποίησης. Πρόκειται για το σημείο όπου λαμβάνεται η απόφαση για την ομαδοποίηση των σημείων I και J που βρέθηκε από τα παραπάνω βήματα ότι έχουν την ελάχιστη απόσταση μεταξύ τους. Υπάρχουν τέσσερις περιπτώσεις:

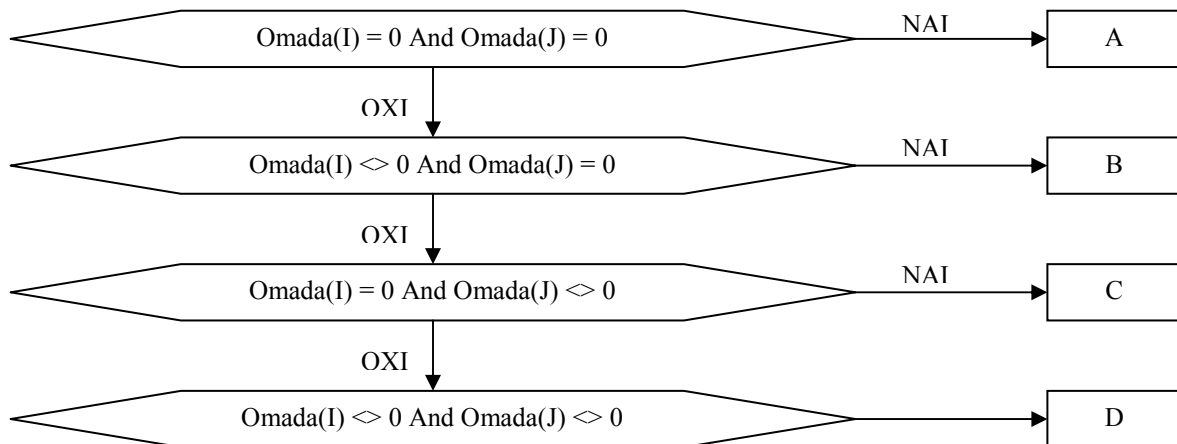
A: Να μην ανήκει κανένα από τα I, J σε ομάδα.

B: να ανήκει το I σε ομάδα και το J όχι.

C: να ανήκει το J σε ομάδα και το I όχι.

D: να ανήκουν και τα δύο σημεία σε ομάδα.

Στην τελευταία περίπτωση είναι ευνόητο ότι αποκλείεται να ανήκουν και τα δύο σημεία στην ίδια ομάδα αφού εάν συνέβαινε αυτό, όπως έχουμε εξηγήσει παραπάνω, η απόσταση μεταξύ τους θα ήταν μηδέν.



Σχήμα 3.10

3.4.3.1 Περίπτωση Α (I, J μη ομαδοποιημένα)

Η πρώτη περίπτωση είναι κανένα από τα I και J να μη ανήκει ήδη σε κάποια ομάδα. Παρ' ότι είναι βέβαιο ότι στην πρώτη επανάληψη της συνολικής διαδικασίας τα I και J είναι αμοιβαία πλησιέστερα (αφού στην πρώτη επανάληψη η ελάχιστη απόσταση αντιστοιχεί σε πραγματική απόσταση), αυτό δεν μπορεί να εξασφαλιστεί για τις επόμενες επαναλήψεις. Έτσι καταφεύγουμε στην εύρεση του πλησιέστερου τόσο του I όσο και του J. Αυτό γίνεται με τη χρήση της συνάρτησης (Function) Find_kontino(point) η οποία βρίσκει το πραγματικά πλησιέστερο του σημείου "point". Το αποτέλεσμα της συνάρτησης το αποθηκεύουμε στη μεταβλητή PlhsiesteroI ή PlhsiesteroJ ανάλογα με το αν την είχαμε καλέσει για εύρεση του πλησιέστερου σημείου του I ή του J.

Στη συνέχεια εφ' όσον η Μέγιστη Ακτίνα (2ο κριτήριο τερματισμού της ομαδοποίησης) δεν είναι μεγαλύτερη από αυτή που έχει τεθεί από το χρήστη διακρίνουμε τέσσερις περιπτώσεις:

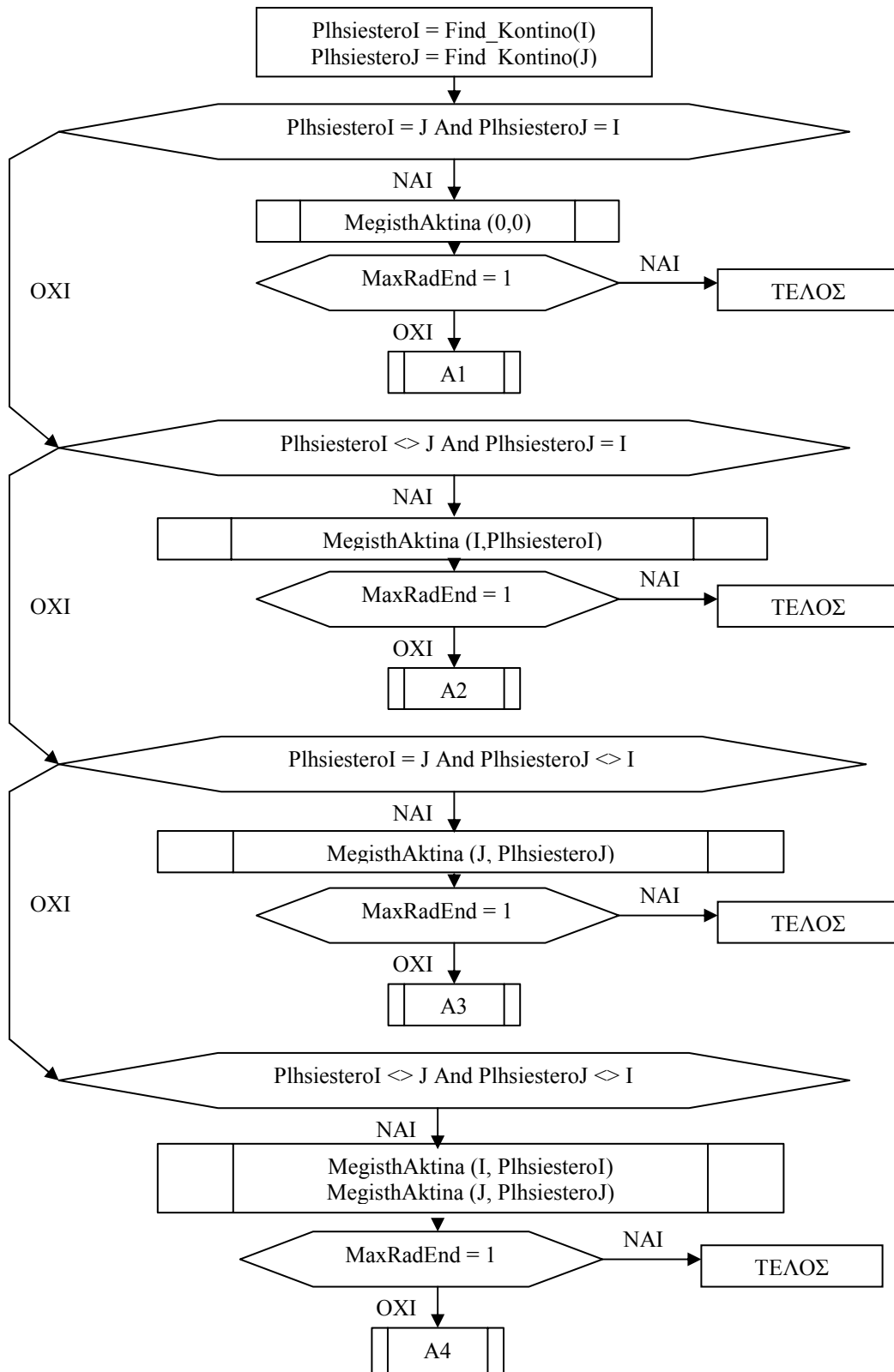
A1: Τα I και J είναι αμοιβαία πλησιέστερα.

A2: Το πλησιέστερο του J είναι το I αλλά το πλησιέστερο του I δεν είναι το J.

A3: Το πλησιέστερο του I είναι το J αλλά το πλησιέστερο του J δεν είναι το I.

A4: Ούτε του I το πλησιέστερο είναι το J, ούτε του J το πλησιέστερο είναι το I.

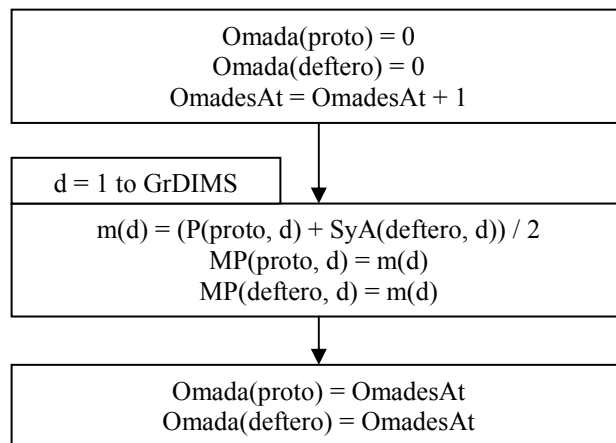
Η διαδικασία αυτή φαίνεται στο Σχήμα 3.11.



Σχήμα 3.11: Διάγραμμα ροής για την περίπτωση που κανένα από τα I και J δεν ανήκουν σε κάποια ομάδα

Συνάρτηση A1(proto,deftero)

Στην περίπτωση που τα σημεία I και J είναι αμοιβαία πλησιέστερα καλούμε τη συνάρτηση A1 με παραμέτρους τα I και J - δηλαδή $A1(I,J)$ - προκειμένου ο αλγόριθμος να δημιουργήσει μία ομάδα και να τα τοποθετήσει σ' αυτήν. Οι ομάδες δημιουργούνται από την αύξηση κατά ένα του μετρητή OmadesAt (πλήθος ομάδων). Παράλληλα ενημερώνουμε τον πίνακα MP() με τις τροποποιημένες προβολές στους νέους άξονες, των σημείων I και J. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.12.

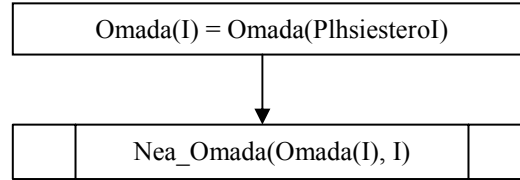


Σχήμα 3.12: Συνάρτηση A1(proto, deftero)

Συνάρτηση A2

Στην περίπτωση που το πλησιέστερο του J είναι το I αλλά το πλησιέστερο του I δεν είναι το J, το σημείο I προσχωρεί στην ομάδα του πλησιέστερού του, δηλαδή στην ομάδα του σημείου PlhsiesteroI. Η βεβαιότητα ότι το πλησιέστερο του σημείου I ανήκει ήδη σε ομάδα προέρχεται από το γεγονός ότι εάν αυτό δεν συνέβαινε, τότε στο στάδιο της εύρεσης της ελάχιστης απόστασης θα είχε βρεθεί η απόσταση του σημείου I και του πλησιέστερού του αντί αυτής του I και του J.

Ταυτόχρονα, καλούμε τη συνάρτηση $Nea_Omada(NrOmadas, PJWPM)$ με παραμέτρους $Omada(I)$ και I προκειμένου να ενημερωθούν οι πίνακες $PM()$, $PLHSIESTERA()$. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.13.

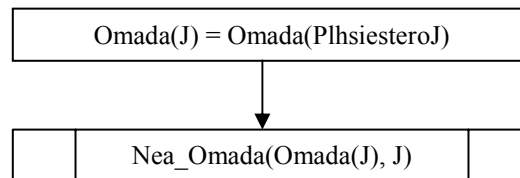


Σχήμα 3.13: Συνάρτηση A2

Συνάρτηση A3

Πρόκειται για τη αντίθετη περίπτωση όπου το πλησιέστερο του I είναι το J αλλά το πλησιέστερο του J δεν είναι το I . Έτσι το σημείο J προσχωρεί στην ομάδα του πλησιέστερού του, δηλαδή στην ομάδα του σημείου $PlhsiesteroJ$.

Ταυτόχρονα, καλούμε τη συνάρτηση $Nea_Omada(NrOmadas, PJWPM)$ με παραμέτρους $Omada(J)$ και J προκειμένου να ενημερωθούν οι πίνακες $PM()$, $PLHSIESTERA()$. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.14.

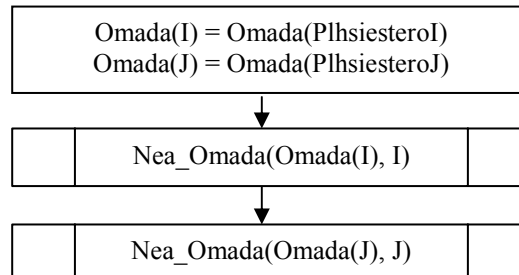


Σχήμα 3.14: Συνάρτηση A3

Συνάρτηση A4

Στην περίπτωση που ούτε του I το πλησιέστερο είναι το J , ούτε του J το πλησιέστερο είναι το I , τόσο το I όσο και το J προσχωρούν στην ομάδα του πλησιέστερου σε αυτά σημείο δηλαδή στις ομάδες των σημείων $PlhsiesteroI$ και $PlhsiesteroJ$ αντίστοιχα.

Επίσης καλούμε δύο φορές τη συνάρτηση `Nea_Omada` για λόγους αντίστοιχους με παραπάνω. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.15.



Σχήμα 3.15: Συνάρτηση A4

4.3.3.2 Περίπτωση Β (I ομαδοποιημένο, J μη ομαδοποιημένο)

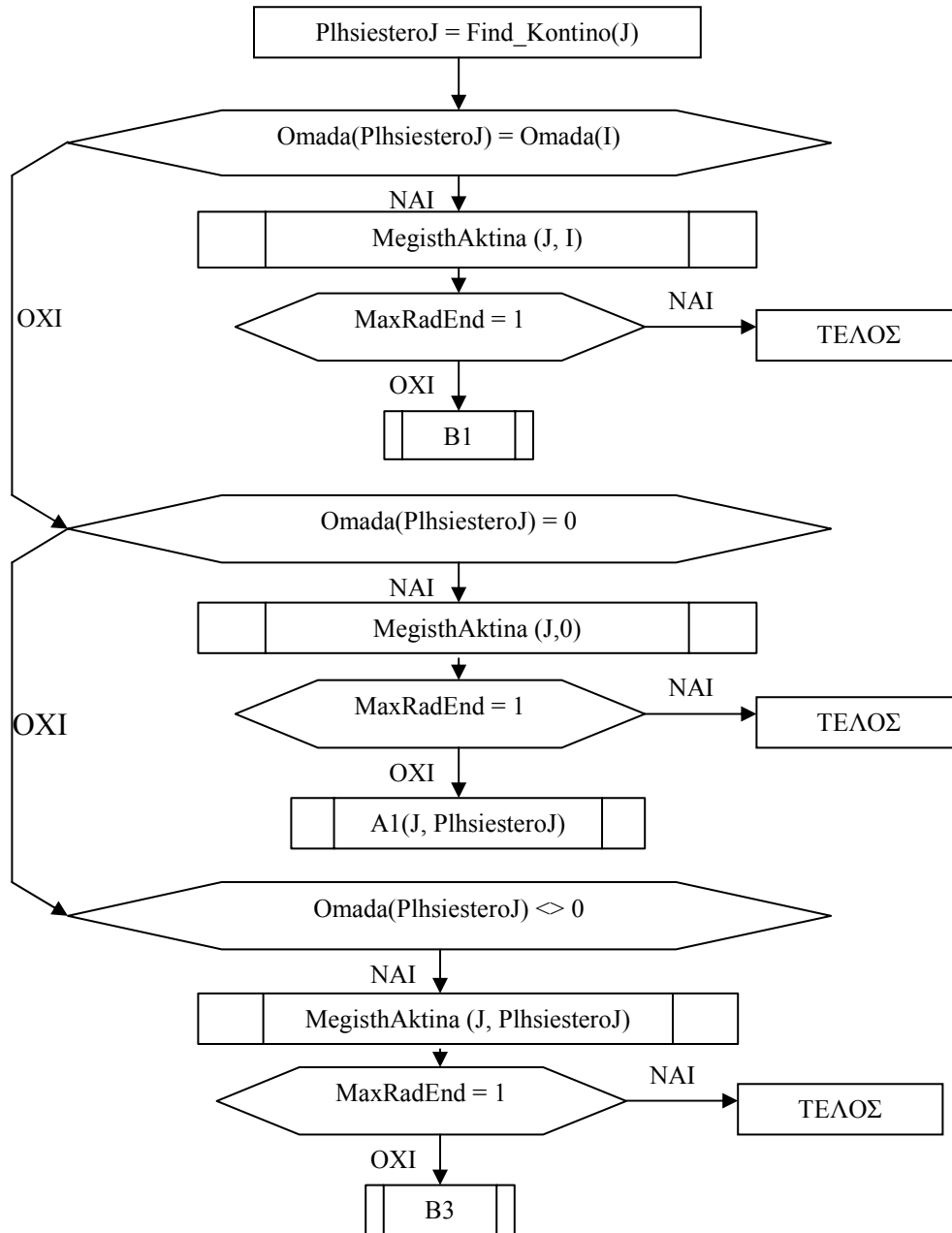
Στην περίπτωση που το I ανήκει ήδη σε ομάδα αλλά το J δεν ανήκει σε ομάδα χρειάζεται η εύρεση του πλησιέστερου του σημείου J. Στη συνέχεια εφ' όσον η Μέγιστη Ακτίνα δεν είναι μεγαλύτερη από αυτή που έχει τεθεί από το χρήστη διακρίνουμε τρεις περιπτώσεις :

B1: Η ομάδα του πλησιέστερου του σημείου J είναι ίδια με την ομάδα του σημείου I.

B2: Το πλησιέστερο του σημείου J δεν ανήκει σε ομάδα.

B3: Το πλησιέστερο του σημείου J ανήκει σε άλλη ομάδα από την ομάδα του I.

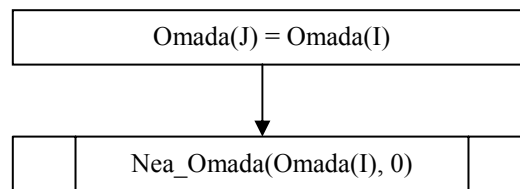
Η διαδικασία αυτή φαίνεται στο Σχήμα 3.16.



Σχήμα 3.16: Διάγραμμα ροής για την περίπτωση που το I ανήκει σε ομάδα ενώ το J δεν ανήκει σε ομάδα

Συνάρτηση B1

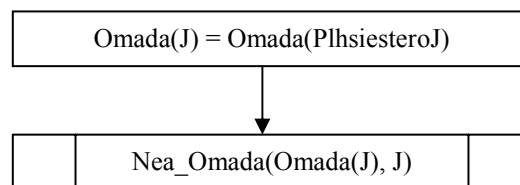
Στην περίπτωση που το πλησιέστερο του σημείου J ανήκει στην ομάδα του I, το σημείο J προσχωρεί στην ομάδα του σημείου I. Η συνάρτηση *Nea_Omada* καλείται με παραμέτρους *Omada(I)* προκειμένου να ενημερωθεί ο πίνακας *MP()* και μηδέν προκειμένου να μην ενημερωθεί ο πίνακας *PLHSIESTERA()* δεδομένου ότι το σημείο J προσχωρεί στην ομάδα λόγω ελάχιστης απόστασης και όχι λόγω του ότι είναι πλησιέστερο με κάποιο άλλο σημείο. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.17.

**Σχήμα 3.17: Συνάρτηση B1**

Στην περίπτωση που το πλησιέστερο του σημείου J δεν ανήκει σε ομάδα, καλείται η συνάρτηση *A1* με παραμέτρους J και *PlhsiesteroJ* ώστε να δημιουργηθεί μία νέα ομάδα που θα τα περιλάβει.

Συνάρτηση B3

Αν τίποτα από τα παραπάνω δεν συμβαίνει, τότε το J προσχωρεί στην ομάδα του πλησιέστερου σε αυτό σημείου και καλείται η συνάρτηση *Nea_Omada* με παραμέτρους *Omada(J)* προκειμένου να ενημερωθεί ο πίνακας *MP()* και J ώστε να ενημερωθεί ο πίνακας *PLHSIESTERA()* με το σημείο J που προσχώρησε στην ομάδα του πλησιέστερου σε αυτό σημείου. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.18.

**Σχήμα 3.18: Συνάρτηση B3**

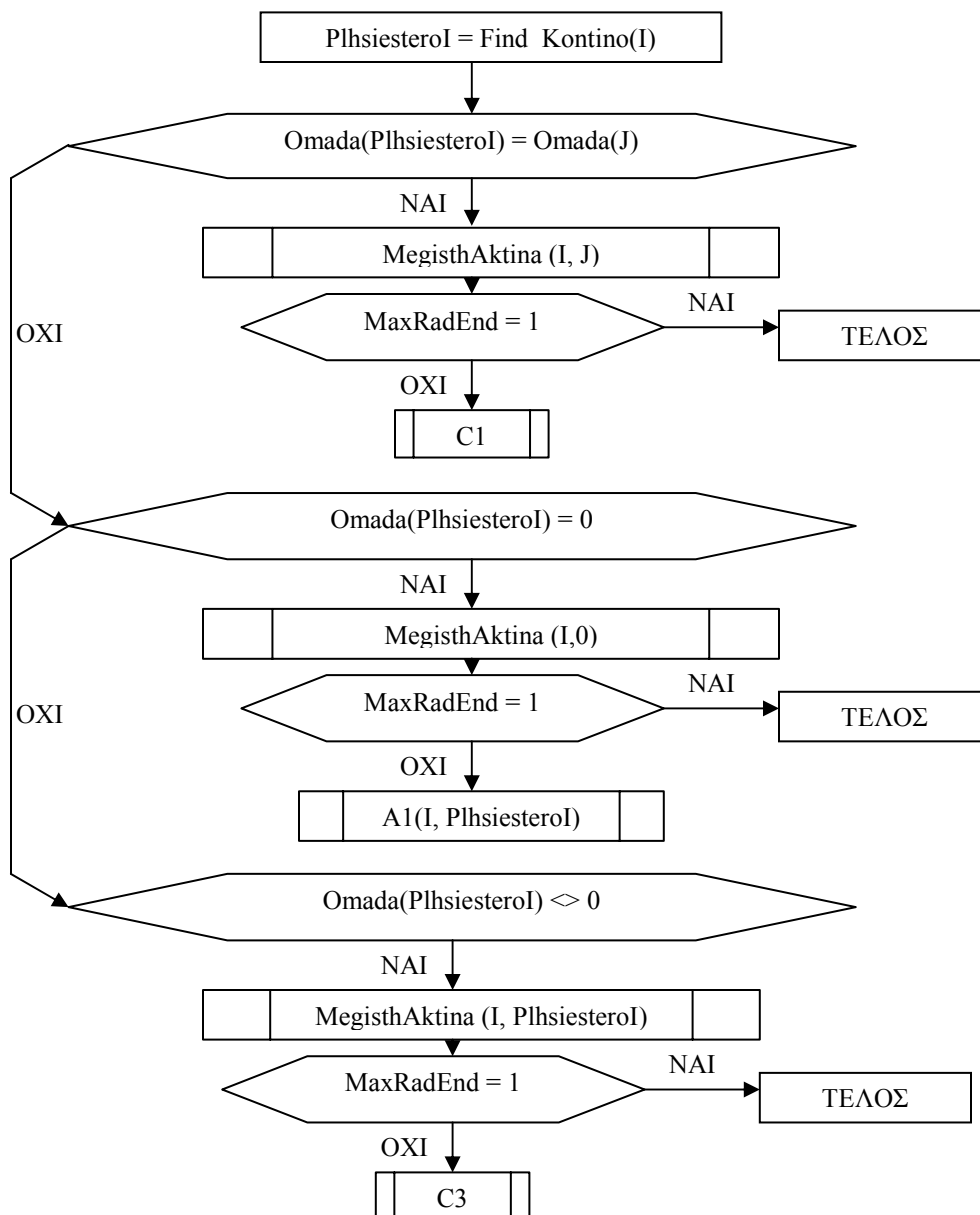
3.4.3.3 Περίπτωση C (I μη ομαδοποιημένο, J ομαδοποιημένο)

Στην περίπτωση που το J ανήκει ήδη σε ομάδα αλλά το I δεν ανήκει σε ομάδα χρειάζεται η εύρεση του πλησιέστερου του σημείου I. Στη συνέχεια εφ' όσον η Μέγιστη Ακτίνα δεν είναι μεγαλύτερη από αυτή που έχει τεθεί από το χρήστη διακρίνουμε τρεις περιπτώσεις:

C1: Η ομάδα του πλησιέστερου του σημείου I είναι ίδια με την ομάδα του σημείου J.

C2: Το πλησιέστερο του σημείου I δεν ανήκει σε ομάδα.

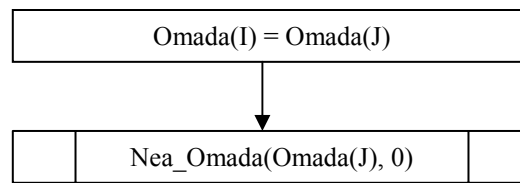
C3: Το πλησιέστερο του σημείου I ανήκει σε άλλη ομάδα από αυτή του J.



Σχήμα 3.19: Διάγραμμα ροής για την περίπτωση που το J ανήκει σε ομάδα ενώ το I δεν ανήκει σε ομάδα

Συνάρτηση C1

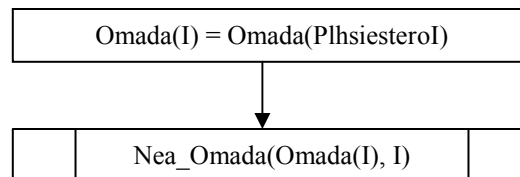
Στην περίπτωση που το πλησιέστερο του σημείου I ανήκει στην ομάδα του J, το σημείο I προσχωρεί στην ομάδα του σημείου J. Η συνάρτηση *Nea_Omada* καλείται με παραμέτρους *Omada(J)* προκειμένου να ενημερωθεί ο πίνακας *MP()* και μηδέν προκειμένου να μην ενημερωθεί ο πίνακας *PLHSIESTERA()* δεδομένου ότι το σημείο I προσχωρεί στην ομάδα λόγω ελάχιστης απόστασης και όχι λόγω του ότι είναι πλησιέστερο με κάποιο άλλο σημείο. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.20.

**Σχήμα 3.20: Συνάρτηση C1**

Στην περίπτωση που το πλησιέστερο του σημείου I δεν ανήκει σε ομάδα, καλείται η συνάρτηση *A1* με παραμέτρους *I* και *PlhsiesteroI* ώστε να δημιουργηθεί μία νέα ομάδα που θα τα περιλάβει.

Συνάρτηση C3

Αν τίποτα από τα παραπάνω δεν συμβαίνει, τότε το I προσχωρεί στην ομάδα του πλησιέστερου σε αυτό σημείου και καλείται η συνάρτηση *Nea_Omada* με παραμέτρους *Omada(I)* προκειμένου να ενημερωθεί ο πίνακας *MP()* και *I* ώστε να ενημερωθεί ο πίνακας *PLHSIESTERA()* με το σημείο I που προσχώρησε στην ομάδα του πλησιέστερου σε αυτό σημείου. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.21.

**Σχήμα 3.21: Συνάρτηση C3**

3.4.3.4 Περίπτωση D (I, J ανήκουν σε διαφορετικές ομάδες)

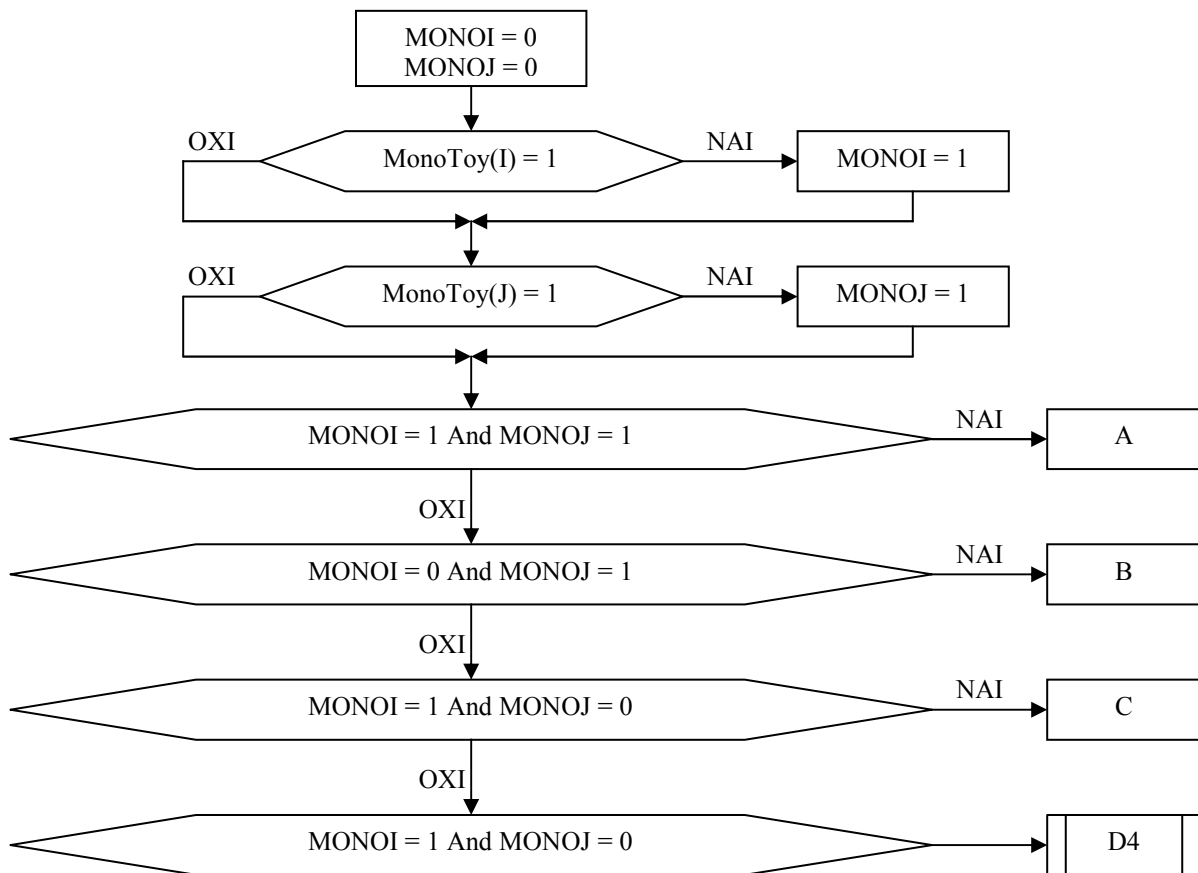
Η τελευταία περίπτωση είναι τόσο το I όσο και το J να ανήκουν ήδη σε κάποια ομάδα. Σ' αυτή την περίπτωση όπως είπαμε και παραπάνω, ενώνουμε τις δύο ομάδες. Όπως όμως έχουμε εξηγήσει, ο αλγόριθμος έχει τη δυνατότητα να δημιουργεί ομάδες με ένα και μόνο σημείο εάν αυτό είναι απομακρυσμένο.

Ας υποθέσουμε ότι επιχειρούμε να ενώσουμε δύο ομάδες και ότι η πρώτη από αυτές περιέχει ένα σημείο σε αντίθεση με τη δεύτερη που περιέχει περισσότερα από ένα (π.χ. τρία σημεία). Εάν ενώσουμε τις ομάδες θα δημιουργηθεί μία ομάδα με τέσσερα σημεία. Τι γίνεται όμως με τη διερεύνηση του πλησιέστερου του σημείου που ήταν μόνο του στην πρώτη ομάδα; Δεδομένου ότι ένα μη ομαδοποιημένο σημείο δεν διαφέρει σε τίποτα από ένα απομακρυσμένο και ομαδοποιημένο μοναχό του σημείο, είμαστε αναγκασμένοι πριν ενώσουμε δύο ομάδες να τις διερευνήσουμε προκειμένου να διαπιστώσουμε πόσα σημεία περιέχουν. Αυτό γίνεται με την κλήση της συνάρτησης `MonoToy(point)` που επιστρέφει ως τιμή της τη μονάδα εφόσον το σημείο "point" βρέθηκε να είναι το μοναδικό στην ομάδα του και το μηδέν εάν το σημείο "point" ανήκει σε ομάδα μαζί με άλλα σημεία.

Διακρίνουμε τέσσερις περιπτώσεις:

- Να είναι τόσο το I όσο και το J τα μοναδικά σημεία στην ομάδα τους. Αυτή είναι η περίπτωση A που περιγράψαμε παραπάνω.
- Να είναι το J μόνο του στην ομάδα του και το I να ανήκει σε ομάδα με περισσότερα από ένα σημεία. Αυτή είναι η περίπτωση B που περιγράψαμε παραπάνω.
- Να είναι το I μόνο του στην ομάδα του και το J να ανήκει σε ομάδα με περισσότερα από ένα σημεία. Αυτή είναι η περίπτωση C που περιγράψαμε παραπάνω.
- Τόσο το I όσο και το J να ανήκουν σε ομάδες με περισσότερα από ένα σημεία. Τότε καλούμε τη συνάρτηση D4.

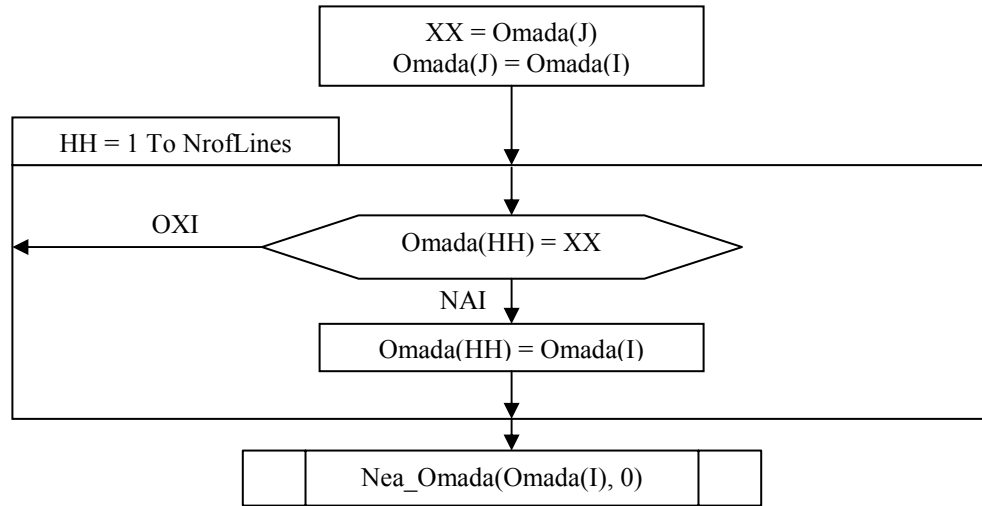
Η διαδικασία αυτή φαίνεται στο Σχήμα 3.22



Σχήμα 3.22: Διάγραμμα ροής για την περίπτωση που τόσο το I όσο και το J να ανήκουν ήδη σε κάποια ομάδα

Συνάρτηση D4

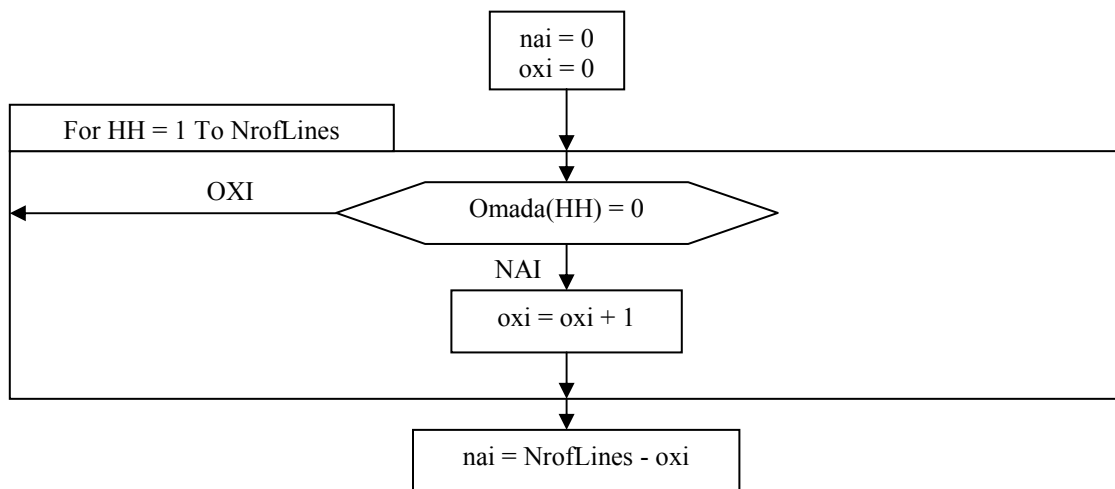
Προκειμένου να ενωθούν οι ομάδες του I και του J πρέπει όλα τα σημεία που ανήκουν στην ομάδα του J να προσχωρήσουν στην ομάδα του I ή το αντίθετο. Αφού τελειώσει αυτή η διαδικασία καλείται η συνάρτηση *Nea_Omada* προκειμένου να ενημερωθεί ο πίνακας *MP()*. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.23.



Σχήμα 3.23: Συνάρτηση D4

3.4.4 Έλεγχος Αριθμού Ομαδοποιημένων Ατόμων

Σ' αυτό το σημείο του αλγορίθμου (Σχήμα 3.24) γίνεται έλεγχος για τον αριθμό των ομαδοποιημένων και των μη ομαδοποιημένων ατόμων την τρέχουσα χρονική στιγμή. Αυτός ο έλεγχος είναι απαραίτητος προκειμένου παρακάτω να διαπιστωθεί εάν έχουμε φτάσει στον επιθυμητό αριθμό ομάδων σύμφωνα με το αντίστοιχο κριτήριο τερματισμού του αλγορίθμου της ομαδοποίησης.



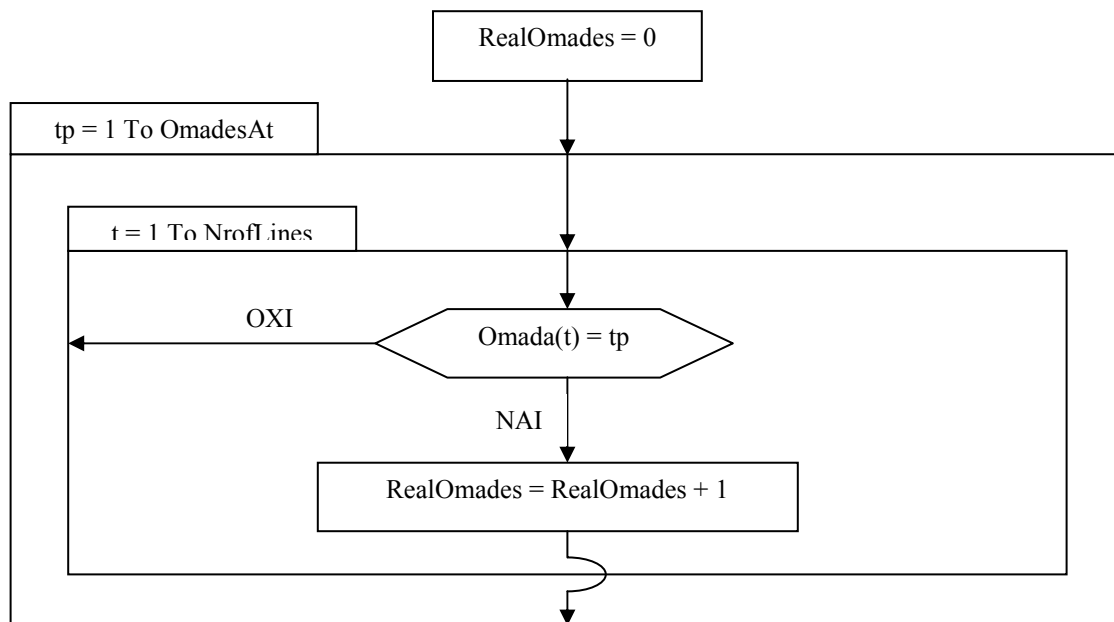
Σχήμα 3.24: Διάγραμμα ροής ελέγχου του αριθμού των ομαδοποιημένων ατόμων

Στη μεταβλητή “naί” αποθηκεύεται ο αριθμός των ομαδοποιημένων ατόμων και στη μεταβλητή “οxί” ο αριθμός των ατόμων που δεν ανήκουν σε καμία ομάδα.

3.4.5 Υπολογισμός Πραγματικού Αριθμού Ομάδων

Λόγω της ένωσης Ομάδων και της συνεχούς μεταβολής του πίνακα Omada(), ο μετρητής OmadesAt (πλήθος ομάδων) δεν μετράει τον πραγματικό αριθμό των ομάδων. Ανατρέχοντας στον πίνακα Omada() μπορούμε να διαπιστώσουμε ποια νούμερα από το 1 έως το OmadesAt αντιστοιχούν πραγματικά σε Ομάδες που περιέχουν έστω και ένα σημείο.

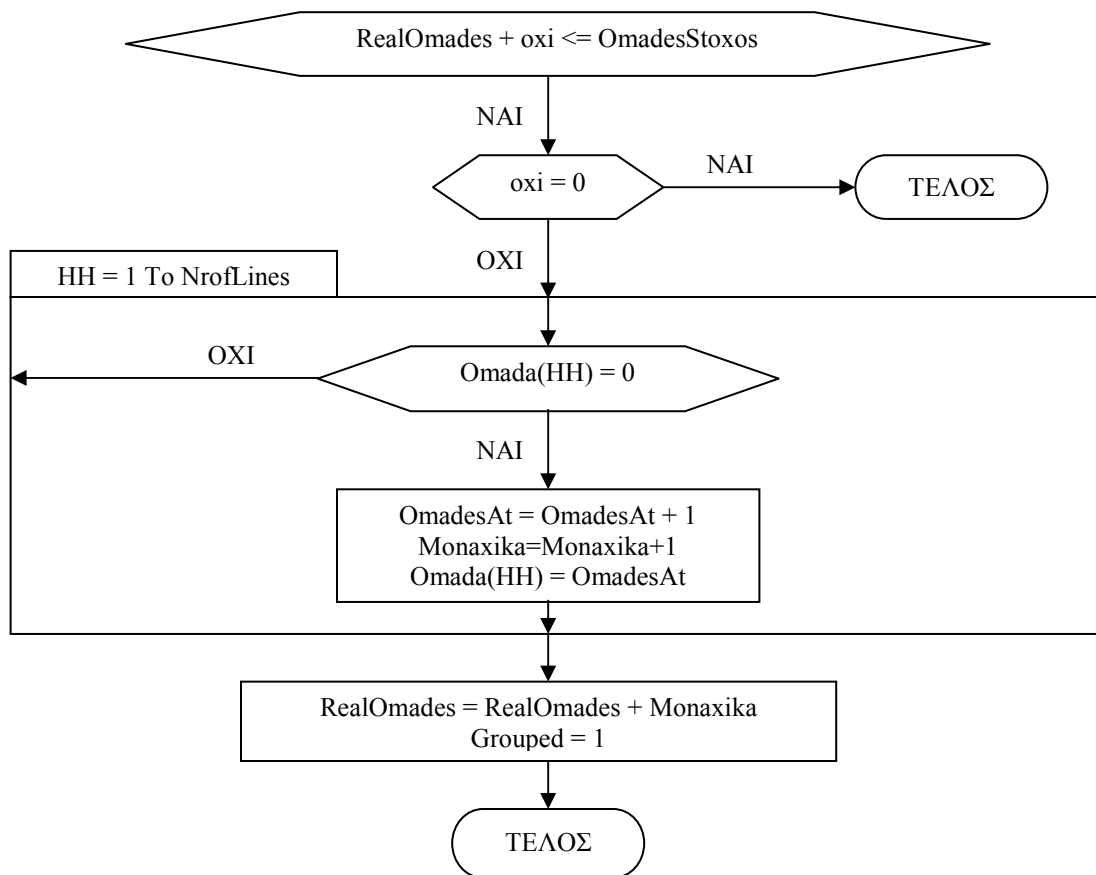
Αυτή η διαδικασία (Σχήμα 3.25) είναι επίσης απαραίτητη για τον έλεγχο του πρώτου κριτηρίου τερματισμού της ομαδοποίησης. Ο πραγματικός αριθμός των ομάδων αποθηκεύεται στη μεταβλητή “RealOmades”.



Σχήμα 3.25: Διάγραμμα ροής υπολογισμού του πραγματικού αριθμού των ομάδων

3.4.6 Έλεγχος Τερματισμού με βάση τον Αριθμό των Ομάδων

Σ' αυτό το σημείο του αλγορίθμου γίνεται ο έλεγχος του πρώτου κριτηρίου τερματισμού του Αλγορίθμου της Ομαδοποίησης (αριθμός ομάδων). Αν ο έλεγχος είναι θετικός, τα σημεία που δεν έχουν ήδη ομαδοποιηθεί μένουν μόνα τους σε μία Ομάδα η οποία δημιουργείται και οι Πραγματικές Ομάδες αυξάνονται κατά τον αριθμό των ατόμων που μένουν μοναχικά (μοναδικά στην ομάδα τους). Η διαδικασία του ελέγχου φαίνεται στο Σχήμα 3.26.



Σχήμα 3.26: Διάγραμμα ροής για τον έλεγχο τερματισμού του αλγορίθμου με βάση τον αριθμό των ομάδων

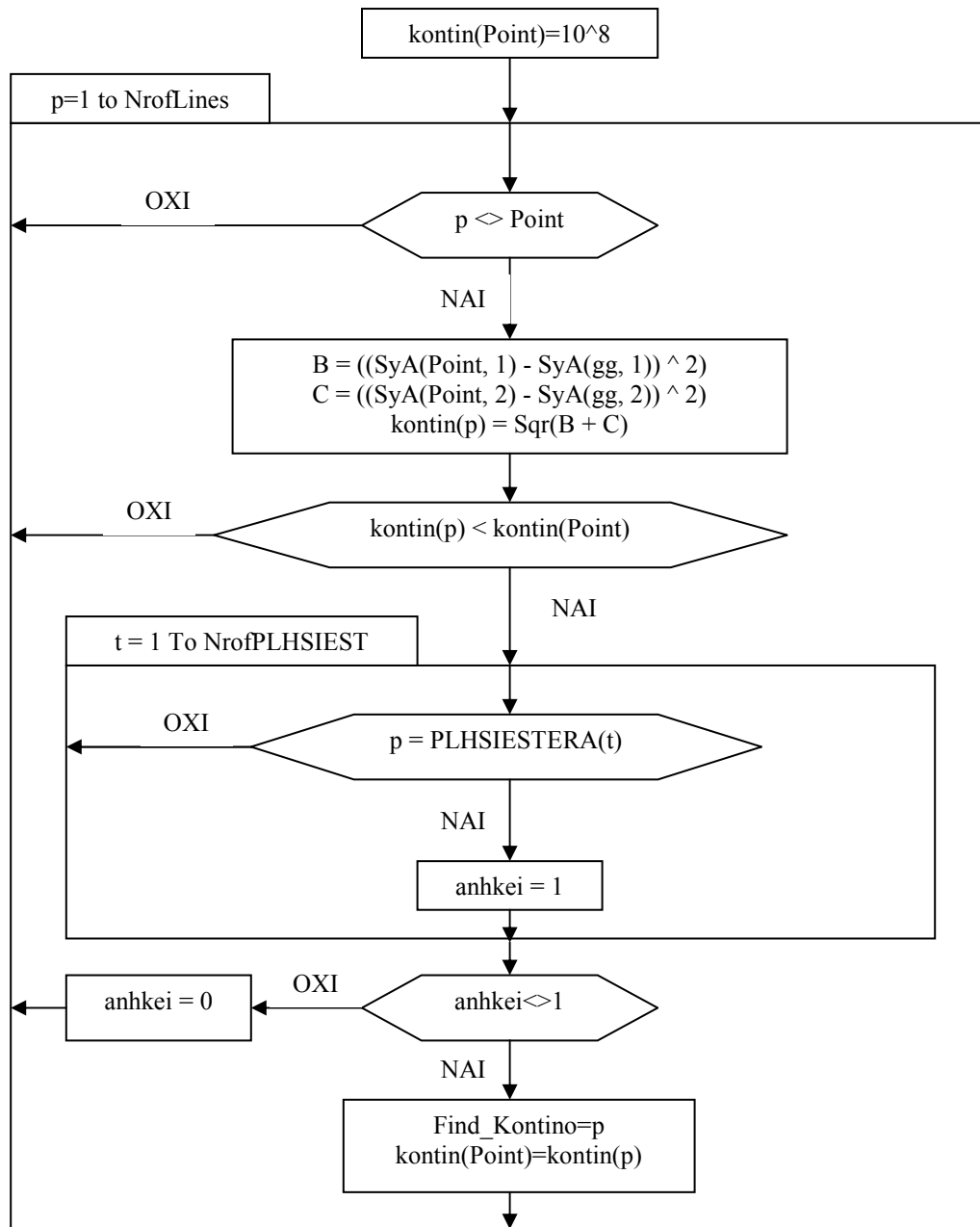
3.4.7 Συναρτήσεις

Παρακάτω παραθέτουμε διάφορες συναρτήσεις που χρησιμοποιούνται κατά τη διαδικασία της ομαδοποίησης.

Συνάρτηση Find_kontino(Point)

Βρίσκει το γεωμετρικά πλησιέστερο σημείο του “Point”. Επιτρεπτά για πλησιέστερα είναι όλα τα σημεία εκτός από αυτά που έχουν στο παρελθόν ομαδοποιηθεί όχι λόγω ελάχιστης απόστασης από κάποιο σημείο αλλά λόγω του ότι με τη σειρά τους ήταν πλησιέστερα κάποιου άλλου σημείου. Με τον τρόπο αυτό αποφεύγεται η ομαδοποίηση σημείων που σχηματίζουν αλυσίδες κάτι που όπως έχουμε ήδη αναφέρει δεν μας εξυπηρετεί κατά τη διαδικασία χαρακτηρισμού των ομάδων που θα δούμε παρακάτω.

Τα αποτελέσματα της συνάρτησης είναι το γεωμετρικά πλησιέστερο σημείο του “Point” που αποθηκεύεται στη μεταβλητή Find_kontino και η απόσταση του σημείου “Point” από το πλησιέστερό του που αποθηκεύεται στον πίνακα-διάνυσμα kontin. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.27

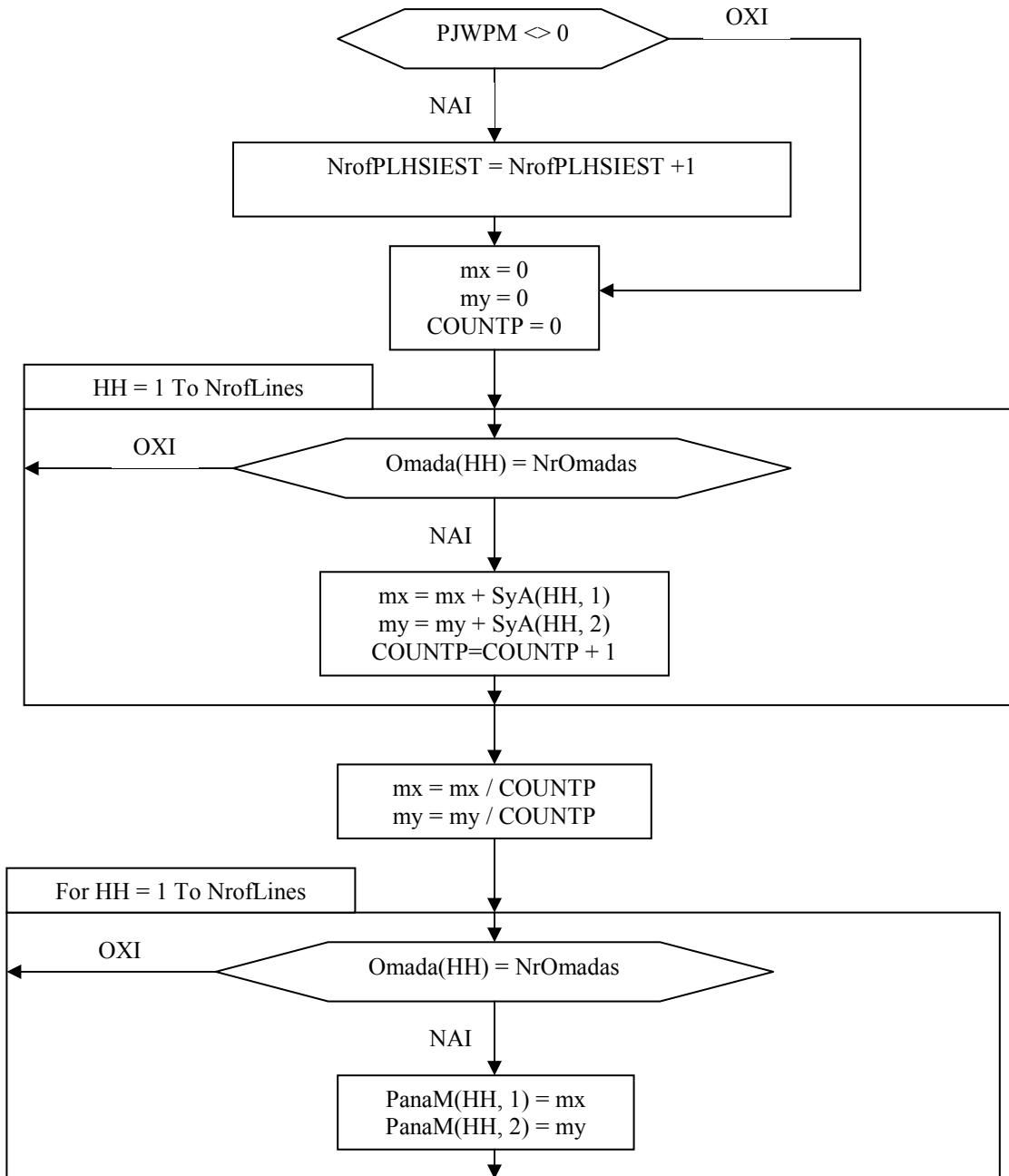


Σχήμα 3.27: Συνάρτηση Find_kontino(Point)

Συνάρτηση *Nea_Omada(NrOmadas, PJWPM)*

Η συνάρτηση αυτή χρησιμοποιείται για τον υπολογισμό του κέντρου βάρους των ομάδων και την ενημέρωση των πινάκων MP() και PLHSIESTERA(). Ο πίνακας-διάνυσμα

PLHSIESTERA() αποθηκεύει τα σημεία που εισέρχονται στη ομάδα τους όχι λόγω ελάχιστης απόστασης (που υπολογίστηκε από τον πίνακα MP()) με κάποιο άλλο σημείο αλλά λόγω του ότι ήταν πλησιέστερα κάποιου άλλου σημείου. Η συνάρτηση καλείται με δύο παραμέτρους. Η πρώτη (“NrOmadas”) αφορά την ομάδα στις οποίας τα άτομα θα δοθούν νέες συντεταγμένες και η δεύτερη (“PJWPM”) αφορά το άτομο που εισέρχεται ως πλησιέστερο σε κάποια ομάδα. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.28.

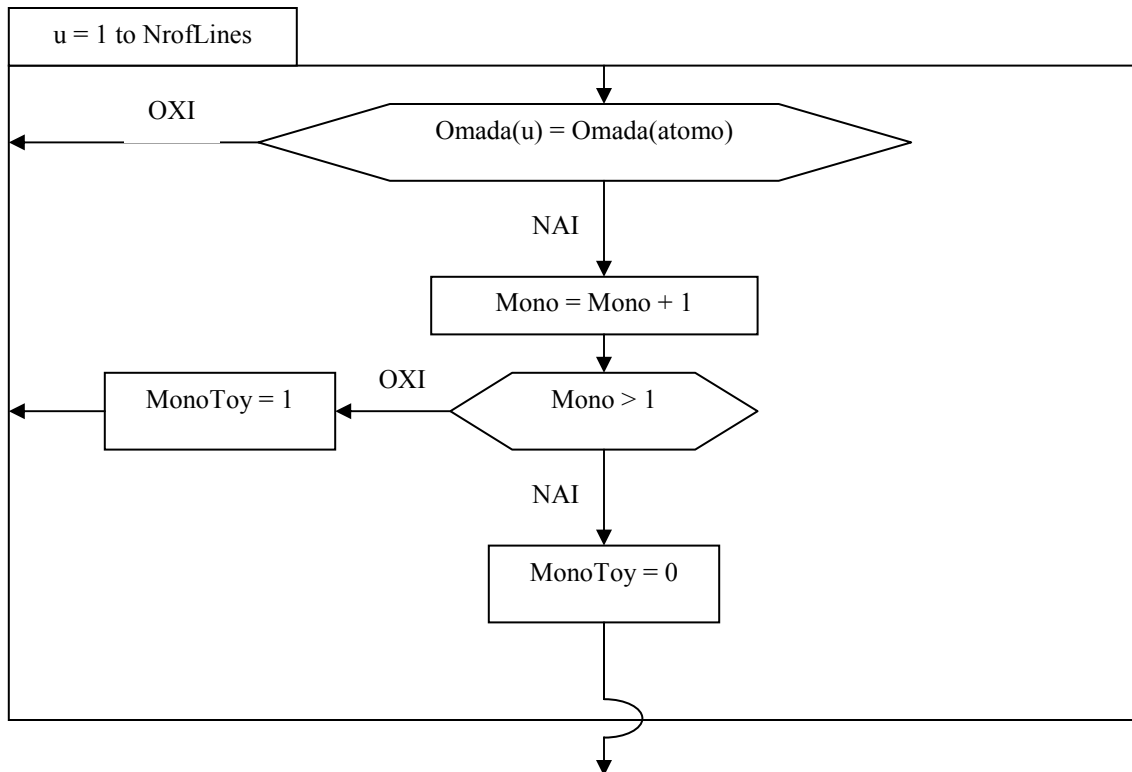


Σχήμα 3.28: Συνάρτηση Nea_Omada(NrOmadas,PJWPM)

Συνάρτηση MonoToy(atomo)

Η συνάρτηση αυτή βρίσκει αν κάποιο σημείο (“*atomo*”) είναι το μοναδικό στην Ομάδα που ανήκει.

Η συνάρτηση επιστρέφει ως αποτέλεσμα την τιμή μηδέν (0) εάν το σημείο δεν είναι μοναδικό στην ομάδα του ή την τιμή ένα (1) εάν το σημείο είναι μοναδικό στην ομάδα του. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.29.

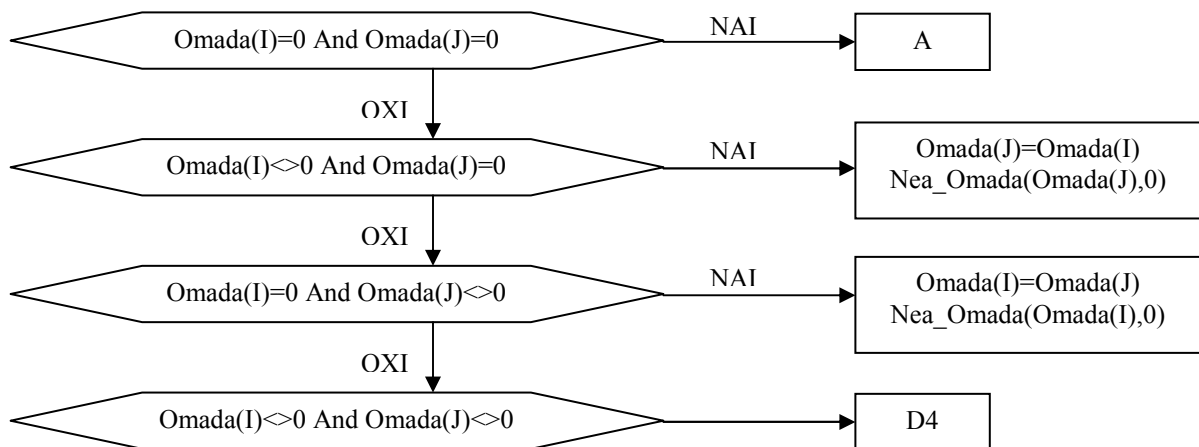


Σχήμα 3.29: Συνάρτηση MonoToy(atomo)

3.5 ΜΙΑ ΑΛΛΗ ΕΚΔΟΧΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ

Στο Σχήμα 3.30 παρουσιάζεται μία άλλη εκδοχή του αλγορίθμου γεωμετρικής ομαδοποίησης που αναπτύχθηκε παραπάνω. Η εκδοχή αυτή που την ονομάζουμε Min Distance είναι ελαφρώς πιο γρήγορη για το λόγο ότι δεν εφαρμόζει τη μέθοδο του Κοντινού Γείτονα. Διαφέρει από την εκδοχή Nearest Neighbour μόνο στο κομμάτι της λήψης αποφάσεων για την ομαδοποίηση των I και J που βρέθηκαν να έχουν τη μικρότερη απόσταση στον πίνακα MP(). Στο Σχήμα 3.30 φαίνεται το διάγραμμα ροής σ' αυτό το σημείο του αλγορίθμου.

Στην εφαρμογή Fuzzy Data Grouping είναι διαθέσιμες και οι δύο εκδοχές του αλγορίθμου. Ωστόσο προτείνεται η χρήση της εκδοχής Min Distance μόνο σε περιπτώσεις που τα δεδομένα σχηματίζουν πυκνό νέφος σημείων στο γράφημα της ΑΚΣ οπότε και δεν υπάρχει κίνδυνος αλυσίδων στα σημεία. Σ' αυτή την περίπτωση οι δύο εκδοχές δίνουν παρόμοια αποτελέσματα.



Σχήμα 3.30

3.6 ΣΥΝΟΨΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ

Συνοψίζοντας μπορούμε να πούμε τα εξής για τον αλγόριθμο της ομαδοποίησης:

- Είναι αρκετά πολύπλοκος ώστε να αντιμετωπίζει τις πιθανές ιδιαιτερότητες που μπορεί να παρουσιάσουν συγκεκριμένα δεδομένα.
- Είναι σχετικά χρονοβόρος. Ωστόσο για ένα υπολογιστή Pentium II 400 MHz οι χρόνοι είναι λογικοί όπως φαίνεται στον Πίνακα 3.16.

Αριθμός Ατόμων	Αριθμός Ομάδων	Χρόνος	
		Nearest Neighbour	Min Distance
100	18	4 sec	4 sec
200	22	22 sec	20 sec
500	24	8:20 min	7 min
1000	26	68 min	56 min

Πίνακας 3.16

ΚΕΦΑΛΑΙΟ 4: ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ ΟΜΑΔΩΝ ΜΕ ΧΡΗΣΗ ΤΗΣ ΘΕΩΡΙΑΣ ΤΩΝ ΑΣΑΦΩΝ ΣΥΝΟΛΩΝ

4.1 ΕΙΣΑΓΩΓΗ

Αρκετά προβλήματα λήψης αποφάσεων είναι ιδιαίτερα πολύπλοκα ειδικά όταν υπάρχει ανάγκη εξήγησης ποσοτικών μεγεθών. Η θεωρία των Ασαφών Συνόλων (Fuzzy Set Theory) σχεδιάστηκε από τον Zadeh τη δεκαετία του 1960 για να αναπαραστήσει την αβεβαιότητα και την ασάφεια, εξασφαλίζοντας τυποποιημένα εργαλεία αντιμετώπισης της ανακρίβειας σε πολλά προβλήματα. Ουσιαστικά δημιουργεί ταξινομήσεις ή ομαδοποιήσεις δεδομένων σε όρια που δεν είναι επακριβώς καθορισμένα. Οποιαδήποτε θεωρία ή μεθοδολογία δίνει σαφείς ορισμούς μπορεί να μετατραπεί σε ασαφή αντικαθιστώντας τα κλασικά σύνολα με ασαφή σύνολα με μεταβλητά όρια.

4.2 ΑΣΑΦΗ ΣΥΝΟΛΑ

Ορισμός: Έστω ότι το σύνολο τιμών της χαρακτηριστικής συνάρτησης ενός συνόλου $A \subseteq E$ δεν είναι το $\{0,1\}$ αλλά το $M[0,1]$ –δηλαδή σε κάθε στοιχείο του συνόλου E αντιστοιχεί ένας αριθμός του διαστήματος $[0,1]$ - τότε το αντίστοιχο σύνολο ονομάζεται Ασαφές (υποσύνολο του βασικού συνόλου E) και σημειώνεται με \tilde{A} .

Έτσι το \tilde{A} είναι το σύνολο των ζευγών $\{(x, \mu_A(x)) : x \in E\}$ όπου η συνάρτηση $\mu: E \rightarrow M$ ονομάζεται Χαρακτηριστική Συνάρτηση Συμμετοχής, ενώ η τιμή $\mu_{\tilde{A}}(x)$ ονομάζεται Βαθμός Συμμετοχής (ή Τιμή Συμμετοχής) του στοιχείου x στο \tilde{A} .

Για παράδειγμα αν $E = \{x_1, x_2, x_3, x_4, x_5\}$ τότε το ασαφές σύνολο

$$\tilde{A} \{(x_1, 0.6), (x_2, 0), (x_3, 0.2), (x_4, 1), (x_5, 0.9)\}$$

αντιστοιχεί στη γνώμη ενός καταναλωτή για την ποιότητα των αγαθών x_1, x_2, x_3, x_4, x_5 .

Στον Πίνακα 4.1 φαίνεται ο βαθμός συμμετοχής των x_i ($i=1, 2, \dots, 6$) στα σύνολα E και \tilde{A} .

	x_1	x_2	x_3	x_4	x_5
E	1	1	1	1	1
\tilde{A}	0.6	0	0.2	1	0.9

Πίνακας 4.1: Βαθμός Συμμετοχής των x_i στα σύνολα E και \tilde{A}

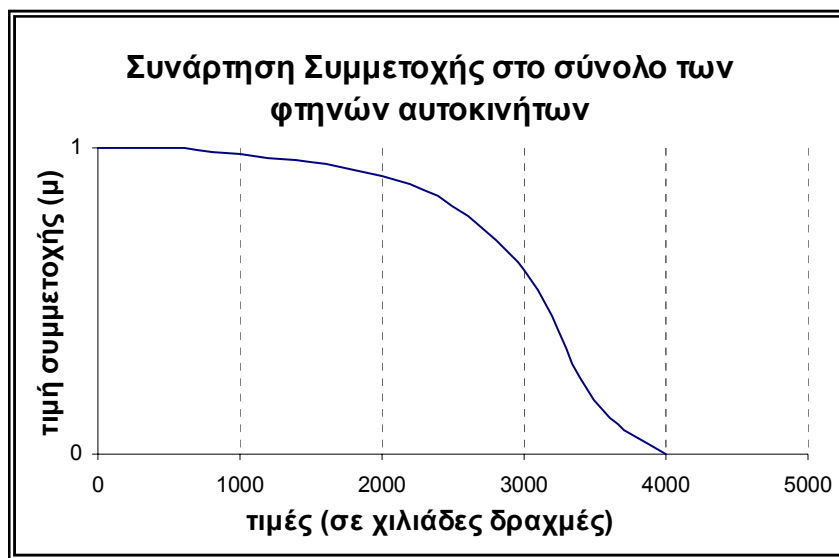
Δηλαδή ο καταναλωτής δίνει βαθμό προτίμησης 0.6 στο αγαθό x_1 , βαθμό προτίμησης μηδέν στο αγαθό x_2 κ.τ.λ.

Για το σύνολο αυτό είναι:

$$x_1 \in_{0,6} A, x_2 \in_0 A \text{ ή } x_2 \notin A, x_3 \in_{0,2} A, x_4 \in_1 A \text{ ή } x_4 \in A, x_5 \in_{0,9} A$$

όπου $x_i \in_{\mu} A$ σημαίνει ότι το στοιχείο x_i ανήκει στο σύνολο A με βαθμό συμμετοχής μ .

Ανάλογη ασάφεια μπορεί να εμφανιστεί σε διάφορες έννοιες. Ας υποθέσουμε ότι κάποιος προτίθεται να αγοράσει αυτοκίνητο και το μόνο κριτήριο που θέτει είναι το αυτοκίνητο να είναι φτηνό. Η έννοια “φτηνό” μπορεί να παρασταθεί ως ασαφές σύνολο σε μία πλειάδα τιμών και εξαρτάται από τον αγοραστή. Για κάποιο αγοραστή η έννοια “φτηνό” ερμηνεύεται όπως φαίνεται στο Σχήμα 4.1.



Σχήμα 4.1: Η ασαφής έννοια “φτηνό”

Από το διάγραμμα της Συνάρτησης Συμμετοχής των αυτοκινήτων στο σύνολο των “φτηνών” αυτοκινήτων συμπεραίνουμε ότι:

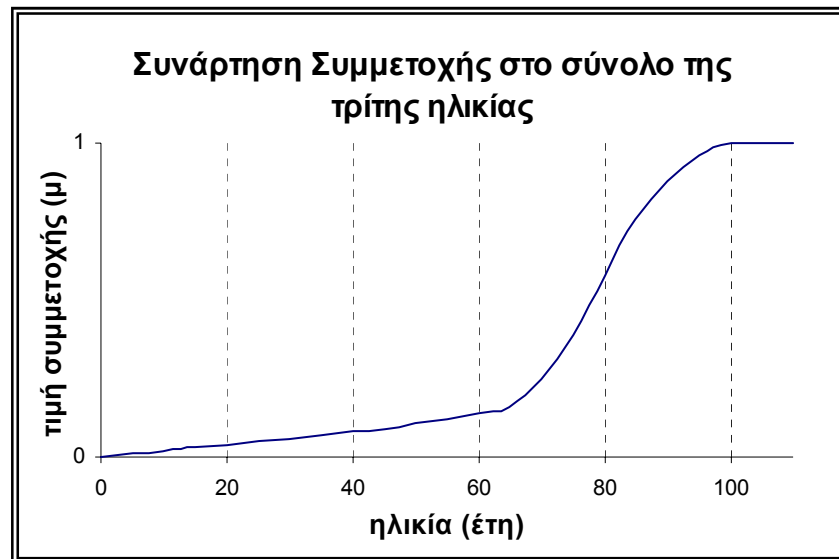
Για τιμές των αυτοκινήτων μικρότερες των 2 εκατομμυρίων, η τιμή δεν δημιουργεί σχεδόν καμία διαφορά στην εκτίμηση του αγοραστή και όλα τα αυτοκίνητα θεωρούνται φτηνά.

Για τιμές των αυτοκινήτων μεταξύ 2 και 3 εκατομμυρίων, η αλλαγή στην τιμή επηρεάζει ασθενώς την εκτίμηση του αγοραστή για φτηνό αυτοκίνητο.

Για τιμές των αυτοκινήτων μεταξύ 3 και 4 εκατομμυρίων, μία μικρή μεταβολή στην τιμή επηρεάζει σαφώς την εκτίμηση του αγοραστή για το ότι το αυτοκίνητο είναι φτηνό.

Για τιμές πάνω από 4 εκατομμύρια το κόστος θεωρείται απλησίαστο για τον αγοραστή.

Τέλος στο Σχήμα 4.2 αποδίδεται την ασαφή έννοια της τρίτης ηλικίας.



Σχήμα 4.2: Η ασαφής έννοια της τρίτης ηλικίας

4.3 ΛΕΚΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ

Μία γλωσσική ή λεκτική κλίμακα (linguistic scale) είναι μία ταξινόμηση του διαστήματος $[0,1]$ σε μία πεπερασμένη ομάδα ασαφών υποσυνόλων

$$D = \{C_0, C_1, C_2, \dots, C_n\}$$

όπου ένας λεκτικός (γλωσσικός) προσδιορισμός (label) αποδίδεται σε κάθε υποσύνολο. Προκειμένου να οριστούν ασαφείς ποσότητες οι οποίες παίρνουν τιμές μέσα στο D , πρέπει να θεωρήσουμε τις παρακάτω συνθήκες κανονικότητας:

i. $C_0 = \{0\}, C_n = \{1\}$

ii. Τα υποσύνολα είναι διατεταγμένα

$$C_0 \leq C_1 \leq C_2 \leq \dots C_{n-1} \leq C_n$$

iii. Ισχύει η συμμετρική ιδιότητα

$$1 - C_i = C_{n-i}, i=0,1,\dots,n$$

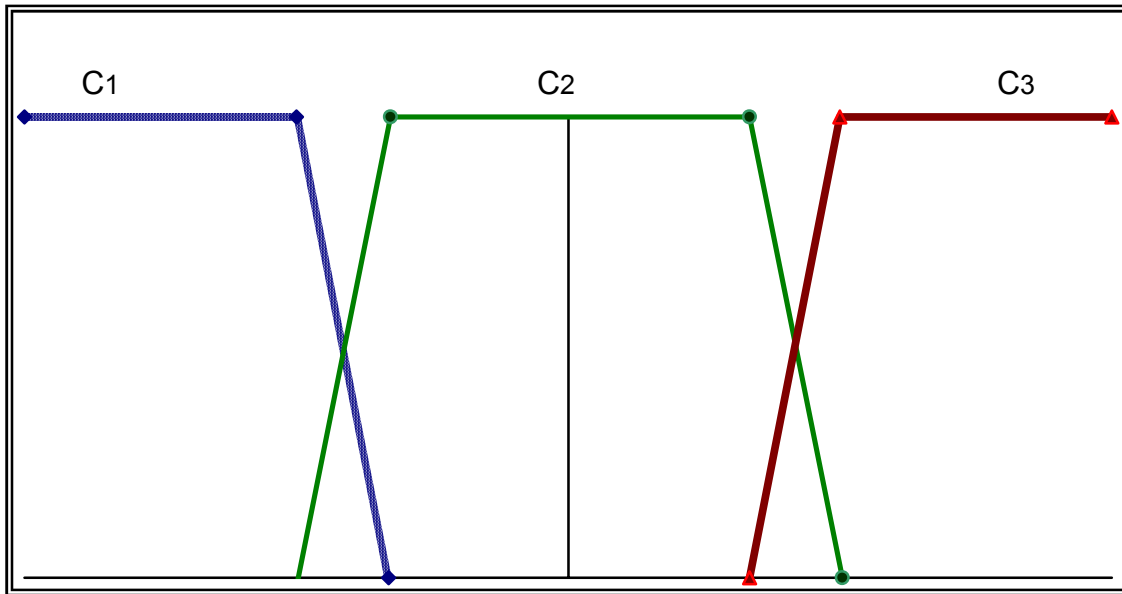
Δεδομένου λοιπόν ότι μπορούμε να ορίσουμε μία λεκτική κλίμακα στο διάστημα $[0,1]$ ομοίως μπορούμε να ορίσουμε μία λεκτική κλίμακα για οποιοδήποτε διάστημα $[a, \beta]$ με κατάλληλο μετασχηματισμό (κανονικοποίηση).

Τα πλεονεκτήματα της χρήσης μίας λεκτικής κλίμακας είναι προφανή. Αντί να δίνουμε αριθμητικό προσδιορισμό σε μία ούτως ή άλλως αριθμητική ποσότητα, μπορούμε να προσδιορίσουμε λεκτικά τη δεδομένη ποσότητα, κάτι που είναι σαφώς πιο προσιτό και κατανοητό. Ιδιαίτερα στην περίπτωση ανάπτυξης λογισμικού όπου η επικοινωνία χρήστη-μηχανής επιδιώκεται σε όσο μεγαλύτερο βαθμό γίνεται, οι λεκτικές μεταβλητές μπορούν να αποδειχθούν ιδιαίτερα χρήσιμες.

Στην περίπτωση μας θεωρούμε τη λεκτική κλίμακα

$$D = \{C_1, C_2, C_3\}$$

με ασαφή υποσύνολα όπως στο Σχήμα 4.3,



Σχήμα 4.3: Ασαφή υποσύνολα

και τις ακόλουθες λεκτικές περιγραφές για τα υποσύνολα

C_1 : Χαμηλή τιμή (Low)

C_2 : Μεσαία τιμή (Medium)

C_3 : Υψηλή τιμή (High)

4.4 ΣΥΝΑΡΤΗΣΕΙΣ ΣΥΜΜΕΤΟΧΗΣ

Η τιμή συμμετοχή κάθε ποσότητας σε ένα ασαφές υποσύνολο ορίζεται από τις Συναρτήσεις Συμμετοχής. Η απλούστερη προσέγγιση είναι η χρήση τραπεζοειδών συναρτήσεων. Τα πλεονεκτήματα χρήσης τραπεζοειδών συναρτήσεων συμμετοχής είναι:

- α) Η ικανοποιητική προσέγγιση της κανονικής κατανομής η οποία θεωρείται δεδομένη από τη χρήση της μεθόδου της Ανάλυσης σε Κύριες Συνιστώσες.
- β) Η προγραμματιστική ευκολία διαχείρισης κάτι που μας ωφελεί τόσο κατά την προγραμματιστική διαδικασία όσο και στον απαραίτητο χρόνο επεξεργασίας.

4.4.1 Κατασκευή των Συναρτήσεων Συμμετοχής.

Όπως έχουμε δει από το Κεφάλαιο 3, μία ομάδα αποτελείται από άτομα (σημεία) που ανήκουν σε μία συγκεκριμένη γεωμετρική περιοχή στο σύστημα των Κυρίων Αξόνων. Για παράδειγμα μικρές τιμές στον Άξονα 1, μεσαίες τιμές στον Άξονα 2, κ.τ.λ. Όμως μέσω της μεθόδου της Ανάλυσης σε Κύριες Συνιστώσες μπορούμε έμμεσα να χρησιμοποιήσουμε τα αρχικά χαρακτηριστικά (μεταβλητές) δεδομένου ότι και αυτά προβάλλονται στους Κύριους Άξονες. Έτσι θα μπορέσουμε κατασκευάζοντας Συναρτήσεις Συμμετοχής για τα ασαφή υποσύνολα που ορίσαμε παραπάνω, να βρούμε τις τιμές συμμετοχής των Ατόμων και κατ' επέκταση και των Ομάδων, και να δώσουμε λεκτικούς χαρακτηρισμούς στα Άτομα και τις Ομάδες σε σχέση με τη θέση τους στο σύστημα των Κυρίων Αξόνων.

Δεδομένου ότι σε κάθε Κύριο Άξονα οι προβολές $P_{i,j}$ των ατόμων έχουν μέση τιμή μηδέν και η διακύμανσή τους δίνεται από τη Σχέση 4.1, άτομα και ομάδες με προβολές κοντά στο μηδέν, χαρακτηρίζονται ως “Μεσαία” ή “Μεσαίες” (“Medium”) στον κάθε Κύριο Άξονα. Με την ίδια διαδικασία άτομα με προβολές κοντά στα άκρα των Κυρίων Αξόνων χαρακτηρίζονται ως “Χαμηλά” (“Low”) ή “Υψηλά” (“High”).

Ωστόσο λόγω της κανονικοποίησης (κανονικότητα των τιμών των προβολών), που είναι βήμα της Ανάλυσης σε Κύριες Συνιστώσες, έχουμε πάντα σημεία ανάμεσα στα άτομα που μπορούν να χαρακτηριστούν με τις δεδομένες λεκτικές μεταβλητές. Γι' αυτά τα άτομα ο χαρακτηρισμός είναι λίγο πιο περίπλοκος.

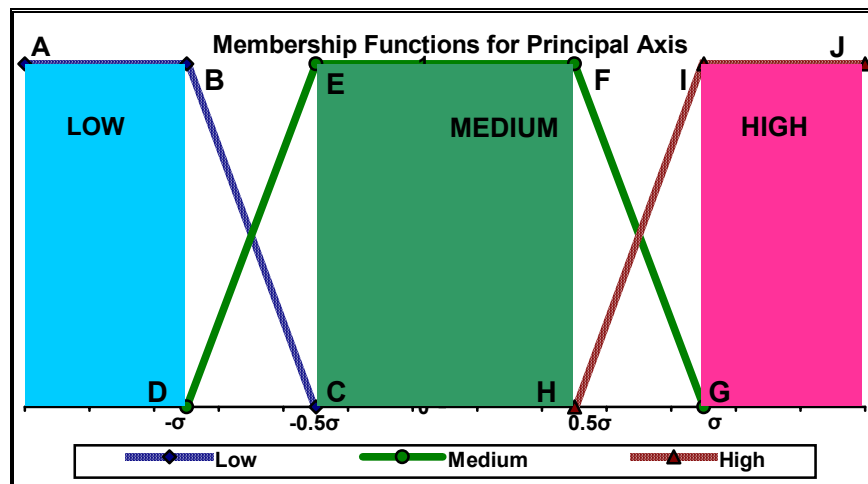
Η κατασκευή των Συναρτήσεων Συμμετοχής στα ασαφή υποσύνολα C_1 , C_2 , C_3 έχει ως εξής:

Σε κάθε Κύριο Άξονα κατασκευάζουμε τρεις σίγουρες περιοχές “Χαμηλή”, “Μεσαία” και “Υψηλή” (λεκτικές μεταβλητές) για τις μικρές (αρνητικές), κοντά στο μηδέν και μεγάλες (θετικές) τιμές προβολών αντίστοιχα. Για την περιοχή που βρίσκεται ανάμεσα

στις σίγουρες περιοχές χρησιμοποιούμε τραπεζοειδείς συναρτήσεις συμμετοχής. Ένα σημαντικό σημείο είναι το γεγονός ότι όλα τα άτομα (σημεία) για κάθε Άξονα έχουν μέση τιμή μηδέν και διασπορά που δίνεται από τη Σχέση (ΑΑ). Είναι λοιπόν λογικό να ορίσουμε τη “Μεσαία” περιοχή γύρω από το μηδέν. Για τις ακραίες τιμές της περιοχής, ένα πολλαπλάσιο της τυπικής απόκλισης σ_j όλων των σημείων στον Κύριο Άξονα μπορεί να χρησιμοποιηθεί ως ρυθμιστής των ορίων. Αυτή δίνεται από την σχέση:

$$\sigma_j = \sqrt{\sigma_j^2} = \sqrt{\sum_{i=1}^N P_{ij}^2} \quad \text{Σχέση 4.1}$$

όπου P_{ij} είναι οι προβολές των N ατόμων στον j Κύριο Άξονα.



Σχήμα 4.4: Συναρτήσεις Συμμετοχής για Κύριο άξονα

Στο Σχήμα 4.4 φαίνονται οι Συναρτήσεις Συμμετοχής για Κύριο Άξονα. Διακρίνονται οι σίγουρες περιοχές (με χαρακτηριστικό χρώμα που θα τις αντιπροσωπεύει στο εξής) και οι περιοχές ασάφειας που βρίσκονται ανάμεσά τους. Όπως φαίνεται οι τιμές ανάμεσα στο διάστημα -0.5σ και 0.5σ θεωρούνται απόλυτα (με τιμή συμμετοχής ένα ή 100%) Μεσαίες. Αντίστοιχα τιμές μικρότερες από $-\sigma$ και μεγαλύτερες από σ θεωρούνται 100% Χαμηλές ή Υψηλές αντίστοιχα. Οι τιμές ανάμεσα στο διαστήματα $(-\sigma, -0.5\sigma)$ ανήκουν

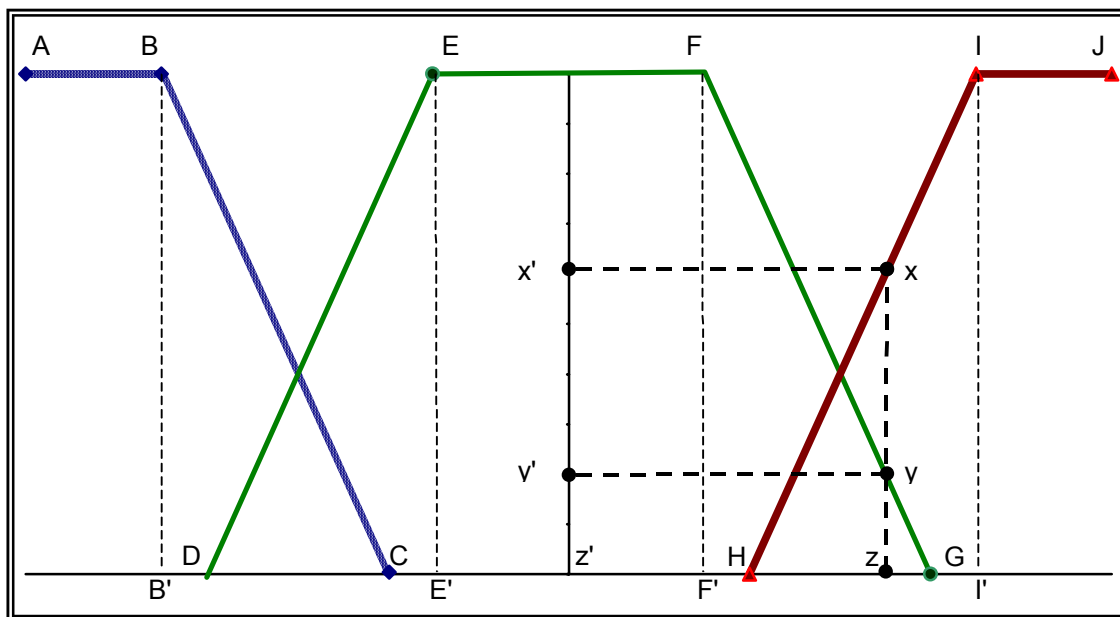
και στα δύο σύνολα Χαμηλό και Μεσαίο με τιμή συμμετοχής που δίνεται από τις γραμμές BC και DE αντίστοιχα ενώ οι τιμές ανάμεσα στο διαστήματα $(0.5\sigma, \sigma)$ ανήκουν και στα δύο σύνολα Μεσαίο και Υψηλό με τιμή συμμετοχής που δίνεται από τις γραμμές FG και HI αντίστοιχα.

4.5 ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ ΑΤΟΜΩΝ

Έχοντας ορίσει τις Συναρτήσεις Συμμετοχής στις τρεις περιοχές για κάθε Κύριο Άξονα μπορούμε να υπολογίσουμε τις τιμές συμμετοχής του κάθε ατόμου (σημείου) στην κάθε περιοχή του κάθε Κύριου Άξονα. Έτσι θα μπορέσουμε να δώσουμε λεκτικό χαρακτηρισμό στα άτομο. Για παράδειγμα το X άτομο θα έχει τιμή συμμετοχής $x\%$ στην “Χαμηλή” (“Low”) περιοχή του Άξονα 1, $xx\%$ στην “Μεσαία” (“Medium”) περιοχή του Άξονα 2. Έτσι έμμεσα θα μπορούμε να πούμε ότι έχει τόσο χαμηλές ή μεσαίες τιμές στα χαρακτηριστικά των αρχικών δεδομένων που αναπαρίστανται στους συγκεκριμένους Κύριους Άξονες.

Η τιμή συμμετοχής μ_{ij}^A , ενός ατόμου i σε μία περιοχή A του Κύριου Άξονα j , υπολογίζεται ως εξής:

Έστω οι συναρτήσεις συμμετοχής που φαίνονται στο Σχήμα 4.5 για τις τρεις περιοχές ενός Κύριου Άξονα j όπου με τα γράμματα A έως J συμβολίζονται τα όρια των περιοχών και έστω ότι η προβολή του ατόμου i στον Κύριο Άξονα j είναι z .



Σχήμα 4.5

Τότε θα ισχύει:

$$\mu_{ij}^L = 0$$

$$\mu_{ij}^M = (y'z') = (yz)$$

$$\mu_{ij}^H = (x'z') = (xz)$$

Από τα όμοια τρίγωνα $(FF'G)$ και (yzG) είναι:

$$\frac{(FF')}{(yz)} = \frac{(F'G)}{(zG)} \xrightarrow{(FF')=1} (yz) = \frac{zG}{F'G} \Rightarrow \mu_{ij}^M = \frac{zG}{F'G}$$

Από τα όμοια τρίγωνα $(II'H)$ και (xzH) είναι:

$$\frac{(II')}{(xz)} = \frac{(I'H)}{(zH)} \xrightarrow{(II')=1} (xz) = \frac{(zH)}{(I'H)} \Rightarrow \mu_{ij}^H = \frac{(zH)}{(I'H)}$$

Για παράδειγμα αν ισχύει:

$F'=0.3$, $H=0.4$, $z=0.7$, $G=0.8$ και $I'=0.9$ τότε

$$\mu_{ij}^M = \frac{zG}{F'G} = \frac{0.1}{0.5} = 0.2$$

$$\mu_{ij}^H = \frac{(zH)}{(I'H)} = \frac{0.3}{0.5} = 0.6$$

Δηλαδή το άτομο i ανήκει στο ασαφές υποσύνολο C_2 (με λεκτικό χαρακτηρισμό “Μεσαίο” ή “Medium”) ή πιο απλά στη “Μεσαία” περιοχή με τιμή συμμετοχής 0.2 και στο ασαφές υποσύνολο C_3 (με λεκτικό χαρακτηρισμό “Υψηλό” ή “High”) ή πιο απλά στην “Υψηλή” περιοχή με τιμή συμμετοχής 0.6.

4.6 ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ ΟΜΑΔΩΝ

Έχοντας υπολογίσει τις τιμές συμμετοχής των ατόμων στις περιοχές του κάθε κύριου άξονα μπορούμε να υπολογίσουμε και τις τιμές συμμετοχής των Ομάδων σ’ αυτούς. Χρησιμοποιούμε το στατιστικό μέσο των τιμών συμμετοχής των ατόμων που απαρτίζουν τις ομάδες σύμφωνα με τη Σχέση 4.2.

$$\mu_{Group}^A = \frac{1}{N_{Group}} \sum_{i=1}^{N_{Group}} \mu_{ij}^A \quad \text{Σχέση 4.2}$$

όπου

μ_{Group}^A είναι η τιμή συμμετοχής της Ομάδας σε κάποια περιοχή (λεκτική μεταβλητή) A (Area),

N_{Group} είναι ο αριθμός των ατόμων που απαρτίζουν την ομάδα και

μ_{ij}^A είναι η τιμή συμμετοχής του ατόμου i στην περιοχή A του άξονα j .

Αυτό μας επιτρέπει να χαρακτηρίσουμε μία ομάδα για παράδειγμα $X_1\%$ “Μεσαία”- $X_2\%$ “Υψηλή” στα χαρακτηριστικά που αντιπροσωπεύει ο Άξονας j .

4.7 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ

Ο ορισμός των Συναρτήσεων Συμμετοχής στις περιοχές του κάθε άξονα είναι υποκειμενικός και η μορφή τους καθορίζει άμεσα όχι τα αποτελέσματα της ομαδοποίησης αλλά το χαρακτηρισμό των ατόμων και των ομάδων που προκύπτουν από τον αλγόριθμο της ομαδοποίησης. Τα χαρακτηριστικά των συναρτήσεων συμμετοχής είναι τα εξής:

4.7.1 Ελαστικότητα των Συναρτήσεων Συμμετοχής στο χαρακτηρισμό ως “Μεσαίο”

Ανάλογα με τις τιμές των ορίων των περιοχών μπορούμε να είμαστε ελαστικοί ή αυστηροί ως προς τον χαρακτηρισμό των σημείων και των ομάδων ως “Μεσαία”. Έχουμε:

Ελαστικότητα στο χαρακτηρισμό “Μεσαία”: Όταν ορίζουμε μεγάλη περιοχή γύρω από το μηδέν για την οποία τα άτομα και οι ομάδες θεωρούνται “Μεσαία” και αντίστοιχα μικρές περιοχές στα άκρα των Κυρίων Αξόνων για τις οποίες τα άτομα και οι ομάδες θεωρούνται “Χαμηλά” ή “Υψηλά”.

Αυστηρότητα στο χαρακτηρισμό “Μεσαία”: Όταν ορίζουμε μικρή περιοχή γύρω από το μηδέν για την οποία τα άτομα και οι ομάδες θεωρούνται “Μεσαία” και αντίστοιχα μεγάλες περιοχές στα άκρα των Κυρίων Αξόνων για τις οποίες τα άτομα και οι ομάδες θεωρούνται “Χαμηλά” ή “Υψηλά”.

4.7.2 Συμμετρία των Συναρτήσεων Συμμετοχής

Ένα σημαντικό θέμα για τις Συναρτήσεις Συμμετοχής είναι η συμμετρία. Στο παράδειγμα που δώσαμε στον υπολογισμό των τιμών συμμετοχής των ατόμων, βρήκαμε ότι το άτομο ανήκει στην “Μεσαία” περιοχή με τιμή συμμετοχής 0.2 και συγχρόνως στην “Υψηλή” περιοχή με τιμή συμμετοχής 0.6. Θεωρητικά το άτομο ανήκει και στην “Χαμηλή” περιοχή με τιμή συμμετοχής του είναι μηδέν. Δηλαδή ισχύει:

$$\mu_{ij}^L + \mu_{ij}^M + \mu_{ij}^H = 0.8$$

Συμμετρικός θεωρείται ένας κανόνας (συναρτήσεις συμμετοχής για τις τρεις περιοχές ενός Κύριου Άξονα) όταν ισχύει η Σχέση 4.3.

$$\mu_{ij}^L + \mu_{ij}^M + \mu_{ij}^H = 1 \quad \text{Σχέση 4.3}$$

ενώ αντίστροφα μη Συμμετρικός θεωρείται ένας κανόνας όταν ισχύει η Σχέση 4.4

$$\mu_{ij}^L + \mu_{ij}^M + \mu_{ij}^H < 1 \quad \text{Σχέση 4.4}$$

4.7.3 Περιορισμοί των Ορίων των Συναρτήσεων Συμμετοχής

Από τους παραπάνω ορισμούς και προκειμένου να έχει νόημα ο ορισμός συναρτήσεων συμμετοχής προκύπτει ο περιορισμός της Σχέσης 4.5 για τις τιμές συμμετοχής στις τρεις περιοχές του κάθε Κύριου Άξονα:

$$0 < \sum_{A=1}^3 \mu_{ij}^{C_A} \leq 1 \quad \text{Σχέση 4.5}$$

Από αυτό τον περιορισμό προκύπτει ότι τα όρια των τριών περιοχών όπως φαίνονται στο Σχήμα 4.5 για κάθε Κύριο Άξονα πρέπει να ικανοποιούν τις εξής ανισότητες:

$D \geq B$, δηλαδή το κάτω αριστερό όριο της “Μεσαίας” περιοχής πρέπει να έχει τιμή μεγαλύτερη ή ίση του πάνω δεξιού ορίου της “Χαμηλής” περιοχής.

$E \geq C$, δηλαδή το πάνω αριστερό όριο της “Μεσαίας” περιοχής πρέπει να έχει τιμή μεγαλύτερη ή ίση του κάτω δεξιού ορίου της “Χαμηλής” περιοχής.

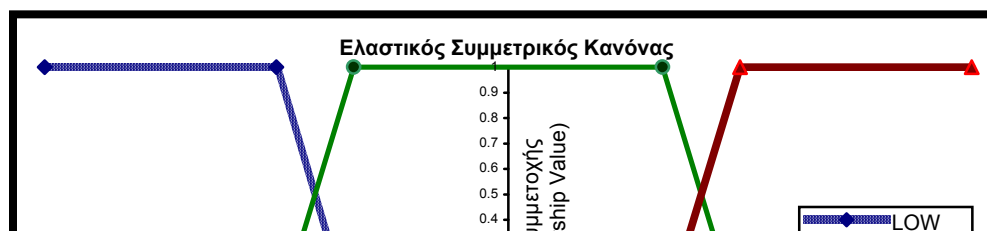
$F \leq H$, δηλαδή το πάνω δεξιό όριο της “Μεσαίας” περιοχής πρέπει να έχει τιμή μικρότερη ή ίση του κάτω αριστερού ορίου της “Υψηλής” περιοχής.

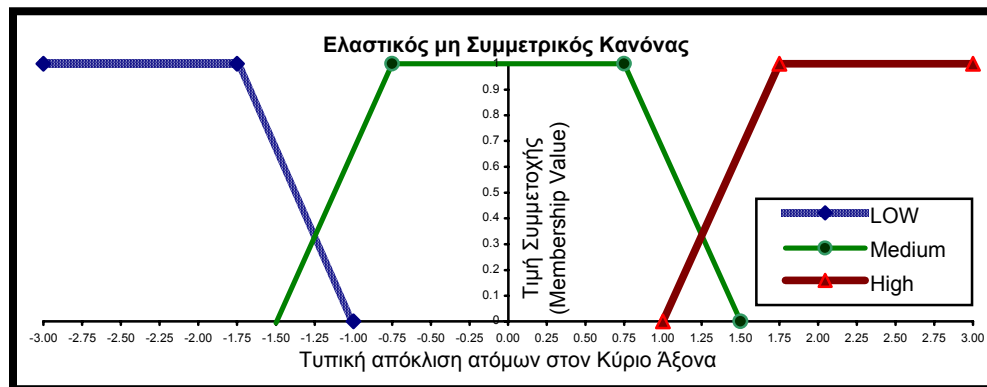
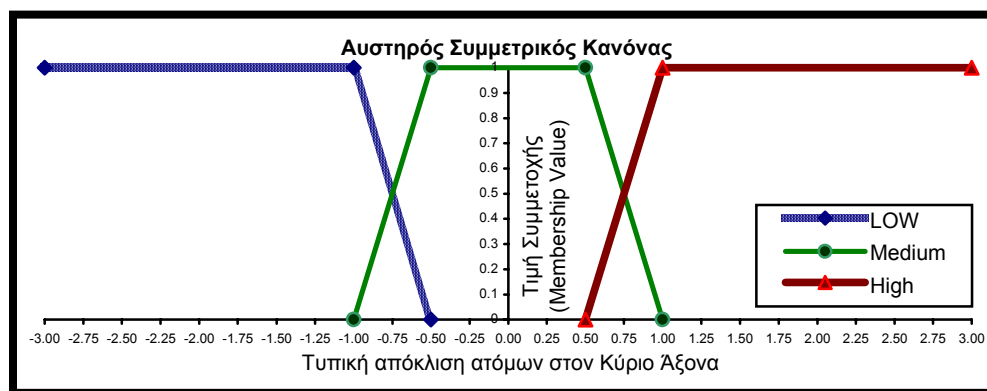
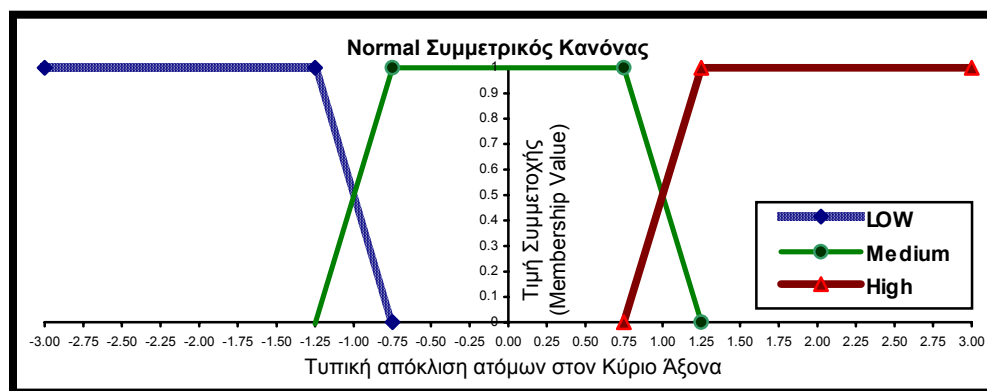
$G \leq I$, δηλαδή το κάτω δεξιό όριο της “Μεσαίας” περιοχής πρέπει να έχει τιμή μικρότερη ή ίση του πάνω αριστερού ορίου της “Υψηλής” περιοχής.

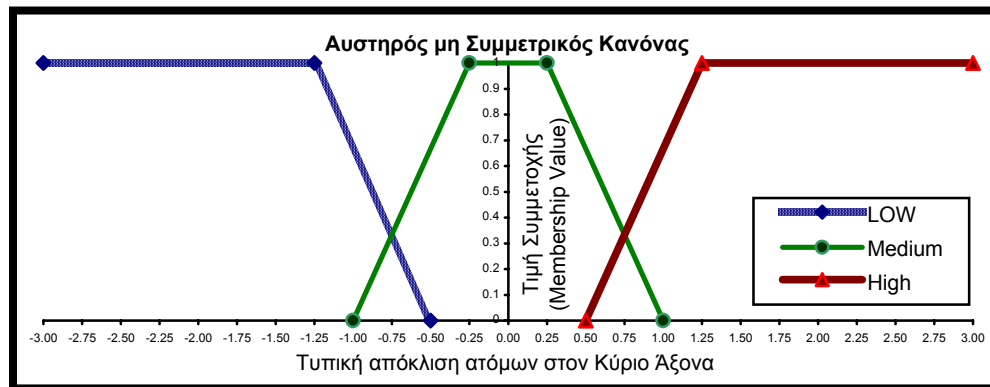
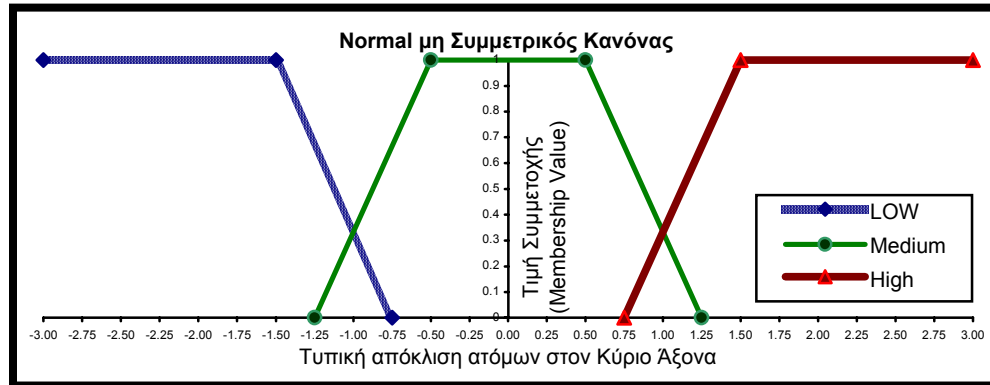
4.8 ΠΡΟΤΕΙΝΟΜΕΝΟΙ ΚΑΝΟΝΕΣ (Standard Rules)

Οι τιμές των ορίων των συναρτήσεων συμμετοχής των ατόμων στις περιοχές του κάθε Κύριου Άξονα ορίζονται στην εφαρμογή Fuzzy Data Grouping από το χρήστη. Για λόγους όμως ευκολίας και ευχρηστίας όσο και για το λόγο ότι ίσως κάποιοι χρήστες θεωρήσουν δύσκολη την επιλογή των ορίων, προτείνονται έξι έτοιμοι κανόνες. Οι τρεις πρώτοι είναι Συμμετρικοί και οι τρεις επόμενοι μη Συμμετρικοί.

Οι κανόνες υπολογίζουν τα όρια των περιοχών σε κάθε Κύριο Άξονα ως πολλαπλάσιο της τυπικής απόκλισης όλων των ατόμων στον Κύριο Άξονα. Έτσι ο χρήστης μπορεί να εφαρμόσει αναλογικά τον ίδιο κανόνα στον κάθε Άξονα ή να επιλέξει διαφορετικούς για κάθε Άξονα ανάλογα με την επιθυμία του. Παρακάτω παραθέτουμε τους κανόνες.







Αν υποθέσουμε ότι τα δεδομένα ακολουθούν την κανονική κατανομή σε ένα άξονα τότε μπορούμε να υπολογίσουμε τον αναμενόμενο αριθμό ατόμων για κάθε περιοχή του άξονα.

Στον Πίνακα 4.2 με “ C_M ” συμβολίζουμε τον αναμενόμενο αριθμό ατόμων που θα χαρακτηριστούν μεσαία με τιμή συμμετοχής 1 (ή 100% όπως θα λέμε στο εξής) και με “ F_M ” τον αναμενόμενο αριθμό ατόμων που όντας βρισκόμενα στις περιοχές ανάμεσα στη σίγουρα Μεσαία και τις σίγουρες Χαμηλή και Υψηλή, θα έχουν μεγαλύτερη τιμή συμμετοχής στη Μεσαία περιοχή.

Κανόνας		C_M (%)	F_M (%)	C_{LH} (%)	F_{LH} (%)
Συμμετρικός	Ελαστικός	68,26	10	13,36	8,38
	Normal	54,68	13,58	21,74	10
	Αυστηρός	38,3	16,38	31,74	13,58
Μη Συμμετρικός	Ελαστικός	54,68	23,58	9,9	11,84
	Normal	38,3	29,96	13,36	18,38
	Αυστηρός	19,74	34,94	21,74	23,58

Πίνακας 4.2: Επί τοις εκατό αναμενόμενος αριθμός ατόμων στις περιοχές

Αντίστοιχα με “ C_{LH} ” συμβολίζουμε τον αναμενόμενο αριθμό ατόμων που θα βρίσκονται 100% στις σίγουρες περιοχές Χαμηλή και Υψηλή (δηλαδή με τιμή συμμετοχής 1) και με “ F_{LH} ” το άθροισμα των ατόμων που όντας βρισκόμενα στις ίδιες παραπάνω ασαφείς περιοχές έχουν μεγαλύτερη τιμή συμμετοχής στις ακραίες περιοχές.

ΚΕΦΑΛΑΙΟ 5: Η ΕΦΑΡΜΟΓΗ ΟΜΑΔΟΠΟΙΗΣΗΣ ΜΕ ΑΣΑΦΗ ΛΟΓΙΚΗ (FUZZY DATA GROUPING)

Σ' αυτό το κεφάλαιο θα προσπαθήσουμε να περιγράψουμε το σύστημα επικοινωνίας (interface) της εφαρμογής Fuzzy Data Grouping με το χρήστη. Γνωρίζοντας ότι η χρησιμότητα του εκάστοτε λογισμικού έγκειται σε μεγάλο βαθμό στο φιλικό του περιβάλλον αλλά και στη δυνατότητα του χρήστη να εκμεταλλευτεί τις δυνατότητές του, θα προσπαθήσουμε να είμαστε όσο πιο αναλυτικοί και κατανοητοί γίνεται.

5.1 Η ΓΛΩΣΣΑ ΑΝΑΠΤΥΞΗΣ

Η Εφαρμογή Fuzzy Data Grouping (FuDaGr) αναπτύχθηκε σε περιβάλλον Windows με τη γλώσσα προγραμματισμού Visual Basic Ver. 6 (SP3). Η ανάπτυξη λογισμικού με τη Visual Basic έχει μερικά βασικά πλεονεκτήματα:

- Είναι βέβαιο ότι κάθε λογισμικό που αναπτύσσεται, έχει ως στόχο την κάλυψη κάποιων αναγκών. Βασική προϋπόθεση ώστε αυτός ο στόχος να επιτευχθεί είναι η δυνατότητα χρήσης του από όσο το δυνατό γίνεται περισσότερους χρήστες. Η καθιέρωση του περιβάλλοντος Windows στο 90% περίπου των Ηλεκτρονικών Υπολογιστών αποτελεί σήμερα γεγονός και η ανάπτυξη λογισμικού σε οποιοδήποτε άλλο περιβάλλον αποτελεί τροχοπέδη για οποιονδήποτε θα ήθελε να χρησιμοποιήσει το λογισμικό αυτό. Η Visual Basic εξασφαλίζει συμβατότητα με Windows 95/98, Windows NT4 και Windows 2000.
- Σημαντικό πλεονέκτημα αποτελεί επίσης η δυνατότητα σχεδίασης της αλληλεπίδρασης χρήστη-μηχανής (interface) με τους άτυπους κανόνες που ουσιαστικά έχουν επιβληθεί από το περιβάλλον Windows. Μία τέτοια σχεδίαση (Windows Standards) εξασφαλίζει ότι ένα μεγάλο πλήθος χρηστών θα εξοικειωθεί γρηγορότερα με τη χρήση του λογισμικού κάτι που αυξάνει το βαθμό χρησιμότητάς του.
- Η δυνατότητα επικοινωνίας με ευρέως διαδεδομένες εφαρμογές όπως το Microsoft Excel και η Microsoft Access (βάση δεδομένων) λύνει πολλά από τα προβλήματα που θα προέκυπταν σε περίπτωση ανάπτυξης λογισμικού σε άλλο περιβάλλον.

5.2 ΑΠΑΙΤΗΣΕΙΣ

Η Εφαρμογή Fuzzy Data Grouping μπορεί να εγκατασταθεί σε οποιονδήποτε Ηλεκτρονικό Υπολογιστή αρκεί να ικανοποιούνται οι παρακάτω απαιτήσεις:

1. Λειτουργία σε περιβάλλον Windows 95 ή νεότερο.
2. Επεξεργαστής Pentium 166 MHz τουλάχιστον.
3. Ο Σκληρός Δίσκος του H/Y να έχει ελεύθερο χώρο αποθήκευσης 10 Mb.

Αυτό μπορεί να γίνει μέσω των αρχείων εγκατάστασης τα οποία δημιουργήθηκαν μέσω του εργαλείου Package & Deployment Wizard του Microsoft Visual Studio 6.0. Τα αρχεία εγκατάστασης έχουν μέγεθος περίπου 7 Mb ενώ το εκτελέσιμο αρχείο της εφαρμογής (fudagr.exe) είναι περίπου 1.1 Mb.

5.3 ΧΡΗΣΗ ΑΡΧΕΙΩΝ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

Η Εφαρμογή Fuzzy Data Grouping (FuDaGr) αποθηκεύει τα δεδομένα σε δύο διαφορετικούς πίνακες στο ίδιο αρχείο (βάση δεδομένων), τύπου .mdb, συμβατό με την Microsoft Access των Windows.

Για την αποθήκευση του Αρχικού Πίνακα Δεδομένων της ΑΚΣ χρησιμοποιείται ο πίνακας “Individuals_Variables_Values” και για την αποθήκευση των σημείων ορισμού των Συναρτήσεων Συμμετοχής ο πίνακας “Principal_Axes_MMF”.

Η δυσκολία που προκύπτει από τη χρήση αρχείων τύπου .mdb είναι το γεγονός ότι κατά τη διαδικασία ανάκτησης των δεδομένων από τη βάση δεδομένων είναι απαραίτητος ο έλεγχος της μορφής των πινάκων ώστε να εξασφαλίζεται η ομαλή λειτουργία της Εφαρμογής. Ο έλεγχος αυτός σε πολλές περιπτώσεις μπορεί να γίνει εξαιρετικά δύσκολος. Δεδομένου όμως ότι ένας τέτοιου είδους έλεγχος ξεφεύγει από τους σκοπούς αυτής της εργασίας, κατά την ανάκτηση των δεδομένων πραγματοποιούνται στοιχειώδεις έλεγχοι και η ομαλή λειτουργία της Εφαρμογής επαφίεται στους χρήστες οι οποίοι θα πρέπει να ακολουθούν συγκεκριμένα πρότυπα κατά τη διαδικασία δημιουργίας της βάσης με τα δεδομένα τους όταν επιθυμούν να τη δημιουργήσουν μέσω της εφαρμογής Microsoft Access. Ωστόσο η διαδικασία δημιουργίας της απαραίτητης βάσης δεδομένων είναι σχετικά απλή υπόθεση όταν γίνεται μέσα από την εφαρμογή FuDaGr όπως θα δούμε παρακάτω. Σε περίπτωση που οι πίνακες δεδομένων είναι ογκώδεις, και είναι ήδη αποθηκευμένοι σε ηλεκτρονική μορφή σε άλλη εφαρμογή των Windows (π.χ. Microsoft Word ή Excel), είναι δυνατή η μεταφορά τους με επίσης απλή διαδικασία.

5.3.1 Ο Αρχικός Πίνακας Δεδομένων (Individuals_Variables_Values)

Ο Πίνακας 5.1 δείχνει τη μορφή του Αρχικού Πίνακα Δεδομένων. Πρόκειται δηλαδή για ένα πίνακα $N+1 \times K+1$ όπου N είναι ο αριθμός των Ατόμων και K είναι ο αριθμός των Χαρακτηριστικών (Μεταβλητών). Στην πρώτη γραμμή γράφονται τα ονόματα των χαρακτηριστικών και στην πρώτη στήλη γράφονται τα ονόματα των ατόμων. Τα Individual1, Individual1,... είναι οι εγγραφές (Records) ενώ τα Variable1, Variable1,... είναι τα πεδία (Fields) της Access.

Individual_Name	Variable1	Variable2	Variable3	...	VariableK
Individual1	(value)	(value)	(value)	...	(value)
Individual2	(value)	(value)	(value)	...	(value)
Individual3	(value)	(value)	(value)	...	(value)
...
...
...
IndividualN	(value)	(value)	(value)	...	(value)

Πίνακας 5.1: Η μορφή του Αρχικού Πίνακα Δεδομένων της βάσης δεδομένων Access

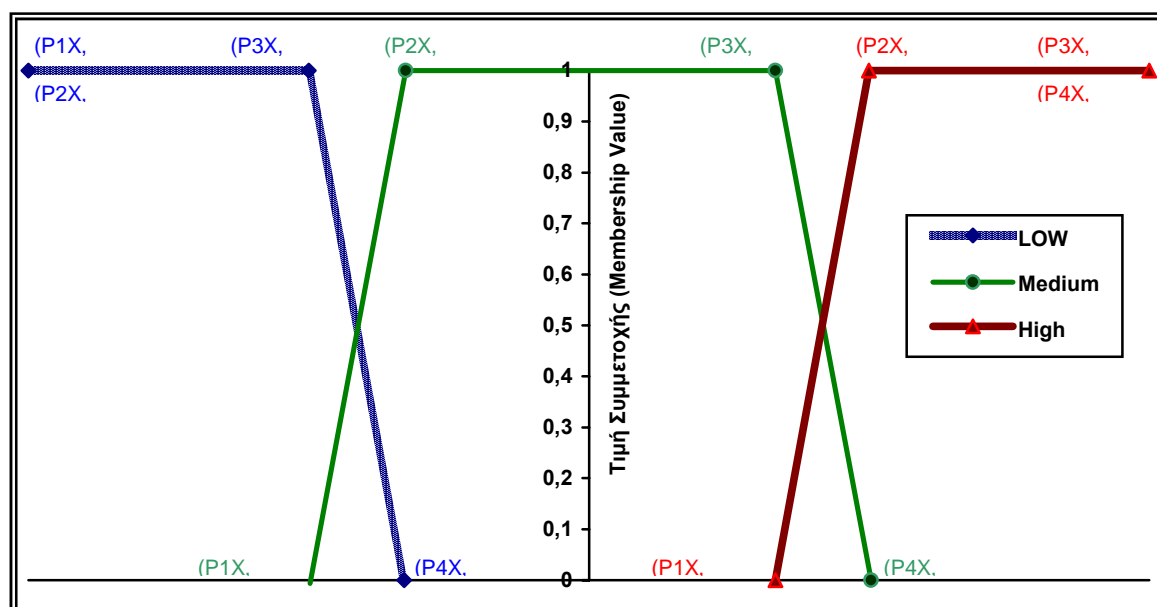
5.3.2 Ο Πίνακας των Συναρτήσεων Συμμετοχής (Principal_Axes_MMF)

Ο Πίνακας των Συναρτήσεων Συμμετοχής περιέχει τις απαραίτητες τιμές για τον ορισμό των συναρτήσεων συμμετοχής. Η μορφή του είναι αυτή του Πίνακα 5.2.

Variable	Area	Point1X	Point1Y	Point2X	Point2Y	Point3X	Point3Y	Point4X	Point4Y
Axis1	Low	(value)	1	(value)	1	(value)	1	(value)	0
Axis1	Medium	(value)	0	(value)	1	(value)	1	(value)	0
Axis1	High	(value)	0	(value)	1	(value)	1	(value)	1
Axis2	Low	(value)	1	(value)	1	(value)	1	(value)	0
Axis2	Medium	(value)	0	(value)	1	(value)	1	(value)	0
Axis2	High	(value)	0	(value)	1	(value)	1	(value)	1
...
...
...
AxisK	Low	(value)	1	(value)	1	(value)	1	(value)	0
AxisK	Medium	(value)	0	(value)	1	(value)	1	(value)	0
AxisK	High	(value)	0	(value)	1	(value)	1	(value)	1

Πίνακας 5.1: Η μορφή του Πίνακα Συναρτήσεων Συμμετοχής της βάσης δεδομένων Access

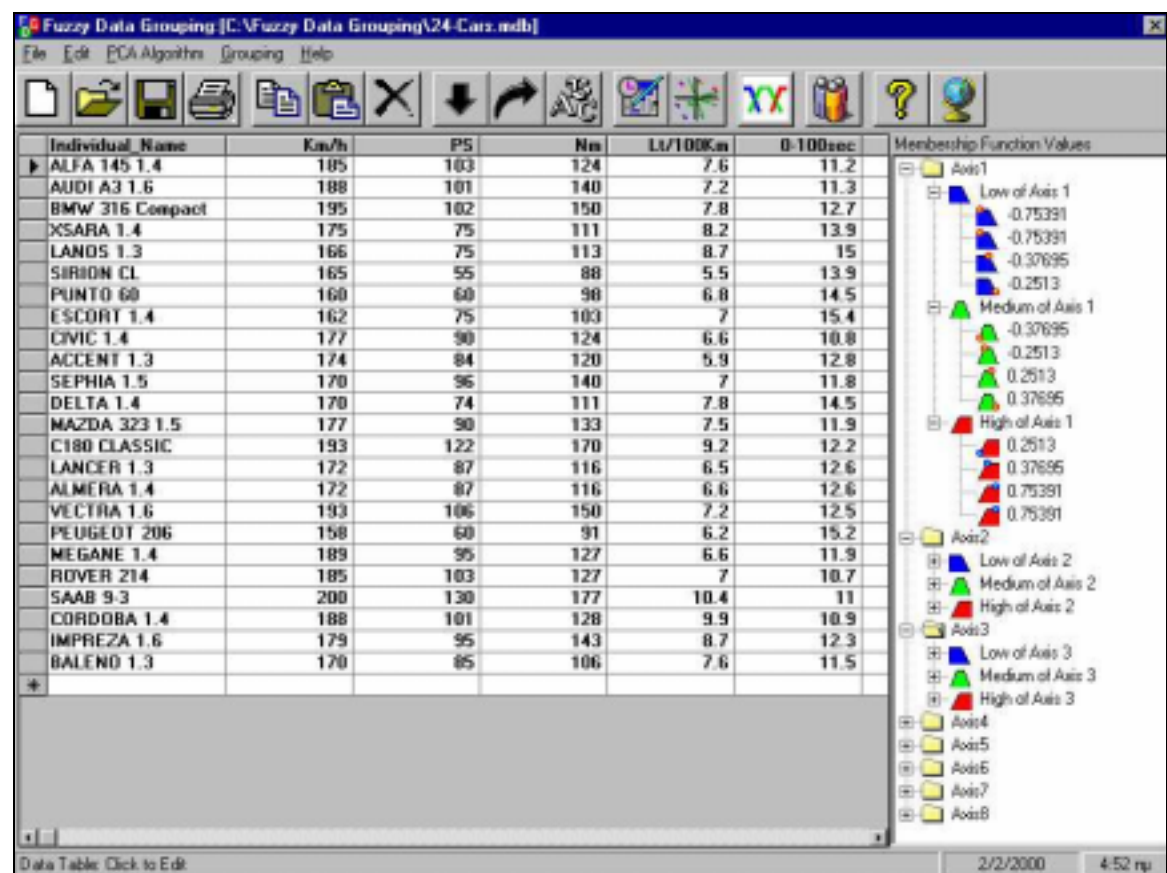
Δεδομένου ότι οι Συναρτήσεις Συμμετοχής είναι τραπεζοειδείς, τέσσερα σημεία είναι απαραίτητα για τον ορισμό της κάθε μίας από τις τρεις περιοχές (Low, Medium, High). Επομένως για κάθε Νέο Άξονα για την απεικόνιση των τεσσάρων σημείων στο επίπεδο είναι απαραίτητα 24 σημεία, οκτώ για κάθε περιοχή. Τα σημεία αυτά για ένα άξονα φαίνονται στο Σχήμα 5.1, όπου για λόγους ευκρίνειας ένα ζεύγος τιμών παριστάνεται για παράδειγμα με (P1X, P1Y) αντί για (Point1X, Point1Y). Ωστόσο επειδή το κάτω αριστερό σημείο και το κάτω δεξιό σημείο της Χαμηλής και της Υψηλής περιοχής αντίστοιχα δεν ορίζονται, το πρώτο σημείο για την Χαμηλή περιοχή συμπίπτει με το δεύτερο και το τρίτο σημείο της Υψηλής περιοχής συμπίπτει με το τέταρτο.



Το μέγεθος της βάσης δεδομένων εξαρτάται από τον αριθμό των ατόμων και των αριθμό των χαρακτηριστικών. Στην εφαρμογή Fuzzy Data Grouping δεν υπάρχει περιορισμός για το μέγεθος της αφού αυτό συμπίπτει με τα όρια της Access που είναι πρακτικά η χωρητικότητα του σκληρού δίσκου του συστήματος. Πρέπει επίσης να τονίσουμε ότι ενώ τα δεδομένα του Αρχικού Πίνακα Δεδομένων πρέπει να εισαχθούν από το χρήστη πριν γίνει εφαρμογή του αλγορίθμου της ΑΚΣ, τα στοιχεία του Πίνακα των Συναρτήσεων Συμμετοχής μπορούν να τεθούν μετά την εφαρμογή του αλγορίθμου.

5.4 Η ΑΡΧΙΚΗ ΦΟΡΜΑ

Η Αρχική Φόρμα (βασικό παράθυρο της εφαρμογής) έχει τη μορφή του Σχήματος 5.2. Ήδη σε αυτό το στάδιο ο χρήστης έχει ανακτήσει δεδομένα από ένα αρχείο τύπου .mdb χρησιμοποιώντας το δεύτερο κουμπί της εργαλειοθήκης. Τα αντικείμενα που διακρίνονται πάνω στη φόρμα είναι ο πίνακας δεδομένων, το δέντρο των ορίων των συναρτήσεων συμμετοχής (Membership Function Values), τα μενού, η εργαλειοθήκη και η γραμμή κατάστασης.



Σχήμα 5.1: Η Αρχική Φόρμα του της εφαρμογής Fuzzy Data Grouping

Παρακάτω θα εξετάσουμε λεπτομερώς την παραπάνω φόρμα επεξηγώντας τα αντικείμενα που βρίσκονται πάνω της και τη χρήση τους.

5.4.1 Ο Πίνακας Δεδομένων

Πρόκειται για τον αρχικό πίνακα δεδομένων της ΑΚΣ. Οι τιμές στα πεδία αλλάζουν με απλή επιλογή του αντίστοιχου κελιού με το ποντίκι (mouse). Σημαντικό είναι να αναφερθεί ότι εξαιτίας της φύσης




της βάσης δεδομένων και της άμεσης αλληλεπίδρασής της με τον πίνακα δεδομένων, οποιαδήποτε αλλαγή τιμής ή ονόματος ατόμου καταχωρείται αυτόματα στη βάση δεδομένων χωρίς δυνατότητα αναιρέσης. Τα δεδομένα αποθηκεύονται στον πίνακα “Individuals_Variables_Values” της βάσης δεδομένων Access.

5.4.2 Το Δέντρο των Ορίων των Συναρτήσεων Συμμετοχής

Η δενδρική αναπαράσταση είναι αρκετά συνήθης στις εφαρμογές των Windows. Η χρήση της για την αναπαράσταση των ορίων των συναρτήσεων συμμετοχής θεωρείται ιδανική αφού έτσι επιτυγχάνεται η εύκολη κατανόηση και περιήγηση στις τιμές του πίνακα “Principal_Axes_MMF” της βάσης δεδομένων. Στο δέντρο αυτό δεν είναι δυνατή η αλλαγή των ορίων των συναρτήσεων συμμετοχής.

Το δέντρο έχει διάρθρωση τριών επιπέδων. Στο πρώτο επίπεδο φαίνονται οι κύριοι άξονες, στο δεύτερο οι τρεις περιοχές του κάθε άξονα και στο τρίτο οι τιμές που απαιτούνται για τον ορισμό των περιοχών. Στο τρίτο επίπεδο φαίνονται μόνο οι συντεταγμένες στον Χ άξονα των σημείων ορισμού των συναρτήσεων συμμετοχής αφού οι συντεταγμένες του άξονα Υ είναι δεδομένες, όπως φαίνεται από τον Πίνακα 5.2. Εξάλλου για λόγους που εξηγήθηκαν παραπάνω, το πάνω αριστερό όριο και το πάνω δεξιό όριο φαίνονται στο δέντρο δύο φορές για την Χαμηλή και την Υψηλή περιοχή αντίστοιχα. Θεωρητικά τα όρια αυτά είναι $-\infty$ για τη Χαμηλή και $+\infty$ για την Υψηλή περιοχή. Πρακτικά εμφανίζονται οι τιμές -3σ και $+3\sigma$ όπου σ η τυπική απόκλιση των ατόμων στον κάθε κύριο άξονα της ΑΚΣ.

Για την ευκολότερη κατανόηση των τιμών που αντιστοιχούν σε κάθε όριο και για κάθε περιοχή, χρησιμοποιούνται οι παρακάτω γραφικοί συμβολισμοί οι οποίοι ισχύουν για ολόκληρη την εφαρμογή. Κάθε περιοχή συμβολίζεται με ένα χρώμα και ένα γεωμετρικό σχήμα. Αυτά φαίνονται στον Πίνακα 5.3.

Περιοχή	Χρώμα	Γεωμετρικό σχήμα
Χαμηλή	Μπλε	
Μεσαία	Πράσινο	
Υψηλή	Κόκκινο	

Πίνακας 5.3: Χρωματικοί και γεωμετρικοί συμβολισμοί περιοχών

Τα γεωμετρικά σχήματα που χρησιμοποιούνται για το συμβολισμό των ορίων των περιοχών φαίνονται στον Πίνακα 5.4.

Περιοχή	Κάτω Αριστερό Σημείο	Πάνω Αριστερό Σημείο	Πάνω Δεξιό Σημείο	Κάτω Δεξιό Σημείο
Χαμηλή	δεν ορίζεται ($-\infty$)			
Μεσαία				
Υψηλή				δεν ορίζεται ($+\infty$)

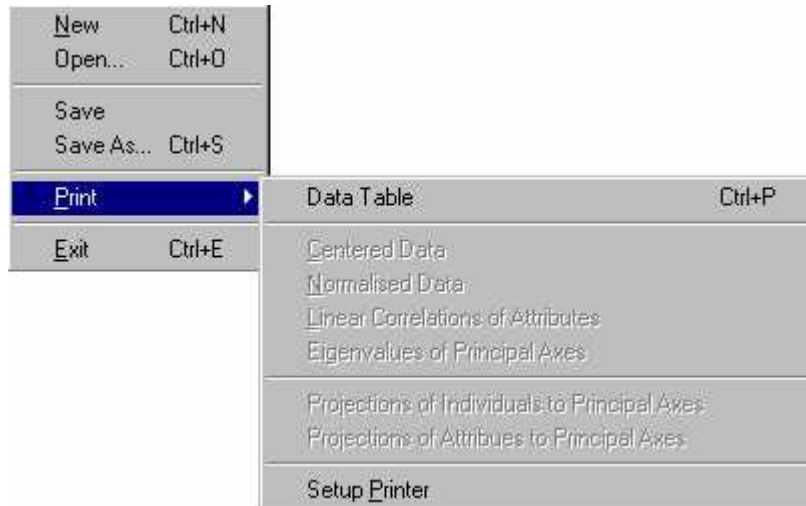
Πίνακας 5.4: Γεωμετρικοί συμβολισμοί ορίων των Συναρτήσεων Συμμετοχής

5.4.3 Τα Μενού της Αρχικής Φόρμας

Παρακάτω παρουσιάζεται η μορφή των μενού της Αρχικής Φόρμας και εξηγούνται οι διάφορες επιλογές (εντολές του μενού).

Το Μενού File

Το μενού αυτό (Σχήμα 5.3) δίνει τη δυνατότητα διαχείρισης του αρχείου της βάσης δεδομένων.

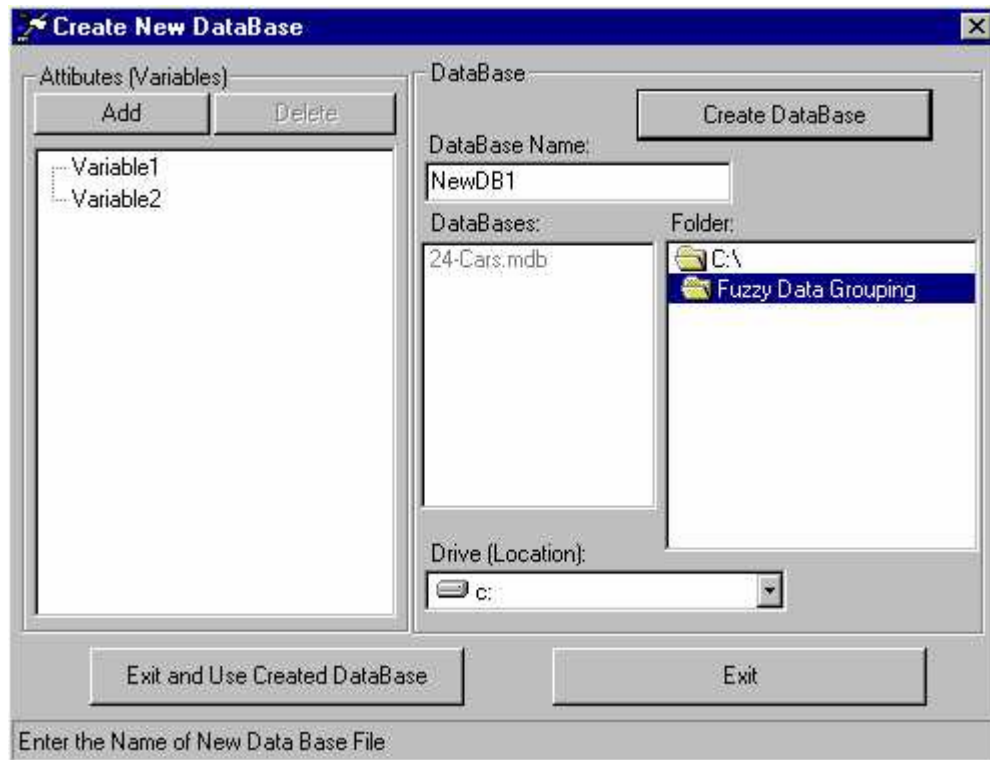


Σχήμα 5.1: Το Μενού File

Οι εντολές παρέχουν τις εξής δυνατότητες:

New...

Με αυτή την επιλογή δημιουργείται μία νέα βάση δεδομένων. Προκειμένου να γίνει αυτό εμφανίζεται η φόρμα του Σχήματος 5.4.



Σχήμα 5.1: Η Φόρμα Create New DataBase

Η επιλογή του ονόματος της Βάσης Δεδομένων που πρόκειται να δημιουργηθεί φαίνεται στο πεδίο με όνομα DataBase Name μέσα στο πλαίσιο “DataBase”. Το προτεινόμενο όνομα είναι “NewDB1” αλλά όπως είναι ευνόητο ο χρήστης μπορεί να επιλέξει το όνομα που ο ίδιος επιθυμεί. Επίσης μπορεί να επιλεγεί η τοποθεσία (path) από τα υπόλοιπα αντικείμενα που βρίσκονται μέσα στο πλαίσιο με όνομα “DataBase” και συγκεκριμένα από τα “Folder” και “Drive (Location)”.

Στο πλαίσιο με όνομα “Attributes(Variables)” γίνεται η επιλογή του αριθμού των χαρακτηριστικών (μεταβλητών) που ο χρήστης επιθυμεί να έχει ο Αρχικός Πίνακας Δεδομένων. Δηλαδή τον αριθμό των στηλών. Με τη χρήση του κουμπιού «Add», προστίθεται ένα χαρακτηριστικό, ενώ με τη χρήση του κουμπιού «Delete» αφαιρείται ένα χαρακτηριστικό. Αλλαγή στο όνομα του χαρακτηριστικού γίνεται με το πάτημα του δεξιού πλήκτρου του ποντικιού σε ένα ήδη επιλεγμένο χαρακτηριστικό. Ο ελάχιστος επιτρεπτός αριθμός χαρακτηριστικών είναι δύο προκειμένου να έχει νόημα η εφαρμογή της ΑΚΣ.

Η Βάση Δεδομένων δημιουργείται μόνο εφ’ όσον πατηθεί το κουμπί “Create DataBase”. Στη συνέχεια μπορεί να γίνει χρήση της νέας Βάσης Δεδομένων στην εφαρμογή με το πάτημα του κουμπιού “Exit and Use Created DataBase”. Το κουμπί “Exit” χρησιμοποιείται για επιστροφή στην Αρχική Φόρμα χωρίς τη δημιουργία νέου Αρχείου Βάσης Δεδομένων.

Open...

Με αυτή την επιλογή γίνεται ανάκτηση δεδομένων από μία ήδη υπάρχουσα βάση δεδομένων τύπου .mdb. Το παράθυρο επιλογής του αρχείου της βάσης δεδομένων είναι το αντίστοιχο παράθυρο (“Ανοιγμα” ή “Open”) που έχουμε συνηθίσει να χρησιμοποιούμε στα Windows. Η εφαρμογή καλεί ακριβώς την ίδια φόρμα του λειτουργικού συστήματος.

Save

Η επιλογή αυτή δεν έχει ουσιαστικό νόημα δεδομένου ότι η βάση δεδομένων ούτως ή άλλως ενημερώνεται αυτόματα σε κάθε αλλαγή που συμβαίνει. Η ύπαρξη αυτής της επιλογής είναι κατά κάποιο τρόπο περιττή εξυπηρετώντας ωστόσο τις περιπτώσεις που ένας χρήστης δεν γνωρίζει τη φύση της βάσης δεδομένων Access.

Save As...

Η επιλογή αυτή δίνει τη δυνατότητα να αποθήκευσης των δεδομένων με διαφορετικό όνομα από αυτό που έχει μέχρι τώρα η βάση δεδομένων. Το παράθυρο επιλογής του νέου ονόματος είναι των Windows (“Αποθήκευση ως” ή “Save As”).

Print

Η επιλογή αυτή μας παρέχει τη δυνατότητα εκτύπωσης των δεδομένων.

Υπάρχει η δυνατότητα εκτύπωσης διαφόρων πινάκων της ΑΚΣ. Αυτοί είναι:

Data Table

Εκτύπωση του Αρχικού Πίνακα δεδομένων.

Εφ’ όσον έχει εφαρμοστεί ο Αλγόριθμος της Ανάλυσης σε Κύριες Συνιστώσες, ενεργοποιούνται οι παρακάτω εντολές.

Centered Data

Εκτύπωση του Κεντροποιημένου Πίνακα Δεδομένων

Normalized Data

Εκτύπωση του Ισοσταθμισμένου Πίνακα Δεδομένων.

Linear Correlations of Attributes

Εκτύπωση του Πίνακα Συσχετίσεων μεταξύ των Χαρακτηριστικών.

Eigenvalues of Principal Axes

Εκτύπωση του Πίνακα των Ιδιοτιμών των Νέων Αξόνων.

Projections of Individuals to Principal Axes

Εκτύπωση του Πίνακα Προβολών των Ατόμων στους Νέους Άξονες.

Projections of Attributes to Principal Axes

Εκτύπωση του Πίνακα Προβολών των Χαρακτηριστικών στους Νέους Άξονες.

Setup Printer

Ρύθμιση του Εκτυπωτή.

To Μενού Edit

Με το Μενού Edit, είναι δυνατή η διαχείριση του Αρχικού Πίνακα Δεδομένων και η δημιουργία νέας Βάσης Δεδομένων. Οι εντολές του μενού Edit (Σχήμα 5.5) και η αναλυτική επεξήγηση των εντολών δίνεται παρακάτω.

<u>C</u> opy	Ctrl+C
<u>P</u> aste	Ctrl+V
<u>D</u> elete	Del
A <u>dd</u> Variable	Ctrl+A
D <u>elete</u> Variable	Ctrl+D
R <u>ename</u> Variable	Ctrl+R
A <u>lter</u> DataBase	

Σχήμα 5.1: Το Μενού Edit

Copy

Αντιγραφή της επιλεγμένης εγγραφής (ή εγγραφών) στο Πρόχειρο (Clipboard).

Paste

Επικόλληση της εγγραφής (ή εγγραφών) που βρίσκονται στο Πρόχειρο. Οι νέες εγγραφές προστίθενται στο τέλος του Αρχικού Πίνακα Δεδομένων. Οι εγγραφές που βρίσκονται στο Πρόχειρο μπορεί να έχουν προέλθει από οποιαδήποτε εφαρμογή που λειτουργεί σε περιβάλλον Windows, αρκεί να έχουν την κατάλληλη μορφή, δηλαδή τον ίδιο αριθμό πεδίων.

Delete

Διαγραφή της επιλεγμένης εγγραφής (ή εγγραφών) από τον Αρχικό Πίνακα Δεδομένων.

Add Variable

Προσθήκη Χαρακτηριστικού (Μεταβλητή) στον Αρχικό Πίνακα Δεδομένων. Η νέα στήλη προστίθεται αμέσως μετά την τελευταία στήλη του Αρχικού Πίνακα Δεδομένων, ενώ παράλληλα ενημερώνεται τόσο ο Πίνακας των Συναρτήσεων Συμμετοχής της βάσης δεδομένων προσθέτοντας πεδία για τον “νέο” Νέο Άξονα που προκύπτει, όσο και το δέντρο των ορίων των συναρτήσεων συμμετοχής προσθέτοντας ένα “κλαδί”. Για το δέντρο αυτό θα μιλήσουμε παρακάτω.

Delete Variable

Διαγραφή του επιλεγμένου Χαρακτηριστικού (Μεταβλητή) από τον Αρχικό Πίνακα Δεδομένων. Παράλληλα με τη μείωση των Χαρακτηριστικών και κατά συνέπεια και των Νέων Αξόνων, διαγράφονται και από τον Πίνακα των Συναρτήσεων Συμμετοχής της βάσης δεδομένων τα αντίστοιχα με τη διαγραμμένη στήλη πεδία και από το δέντρο των ορίων των συναρτήσεων συμμετοχής ένα “κλαδί”. Επίσης μηδενίζονται όλες οι τιμές των ορίων των συναρτήσεων συμμετοχής δεδομένου ότι

λόγω της διαγραφής μίας στήλης από τα δεδομένα και της μείωσης των κυρίων αξόνων της ΑΚΣ αλλάζει και η τυπική απόκλιση των σημείων στους άξονες.

Rename Variable

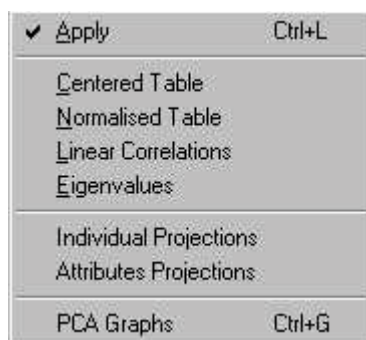
Μετονομασία του επιλεγμένου Χαρακτηριστικού (Μεταβλητή).

Alter Database

Αλλαγή της δομής της βάσης δεδομένων. Ουσιαστικά πρόκειται για αναδιοργάνωση των πεδίων των χαρακτηριστικών του αρχικού πίνακα δεδομένων, όπως είναι ο αριθμός τους και η ονομασία τους. Προκειμένου να γίνει αυτό εμφανίζεται η φόρμα του Σχήματος 5.4.

To Μενού PCA Algorithm

Με το Μενού PCA Algorithm είναι δυνατή η εφαρμογή του αλγορίθμου της Ανάλυσης σε Κύριες Συνιστώσες στα δεδομένα και η λήψη των αποτελεσμάτων που προκύπτουν από αυτόν. Οι εντολές του μενού PCA Algorithm (Σχήμα 5.6) και η αναλυτική επεξήγηση των εντολών δίνεται παρακάτω.



Σχήμα 5.1: Το Μενού PCA Algorithm

Apply

Εφαρμογή του Αλγορίθμου Ανάλυσης σε Κύριες Συνιστώσες στα δεδομένα. Προϋπόθεση είναι η ύπαρξη δεδομένων. Δηλαδή πρέπει να έχει προηγηθεί είτε ανάκτηση δεδομένων από Αρχείο Βάσης Δεδομένων είτε δημιουργία Βάσης Δεδομένων και συμπλήρωση του Αρχικού Πίνακα για τουλάχιστον πέντε Άτομα.

Centered Table

Εμφάνιση του Πίνακα Ανάλυσης και Αποτελεσμάτων με απεικόνιση του Κεντροποιημένου Πίνακα Δεδομένων

Normalized Table

Εμφάνιση του Πίνακα Ανάλυσης και Αποτελεσμάτων με απεικόνιση του Ισοσταθμισμένου Πίνακα Δεδομένων

Linear Correlations

Εμφάνιση του Πίνακα Ανάλυσης και Αποτελεσμάτων με απεικόνιση του Πίνακα Συσχετίσεων των Χαρακτηριστικών.

Eigenvalues

Εμφάνιση του Πίνακα Ανάλυσης και Αποτελεσμάτων με απεικόνιση του Πίνακα Ιδιοτιμών των Νέων Αξόνων.

Individual Projections

Εμφάνιση του Πίνακα Ανάλυσης και Αποτελεσμάτων με απεικόνιση του Πίνακα Προβολών των Ατόμων στους Νέους Άξονες.

Attributes Projections

Εμφάνιση του Πίνακα Ανάλυσης και Αποτελεσμάτων με απεικόνιση του Πίνακα Προβολών των Χαρακτηριστικών στους Νέους Άξονες.

PCA Graphs

Εμφάνιση του Γραφήματος της Ανάλυσης σε Κύριες Συνιστώσες

Οι Πίνακες της Ανάλυσης και Αποτελεσμάτων

Πρόκειται για μία ξεχωριστή Φόρμα που χρησιμεύει στην απεικόνιση των Πινάκων της Ανάλυσης σε Κύριες Συνιστώσες αλλά και σε απεικόνιση διάφορων άλλων αποτελεσμάτων όπως θα δούμε στη συνέχεια. Η μορφή του ενδεικτικά για την περίπτωση του Πίνακα Προβολών των Ατόμων στους Νέους Άξονες φαίνεται στο Σχήμα 5.7.

Analysis and Results Tables								
Projections of Individuals to Principal Axes								
Indiv. \ Ax.	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8
ALFA 145 1.4	.169	-.196	-.035	.090	.075	-.129	.019	.000
AUDI A3 1.6	.343	-.150	.092	-.073	-.006	.056	-.038	.042
BMW 316 Compact	.461	.045	.111	.059	.110	.119	.004	-.004
XSARA 1.4	-.223	.179	-.131	.029	.082	.030	-.043	.016
LANDO 1.3	-.344	.330	-.216	.109	-.027	.048	.042	-.008
SIRION CL	-.892	-.086	.082	-.148	.148	.030	-.002	-.053
PUNTO 60	-.696	.140	-.043	-.085	-.041	.069	-.022	.008
ESCORT 1.4	-.387	.362	.118	.027	-.051	-.072	.016	.071
CIVIC 1.4	-.013	-.270	.052	-.080	-.051	-.059	-.062	-.021
ACCENT 1.3	-.322	-.172	.113	.014	-.054	.115	.046	.024
SEPHIA 1.5	.012	-.131	.016	.077	-.247	.021	.029	-.051
DELTA 1.4	-.239	.286	-.006	.105	.027	-.060	-.056	-.007
MAZDA 323 1.5	.091	-.055	-.004	-.011	-.081	.006	-.052	-.026
C180 CLASSIC	1.042	.275	.131	-.152	.029	-.073	.048	-.050
LANCER 1.3	-.252	-.113	.061	.003	-.025	-.029	.048	-.005
ALMERA 1.4	-.163	-.054	.093	.043	-.061	-.069	-.016	.026
VECTRA 1.6	.477	.008	.223	.210	.049	.008	-.012	-.015
PEUGEOT 206	-.785	.200	.091	-.097	.038	-.031	.034	-.015
MEGANE 1.4	.091	-.191	.133	.049	.125	.014	-.007	.014

Σχήμα 5.1: Η Φόρμα του Πίνακα της Ανάλυσης και Αποτελεσμάτων

Τα δύο κουμπιά πάνω αριστερά έχουν την εξής χρησιμότητα:



Copy

Αντιγραφή των δεδομένων του πίνακα στο Πρόχειρο (Clipboard) για μεταφορά τους σε άλλη εφαρμογή των Windows.



Select Table

Επιλογή απεικόνισης άλλου πίνακα δεδομένων χωρίς το κλείσιμο της Φόρμας των Πινάκων Ανάλυσης και Αποτελεσμάτων. Το κουμπί αυτό αντιστοιχεί στο μενού του Σχήματος 5.8, όπου οι εντολές είναι ανάλογες με αυτές που έχουν περιγραφεί παραπάνω.

Centered Data
Normalised Data
Linear Correlations of Attributes
Eigenvalues of Principal Axes
Projections of Individuals to Principal Axes
Projections of Attributes to Principal Axes

Σχήμα 5.2: Το Μενού Select Table

To Μενού Grouping

Οι εντολές του μενού Grouping (Σχήμα 5.9) και η αναλυτική επεξήγησή τους δίνεται παρακάτω.



Σχήμα 5.1: To Μενού Grouping

Grouping...

Εμφάνιση της Φόρμας του Γραφήματος της Ανάλυσης σε Κύριες Συνιστώσες και συγχρόνως της Εργαλειοθήκης της Ομαδοποίησης. Γι' αυτά θα μιλήσουμε αναλυτικότερα παρακάτω.

MV's Table

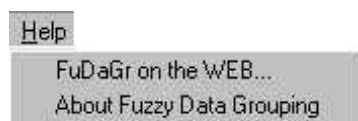
Εμφάνιση του Πίνακα Ανάλυσης και Αποτελεσμάτων με απεικόνιση της μεγαλύτερης από τις τιμές συμμετοχής σε κάποια περιοχή του κάθε ατόμου σε όλους τους κύριους άξονες. Η περιοχή φαίνεται από το χαρακτηριστικό χρώμα των πεδίων.

Membership Functions

Εμφάνιση της Φόρμας διαχείρισης των Συναρτήσεων Συμμετοχής η οποία θα αναλυθεί λεπτομερώς παρακάτω.

To Μενού Help

Οι εντολές του μενού Help (Σχήμα 5.10) και η αναλυτική επεξήγησή τους δίνεται παρακάτω.



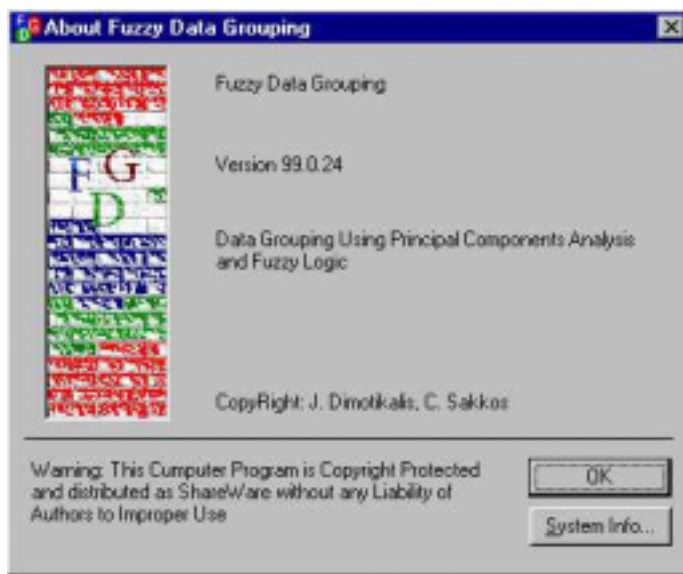
Σχήμα 5.2: To Μενού Help

FuDaGr on the WEB...

Σύνδεση στις σελίδες της εφαρμογής στο INTERNET.

About Fuzzy Data Grouping

Εμφάνιση της Φόρμας About Fuzzy Data Grouping (Πληροφορίες για το Fuzzy Data Grouping) η οποία είναι αυτή του Σχήματος 5.11.



Σχήμα 5.3: Η Φόρμα About Fuzzy Data Grouping

5.4.4 Η Εργαλειοθήκη

Τα κουμπιά της εργαλειοθήκης της Αρχικής Φόρμας είναι τα εξής:

	<i>New DataBase File</i> : Παραπέμπει στο μενού File→New
	<i>Open DataBase File</i> : Παραπέμπει στο μενού File→Open...
	<i>Save As...</i> : Παραπέμπει στο μενού File→Save As...
	<i>Print</i> : Εμφάνιση του μενού του Σχήματος 5.3. Οι εντολές είναι αυτές που έχουν περιγραφεί παραπάνω.
	<i>Copy Record(s)</i> : Παραπέμπει στο μενού Edit→Copy
	<i>Paste Record(s)</i> : Παραπέμπει στο μενού Edit→Paste
	<i>Delete Record(s)</i> : Παραπέμπει στο μενού Edit→Delete
	<i>Add Attribute (Variable)</i> : Παραπέμπει στο μενού Edit→Add Attribute (Variable)
	<i>Delete Attribute (Variable)</i> : Παραπέμπει στο μενού Edit→Delete
	<i>Rename Attribute (Variable)</i> : Παραπέμπει στο μενού Edit→Rename Attribute (Variable)
	<i>PCA Solve</i> : Παραπέμπει στο μενού PCA Algorithm→Apply
	<i>PCA Graph</i> : Παραπέμπει στο μενού PCA Algorithm→PCA Graph



Membership Function Graphs: Παραπέμπει στο μενού Grouping→Membership Functions

Grouping...: Παραπέμπει στο μενού Grouping→Grouping...

Help: Εμφάνιση της φόρμας ταυτότητας του Fuzzy Data Grouping

FuDaGr on the WEB...: Παραπέμπει στο μενού Help→FuDaGr on the WEB...

5.4.5 Η Γραμμή Κατάστασης

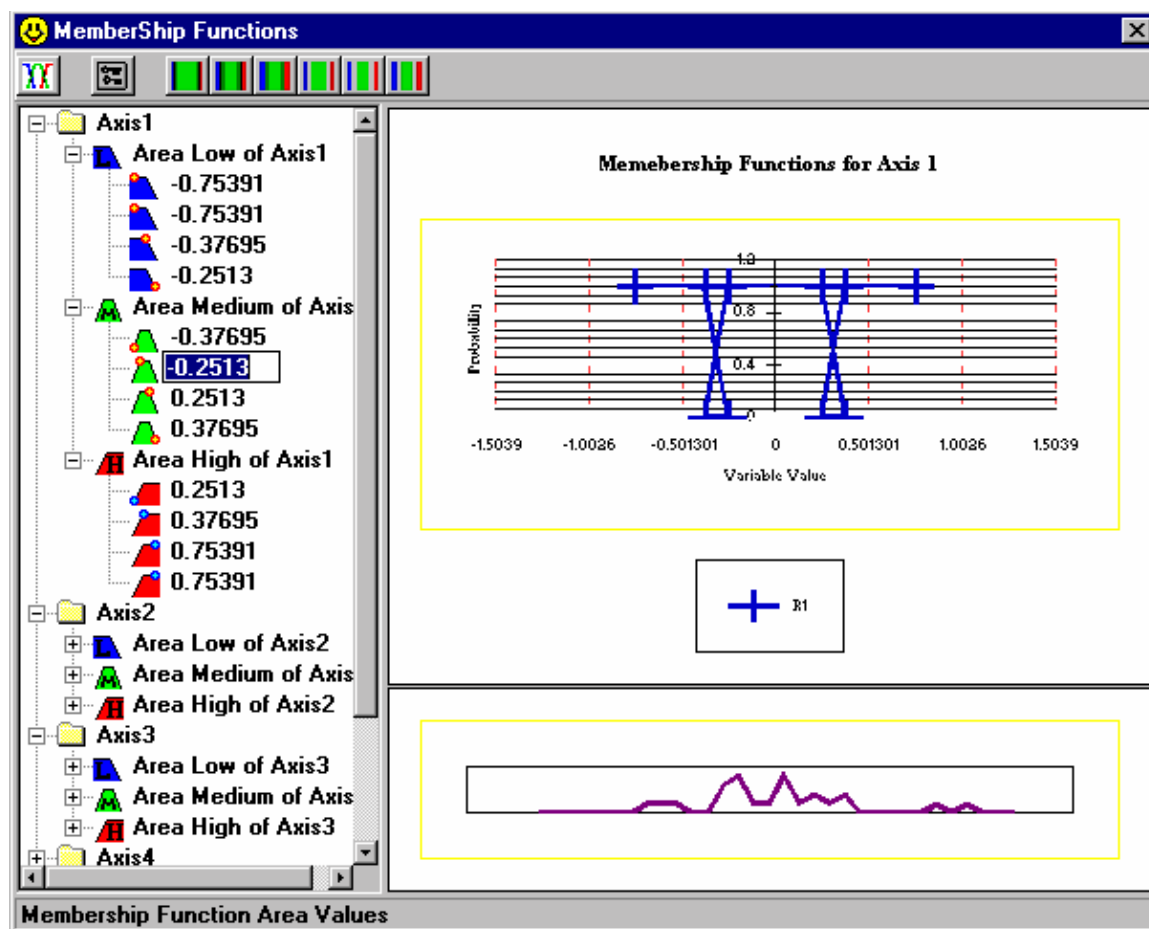
Η γραμμή κατάστασης στο κάτω μέρος της οθόνης (Σχήμα 5.12) παρέχει πληροφορίες και διευκρινήσεις ανάλογα με τη θέση του δείκτη του ποντικού στην οθόνη ή την επιλογή αντικειμένων.

Data Table: Click to Edit	2/2/2000	4:52 πμ
---------------------------	----------	---------

Σχήμα 5.4: Η Γραμμή Κατάστασης της αρχικής φόρμας

5.5 Η ΦΟΡΜΑ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ

Από τη φόρμα αυτή είναι δυνατή η επιλογή ή η αλλαγή των ορίων των περιοχών για κάθε κύριο άξονα. Δηλαδή η επιλογή ή αλλαγή των “κανόνων”. Η μορφή της φόρμας αυτής φαίνεται στο Σχήμα 5.13.



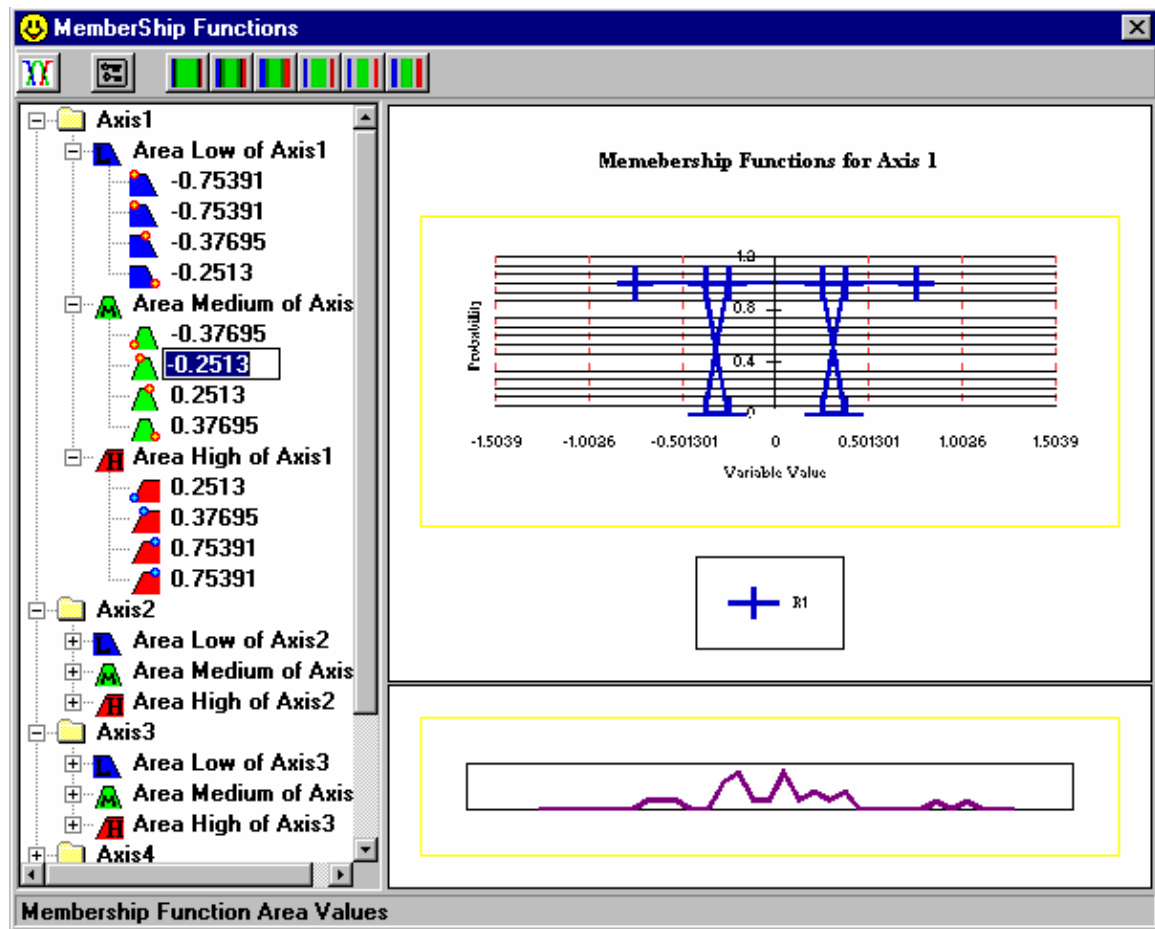
Σχήμα 5.5: Η Φόρμα των Συναρτήσεων Συμμετοχής

5.5.1 Το Δέντρο των Ορίων των Συναρτήσεων Συμμετοχής

Στο αριστερό μέρος της φόρμας φαίνεται η δενδρική μορφή των ορίων των συναρτήσεων συμμετοχής. Το δένδρο αυτό είναι το ίδιο με αυτό της αρχικής φόρμας με μία διαφορά. Είναι δυνατή η αλλαγή των τιμών των ορίων, κάνοντας κλικ πάνω σε μία ήδη επιλεγμένη τιμή. Στο Σχήμα 5.13 φαίνεται η διαδικασία αλλαγής του πάνω αριστερού ορίου της μεσαίας περιοχής του πρώτου κύριου άξονα. Ιδιαίτερη προσοχή απαιτείται κατά τη διαδικασία αλλαγής των τιμών διότι δεν υπάρχει η επιλογή της αναίρεσης. Ο πίνακας των συναρτήσεων συμμετοχής της βάσης δεδομένων ("Principal_Axes_MMF") ενημερώνεται αυτόματα με κάθε αλλαγή τιμής.

5.5 Η ΦΟΡΜΑ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ

Από τη φόρμα αυτή είναι δυνατή η επιλογή ή η αλλαγή των ορίων των περιοχών για κάθε κύριο άξονα. Δηλαδή η επιλογή ή αλλαγή των “κανόνων”. Η μορφή της φόρμας αυτής φαίνεται στο Σχήμα 5.13.



Σχήμα 5.13: Η Φόρμα των Συναρτήσεων Συμμετοχής

5.5.1 Το Δέντρο των Ορίων των Συναρτήσεων Συμμετοχής

Στο αριστερό μέρος της φόρμας φαίνεται η δενδρική μορφή των ορίων των συναρτήσεων συμμετοχής. Το δένδρο αυτό είναι το ίδιο με αυτό της αρχικής φόρμας με μία διαφορά. Είναι δυνατή η αλλαγή των τιμών των ορίων, κάνοντας κλικ πάνω σε μία ήδη επιλεγμένη τιμή. Στο Σχήμα 5.13 φαίνεται η διαδικασία αλλαγής του πάνω

αριστερού ορίου της μεσαίας περιοχής του πρώτου κύριου άξονα. Ιδιαίτερη προσοχή απαιτείται κατά τη διαδικασία αλλαγής των τιμών διότι δεν υπάρχει η επιλογή της αναίρεσης. Ο πίνακας των συναρτήσεων συμμετοχής της βάσης δεδομένων (“Principal_Axes_MMF”) ενημερώνεται αυτόματα με κάθε αλλαγή τιμής.

5.5.2 Το Γράφημα των Συναρτήσεων Συμμετοχής

Στο δεξιό μέρος της φόρμας φαίνεται η μορφή των συναρτήσεων συμμετοχής. Στο Σχήμα 5.13 φαίνονται οι συναρτήσεις συμμετοχής του πρώτου κύριου άξονα. Για απεικόνιση των συναρτήσεων των άλλων κυρίων αξόνων υπάρχει κουμπί στην εργαλειοθήκη την οποία θα αναλύσουμε παρακάτω.

5.5.3 Το Γράφημα της Κατανομής των Ατόμων

Κάτω από το γράφημα των συναρτήσεων συμμετοχής υπάρχει ένα διάγραμμα που αποτυπώνει την κατανομή των ατόμων(σημείων) στον άξονα του οποίου φαίνονται οι συναρτήσεις συμμετοχής. Η κλίμακα σ’ αυτό το γράφημα είναι η ίδια με την κλίμακα στον οριζόντιο άξονα του επάνω γραφήματος. Η κατανομή των σημείων σε κάποιο κύριο άξονα μπορεί να αποδειχθεί χρήσιμη στη σωστή επιλογή του κανόνα που θα εφαρμοστεί στον άξονα, εφ’ όσον θα είναι γνωστό εκ των προτέρων πόσα περίπου σημεία θα υπάρχουν σε κάθε μία από τις τρεις περιοχές.

5.5.4 Η Εργαλειοθήκη

Τα κουμπιά της εργαλειοθήκης της Φόρμας των Συναρτήσεων Συμμετοχής είναι τα εξής:



Redraw Membership Graph: Επανασχεδίαση ή σχεδίαση του γραφήματος των συναρτήσεων συμμετοχής για τον επιλεγμένο από το δέντρο άξονα. Το ίδιο αποτέλεσμα έχει και το διπλό κλικ πάνω στο γράφημα.



Soft Symmetrical Rule, Normal Symmetrical Rule, Hard Symmetrical

Rule: Πρόκειται για τους τρεις συμμετρικούς κανόνες για τους οποίους μιλήσαμε στο Κεφάλαιο 4.



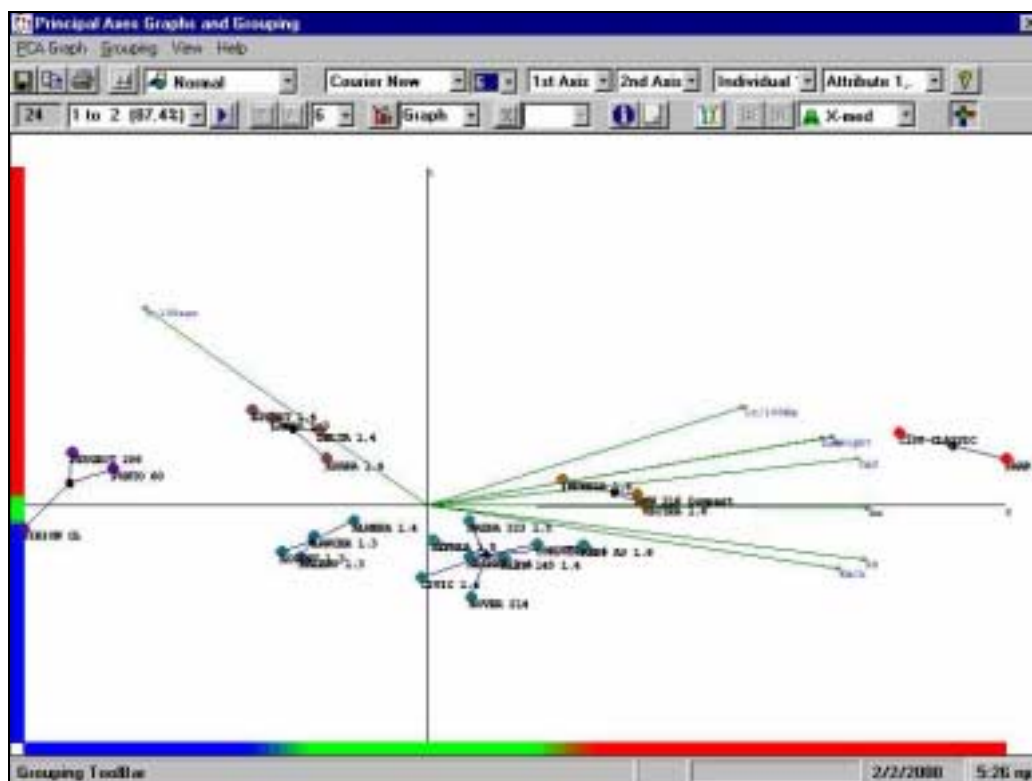
Soft no-Symmetrical Rule, Normal no-Symmetrical Rule, Hard no-Symmetrical Rule: Πρόκειται για τους τρεις μη συμμετρικούς κανόνες για τους οποίους μιλήσαμε στο Κεφάλαιο 4.



Apply Rule to All Axes: Εφαρμογή του ίδιου επιλεγμένου κανόνα σε όλους τους κύριους άξονες.

5.6 Η ΦΟΡΜΑ ΤΟΥ ΓΡΑΦΗΜΑΤΟΣ ΤΗΣ ΑΚΣ ΚΑΙ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ

Στο Σχήμα 5.14 φαίνεται η μορφή της φόρμας του Γραφήματος της Ανάλυσης σε Κύριες Συνιστώσες. Εκτός από τα άτομα πάνω στο γράφημα φαίνονται και οι έξι ομάδες που έχουν σχηματιστεί από την εφαρμογή του αλγορίθμου της ομαδοποίησης.



Σχήμα 5.14: Η Φόρμα του Γραφήματος της ΑΚΣ και της Ομαδοποίησης

Θα αναλύσουμε τη φόρμα εξετάζοντας τις δυνατότητες που παρέχει το κάθε αντικείμενό της στο χρήστη. Δηλαδή τα μενού, τις δύο εργαλειοθήκες που φαίνονται και τη γραμμή κατάστασης.

5.6.1 Τα Μενού

Το Μενού PCA Graph (Γράφημα Ανάλυσης σε Κύριες Συνιστώσες)

Με το μενού αυτού ο χρήστης έχει τη δυνατότητα να αποθηκεύσει, να εκτυπώσει ή να μεταφέρει το γράφημα σε κάποια άλλη εφαρμογή Windows. Το μενού φαίνεται στο Σχήμα 5.15 ενώ οι εντολές του εξηγούνται παρακάτω.

Save Chart	Ctrl+S
Copy Chart to Clipboard	Ctrl+C
Print...	Ctrl+P
Help	F1

Σχήμα 5.15: Το Μενού PCA Graph

Save Chart

Αποθήκευση του γραφήματος ΑΚΣ σε αρχείο τύπου .bmp.

Copy Chart to Clipboard

Αντιγραφή του γραφήματος ΑΚΣ στο πρόχειρο. Με αυτό τον τρόπο μπορεί να μεταφερθεί το γράφημα σε άλλες εφαρμογές των Windows.

Print...

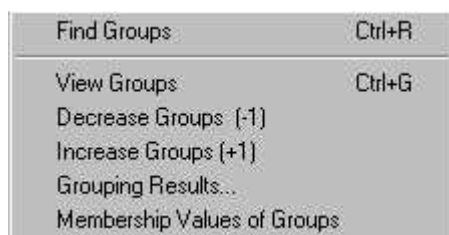
Εκτύπωση του γραφήματος ΑΚΣ.

Help

Εμφάνιση της φόρμας ταυτότητας της εφαρμογής

Το Μενού Grouping

Μέσο του μενού Grouping ο χρήστης μπορεί να ελέγχει τον αλγόριθμο της ομαδοποίησης και να λαμβάνει τα αποτελέσματά του. Το μενού φαίνεται στο Σχήμα 5.16 ενώ οι εντολές του εξηγούνται παρακάτω.



Σχήμα 5.16: Το Μενού Grouping

Find Groups

Εφαρμογή του αλγορίθμου της ομαδοποίησης. Μετά το πέρας της διαδικασίας εμφανίζεται μήνυμα ολοκλήρωσης του αλγορίθμου που αναφέρει τον αριθμό των ομάδων που δημιουργήθηκαν. Ως κριτήριο τερματισμού του αλγορίθμου θεωρείται μόνο ο αριθμός των ομάδων. Το κριτήριο της μέγιστης απόστασης μεταξύ ατόμων της ίδιας ομάδας ενεργοποιείται από την εργαλειοθήκη ή από τη Φόρμα “Options” (“Επιλογές”), την οποία θα αναλύσουμε παρακάτω.

View Groups

Εμφάνιση των ομάδων εφ’ όσον έχει εφαρμοστεί ο αλγόριθμος της ομαδοποίησης είτε στο γράφημα είτε σε πίνακα. Η επιλογή αυτή γίνεται από την εργαλειοθήκη όπως θα δούμε παρακάτω.

Decrease Groups (-1)

Περιορισμός των ήδη υπαρχόντων ομάδων κατά μία. Οι ομάδες δεν μπορούν να περιοριστούν στην περίπτωση που είναι δύο, καθώς η ομαδοποίηση όλων των ατόμων στην ίδια ομάδα είναι εκφυλισμένη μορφή ομαδοποίησης και δεν έχει νόημα.

Increase Groups (+1)

Αύξηση των ήδη υπαρχόντων ομάδων κατά μία. Οι ομάδες δεν μπορούν να αυξηθούν στην περίπτωση που είναι N-1 (π.χ. αν οι ομάδες είναι επτά και τα άτομα οκτώ), καθώς η ομαδοποίηση των ατόμων σε διαφορετική ομάδα το κάθε ένα είναι επίσης εκφυλισμένη μορφή ομαδοποίησης και δεν έχει νόημα.

Grouping Results...

Εμφάνιση της φόρμας των αναλυτικών αποτελεσμάτων της ομαδοποίησης η οποία θα αναλυθεί παρακάτω.

Membership Values of Groups

Εμφάνιση της φόρμας των πινάκων ανάλυσης και αποτελεσμάτων (“Analysis and Results Table”) με απεικόνιση των τιμών συμμετοχής της κάθε ομάδας στις τρεις περιοχές των δύο πρώτων κυρίων αξόνων.

To Menu View

Με το μενού View ο χρήστης έχει τη δυνατότητα να επιλέξει τη μορφή του γραφήματος ΑΚΣ και να ρυθμίσει διάφορες λεπτομέρειες που αφορούν τον αλγόριθμο της ομαδοποίησης. Το μενού φαίνεται στο Σχήμα 5.17 ενώ οι εντολές του εξηγούνται παρακάτω.



Σχήμα 5.17: Το Μενού View

Chart Title

Εμφάνιση ή απόκρυψη τίτλου στο γράφημα της ΑΚΣ. Ο τίτλος αναφέρει τους δύο κύριους άξονες που χρησιμοποιούνται στο γράφημα και το αθροιστικό ποσοστό της πληροφορίας του αρχικού πίνακα δεδομένων που αντιπροσωπεύουν.

Fuzzy Rulers

Εμφάνιση ή απόκρυψη των Fuzzy Ruler. Πρόκειται για δύο οδηγούς (ένα για κάθε κύριο άξονα του γραφήματος) που απεικονίζουν χρωματικά τα όρια των τριών περιοχών για κάθε άξονα. Έτσι στα “καθαρά” χρώματα μπλε, πράσινο και κόκκινο αντιστοιχούν οι σίγουρες περιοχές “Χαμηλή”, “Μεσαία” και “Υψηλή” αντίστοιχα ενώ οι πιο σκούροι χρωματισμοί ανάμεσα αντιστοιχούν στις ασαφείς περιοχές.

Normal

Επανασχεδιασμός του γραφήματος ΑΚΣ με ίδια κλίμακα για κάθε κύριο άξονα.

Extended

Επανασχεδιασμός του γραφήματος ΑΚΣ με διαφορετική κλίμακα για κάθε κύριο άξονα. Η κλίμακα είναι τέτοια ώστε μέσα στα όρια του γραφήματος να περιέχονται τόσο όλα τα άτομα όσο και όλα τα διανύσματα των χαρακτηριστικών.

All Individuals

Επανασχεδιασμός του γραφήματος ΑΚΣ με τέτοια κλίμακα σε κάθε κύριο άξονα ώστε στα όρια του γραφήματος να περιέχονται όλα τα άτομα. Κάποια από τα διανύσματα χαρακτηριστικών μπορεί να μην φαίνονται.

User Zoom

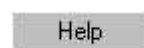
Επανασχεδιασμός του γραφήματος ΑΚΣ με μεγέθυνση στην περιοχή που ορίζει ο χρήστης. Η επιλογή της περιοχής γίνεται εύκολα με σύρσιμο του ποντικιού όπως σε πολλές εφαρμογές των Windows.

Toolbars → Grouping

Εμφάνιση ή απόκρυψη της εργαλειοθήκης της ομαδοποίησης. Η εργαλειοθήκη του γραφήματος ΑΚΣ είναι σταθερή και θεωρείται ότι δεν υπάρχει λόγος απόκρυψής της.

Options...

Εμφάνιση της φόρμας “Options” για τη ρύθμιση των παραμέτρων της ομαδοποίησης. Τη φόρμα αυτή θα την αναλύσουμε παρακάτω.

Το Μενού Help

Σχήμα 5.18

Με την εντολή *Help* (Σχήμα 5.18) εμφανίζεται φόρμα ταυτότητας της εφαρμογής.

5.6.2 Οι Εργαλειοθήκες**Η Εργαλειοθήκη του γραφήματος ΑΚΣ**

Save Chart (.bmp):* Παραπέμπει στο μενού PCA Graph → Save Chart



Copy Chart to Clipboard: Παραπέμπει στο μενού PCA Graph → Copy Chart to Clipboard



Print Graph: Παραπέμπει στο μενού PCA Graph → Print...



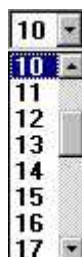
Show Grouping Toolbar: Παραπέμπει στο μενού View → Toolbars → Grouping



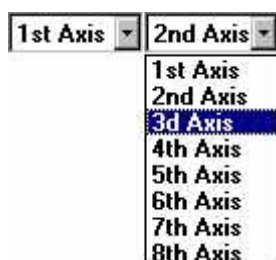
Graph View: Παραπέμπει στο μενού View → Normal, Extended, All Individuals και User Zoom



Graph Font Name: Επιλογή γραμματοσειράς για το γράφημα ΑΚΣ

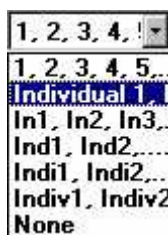


Graph Font Size: Επιλογή του μεγέθους της γραμματοσειράς για το γράφημα ΑΚΣ

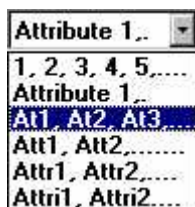


Horizontal X Axis, Vertical Y Axis: Επιλογή του κύριου άξονα που θα αναπαρίσταται στον X ή στον Y άξονα του γραφήματος.

Οι δύο άξονες πρέπει να είναι διαφορετικοί γιατί δεν έχει νόημα σε ένα επίπεδο γράφημα ΑΚΣ ο κάθετος και ο οριζόντιος άξονας να είναι οι ίδιοι κύριοι άξονες.



Individuals View Pattern: Επιλογή της μορφής εμφάνισης των ονομάτων των ατόμων στο γράφημα. Ουσιαστικά η επιλογή είναι πόσα γράμματα από το όνομα θα εμφανίζονται. Βοηθάει σε περιπτώσεις που τα ονόματα των ατόμων έχουν πολλούς χαρακτήρες ή τα άτομα είναι πολλά με αποτέλεσμα να μειώνεται η ευκρίνεια του γραφήματος.



Attributes View Pattern: Επιλογή της μορφής εμφάνισης των ονομάτων των χαρακτηριστικών. Έχει ανάλογη χρήση με το αντικείμενο “Individuals View Pattern”.



Help: Παραπέμπει στο μενού Help→Help

Η Εργαλειοθήκη της Ομαδοποίησης

24

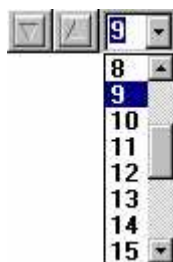
Nr. Of Individuals: Σ' αυτό το πεδίο εμφανίζεται ο αριθμός των ατόμων, δηλαδή ο αριθμός των γραμμών του αρχικού πίνακα δεδομένων.

1 to 2 (87.4%)
1 to 1 (75.4%)
1 to 2 (87.4%)
1 to 3 (93.6%)
1 to 4 (96.1%)
1 to 5 (98.3%)
1 to 6 (99.3%)
1 to 7 (99.7%)
1 to 8 (100.%)

Axes Used: Από αυτό το αντικείμενο η χρήστης μπορεί να επιλέξει τον αριθμό των αξόνων στους οποίους θα γίνει η ομαδοποίηση. Το ποσοστό που αναγράφεται σε παρένθεση αντιστοιχεί στο αθροιστικό ποσοστό της πληροφορίας που αναπαρίσταται στο σύνολο των επιλεγμένων αξόνων. Αν επιλεγούν όλοι οι άξονες έχουμε το 100% της πληροφορίας, κάτι όμως που έχει αρνητική επίδραση στο χρόνο εφαρμογής του αλγορίθμου της ομαδοποίησης. Επίσης οποιαδήποτε άλλη επιλογή εκτός της χρήσης των δύο πρώτων κύριων αξόνων έχει ως αποτέλεσμα την αδυναμία της σωστής εμφάνισης των ομάδων στο γράφημα αφού έχουμε προβολή κ-διάστατης ομαδοποίησης σε διδιάστατο γράφημα, όπου $2 < k \leq K$ (κ επιλεγμένοι άξονες από το σύνολο των K κύριων αξόνων).



Start Grouping Process: Παραπέμπει στο μενού Grouping→ Find Groups



Decrease Groups (-1): Παραπέμπει στο μενού Grouping→ Decrease Groups (-1)

Increase Groups (+1): Παραπέμπει στο μενού Grouping→ Increase Groups (+1)

Customize Nr. of Groups: Από αυτό το αντικείμενο ο χρήστης επιλέγει τον ακριβή αριθμό των ομάδων που επιθυμεί. Ο επιτρεπτός αριθμός των ομάδων είναι από 2 έως N-1 όπου N ο αριθμός των ατόμων.



View Current Groups: Παραπέμπει στο μενού Grouping→ View Groups

Group View: Από αυτό το αντικείμενο ο χρήστης επιλέγει τη μορφή εμφάνισης των ομάδων μετά το πάτημα του κουμπιού

“View Current Groups”. Η εμφάνιση μπορεί να γίνει είτε στο γράφημα είτε στον πίνακα της φόρμας «Analysis and Results Table».



Activate Max. Radius Criterion: Ενεργοποίηση ή απενεργοποίηση του κριτηρίου τερματισμού του αλγορίθμου της ομαδοποίησης με βάση τη μέγιστη απόσταση μεταξύ ατόμων της ίδιας ομάδας.

Max. Radius: Επιλογή του ορίου της μέγιστης απόστασης μεταξύ ατόμων της ίδιας ομάδας για το κριτήριο τερματισμού του αλγορίθμου της ομαδοποίησης. Η απόσταση εμφανίζεται ως ποσοστό επί του κύριου άξονα που εμφανίζεται στον Χ άξονα του γραφήματος. Με την ενεργοποίηση του κριτηρίου από το κουμπί “Activate Max. Radius” στο πεδίο του αντικειμένου εμφανίζεται η τωρινή μέγιστη απόσταση εφ’ όσον έχει γίνει ομαδοποίηση. Ο χρήστης μπορεί να επιλέξει σε ποσοστό ανάμεσα στο 1% και το 50%. Για παράδειγμα επιλογή ποσοστού 15% σημαίνει τερματισμό του αλγορίθμου της ομαδοποίησης εάν η μέγιστη απόσταση μεταξύ ατόμων της ίδιας ομάδας πάει να ξεπεράσει το 15% του μήκους του άξονα Χ. Με αυτό τον τρόπο ο χρήστης μπορεί να ελέγξει την “συνοχή” ή “ομοιογένεια” των ομάδων.



Grouping Results...: Παραπέμπει στο μενού Grouping→ Grouping Results...



Membership Values of Groups: Παραπέμπει στο μενού Grouping→Membership Values of Groups.



Membership Functions Graphs: Εμφάνιση της φόρμας των Συναρτήσεων Συμμετοχής.



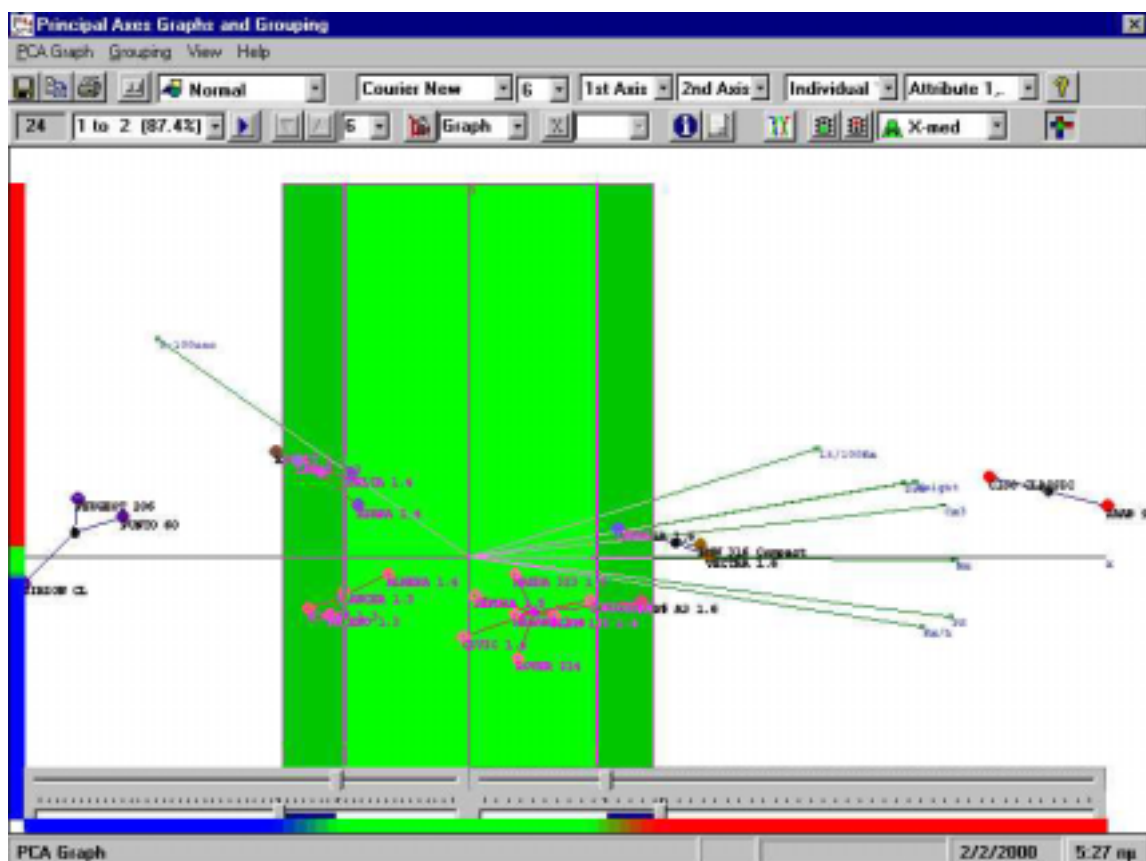
Accept New Area Definition: Αποδοχή των νέων ορίων των περιοχών που καθορίστηκαν με χρήση του αντικειμένου “Define Fuzzy Area”.

Reject New Area Definition: Απόρριψη των νέων ορίων των περιοχών που καθορίστηκαν με χρήση του αντικειμένου “Define Fuzzy Area”.

Define Fuzzy Area: Με αυτό το αντικείμενο ο χρήστης έχει τη δυνατότητα καθορισμού ή αλλαγής των ορίων των συναρτήσεων συμμετοχής για τους δύο κύριους άξονες του γραφήματος ΑΚΣ. Βοηθάει τον χρήστη ώστε να έχει μία εποπτική εικόνα των ορίων ώστε ο καθορισμός τους να γίνεται πιο εύκολα δεδομένου ότι μπορεί ο χρήστης να διαπιστώσει άμεσα πόσα και ποια άτομα βρίσκονται στις σίγουρες ή ασαφείς περιοχές. Στο Σχήμα 5.19 φαίνεται η μορφή της φόρμας όταν επιχειρείται αλλαγή στα όρια των Συναρτήσεων Συμμετοχής στη Μεσαία περιοχή του X άξονα.



Show Fuzzy Rulers: Παραπέμπει στο μενού View→Fuzzy Rulers

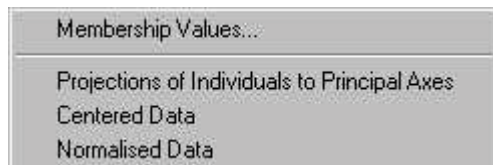


Σχήμα 5.19: Η φόρμα της ΑΚΣ και της ομαδοποίησης κατά τη διαδικασία αλλαγής των ορίων των Συναρτήσεων Συμμετοχής στη Μεσαία περιοχή του Χ άξονα

5.6.3 Pop Up Μενού και Γραμμή Κατάστασης (Status Bar)

Συνήθη σε πολλές εφαρμογές των Windows είναι τα λεγόμενα Pop Up Menus. Πρόκειται για μενού που εμφανίζονται συνήθως με το πάτημα του δεξιού πλήκτρου του ποντικιού στη θέση που το πλήκτρο πατήθηκε και δίνουν διάφορες επιλογές στο χρήστη ανάλογα με τη θέση του δείκτη του ποντικιού και το αντικείμενο που βρίσκεται από κάτω. Τα Pop Up μενού αυξάνουν την ευχρηστία των εφαρμογών. Διάφορα Pop Up μενού εμφανίζονται στην εφαρμογή Fuzzy Data Grouping ανάλογα με τη θέση του δείκτη του ποντικιού πάνω στο γράφημα της ΑΚΣ. Ανάλογα, διάφορα μηνύματα εμφανίζονται και στη γραμμή κατάστασης η οποία βρίσκεται στο κάτω μέρος της οθόνης και παρέχει χρήσιμες πληροφορίες. Για την εμφάνιση των μηνυμάτων στην γραμμή κατάστασης δεν χρειάζεται πάτημα πλήκτρου του ποντικιού αλλά αρκεί ο δείκτης του ποντικιού να βρίσκεται σε κάποιο συγκεκριμένο σημείο του γραφήματος.

- Εάν το πάτημα του δεξιού πλήκτρου του ποντικιού γίνει όταν ο δείκτης του ποντικιού είναι κοντά σε ένα σημείο που αναπαριστά άτομο τότε εμφανίζεται το μενού του Σχήματος 5.20. Οι εντολές του εξηγούνται παρακάτω.



Σχήμα 5.20: Pop Up μενού στο Γράφημα της ΑΚΣ

Membership Values...

Εμφάνιση της φόρμας των πινάκων ανάλυσης και αποτελεσμάτων (“Analysis and Results Table”) με απεικόνιση των τιμών συμμετοχής του ατόμου σε κάθε περιοχή όλων των κύριων αξόνων.

Projections of Individuals to Principal Axes

Αντιστοιχεί στο μενού της αρχικής φόρμας PCA Algorithm→Individual Projections.

Centered Data

Αντιστοιχεί στο μενού της αρχικής φόρμας PCA Algorithm→Centered Table.

Normalized Data

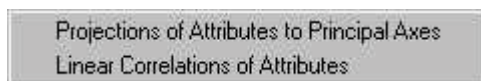
Αντιστοιχεί στο μενού της αρχικής φόρμας PCA Algorithm→Normalized Table.

Το μήνυμα που εμφανίζεται στη γραμμή κατάστασης είναι:

“όνομα ατόμου”(αρ. γραμμής): $Lx = \mu_{i1}^L$ $Mx = \mu_{i1}^M$ $Hx = \mu_{i1}^H$, $Ly = \mu_{i2}^L$ $My = \mu_{i2}^M$ $Hy = \mu_{i2}^H$

δηλαδή εμφανίζονται οι τιμές συμμετοχής του ατόμου στις τρεις περιοχές των δύο πρώτων κυρίων αξόνων.

- Εάν το πάτημα του δεξιού πλήκτρου του ποντικιού γίνει όταν ο δείκτης του ποντικιού είναι κοντά σε ένα σημείο που αναπαριστά χαρακτηριστικό τότε εμφανίζεται το μενού του Σχήματος 5.21. Οι εντολές εξηγούνται παρακάτω.



Σχήμα 5.21: Pop Up μενού στο Γράφημα της ΑΚΣ

Projections of Attributes to Principal Axes

Αντιστοιχεί στο μενού της αρχικής φόρμας PCA Algorithm → Projections of Attributes to Principal Axes.

Linear Correlations of Attributes

Αντιστοιχεί στο μενού της αρχικής φόρμας PCA Algorithm → Linear Correlations of Attributes.

Το μήνυμα που εμφανίζεται στη γραμμή κατάστασης είναι:

Attribute “όνομα χαρακτηριστικού”: X% on Axis 1, Y% on Axis 2

δηλαδή εμφανίζεται το όνομα του χαρακτηριστικού και σε τι ποσοστό αναπαρίσταται στους δύο πρώτους κύριους άξονες. Σε περίπτωση που το χαρακτηριστικό έχει αρνητική προβολή σε κάποιο άξονα εμφανίζεται το διακριτικό “(Negative)”.

- Εάν το πάτημα του δεξιού πλήκτρου του ποντικιού γίνει όταν ο δείκτης του ποντικιού είναι κοντά σε ένα άξονα του γραφήματος τότε εμφανίζεται το μενού του Σχήματος 5.22. Οι εντολές εξηγούνται παρακάτω.



Σχήμα 5.22: Pop Up μενού στο Γράφημα της ΑΚΣ

Membership Functions

Αντιστοιχεί στο μενού της αρχικής φόρμας Grouping→Membership Functions.

Eigenvalues of Principal Axes

Αντιστοιχεί στο μενού της αρχικής φόρμας PCA Algorithm→Eigenvalues.

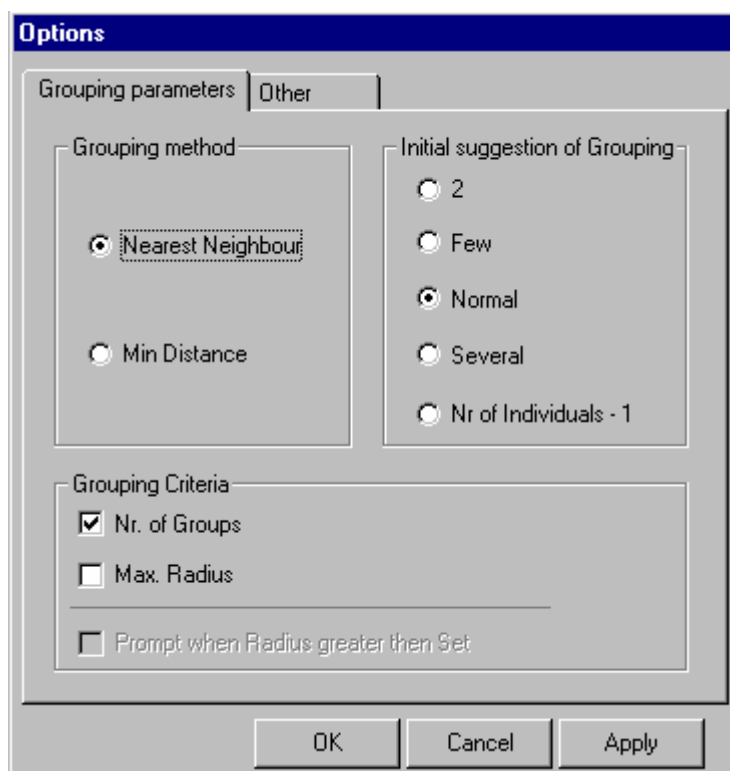
Το μήνυμα που εμφανίζεται στη γραμμή κατάστασης είναι:

Axis X (ή Axis Y): Attribute A (A%), Attribute B (B%),...

δηλαδή εμφανίζεται ο άξονας στον οποίο ο δείκτης του ποντικιού βρίσκεται κοντά και τα χαρακτηριστικά που αναπαρίστανται στον άξονα με το ποσοστό αναπαράστασής τους σε παρένθεση.

5.7 Η ΦΟΡΜΑ OPTIONS

Από τη φόρμα Options είναι δυνατή η αλλαγή παραμέτρων που αφορούν την ομαδοποίηση αλλά και άλλων ρυθμίσεων. Υπάρχουν δύο ομάδες επιλογών. Στο Σχήμα 5.23 φαίνονται οι εντολές των παραμέτρων της ομαδοποίησης.



Σχήμα 5.23: Οι παράμετροι της Ομαδοποίησης

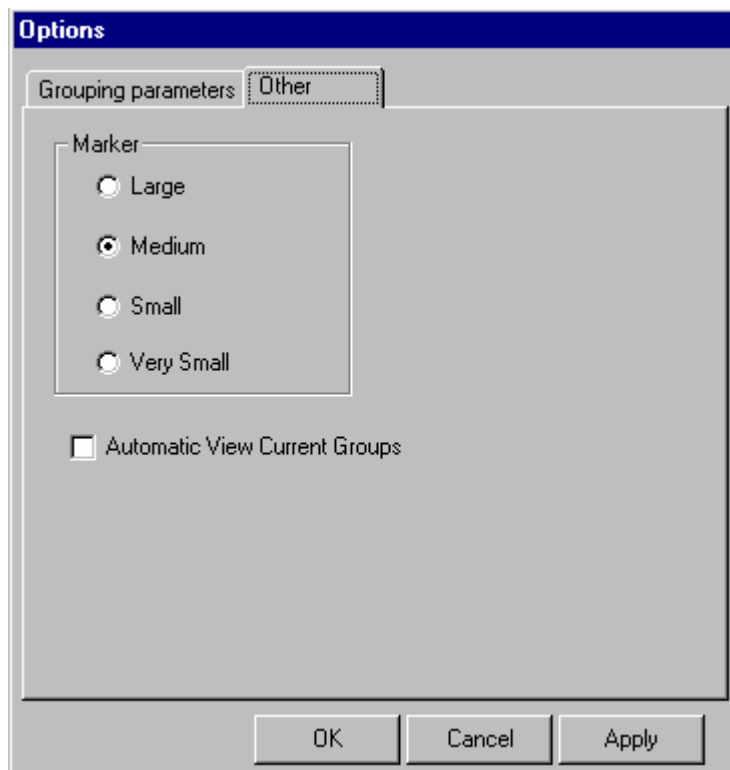
Στο πλαίσιο “Grouping Method” δίνεται η δυνατότητα επιλογής ανάμεσα στις δύο εκδοχές του αλγορίθμου της ομαδοποίησης. Η εκδοχή Nearest Neighbour είναι η πλήρης εφαρμογή του αλγορίθμου ενώ η επιλογή Min Distance είναι η εκδοχή που δεν γίνεται η εξέταση του κοντινότερου του κάθε ατόμου πριν προσχωρήσει σε μία ομάδα. Η εκδοχή Min Distance είναι ελαφρώς γρηγορότερη όμως θα πρέπει να προτιμάται μόνο σε περιπτώσεις πυκνού νέφους σημείων οπότε και τα αποτελέσματα δεν διαφέρουν σημαντικά από την εκδοχή Nearest Neighbour.

Στο πλαίσιο “Initial Suggestion of Grouping” είναι δυνατή η ρύθμιση του κριτηρίου τερματισμού του αλγορίθμου της ομαδοποίησης που αφορά τον αριθμό των ομάδων. Έτσι οι εντολές “2” και “Nr. of Individuals-1” αντιστοιχούν σε τερματισμό του αλγορίθμου όταν δημιουργηθούν δύο ομάδες ή όταν δημιουργηθεί η πρώτη ομάδα αντίστοιχα. Οι ενδιάμεσες εντολές “Few”, “Normal” και “Several” αντιστοιχούν σε αριθμούς ομάδων σε συνάρτηση με τον αριθμό των ατόμων. Οι συναρτήσεις έχουν επιλεγεί αυθαίρετα και έχουν αναφερθεί στο Κεφάλαιο 3.

Στο πλαίσιο “Grouping Criteria” επιλέγονται τα κριτήρια τερματισμού της ομαδοποίησης. Τα κριτήρια όπως είναι ήδη γνωστό από το Κεφάλαιο 3 είναι ο αριθμός των ομάδων (“Nr. of Groups”) και η μέγιστη απόσταση μεταξύ ατόμων της ομάδας (“Max. Radius”). Είναι δυνατή η επιλογή ενός μόνο από τα κριτήρια, και των δύο μαζί ή κανενός. Στην περίπτωση επιλογής και των δύο κριτηρίων ο αλγόριθμος της ομαδοποίησης σταματάει όταν τουλάχιστον ένα από τα κριτήρια ικανοποιείται. Στην περίπτωση που δεν επιλεγεί κανένα ο αλγόριθμος σταματάει όταν οι ομάδες γίνουν δύο. Όταν έχει επιλεγεί το κριτήριο “Max. Radius” είναι δυνατή και η επιλογή “Prompt when Radius greater than Set”. Η επιλογή αυτής της παραμέτρου έχει σαν αποτέλεσμα την εμφάνιση μηνύματος προειδοποίησης στην οθόνη, κατά τη διαδικασία περιορισμού των ομάδων, όταν η μέγιστη απόσταση μεταξύ ατόμων της ίδιας ομάδας είναι μεγαλύτερη από αυτή που αναγράφεται στο αντικείμενο Max. Radius της φόρμας του γραφήματος της ΑΚΣ.

Στο Σχήμα 5.24 φαίνονται οι υπόλοιπες ρυθμίσεις της φόρμας (“Other”). Στο πλαίσιο “Marker” δίνεται η δυνατότητα επιλογής του μεγέθους της κηλίδας των ατόμων όταν μετά την εφαρμογή του αλγορίθμου της ομαδοποίησης επιλεγεί η εμφάνιση των ομάδων στο γράφημα. Μικρότερες κηλίδες χρησιμοποιούνται όταν το νέφος των σημείων είναι πυκνό.

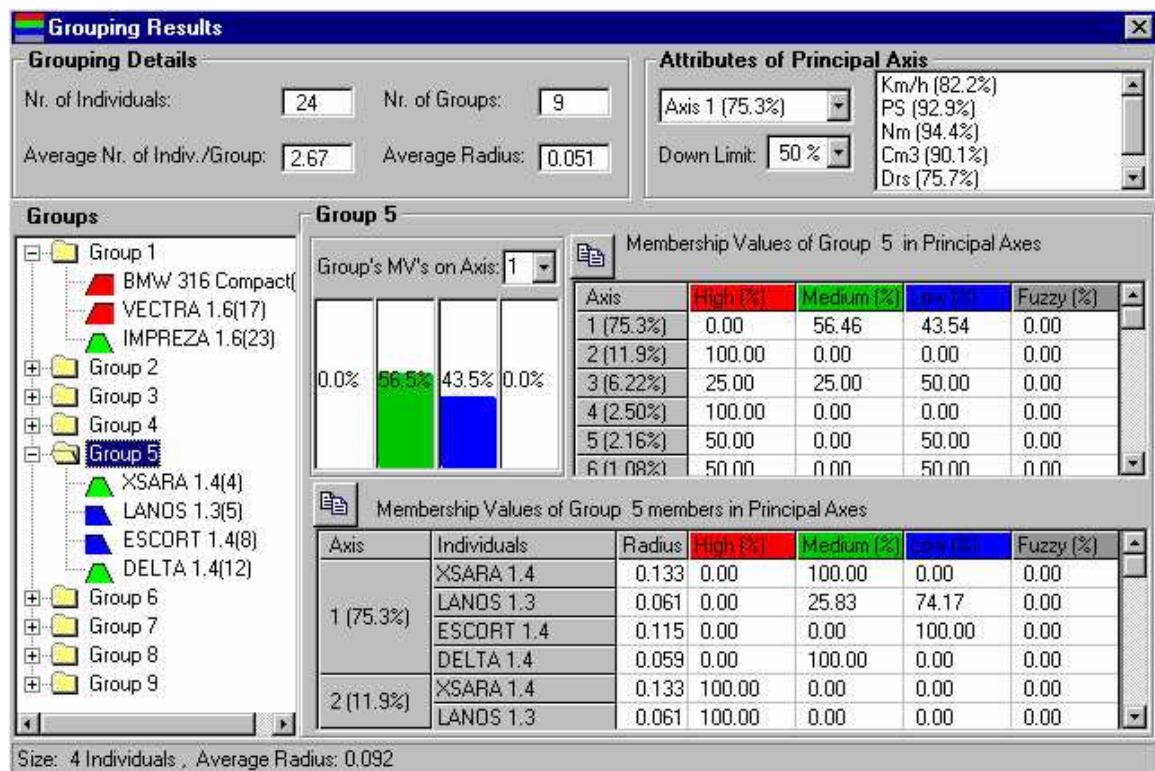
Η επιλογή Automatic View Current Groups έχει ως αποτέλεσμα την αυτόματη εμφάνιση των ομάδων κάθε φορά που ο αριθμός τους αλλάζει.



Σχήμα 5.24: Ρυθμίσεις για το Γράφημα της ΑΚΣ

5.8 Η ΦΟΡΜΑ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ

Από τη φόρμα αυτή είναι δυνατή η περιήγηση στο σύνολο των αποτελεσμάτων της ομαδοποίησης. Απεικονίζει στοιχεία τόσο για τις ομάδες που έχουν ήδη σχηματιστεί όσο και για τα άτομα από τα οποία αυτές συνίστανται αναλυτικά. Η μορφή της φόρμας είναι αυτή του Σχήματος 5.25.



Σχήμα 5.25: Η Φόρμα των Αποτελεσμάτων της Ομαδοποίησης

Στη φόρμα διακρίνονται τα εξής αντικείμενα:

5.8.1 Λεπτομέρειες της Ομαδοποίησης (Grouping Details)

Σ' αυτό το πλαίσιο της φόρμας (Σχήμα 5.26) αναφέρονται κάποιες λεπτομέρειες της ομαδοποίησης. Αυτές είναι:

Grouping Details			
Nr. of Individuals:	24	Nr. of Groups:	9
Average Nr. of Indiv./Group:	2.67	Average Radius:	0.046

Σχήμα 5.26: Το πλαίσιο Grouping Details

Nr. of Individuals: Ο αριθμός των ατόμων δηλαδή ο αριθμός των γραμμών του αρχικού πίνακα δεδομένων της ΑΚΣ.

Nr. of Groups: Ο αριθμός των ομάδων που έχουν δημιουργηθεί από τον αλγόριθμο της ομαδοποίησης.

Average Nr. of Indiv./Group: Ο μέσος αριθμός ατόμων ανά ομάδα. Η τιμή αυτή δίνεται από τη Σχέση 5.1.

$$\text{Average Nr. of Indiv./Group} = \frac{\text{Nr. of Individuals}}{\text{Nr. of Groups}} \quad \text{Σχέση 5.1}$$

Average Radius: Η μέση τιμή των αποστάσεων των ατόμων από το κέντρο της ομάδας τους. Η τιμή αυτή δίνεται από τη Σχέση 5.2.

$$\text{Average Radius} = \frac{\sum_{g=1}^{Nr.ofGroups} \sum_{\alpha=1}^{N_{Group}} D_{\alpha, CG_g}}{N} \quad \text{Σχέση 5.2}$$

όπου CG_g είναι το κέντρο της ομάδας g , D_{α, CG_g} είναι η απόσταση του ατόμου α από το κέντρο της ομάδας του (ομάδα g), N_{Group} ο αριθμός των ατόμων της ομάδας g και N ο συνολικός αριθμός των ατόμων.

Η τιμή αυτής της παράστασης αυξάνεται με τον περιορισμό των ομάδων και αποτελεί κριτήριο “συνοχής” ή “συνεκτικότητας” των ομάδων.

5.8.2 Χαρακτηριστικά των Κυρίων Αξόνων (Attributes of Principal Axis)

Σ' αυτό το πλαίσιο (Σχήμα 5.27) αναφέρονται τα χαρακτηριστικά που αναπαρίστανται στον κάθε κύριο άξονα. Η επιλογή του άξονα γίνεται από το αντικείμενο που αναφέρει ποιος είναι ο άξονας και σε παρένθεση το ποσοστό της πληροφορίας του αρχικού πίνακα δεδομένων της ΑΚΣ που αναπαριστά.



Σχήμα 5.27: Το πλαίσιο Attributes of Principal Axis με αναπαράσταση των χαρακτηριστικών που αντιπροσωπεύει ο πρώτος κύριος άξονας

Στη λίστα που φαίνεται δεξιά αναγράφονται τα χαρακτηριστικά που αντιπροσωπεύει ο επιλεγμένος άξονας και σε τι ποσοστό τα αντιπροσωπεύει. Σε περίπτωση που η προβολή του χαρακτηριστικού στον επιλεγμένο κύριο άξονα είναι αρνητική δίπλα στο χαρακτηριστικό εμφανίζεται το διακριτικό “(Negative)”. Στη λίστα αυτή αναγράφονται μόνο τα χαρακτηριστικά που αντιπροσωπεύονται στον κύριο άξονα σε ποσοστό μεγαλύτερο του επιλεγμένου από το αντικείμενο “Down Limit”. Έτσι για παράδειγμα εάν το επιλεγμένο ποσοστό ήταν 90%, τα χαρακτηριστικά που θα αναγράφονταν θα ήταν “PS (92,9%)”, “Nm (94,4%)” και “Cm3 (90,1%)” ενώ εάν είχε επιλεγεί ο δεύτερος κύριος άξονας το πλαίσιο θα είχε τη μορφή του Σχήματος 5.28.



Σχήμα 5.28: Το πλαίσιο Attributes of Principal Axis με αναπαράσταση των χαρακτηριστικών που αντιπροσωπεύει ο δεύτερος κύριος άξονας

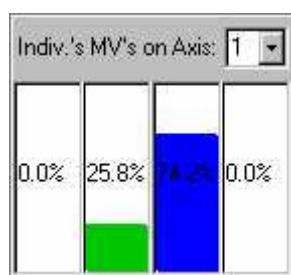
5.8.3 Το Δέντρο των Ομάδων

Στα αριστερά της φόρμας φαίνεται η δενδρική αναπαράσταση των ομάδων. Τα άτομα από τα οποία αποτελείται κάθε ομάδα φαίνονται αναπτύσσοντας τα “κλαδιά” που έχουν το σύμβολο του φακέλου. Τα γεωμετρικά σχήματα που φαίνονται αριστερά από τα ονόματα των ατόμων φανερώνουν την υψηλότερη τιμή συμμετοχής στον πρώτο κύριο άξονα ο οποίος μας δίνει και τη μέγιστη δυνατή πληροφορία.

Το πλεονέκτημα της χρήσης δενδρικής μορφής για τις ομάδες εκτός από την ταξινόμηση που επιτυγχάνει είναι η ευκολία επιλογής της ομάδας ή του ατόμου για το οποίο ο χρήστης επιθυμεί να έχει αναλυτικότερα αποτελέσματα. Έτσι τόσο το ραβδόγραμμα και οι δύο πίνακες που φαίνονται στη φόρμα όσο και η γραμμή κατάστασης στο κάτω μέρος της φόρμας διαμορφώνονται (παίρνουν τιμές) ανάλογα με την επιλογή του χρήστη στο δέντρο.

5.8.4 Το Ραβδόγραμμα και οι Πίνακες

Στο Σχήμα 5.29 φαίνεται σε μορφή ραβδογράμματος οι τιμές συμμετοχής της επιλεγμένης ομάδας ή του επιλεγμένου ατόμου στις τρεις περιοχές του κύριου άξονα που επιλέγεται από το αντικείμενο “Indiv.’s MV’s on Axis”, πάνω από το Ραβδόγραμμα.



Σχήμα 5.29: Ραβδόγραμμα τιμών συμμετοχής στον πρώτο κύριο άξονα

Οι τιμές συμμετοχής εμφανίζονται σε μορφή ποσοστού για να είναι πιο άνετη η παρατήρησή τους. Η τέταρτη ράβδος αντιστοιχεί στο υπόλοιπο των τιμών συμμετοχής των τριών περιοχών μέχρι τη μονάδα σε περίπτωση που ο κανόνας που έχει εφαρμοστεί στον κύριο άξονα είναι μη συμμετρικός.

Στον πίνακα του Σχήματος 5.30 φαίνονται οι τιμές συμμετοχής της επιλεγμένης ομάδας ή του επιλεγμένου ατόμου στις τρεις περιοχές του κάθε κύριου άξονα. Η τέταρτη στήλη αντιστοιχεί όπως και προηγουμένως στο υπόλοιπο των τιμών συμμετοχής των τριών περιοχών μέχρι τη μονάδα σε περίπτωση που ο κανόνας που έχει εφαρμοστεί στον κύριο άξονα είναι μη συμμετρικός. Σε περίπτωση συμμετρικού κανόνα οι τιμές της στήλης αυτής είναι όλες μηδέν.


Membership Values of Group 5 in Principal Axes				
Axis	High (%)	Medium (%)	Low (%)	Fuzzy (%)
1 (75.3%)	0.00	56.46	43.54	0.00
2 (11.9%)	100.00	0.00	0.00	0.00
3 (6.22%)	25.00	25.00	50.00	0.00
4 (2.50%)	100.00	0.00	0.00	0.00
5 (2.16%)	50.00	0.00	50.00	0.00
6 (1.08%)	50.00	0.00	50.00	0.00

Σχήμα 5.30: Πίνακας τιμών συμμετοχής με εφαρμογή συμμετρικού κανόνα

Στον πίνακα του Σχήματος 5.31 φαίνονται οι τιμές συμμετοχής όλων των ατόμων της επιλεγμένης ομάδας σε κάθε περιοχή του κάθε κύριου άξονα. Σε περίπτωση επιλογής ατόμου στο δέντρο των ομάδων εμφανίζονται οι τιμές συμμετοχής όλων των ατόμων που ανήκουν στην ίδια ομάδα με το επιλεγμένο άτομο.

Membership Values of Group 5 Members in Principal Axes						
Axis	Individuals	Radius	High (%)	Medium (%)	Low (%)	Fuzzy (%)
1 (75.3%)	XSARA 1.4	0.133	0.00	100.00	0.00	0.00
	LANOS 1.3	0.061	0.00	25.83	74.17	0.00
	ESCORT 1.4	0.115	0.00	0.00	100.00	0.00
	DELTA 1.4	0.059	0.00	100.00	0.00	0.00
2 (11.9%)	XSARA 1.4	0.133	100.00	0.00	0.00	0.00
	LANOS 1.3	0.061	100.00	0.00	0.00	0.00

Σχήμα 5.31: Πίνακας τιμών συμμετοχής με εφαρμογή συμμετρικού κανόνα

Σε κάθε περίπτωση ανάλογα με την επιλογή στο δέντρο των ομάδων οι πίνακες έχουν τίτλο για το περιεχόμενό τους ενώ είναι δυνατή η μεταφορά των δεδομένων τους στο Πρόχειρο (Clipboard) και στη συνέχεια σε κάποια άλλη εφαρμογή των Windows με χρήση του κουμπιού  που βρίσκεται πάνω από τους πίνακες.

5.8.5 Η Γραμμή Κατάστασης (Status Bar)

Η γραμμή κατάστασης παρέχει διάφορες πληροφορίες ανάλογα με την επιλογή στο δέντρο των ομάδων.

Εάν έχει επιλεγεί ομάδα, το μήνυμα που εμφανίζεται είναι το εξής:

“Nr. of Individuals: #, Average Radius: #”

δηλαδή ο αριθμός των ατόμων της επιλεγμένης ομάδας και η μέση απόσταση των ατόμων που ανήκουν στην ομάδα από το κέντρο της. Η τιμή αυτή δίνεται από τη Σχέση 5.3.

$$\text{Average Radius} = \frac{\sum_{\alpha=1}^{N_{Group}} D_{\alpha, CG_g}}{N_{Group}} \quad \text{Σχέση 5.3}$$

Εάν έχει επιλεγεί άτομο, το μήνυμα που εμφανίζεται είναι το εξής:

“Radius: #”

δηλαδή η απόσταση του ατόμου από το κέντρο της ομάδας του. Η τιμή αυτή είναι η D_{α, CG_g} .

ΚΕΦΑΛΑΙΟ 6: ΜΙΑ ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΕΘΟΔΟΥ ΜΕ ΤΟ ΠΡΟΓΡΑΜΜΑ

Έχοντας παρουσιάσει τη μεθοδολογική προσέγγιση στα προηγούμενα κεφάλαια, σ' αυτό το κεφάλαιο θα παρουσιάσουμε μία εφαρμογή του προγράμματος Fuzzy Data Grouping σε πραγματικά δεδομένα. Στόχος μας είναι να διερευνήσουμε την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας ομαδοποίησης και χαρακτηρισμού των ομάδων καθώς και της εφαρμογής που υλοποιήθηκε.

6.1 ΑΡΧΙΚΑ ΔΕΔΟΜΕΝΑ

Ο αρχικός πίνακας δεδομένων αφορά 24 μοντέλα αυτοκινήτων τα οποία περιγράφονται από 8 τεχνικά χαρακτηριστικά. Τα μοντέλα αυτοκινήτων που επιλέχθηκαν ήταν όλα διαθέσιμα στην Ελληνική αγορά το καλοκαίρι του 1999. Η επιλογή του ενός μοντέλου από κάθε μάρκα έγινε αυθαίρετα με μοναδική προϋπόθεση να έχει στον εξοπλισμό του υδραυλικό τιμόνι και κλιματισμό. Σε αντίθετη περίπτωση προστέθηκε στην τιμή του το αντίστοιχο ποσό.

Ο Πίνακας 6.1 περιλαμβάνει τα 24 μοντέλα αυτοκινήτων με τα 8 τεχνικά χαρακτηριστικά που χρησιμοποιήθηκαν. Τα τεχνικά χαρακτηριστικά φαίνονται στον Πίνακα 6.2.

Αρχικός Πίνακας Δεδομένων								
Μοντέλα/Χαρακτηριστικά	Km/h	PS	Nm	Lt./100 Km	0-100 (sec)	cm ³	Drs	Weight
ALFA 145 1.4	185	103	124	7.6	11.2	1370	4790	1135
AUDI A3 1.6	188	101	140	7.2	11.3	1595	6715	1090
BMW 316 Compact	195	102	150	7.8	12.7	1596	6822	1175
XSARA 1.4	175	75	111	8.2	13.9	1360	4298	1020
LANOS 1.3	166	75	113	8.7	15	1349	3075	1005
SIRION CL	165	55	88	5.5	13.9	989	3390	810
PUNTO 60	160	60	98	6.8	14.5	1242	3280	865
ESCORT 1.4	162	75	103	7	15.4	1391	4761	1080
CIVIC 1.4	177	90	124	6.6	10.8	1396	5155	1030
ACCENT 1.3	174	84	120	5.9	12.8	1341	3298	930
SEPHIA 1.5	170	96	140	7	11.8	1498	3598	1055
DELTA 1.4	170	74	111	7.8	14.5	1372	4295	1120
MAZDA 323 1.5	177	90	133	7.5	11.9	1498	5190	1075
C180 CLASSIC	193	122	170	9.2	12.2	1799	11650	1350
LANCER 1.3	172	87	116	6.5	12.6	1299	4033	1000
ALMERA 1.4	172	87	116	6.6	12.6	1392	4233	1070
VECTRA 1.6	193	106	150	7.2	12.5	1598	5890	1286
PEUGEOT 206	158	60	91	6.2	15.2	1124	4000	910
MEGANE 1.4	189	95	127	6.6	11.9	1390	5170	1090
ROVER 214	185	103	127	7	10.7	1396	4410	1015
SAAB 9-3	200	130	177	10.4	11	1985	11150	1300
CORDOBA 1.4	188	101	128	9.9	10.9	1391	4195	1025
IMPREZA 1.6	179	95	143	8.7	12.3	1597	5830	1125
BALENO 1.3	170	85	106	7.6	11.5	1298	3798	903

Πίνακας 6.1: Αρχικός Πίνακας Δεδομένων

Τεχνικό χαρακτηριστικό	Περιγραφή
Km/h	Τελική ταχύτητα του μοντέλου σε χιλιόμετρα ανά ώρα.
PS	Ισχύς του κινητήρα σε ίππους.
Nm	Ροπή στρέψης του κινητήρα.
Lt./100 Km	Κατανάλωση του μοντέλου σε λίτρα ανά 100 χιλιόμετρα.
0-100 (sec)	Επιτάχυνση του μοντέλου από ακίνητη θέση ως τα 100 χιλιόμετρα σε δευτερόλεπτα.
cm³	Κυβισμός του κινητήρα.
Drs.	Κόστος του μοντέλου σε δραχμές.
Weight	Βάρος του μοντέλου σε κιλά.

Πίνακας 6.2: Τεχνικά χαρακτηριστικά αυτοκινήτων

6.2 ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΝΑΛΥΣΗΣ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ

Εφαρμόζοντας στα παραπάνω δεδομένα τον αλγόριθμο της ΑΚΣ με την εντολή PCA Algorithm→Apply από τα μενού της αρχικής φόρμας προκύπτουν οι πίνακες αποτελεσμάτων της Ανάλυσης σε Κύριες Συνιστώσες.

6.2.1 Κεντροποιημένος Πίνακας Δεδομένων.

Με την εντολή PCA Algorithm→Centered Table από τα μενού της αρχικής φόρμας λαμβάνουμε τον Πίνακα 6.3. Ο Κεντροποιημένος Πίνακας χρησιμοποιείται για την εύρεση του Ισοσταθμισμένου Πίνακα Δεδομένων στο επόμενο βήμα της ΑΚΣ.

Κεντροποιημένος Πίνακας Δεδομένων								
Indiv.\Attr.	Km/h	PS	Nm	Lt/100 Km	0-100 (sec)	cm ³	Drs	Weight
ALFA 145 1.4	7.375	13.38	-1.25	0.12	-1.43	-57.75	-336.08	74
AUDI A3 1.6	10.375	11.38	14.75	-0.28	-1.33	167.25	1588.92	29
BMW 316 Compact	17.375	12.38	24.75	0.32	0.07	168.25	1695.92	114
XSARA 1.4	-2.625	-14.63	-14.25	0.72	1.27	-67.75	-828.08	-41
LANOS 1.3	-11.625	-14.63	-12.25	1.22	2.37	-78.75	-2051.08	-56
SIRION CL	-12.625	-34.63	-37.25	-1.98	1.27	-438.75	-1736.08	-251
PUNTO 60	-17.625	-29.63	-27.25	-0.68	1.87	-185.75	-1846.08	-196
ESCORT 1.4	-15.625	-14.63	-22.25	-0.48	2.77	-36.75	-365.08	19
CIVIC 1.4	-0.625	0.38	-1.25	-0.88	-1.83	-31.75	28.92	-31
ACCENT 1.3	-3.625	-5.63	-5.25	-1.58	0.17	-86.75	-1828.08	-131
SEPHIA 1.5	-7.625	6.38	14.75	-0.48	-0.83	70.25	-1528.08	-6
DELTA 1.4	-7.625	-15.63	-14.25	0.32	1.87	-55.75	-831.08	59
MAZDA 323 1.5	-0.625	0.38	7.75	0.02	-0.73	70.25	63.92	14
C180 CLASSIC	15.375	32.38	44.75	1.72	-0.43	371.25	6523.92	289
LANCER 1.3	-5.625	-2.63	-9.25	-0.98	-0.03	-128.75	-1093.08	-61
ALMERA 1.4	-5.625	-2.63	-9.25	-0.88	-0.03	-35.75	-893.08	9
VECTRA 1.6	15.375	16.38	24.75	-0.28	-0.13	170.25	763.92	225
PEUGEOT 206	-19.625	-29.63	-34.25	-1.28	2.57	-303.75	-1126.08	-151
MEGANE 1.4	11.375	5.38	1.75	-0.88	-0.73	-37.75	43.92	29
ROVER 214	7.375	13.38	1.75	-0.48	-1.93	-31.75	-716.08	-46
SAAB 9-3	22.375	40.38	51.75	2.92	-1.63	557.25	6023.92	239
CORDOBA 1.4	10.375	11.38	2.75	2.42	-1.73	-36.75	-931.08	-36
IMPREZA 1.6	1.375	5.38	17.75	1.22	-0.33	169.25	703.92	64
BALENO 1.3	-7.625	-4.63	-19.25	0.12	-1.13	-129.75	-1328.08	-158

Πίνακας 6.3: Κεντροποιημένος Πίνακας Δεδομένων

6.2.2 Ισοσταθμισμένος Πίνακας Δεδομένων

Με την εντολή PCA Algorithm→Normalized Table από τα μενού της αρχικής φόρμας λαμβάνουμε τον Πίνακα 6.4. Ο Κεντροποιημένος Πίνακας χρησιμοποιείται για την εύρεση του Πίνακα Συσχετίσεων των Χαρακτηριστικών.

Ισοσταθμισμένος Πίνακας Δεδομένων								
Indiv.\Attr.	Km/h	PS	Nm	Lt/100 Km	0-100 (sec)	cm ³	Drs	Weight
ALFA 145 1.4	0.130	0.152	-0.012	0.021	-0.202	-0.059	-0.032	0.117
AUDI A3 1.6	0.183	0.129	0.137	-0.048	-0.188	0.171	0.152	0.046
BMW 316 Compact	0.306	0.141	0.229	0.055	0.010	0.172	0.162	0.180
XSARA 1.4	-0.046	-0.166	-0.132	0.124	0.180	-0.069	-0.079	-0.065
LANOS 1.3	-0.205	-0.166	-0.113	0.210	0.335	-0.080	-0.196	-0.089
SIRION CL	-0.222	-0.394	-0.345	-0.341	0.180	-0.448	-0.166	-0.397
PUNTO 60	-0.310	-0.337	-0.252	-0.117	0.264	-0.190	-0.176	-0.310
ESCORT 1.4	-0.275	-0.166	-0.206	-0.083	0.392	-0.038	-0.035	0.030
CIVIC 1.4	-0.011	0.004	-0.012	-0.151	-0.259	-0.032	0.003	-0.049
ACCENT 1.3	-0.064	-0.064	-0.049	-0.272	0.024	-0.089	-0.175	-0.207
SEPHIA 1.5	-0.134	0.073	0.137	-0.083	-0.117	0.072	-0.146	-0.009
DELTA 1.4	-0.134	-0.178	-0.132	0.055	0.264	-0.057	-0.079	0.093
MAZDA 323 1.5	-0.011	0.004	0.072	0.004	-0.103	0.072	0.006	0.022
C180 CLASSIC	0.271	0.368	0.414	0.296	-0.061	0.379	0.623	0.458
LANCER 1.3	-0.099	-0.030	-0.086	-0.169	-0.004	-0.132	-0.104	-0.097
ALMERA 1.4	-0.099	-0.030	-0.086	-0.151	-0.004	-0.037	-0.085	0.014
VECTRA 1.6	0.271	0.186	0.229	-0.048	-0.018	0.174	0.073	0.356
PEUGEOT 206	-0.346	-0.337	-0.317	-0.220	0.363	-0.310	-0.108	-0.239
MEGANE 1.4	0.200	0.061	0.016	-0.151	-0.103	-0.039	0.004	0.046
ROVER 214	0.130	0.152	0.016	-0.083	-0.273	-0.032	-0.068	-0.073
SAAB 9-3	0.394	0.459	0.479	0.503	-0.230	0.570	0.575	0.378
CORDOBA 1.4	0.183	0.129	0.025	0.417	-0.244	-0.038	-0.089	-0.057
IMPREZA 1.6	0.024	0.061	0.164	0.210	-0.047	0.173	0.067	0.101
BALENO 1.3	-0.134	-0.053	-0.178	0.021	-0.160	-0.133	-0.127	-0.250

Πίνακας 6.4: Ισοσταθμισμένος Πίνακας Δεδομένων

6.2.3 Πίνακας Συσχετίσεων των Χαρακτηριστικών

Με την εντολή PCA Algorithm→Linear Correlations από τα μενού της αρχικής φόρμας λαμβάνουμε τον Πίνακα 6.5. Χαρακτηριστικά με συσχέτιση κοντά στη μονάδα ονομάζονται συσχετισμένα χαρακτηριστικά. Αυτό σημαίνει ότι αν κάποιο μοντέλο έχει για παράδειγμα μικρή τιμή σε κάποιο χαρακτηριστικό αναμένουμε να έχει χαμηλή τιμή και στα χαρακτηριστικά με τα οποία το συγκεκριμένο χαρακτηριστικό είναι ισχυρά συσχετισμένο. Αντίστοιχα συσχετίσεις με τιμή κοντά στο μείον ένα (-1) φανερώνουν αρνητική συσχέτιση ενώ συσχετίσεις με τιμή κοντά στο μηδέν φανερώνουν ασυσχέτιστα χαρακτηριστικά.

Πίνακας Συσχετίσεων των Χαρακτηριστικών								
Χαρακτ.\Χαρακτ.	Km/h	PS	Nm	Lt/100Km	0-100sec	cm ³	Drs	Weight
Km/h	1.000	0.902	0.864	0.530	-0.699	0.769	0.724	0.764
PS	<u>0.902</u>	1.000	0.935	0.596	-0.750	0.879	0.767	0.822
Nm	0.864	<u>0.935</u>	1.000	0.614	-0.594	0.952	0.830	0.867
Lt/100Km	0.530	0.596	0.614	1.000	-0.258	0.686	0.588	0.589
0-100sec	-0.699	<u>-0.750</u>	-0.594	<u>-0.258</u>	1.000	-0.462	-0.352	-0.326
Cm3	0.769	0.879	<u>0.952</u>	0.686	-0.462	1.000	0.856	0.886
Drs	0.724	0.767	0.830	0.588	-0.352	0.856	1.000	0.816
Weight	0.764	0.822	0.867	0.589	-0.326	0.886	0.816	1.000

Πίνακας 6.5: Πίνακας Συσχετίσεων των Χαρακτηριστικών

Έτσι όπως φαίνεται τα χαρακτηριστικά Τελική ταχύτητα (Km/h), Ιπποδύναμη (PS) και Ροπή στρέψης (Nm) είναι ισχυρά συσχετισμένα με βαθμό συσχέτισης πάνω από 0.9. Τη μεγαλύτερη αρνητική συσχέτιση παρουσιάζουν τα χαρακτηριστικά Ιπποδύναμη και Κατανάλωση (Lt./100 Km) με τιμή -0.75, ενώ τα πλέον ασυσχέτιστα χαρακτηριστικά μεταξύ τους είναι η Επιτάχυνση (0-100 (sec)) και η Κατανάλωση με τιμή συσχέτισης -0.258.

6.2.4 Πίνακας Ιδιοτιμών

Με την εντολή PCA Algorithm → Eigenvalues από τα μενού της αρχικής φόρμας λαμβάνουμε τον Πίνακα 6.6.

Ιδιοτιμές Κυρίων Αξόνων			
Ιδιοτιμή	Τιμή	Ποσοστό %	Αθροιστικό Ποσοστό %
Ιδιοτιμή 1	6.031	75.39	75.39
Ιδιοτιμή 2	0.957	11.96	87.35
Ιδιοτιμή 3	0.497	6.22	93.58
Ιδιοτιμή 4	0.2	2.5	96.08
Ιδιοτιμή 5	0.173	2.16	98.25
Ιδιοτιμή 6	0.086	1.08	99.33
Ιδιοτιμή 7	0.028	0.35	99.69
Ιδιοτιμή 8	0.024	0.3	100

Πίνακας 6.6: Ιδιοτιμές των Κυρίων Αξόνων

Από τον πίνακα των Ιδιοτιμών βλέπουμε το ποσοστό της πληροφορίας του αρχικού πίνακα δεδομένων που εκφράζεται σε κάθε κύριο άξονα της ΑΚΣ. Από τον Πίνακα 6.6 φαίνεται ότι το επίπεδο γράφημα του πρώτου κύριου άξονα με τον δεύτερο κύριο άξονα εκφράζει το $75,39\% + 11,96\% = 87,35\%$ της συνολικής πληροφορίας του αρχικού πίνακα δεδομένων.

6.2.5 Προβολές των Χαρακτηριστικών στους Νέους Άξονες.

Με την εντολή PCA Algorithm→Attributes Projections από τα μενού της αρχικής φόρμας λαμβάνουμε τον Πίνακα 6.7.

Προβολές Χαρακτηριστικών στους Νέους Άξονες								
Χαρ.\Άξ.	Άξονας 1 (75.3%)	Άξονας 2 (11.9%)	Άξονας 3 (6.22%)	Άξονας 4 (2.50%)	Άξονας 5 (2.16%)	Άξονας 6 (1.08%)	Άξονας 7 (0.35%)	Άξονας 8 (0.30%)
Km/h	0.906 (82.2%)	-0.24	0.039	0.094	0.309	0.116	-0.021	0.027
PS	0.964 (92.9%)	-0.2	-0.002	0.051	-0.05	-0.063	0.133	0.05
Nm	0.971 (94.4%)	-0.01	0.087	0.031	-0.122	0.141	0.028	-0.107
Lt/100Km	0.696	0.37	-0.612	0.003	0.042	-0.02	-0.002	-0.013
0-100sec	-0.626	0.745 (55.6%)	0.162	0.087	0.073	0.087	0.06	0.019
cm ³	0.949 (90.1%)	0.179	0.049	-0.003	-0.208	0.091	-0.062	0.088
Drs	0.870 (75.7%)	0.255	0.189	-0.359	0.096	-0.053	0.002	-0.008
Weight	0.893 (79.8%)	0.261	0.221	0.226	0.019	-0.171	-0.044	-0.027

Πίνακας 6.7: Προβολές Χαρακτηριστικών στους Νέους Άξονες

Στον πίνακα των Προβολών των Χαρακτηριστικών φαίνονται εκτός από τις προβολές και το ποσοστό πληροφoρίας που αντιπροσωπεύει ο κάθε κύριος άξονας και το ποσοστό με το οποίο αντιπροσωπεύεται το κάθε χαρακτηριστικό στον κάθε άξονα αρκεί αυτό να είναι πάνω από 50%. Το ποσοστό αυτό αντιστοιχεί στο τετράγωνο της προβολής του χαρακτηριστικού στον νέο άξονα.

Στον Πίνακα 6.8 φαίνονται τα χαρακτηριστικά που εκφράζει ο κάθε κύριος άξονας.

Άξονας και ποσοστό πληροφορίας (%)	Χαρακτηριστικό	% προβολή στον άξονα
Άξονας 1 (75.39)	Τελική ταχύτητα (Km/h)	82.2
	Ιπποδύναμη (PS)	92.9
	Ροπή στρέψης (Nm)	94.4
	Κυβισμός (Cm3)	90.1
	Τιμή (Drs)	75.7
	Βάρος (Kgr)	79.8
Άξονας 2 (11.96)	Επιτάχυνση (0-100 sec)	55.6

Πίνακας 6.8: Χαρακτηριστικά που προβάλλονται στους Νέους Άξονες

Παρατηρούμε ότι στον πρώτο κύριο άξονα αντιπροσωπεύονται σε ικανοποιητικό ποσοστό έξι από τα οκτώ τεχνικά χαρακτηριστικά. Επίσης είναι δυνατό ορισμένα χαρακτηριστικά να μην προβάλλονται σε ικανοποιητικό ποσοστό σε κάποιο κύριο άξονα. Για παράδειγμα η μεγαλύτερη προβολή του χαρακτηριστικού της κατανάλωσης ανά 100 χιλιόμετρα (Lt./100 Km) είναι 0.696 στον πρώτο κύριο άξονα και το τετράγωνό της που αντιστοιχεί στο ποσοστό που αντιπροσωπεύεται το χαρακτηριστικό στον άξονα αυτό είναι 0.4844. Δηλαδή το χαρακτηριστικό αυτό αντιπροσωπεύεται σε ποσοστό μόλις 48,44%.

6.2.6 Προβολές των Ατόμων στους Νέους Άξονες.

Με την εντολή PCA Algorithm→Individuals Projections από τα μενού της αρχικής φόρμας λαμβάνουμε τον Πίνακα 6.9.

Προβολές των Ατόμων στους Νέους Άξονες								
Άτομο\Χαρακτ.	Άξονας 1	Άξονας 2	Άξονας 3	Άξονας 4	Άξονας 5	Άξονας 6	Άξονας 7	Άξονας 8
ALFA 145 1.4	0.169	-0.2	-0.035	0.09	0.075	-0.129	0.019	0
AUDI A3 1.6	0.343	-0.15	0.092	-0.073	-0.006	0.056	-0.038	0.042
BMW 316 Compact	0.461	0.045	0.111	0.059	0.11	0.119	0.004	-0.004
XSARA 1.4	-0.223	0.179	-0.131	0.029	0.082	0.03	-0.043	0.016
LANOS 1.3	-0.344	0.33	-0.216	0.109	-0.027	0.048	0.042	-0.008
SIRION CL	-0.892	-0.09	0.082	-0.148	0.148	0.03	-0.002	-0.053
PUNTO 60	-0.696	0.14	-0.043	-0.085	-0.041	0.069	-0.022	0.008
ESCORT 1.4	-0.387	0.362	0.118	0.027	-0.051	-0.072	0.016	0.071
CIVIC 1.4	-0.013	-0.27	0.052	-0.08	-0.051	-0.059	-0.062	-0.021
ACCENT 1.3	-0.322	-0.17	0.113	0.014	-0.054	0.115	0.046	0.024
SEPHIA 1.5	0.012	-0.13	0.016	0.077	-0.247	0.021	0.029	-0.051
DELTA 1.4	-0.239	0.286	-0.006	0.105	0.027	-0.06	-0.056	-0.007
MAZDA 323 1.5	0.091	-0.06	-0.004	-0.011	-0.081	0.006	-0.052	-0.026
C180 CLASSIC	1.042	0.275	0.131	-0.152	0.029	-0.073	0.048	-0.05
LANCER 1.3	-0.252	-0.11	0.061	0.003	-0.025	-0.029	0.048	-0.005
ALMERA 1.4	-0.163	-0.05	0.093	0.043	-0.061	-0.069	-0.016	0.026
VECTRA 1.6	0.477	0.008	0.223	0.21	0.049	0.008	-0.012	-0.015
PEUGEOT 206	-0.785	0.2	0.091	-0.097	0.038	-0.031	0.034	-0.015
MEGANE 1.4	0.091	-0.19	0.133	0.049	0.125	0.014	-0.007	0.014
ROVER 214	0.097	-0.35	-0.025	0.01	0.013	-0.004	0.039	0.032
SAAB 9-3	1.278	0.176	-0.096	-0.145	-0.027	0.036	0.001	0.041
CORDOBA 1.4	0.242	-0.15	-0.449	0.053	0.107	-0.007	0.017	-0.017
IMPREZA 1.6	0.297	0.1	-0.109	0.012	-0.09	0.029	-0.036	-0.036
BALENO 1.3	-0.281	-0.19	-0.206	-0.101	-0.041	-0.049	0.001	0.035

Πίνακας 6.9: Προβολές των μοντέλων στους Νέους Άξονες

Από τα δεδομένα του πίνακα προβολών των ατόμων και του πίνακα προβολών των χαρακτηριστικών στους νέους άξονες μπορεί να κατασκευαστεί το γράφημα της Ανάλυσης σε Κύριες Συνιστώσες για το οποίο θα μιλήσουμε παρακάτω. Επιπλέον ο παραπάνω πίνακας χρησιμοποιείται για τη γεωμετρική ομαδοποίηση των μοντέλων.

6.3 ΟΡΙΣΜΟΣ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ

Προκειμένου να προχωρήσουμε στην ομαδοποίηση των δεδομένων θα ορίσουμε τις συναρτήσεις συμμετοχής σε κάθε μία από τις τρεις περιοχές (Υψηλή, Μεσαία, Χαμηλή). Ο ορισμός των Συναρτήσεων Συμμετοχής γίνεται από τη φόρμα “Membership Functions Values”.

Παρακάτω θα εφαρμόσουμε τους έξι προτεινόμενους κανόνες (Standard Rules) στα δεδομένα του παραδείγματος. Το γράφημα της κατανομής των ατόμων σε κάθε κύριο άξονα που φαίνεται στην φόρμα Membership Functions κάτω από το γράφημα των συναρτήσεων συμμετοχής μπορεί να μας βοηθήσει στην επιλογή του επιθυμητού κανόνα. Επίσης είναι δυνατή η εφαρμογή οποιουδήποτε κανόνα επιλέγοντας τα όρια ακριβώς εκεί που θέλουμε αρκεί να μην παραβιάζεται η Σχέση 4.5.

Υπενθυμίζουμε ότι συμμετρικός είναι ένας κανόνας (συναρτήσεις συμμετοχής στις τρεις περιοχές) όταν τηρείται η Σχέση 4.3. Για παράδειγμα εάν ένα μοντέλο έχει τιμή συμμετοχής 0.55 ή 55% στη Χαμηλή περιοχή τότε θα έχει τιμή συμμετοχής $100\% - 55\% = 45\%$ στη Μεσαία περιοχή. Αντίθετα μη συμμετρικός είναι ένας κανόνας όταν το άθροισμα των τιμών συμμετοχής για ένα σημείο που βρίσκεται ανάμεσα σε δύο σίγουρες περιοχές δεν είναι 100% αλλά μικρότερο με αποτέλεσμα να παραμένει ένα ποσοστό ασάφειας.

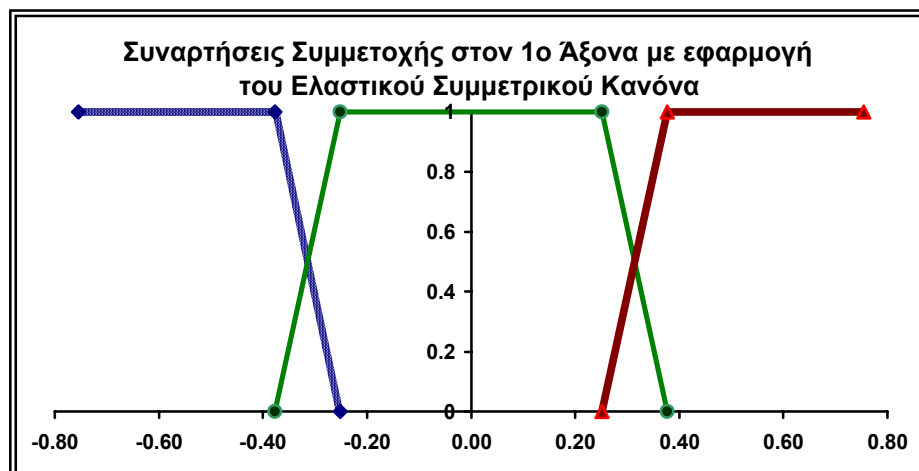
6.3.1 Ελαστικός Συμμετρικός κανόνας

Με την εφαρμογή του Ελαστικού Συμμετρικού κανόνα, οι τιμές των ορίων των συναρτήσεων συμμετοχής, όπως θα φαίνονται στην βάση δεδομένων Access, για τους δύο πρώτους κύριους άξονες θα είναι αυτές του Πίνακα 6.10.

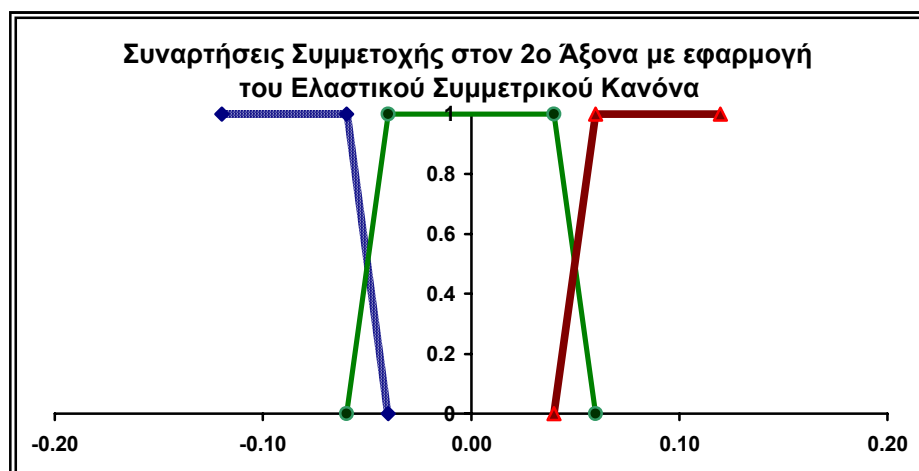
Axis	Area	Point1X	Point1Y	Point2X	Point2Y	Point3X	Point3Y	Point4X	Point4Y
Axis1	Low	-0.75	1.00	-0.75	1.00	-0.38	1.00	-0.25	0.00
	Medium	-0.38	0.00	-0.25	1.00	0.25	1.00	0.38	0.00
	High	0.25	0.00	0.38	1.00	0.75	1.00	0.75	1.00
Axis2	Low	-0.12	1.00	-0.12	1.00	-0.06	1.00	-0.04	0.00
	Medium	-0.06	0.00	-0.04	1.00	0.04	1.00	0.06	0.00
	High	0.04	0.00	0.06	1.00	0.12	1.00	0.12	1.00

Πίνακας 6.10: Όρια Συναρτήσεων Συμμετοχής με εφαρμογή του Ελαστικού Συμμετρικού Κανόνα

Στα Σχήματα 6.1 και 6.2 φαίνονται οι γραφικές παραστάσεις των συναρτήσεων συμμετοχής σ' αυτούς τους άξονες.



Σχήμα 6.1



Σχήμα 6.2

Ιδιαίτερη προσοχή πρέπει να δοθεί στο γεγονός ότι η κλίμακα στον X άξονα είναι διαφορετική στα Σχήματα 6.1 και 6.2. Κι' αυτό γιατί ρυθμιστής των ορίων των συναρτήσεων συμμετοχής είναι η τυπική απόκλιση των μοντέλων στον κύριο άξονα της οποίας η τιμή μειώνεται όσο αυξάνεται ο αριθμός του άξονα. Αυτό ισχύει και για τους παρακάτω κανόνες.

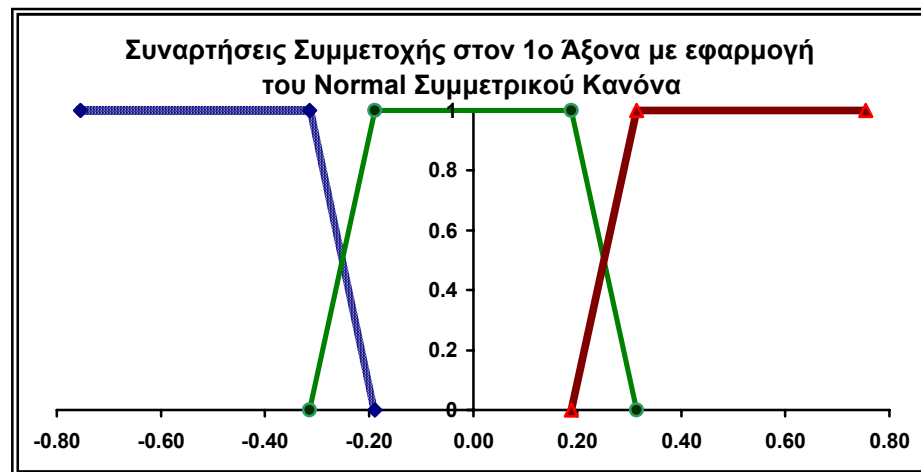
6.3.2 Normal Συμμετρικός κανόνας

Με την εφαρμογή του Normal Συμμετρικού κανόνα, οι τιμές των ορίων των συναρτήσεων συμμετοχής, όπως θα φαίνονται στην βάση δεδομένων Access, για τους δύο πρώτους κύριους άξονες θα είναι αυτές του Πίνακα 6.11.

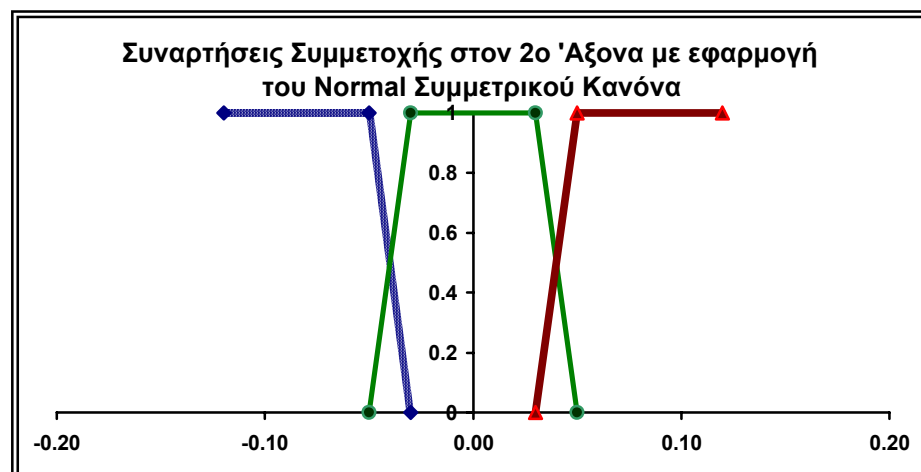
Axis	Area	Point1X	Point1Y	Point2X	Point2Y	Point3X	Point3Y	Point4X	Point4Y
Axis1	Low	-0.75	1.00	-0.75	1.00	-0.31	1.00	-0.19	0.00
	Medium	-0.31	0.00	-0.19	1.00	0.19	1.00	0.31	0.00
	High	0.19	0.00	0.31	1.00	0.75	1.00	0.75	1.00
Axis2	Low	-0.12	1.00	-0.12	1.00	-0.05	1.00	-0.03	0.00
	Medium	-0.05	0.00	-0.03	1.00	0.03	1.00	0.05	0.00
	High	0.03	0.00	0.05	1.00	0.12	1.00	0.12	1.00

Πίνακας 6.11: Όρια Συναρτήσεων Συμμετοχής με εφαρμογή του Normal Συμμετρικού Κανόνα

Στα Σχήματα 6.3 και 6.4 φαίνονται οι γραφικές παραστάσεις των συναρτήσεων συμμετοχής σ' αυτούς τους άξονες.



Σχήμα 6.3



Σχήμα 6.4

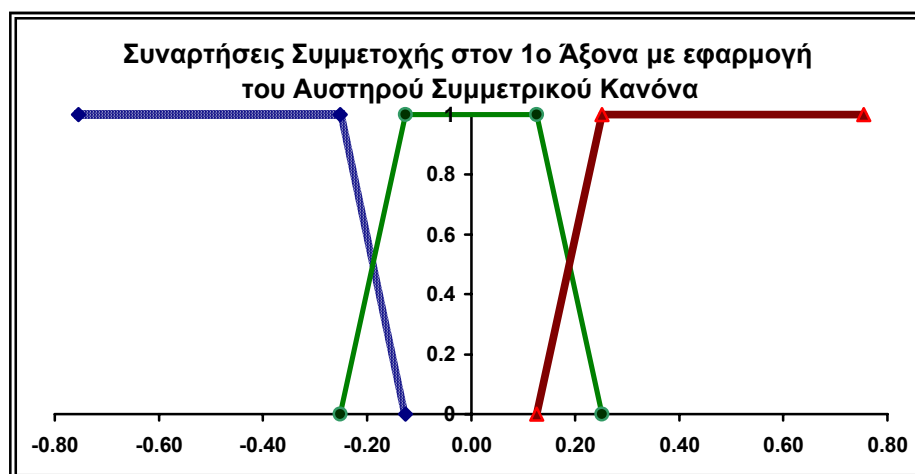
6.3.3 Αυστηρός Συμμετρικός κανόνας

Με την εφαρμογή του Αυστηρού συμμετρικού κανόνα, οι τιμές των ορίων των συναρτήσεων συμμετοχής, όπως θα φαίνονται στην βάση δεδομένων Access, για τους δύο πρώτους κύριους άξονες θα είναι αυτές του Πίνακα 6.12.

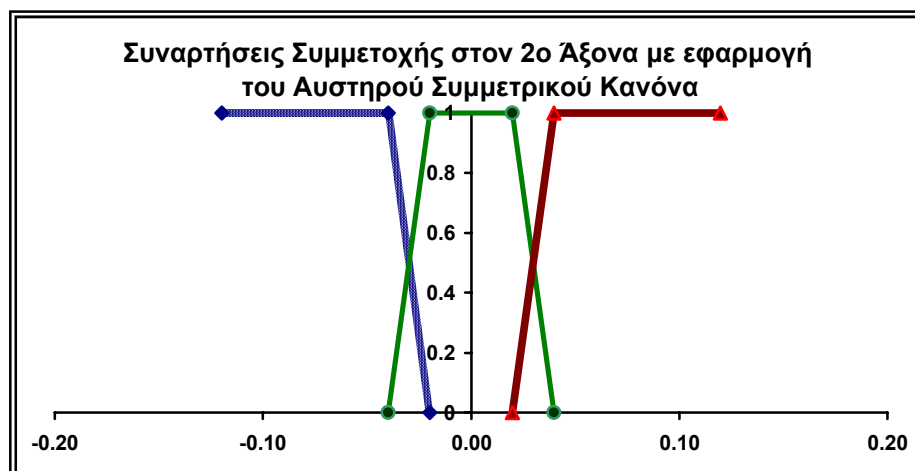
Axis	Area	Point1X	Point1Y	Point2X	Point2Y	Point3X	Point3Y	Point4X	Point4Y
Axis1	Low	-0.75	1.00	-0.75	1.00	-0.25	1.00	-0.13	0.00
	Medium	-0.25	0.00	-0.13	1.00	0.13	1.00	0.25	0.00
	High	0.13	0.00	0.25	1.00	0.75	1.00	0.75	1.00
Axis2	Low	-0.12	1.00	-0.12	1.00	-0.04	1.00	-0.02	0.00
	Medium	-0.04	0.00	-0.02	1.00	0.02	1.00	0.04	0.00
	High	0.02	0.00	0.04	1.00	0.12	1.00	0.12	1.00

Πίνακας 6.12: Όρια Συναρτήσεων Συμμετοχής με εφαρμογή του Αυστηρού Συμμετρικού Κανόνα

Στα Σχήματα 6.5 και 6.6 φαίνονται οι γραφικές παραστάσεις των συναρτήσεων συμμετοχής σ' αυτούς τους άξονες.



Σχήμα 6.5



Σχήμα 6.6

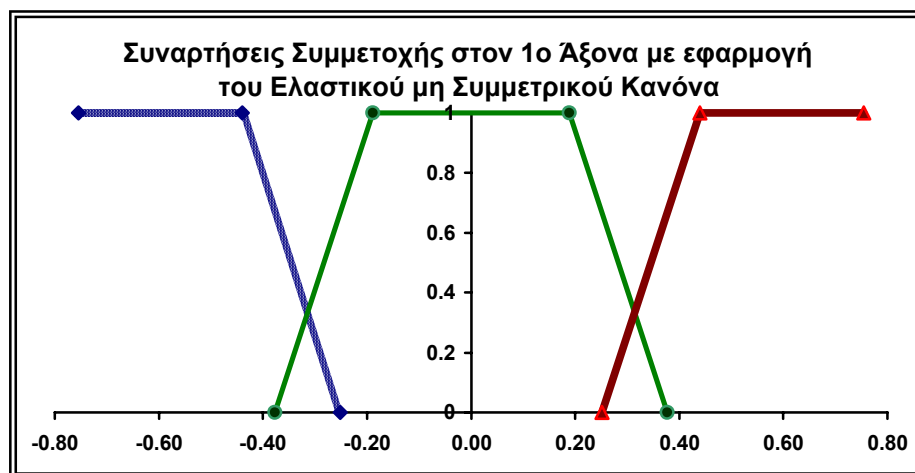
6.3.4 Ελαστικός μη Συμμετρικός κανόνας

Με την εφαρμογή του Ελαστικού μη Συμμετρικού κανόνα, οι τιμές των ορίων των συναρτήσεων συμμετοχής, όπως θα φαίνονται στην βάση δεδομένων Access, για τους δύο πρώτους κύριους άξονες θα είναι αυτές του Πίνακα 6.13.

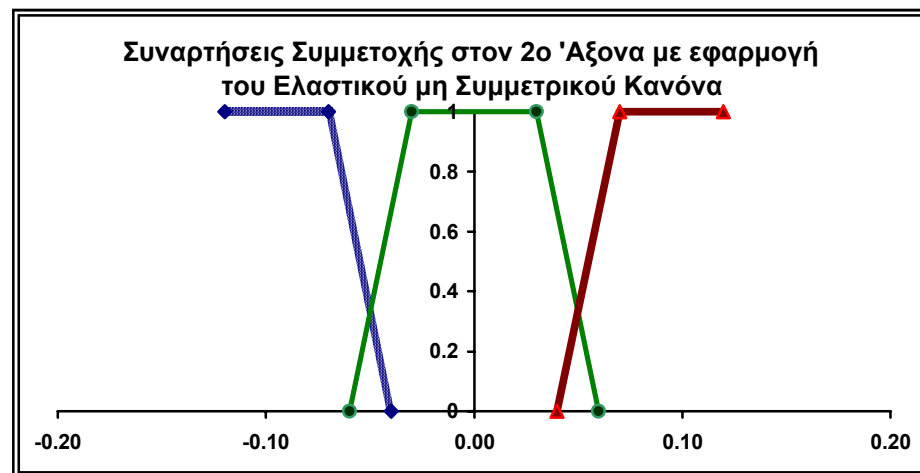
Axis	Area	Point1X	Point1Y	Point2X	Point2Y	Point3X	Point3Y	Point4X	Point4Y
Axis1	Low	-0.75	1.00	-0.75	1.00	-0.44	1.00	-0.25	0.00
	Medium	-0.38	0.00	-0.19	1.00	0.19	1.00	0.38	0.00
	High	0.25	0.00	0.44	1.00	0.75	1.00	0.75	1.00
Axis2	Low	-0.12	1.00	-0.12	1.00	-0.07	1.00	-0.04	0.00
	Medium	-0.06	0.00	-0.03	1.00	0.03	1.00	0.06	0.00
	High	0.04	0.00	0.07	1.00	0.12	1.00	0.12	1.00

Πίνακας 6.13: Όρια Συναρτήσεων Συμμετοχής με εφαρμογή του Ελαστικού μη Συμμετρικού Κανόνα

Στα Σχήματα 6.7 και 6.8 φαίνονται οι γραφικές παραστάσεις των συναρτήσεων συμμετοχής σ' αυτούς τους άξονες.



Σχήμα 6.7



Σχήμα 6.8

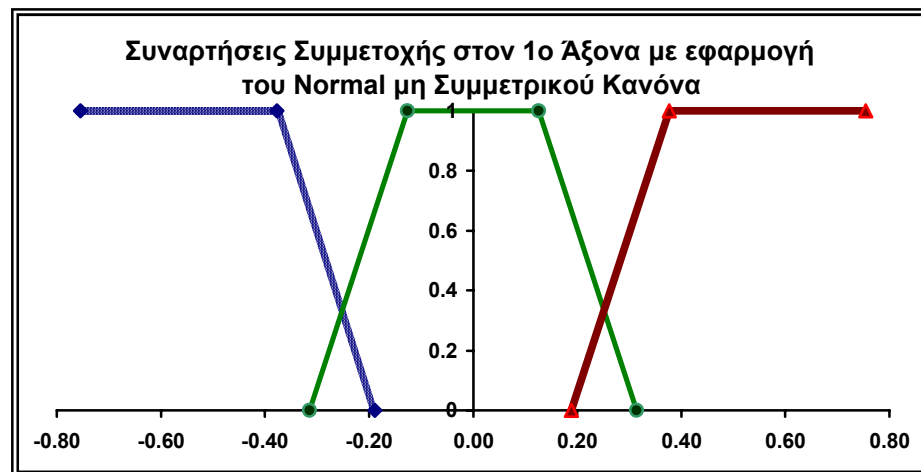
6.3.5 Normal μη Συμμετρικός κανόνας

Με την εφαρμογή του Normal μη συμμετρικού κανόνα, οι τιμές των ορίων των συναρτήσεων συμμετοχής, όπως θα φαίνονται στην βάση δεδομένων Access, για τους δύο πρώτους κύριους άξονες θα είναι αυτές που φαίνονται στον Πίνακα 6.14.

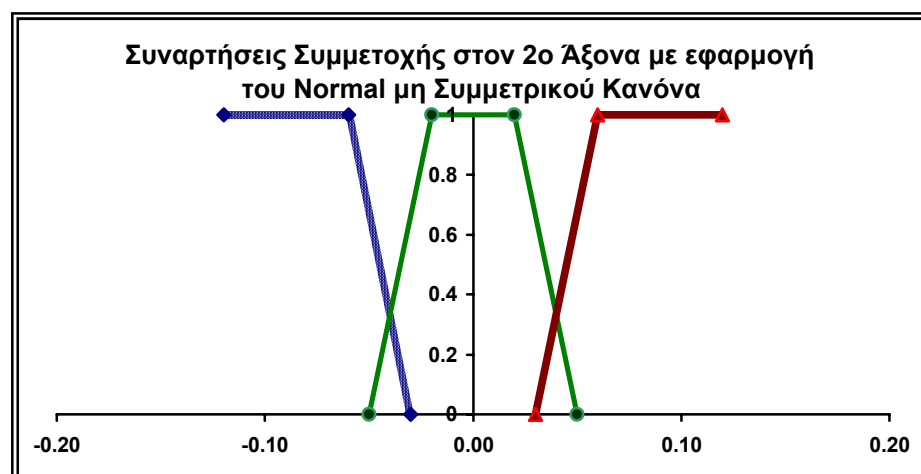
Axis	Area	Point1X	Point1Y	Point2X	Point2Y	Point3X	Point3Y	Point4X	Point4Y
Axis1	Low	-0.75	1.00	-0.75	1.00	-0.38	1.00	-0.19	0.00
	Medium	-0.31	0.00	-0.13	1.00	0.13	1.00	0.31	0.00
	High	0.19	0.00	0.38	1.00	0.75	1.00	0.75	1.00
Axis2	Low	-0.12	1.00	-0.12	1.00	-0.06	1.00	-0.03	0.00
	Medium	-0.05	0.00	-0.02	1.00	0.02	1.00	0.05	0.00
	High	0.03	0.00	0.06	1.00	0.12	1.00	0.12	1.00

Πίνακας 6.14: Όρια Συναρτήσεων Συμμετοχής με εφαρμογή του Normal μη Συμμετρικού Κανόνα

Στα Σχήματα 6.9 και 6.10 φαίνονται οι γραφικές παραστάσεις των συναρτήσεων συμμετοχής σ' αυτούς τους άξονες.



Σχήμα 6.9



Σχήμα 6.10

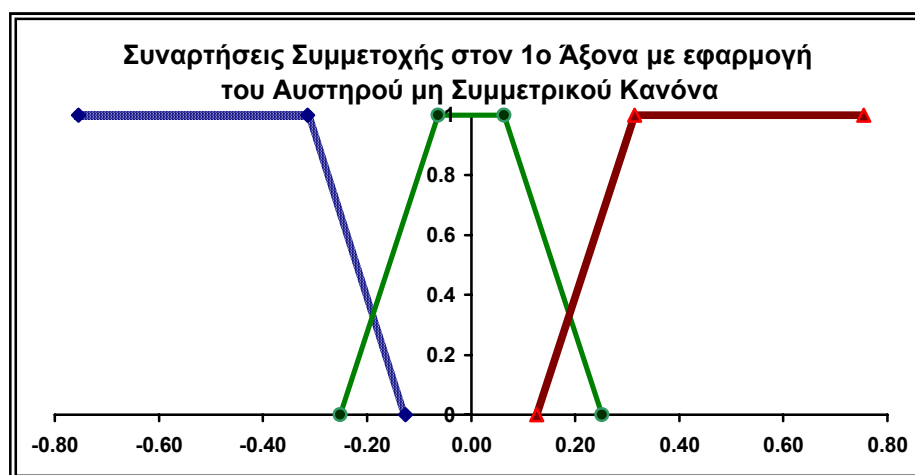
6.3.6 Αυστηρός μη Συμμετρικός κανόνας

Με την εφαρμογή του Αυστηρού μη συμμετρικού κανόνα, οι τιμές των ορίων των συναρτήσεων συμμετοχής, όπως θα φαίνονται στην βάση δεδομένων Access, για τους δύο πρώτους κύριους άξονες θα είναι αυτές του Πίνακα 6.15.

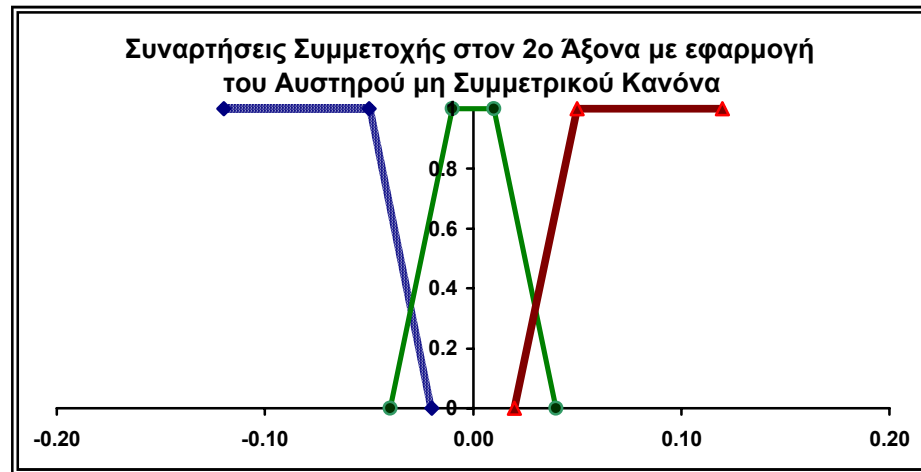
Axis	Area	Point1X	Point1Y	Point2X	Point2Y	Point3X	Point3Y	Point4X	Point4Y
Axis1	Low	-0.75	1.00	-0.75	1.00	-0.31	1.00	-0.13	0.00
	Medium	-0.25	0.00	-0.06	1.00	0.06	1.00	0.25	0.00
	High	0.13	0.00	0.31	1.00	0.75	1.00	0.75	1.00
Axis2	Low	-0.12	1.00	-0.12	1.00	-0.05	1.00	-0.02	0.00
	Medium	-0.04	0.00	-0.01	1.00	0.01	1.00	0.04	0.00
	High	0.02	0.00	0.05	1.00	0.12	1.00	0.12	1.00

Πίνακας 6.15: Όρια Συναρτήσεων Συμμετοχής με εφαρμογή του Αυστηρού μη Συμμετρικού Κανόνα

Στα Σχήματα 6.11 και 6.12 φαίνονται οι γραφικές παραστάσεις των συναρτήσεων συμμετοχής σ' αυτούς τους άξονες.



Σχήμα 6.11



Σχήμα 6.12

6.3.7 Τιμές Συμμετοχής των Ατόμων

Έχοντας εφαρμόσει κάποιο από τους παραπάνω κανόνες για τον ορισμό των συναρτήσεων συμμετοχής στους κύριους άξονες, μπορούμε με την εντολή Grouping→MV's Table από τα μενού της αρχικής φόρμας να λάβουμε τον πίνακα των τιμών συμμετοχής του κάθε ατόμου στον κάθε κύριο άξονα. Στον πίνακα αυτό φαίνεται η μεγαλύτερη από τις τιμές συμμετοχής σε κάποια περιοχή ενώ η περιοχή ορίζεται από το χαρακτηριστικό χρώμα. Στον Πίνακα 6.16 φαίνονται οι τιμές συμμετοχής των ατόμων με την εφαρμογή του Ελαστικού Συμμετρικού Κανόνα.

Τιμές Συμμετοχής των Ατόμων στους Κύριους Άξονες								
Άτομο\Άξονας	Άξονας 1	Άξονας 2	Άξονας 3	Άξονας 4	Άξονας 5	Άξονας 6	Άξονας 7	Άξονας 8
ALFA 145 1.4	1	1	1	1	1	1	1	1
AUDI A3 1.6	0.731	1	1	1	1	1	1	1
BMW 316 Compact	1	0.72	1	1	1	1	1	1
XSARA 1.4	1	1	1	1	1	1	1	1
LANOS 1.3	0.741	1	1	1	1	1	1	1
SIRION CL	1	1	1	1	1	1	1	1
PUNTO 60	1	1	1	1	1	1	1	1
ESCORT 1.4	1	1	1	1	1	1	1	1
CIVIC 1.4	1	1	1	1	1	1	1	1
ACCENT 1.3	0.569	1	1	1	1	1	1	1
SEPHIA 1.5	1	1	1	1	1	1	1	1
DELTA 1.4	1	1	1	1	1	1	1	1
MAZDA 323 1.5	1	0.77	1	0.642	1	1	1	1
C180 CLASSIC	1	1	1	1	1	1	1	1
LANCER 1.3	0.994	1	1	1	1	1	1	1
ALMERA 1.4	1	0.743	1	1	1	1	1	1
VECTRA 1.6	1	1	1	1	1	1	1	1
PEUGEOT 206	1	1	1	1	1	1	1	1
MEGANE 1.4	1	1	1	1	1	1	1	1
ROVER 214	1	1	0.509	0.508	1	0.609	1	1
SAAB 9-3	1	1	1	1	1	1	0.52	1
CORDOBA 1.4	1	1	1	1	1	1	1	1
IMPREZA 1.6	0.635	1	1	0.954	1	1	1	1
BALENO 1.3	0.76	1	1	1	1	1	1	1

Πίνακας 6.1.16: Τιμές Συμμετοχής των μοντέλων στους Κύριους Άξονες

6.4 ΟΜΑΔΟΠΟΙΗΣΗ

Για την ομαδοποίηση των δεδομένων θα κάνουμε χρήση και των δύο εκδοχών του αλγορίθμου της ομαδοποίησης. Σημαντική παράμετρος κατά τη διαδικασία της ομαδοποίησης είναι επίσης οι χρησιμοποιούμενοι κύριοι άξονες, δηλαδή η διάσταση της ομαδοποίησης.

6.4.1 Ομαδοποίηση με χρήση του αλγορίθμου Nearest Neighbour (NN)

Στη συνέχεια παραθέτουμε τις ομάδες που προκύπτουν με τη χρήση της εκδοχής Nearest Neighbour του αλγορίθμου της ομαδοποίησης στους δύο πρώτους κύριους άξονες (87.35% του συνόλου της πληροφορίας) και στους οκτώ κύριους άξονες (100% του συνόλου της πληροφορίας).

Ομαδοποίηση στους δύο πρώτους Κύριους Άξονες

Στους πίνακες 6.17 και 6.18 φαίνονται δύο διαφορετικές ομαδοποιήσεις που προέκυψαν με την ομαδοποίηση στους δύο πρώτους κύριους άξονες.

Ομαδοποίηση με αλγόριθμο NN στους δύο πρώτους Κύριους Άξονες									
Αρ. Ατόμων	Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4	Ομάδα 5	Ομάδα 6	Ομάδα 7	Ομάδα 8	Ομάδα 9
1	BMW	ACCENT	ALFA	PUNTO	XSARA	AUDI	SIRION	C180	SAAB
2	VECTRA	LANCER	CIVIC	PEUGEOT	LANOS				
3	IMPREZA	ALMERA	SEPHIA		ESCORT				
4		BALENO	MAZDA		DELTA				
5			MEGANE						
6			ROVER						
7			CORDOBA						

Πίνακας 6.17: Εννέα ομάδες που προκύπτουν με χρήση του αλγορίθμου Nearest Neighbour στους δύο πρώτους κύριους άξονες

Ομαδοποίηση με αλγόριθμο NN στους δύο πρώτους Κύριους Άξονες					
Αρ. Ατόμων	Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4	Ομάδα 5
1	ACCENT	ALFA	SIRION	XSARA	C180
2	LANCER	AUDI	PUNTO	LANOS	SAAB
3	ALMERA	BMW	PEUGEOT	ESCORT	
4	BALENO	CIVIC		DELTA	
5		SEPHIA			
6		MAZDA			
7		VECTRA			
8		MEGANE			
9		ROVER			
10		CORDOBA			
11		IMPREZA			

Πίνακας 6.18: Πέντε ομάδες που προκύπτουν με χρήση του αλγορίθμου Nearest Neighbour στους δύο πρώτους κύριους άξονες

Στους πίνακες 6.19 και 6.20 φαίνονται οι τιμές συμμετοχής των ομάδων του Πίνακα 6.18 στις τρεις περιοχές με την εφαρμογή του Normal Συμμετρικού Κανόνα και του Normal μη Συμμετρικού Κανόνα αντίστοιχα. Επίσης φαίνεται η μέση απόσταση των μοντέλων από το κέντρο της ομάδας τους (Radius) και σε το ποσοστό επί του μήκους του πρώτου κύριου άξονα.

Τιμές Συμμετοχής με χρήση του Normal Συμμετρικού Κανόνα										
ΟΜΑΔΑ	Αρ. Ατόμων	Radius	Άξονας 1				Άξονας 2			
			High	Medium	Low	Fuzzy	High	Medium	Low	Fuzzy
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1	4	0.071 (3.28%)	0	43.88	56.12	0	0	0	100	0
2	11	0.190 (8.75%)	39.02	60.98	0	0	16.18	11.1	72.73	0
3	3	0.142 (6.52%)	0	0	100	0	66.67	0	33.33	0
4	4	0.092 (4.24%)	0	32.83	67.17	0	100	0	0	0
5	2	0.128 (5.91%)	100	0	0	0	100	0	0	0

Πίνακας 6.19: Τιμές Συμμετοχής με χρήση του Normal Συμμετρικού Κανόνα στις πέντε ομάδες που προέκυψαν από την εφαρμογή του αλγορίθμου Nearest Neighbour στους δύο πρώτους κύριους άξονες.

Τιμές Συμμετοχής με χρήση του Normal μη Συμμετρικού Κανόνα										
ΟΜΑΔΑ	Αρ. Ατόμων	Radius	Άξονας 1				Άξονας 2			
			High (%)	Medium (%)	Low (%)	Fuzzy (%)	High (%)	Medium (%)	Low (%)	Fuzzy (%)
1	4	0.071 (3.28%)	0	32.59	38.58	28.83	0	0	95.73	4.27
2	11	0.190 (8.75%)	33.48	56.74	0	9.78	13.81	10.43	71.34	4.42
3	3	0.142 (6.52%)	0	0	100	0	66.67	0	33.33	0
4	4	0.092 (4.24%)	0	21.89	57.14	20.97	100	0	0	0
5	2	0.128 (5.91%)	100	0	0	0	100	0	0	0

Πίνακας 6.20: Τιμές Συμμετοχής με χρήση του Normal μη Συμμετρικού Κανόνα στις πέντε ομάδες που προέκυψαν από την εφαρμογή του αλγορίθμου Nearest Neighbour στους δύο πρώτους κύριους άξονες.

Από τη σύγκριση των πινάκων 6.19 και 6.20 παρατηρούμε για το χαρακτηρισμό των ομάδων στον πρώτο κύριο άξονα της ΑΚΣ ότι:

- Τόσο η ομάδα 1 όσο και η ομάδα 4 αποτελούνται από μοντέλα που βρίσκονται μεταξύ της Μεσαίας και Χαμηλής περιοχής χωρίς να ανήκουν 100% σε κάποια από τις δύο. Αυτό αποτυπώνεται από το γεγονός της συμμετοχής των ομάδων με υψηλό ποσοστό και στις δύο περιοχές και από το γεγονός ότι με την εφαρμογή του μη συμμετρικού κανόνα και οι δύο ομάδες εμφανίζουν υψηλό ποσοστό ασάφειας στο χαρακτηρισμό τους (28.83% και 20.97% αντίστοιχα).
- Η ομάδα 2 αποτελείται από μοντέλα που ανήκουν λιγότερο στην Υψηλή περιοχή και περισσότερο στην Μεσαία περιοχή. Από το χαμηλό ποσοστό ασάφειας στο χαρακτηρισμό της ομάδας με την εφαρμογή του μη συμμετρικού κανόνα συμπεραίνουμε ότι αρκετά από τα μοντέλα αυτής της ομάδας χαρακτηρίζονται ως 100% Μεσαία στα χαρακτηριστικά που εκφράζει ο πρώτος κύριος άξονας.
- Τα μοντέλα των ομάδων 3 και 5 ανήκουν με ποσοστό 100% στη Χαμηλή και με 100% στην Υψηλή περιοχή αντίστοιχα.

Για το χαρακτηρισμό των ομάδων στον δεύτερο κύριο άξονα της ΑΚΣ παρατηρούμε ότι:

- Τα μοντέλα της ομάδας 1 χαρακτηρίζονται ως 100% Χαμηλά ενώ τα μοντέλα των ομάδων 4 και 5 χαρακτηρίζονται 100% Υψηλά στα χαρακτηριστικά που αντιπροσωπεύει ο δεύτερος κύριος άξονας.
- Εντύπωση προκαλεί η συμμετοχή των μοντέλων της ομάδας 2 και στις τρεις περιοχές όσο και η συμμετοχή των μοντέλων της ομάδας 3 στην Υψηλή και στη Χαμηλή περιοχή. Αυτό παρ' ότι εκ πρώτης όψεως φαίνεται παράλογο έχει την εξής εξήγηση: Η διασπορά σε κάθε κύριο άξονα είναι το άθροισμα των τετραγώνων των προβολών των ατόμων στον άξονα, όπως δείξαμε στη Σχέση 4.1. Εξάλλου η διασπορά των προβολών των ατόμων σε κάθε κύριο άξονα ισούται με την ιδιοτιμή του άξονα αυτού. Έτσι στον πρώτο κύριο άξονα συναντάμε τη μεγαλύτερη διασπορά, στο δεύτερο τη μεγαλύτερη από την υπολειπόμενη κ.τ.λ. Ο αλγόριθμος της γεωμετρικής ομαδοποίησης δεν λαμβάνει υπόψη του το βάρος του κάθε άξονα, δηλαδή το ποσοστό της αρχικής πληροφορίας που αυτός αντιπροσωπεύει. Επομένως δύο σημεία που έχουν μεταξύ τους απόσταση a στον πρώτο άξονα και ομαδοποιούνται στην ίδια ομάδα είναι απίθανο έως αδύνατο να ανήκουν το ένα στη Χαμηλή περιοχή και το άλλο στην Υψηλή. Δηλαδή στον πρώτο άξονα παρατηρούνται πιο εύκολα “καθαρές” ομάδες (τα σημεία τους ανήκουν στην ίδια περιοχή). Αντίθετα δύο σημεία που έχουν μεταξύ τους απόσταση a στο δεύτερο άξονα και ομαδοποιούνται στην ίδια ομάδα ενδέχεται να βρίσκονται σε διαφορετικές περιοχές. Ιδιαίτερα σε περιπτώσεις που οι ομάδες είναι μεγάλες και εκτείνονται σε μεγάλες περιοχές του γραφήματος τελικά μπορεί να χαρακτηρίζονται και με τις τρεις λεκτικές μεταβλητές.

Αυτό ενώ θα μπορούσε να θεωρηθεί ως μειονέκτημα του αλγορίθμου της ομαδοποίησης ουσιαστικά οδηγεί προς τη σωστή κατεύθυνση. Κι' αυτό διότι εάν υποθέσουμε ότι αναπτύσσαμε τεχνικές ώστε να παράγουμε “καθαρές” ομάδες, θα αναγκαζόμασταν να διασπάσουμε μία ομάδα σαν κι αυτή του παραδείγματος. Έτσι ουσιαστικά θα δίνουμε βάρος στον δεύτερο κύριο άξονα όπου κατά κανόνα έχουμε λιγότερα κοινά χαρακτηριστικά για τα σημεία που απαρτίζουν τις ομάδες, ξεχνώντας ότι τα σημεία αυτά έχουν περισσότερα κοινά χαρακτηριστικά στον πρώτο κύριο

άξονα που αντιπροσωπεύει και τη μέγιστη δυνατή πληροφορία του αρχικού πίνακα δεδομένων.

Ομαδοποίηση σε όλους τους Κύριους Άξονες

Στους πίνακες 6.21 και 6.22 φαίνονται δύο διαφορετικές ομαδοποιήσεις που προέκυψαν με την ομαδοποίηση σε όλους τους κύριους άξονες.

Ομαδοποίηση με αλγόριθμο NN σε όλους τους Κύριους Άξονες									
Αρ. Ατόμων	Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4	Ομάδα 5	Ομάδα 6	Ομάδα 7	Ομάδα 8	Ομάδα 9
1	XSARA	PUNTO	ALFA	AUDI	SIRION	C180	SAAB	CORDOBA	IMPREZA 1.6
2	LANOS	PEUGEOT	CIVIC	BMW					
3	ESCORT		ACCENT	VECTRA					
4	DELTA		SEPHIA						
5			MAZDA						
6			LANCER						
7			ALMERA						
8			MEGANE						
9			ROVER						
10			BALENO						

Πίνακας 6.21: Εννέα ομάδες που προκύπτουν με χρήση του αλγορίθμου Nearest Neighbour σε όλους τους κύριους άξονες

Ομαδοποίηση με αλγόριθμο NN σε όλους τους Κύριους Άξονες					
Αρ. Ατόμων	Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4	Ομάδα 5
1	XSARA	SIRION	ALFA	CORDOBA	C180
2	LANOS	PUNTO	AUDI		SAAB
3	ESCORT	PEUGEOT	BMW		
4	DELTA		CIVIC		
5			ACCENT		
6			SEPHIA		
7			MAZDA		
8			LANCER		
9			ALMERA		
10			VECTRA		
11			MEGANE		
12			ROVER		
13			IMPREZA		
14			BALENO		

Πίνακας 6.22: Πέντε ομάδες που προκύπτουν με χρήση του αλγορίθμου Nearest Neighbour σε όλους τους κύριους άξονες

Από τη σύγκριση των ομαδοποιήσεων για τις εννέα ομάδες στους δύο πρώτους κύριους άξονες (87,35%) και σε όλους τους κύριους άξονες (100%) που φαίνονται στους Πίνακες 6.17 και 6.21 αντίστοιχα παρατηρούμε ότι:

- Πέντε από τις εννέα ομάδες που δημιουργήθηκαν αποτελούνται από τα ίδια μοντέλα.
- Από τα μοντέλα της ομάδας 1 του Πίνακα 6.17 αποχωρεί το μοντέλο IMPREZA και τη θέση του στον Πίνακα 6.21 (ομάδα 4) παίρνει το μοντέλο AUDI.
- Τα μοντέλα των ομάδων 2 και 3 της ομαδοποίησης στους δύο πρώτους κύριους άξονες κατανέμονται διαφορετικά στις ομάδες 3 και 8 της ομαδοποίησης σε όλους τους κύριους άξονες με το μοντέλο CORDOBA να είναι μοναδικό στην ομάδα του.

Από τη σύγκριση των ομαδοποιήσεων για τις πέντε ομάδες στους δύο πρώτους κύριους άξονες και σε όλους τους κύριους άξονες που φαίνονται στους Πίνακες 6.18 και 6.22 αντίστοιχα παρατηρούμε ότι:

- Τρεις από τις πέντε ομάδες που δημιουργήθηκαν αποτελούνται από τα ίδια μοντέλα.
- Τα μοντέλα των ομάδων 2 και 3 της ομαδοποίησης στους δύο πρώτους κύριους άξονες κατανέμονται διαφορετικά στις ομάδες 3 και 4 στην ομαδοποίηση σε όλους τους κύριους άξονες με το μοντέλο CORDOBA να είναι μοναδικό στην ομάδα του.

Συμπερασματικά μπορούμε να πούμε ότι η απώλεια πληροφορίας κατά 100%-87.35%=12.65% στην ομαδοποίηση στους δύο πρώτους κύριους άξονες της ΑΚΣ επηρεάζει την ομαδοποίηση με χρήση του αλγορίθμου Nearest Neighbour περισσότερο για ομάδες που κινούνται στη Μεσαία περιοχή των κυρίων αξόνων, χωρίς όμως αυτή η διαφοροποίηση να κρίνεται σημαντική.

Στους πίνακες 6.23 και 6.24 φαίνονται οι τιμές συμμετοχής των ομάδων στις τρεις περιοχές με την εφαρμογή του Normal Συμμετρικού Κανόνα και του Normal μη Συμμετρικού Κανόνα αντίστοιχα.

Τιμές Συμμετοχής με χρήση του Normal Συμμετρικού Κανόνα										
ΟΜΑΔΑ	Αρ. Ατόμων	Radius	Άξονας 1				Άξονας 2			
			High (%)	Medium (%)	Low (%)	Fuzzy (%)	High (%)	Medium (%)	Low (%)	Fuzzy (%)
1	4	0.092 (4.24%)	0	32.83	67.17	0	100	0	0	0
2	3	0.142 (6.52%)	0	0	100	0	66.67	0	33.33	0
3	14	0.249 (11.47%)	27.6	56.36	16.04	0	12.71	8.72	78.57	0
4	1	0.000 (0%)	42.79	57.21	0	0	0	0	100	0
5	2	0.128 (5.91%)	100	0	0	0	100	0	0	0

Πίνακας 6.23: Τιμές Συμμετοχής με χρήση του Normal Συμμετρικού Κανόνα στις πέντε ομάδες που προέκυψαν από την εφαρμογή του αλγορίθμου Nearest Neighbour σε όλους τους κύριους άξονες.

Τιμές Συμμετοχής με χρήση του Normal μη Συμμετρικού Κανόνα										
ΟΜΑΔΑ	Αρ. Ατόμων	Radius	Άξονας 1				Άξονας 2			
			High (%)	Medium (%)	Low (%)	Fuzzy (%)	High (%)	Medium (%)	Low (%)	Fuzzy (%)
1	4	0.092 (4.24%)	0	21.89	57.14	20.97	100	0	0	0
2	3	0.142 (6.52%)	0	0	100	0	66.67	0	33.33	0
3	14	0.249 (11.47%)	24.27	51.17	11.02	13.54	10.85	8.19	76.26	4.69
4	1	0.000 (0%)	28.53	38.14	0	33.33	0	0	100	0
5	2	0.128 (5.91%)	100	0	0	0	100	0	0	0

Πίνακας 6.24: Τιμές Συμμετοχής με χρήση του Normal μη Συμμετρικού Κανόνα στις πέντε ομάδες που προέκυψαν από την εφαρμογή του αλγορίθμου Nearest Neighbour σε όλους τους κύριους άξονες.

Από τη σύγκριση των πινάκων 6.23 και 6.24 παρατηρούμε για το χαρακτηρισμό των ομάδων στον πρώτο κύριο άξονα της ΑΚΣ ότι:

- Τα μοντέλα της ομάδας 1 χαρακτηρίζονται περισσότερο Χαμηλά και λιγότερο Μεσαία στα χαρακτηριστικά που αντιπροσωπεύει ο πρώτος κύριος άξονας. Από το αρκετά υψηλό ποσοστό ασάφειας (20.97%) που παρατηρείται με την εφαρμογή του μη συμμετρικού κανόνα συμπεραίνουμε ότι τα περισσότερα μοντέλα δεν ανήκουν με ποσοστό 100% σε κάποια από τις δύο αυτές περιοχές.
- Η ομάδα 3 που είναι και η μεγαλύτερη είναι πολυσυλλεκτική. Αποτελείται από μοντέλα που ανήκουν κυρίως στην Μεσαία περιοχή. Το ποσοστό ασάφειας (13.54%) που παρατηρείται με την εφαρμογή του μη συμμετρικού κανόνα μας επιτρέπει να συμπεράνουμε ότι υπάρχουν μοντέλα που βρίσκονται στις περιοχές ασάφειας εκατέρωθεν της Μεσαίας περιοχής.
- Το μοντέλο CORDOBA που είναι το μοναδικό της ομάδας 4 βρίσκεται μεταξύ της σίγουρα Υψηλής και της σίγουρα Μεσαίας περιοχής. Μάλιστα από το υψηλό ποσοστό ασάφειας στο χαρακτηρισμό του με την εφαρμογή του μη συμμετρικού κανόνα (33.33%) φαίνεται ότι βρίσκεται στο μέσο των δύο περιοχών.
- Τα μοντέλα των ομάδων 2 και 5 χαρακτηρίζονται ως 100% Χαμηλά και 100% Υψηλά αντίστοιχα στα χαρακτηριστικά που αντιπροσωπεύονται στον πρώτο κύριο άξονα.

Για το χαρακτηρισμό των ομάδων στον δεύτερο κύριο άξονα της ΑΚΣ παρατηρούμε ότι:

- Τα μοντέλα των ομάδων 1 και 5 χαρακτηρίζονται ως 100% Υψηλά ενώ το μοντέλο της ομάδας 4 χαρακτηρίζεται ως 100% Χαμηλό στα χαρακτηριστικά που αντιπροσωπεύονται στον δεύτερο κύριο άξονα.
- Η ομάδα 3 εξακολουθεί να είναι πολυσυλλεκτική και ισχύουν αυτά που είπαμε παραπάνω
- Δύο από τα τρία μοντέλα της ομάδας 2 χαρακτηρίζονται ως 100% Υψηλά και ένα ως 100% Χαμηλό. Στην περίπτωση αυτή ισχύουν όσα είπαμε στην ομαδοποίησης στους δύο πρώτους άξονες.

6.4.2 Ομαδοποίηση με χρήση του αλγορίθμου Min Distance (MD)

Στη συνέχεια παραθέτουμε τις ομάδες που προκύπτουν με τη χρήση της εκδοχής Min Distance του αλγορίθμου της ομαδοποίησης στους δύο πρώτους κύριους άξονες (87.35% του συνόλου της πληροφορίας) και στους οκτώ κύριους άξονες (100% του συνόλου της πληροφορίας).

Ομαδοποίηση στους δύο πρώτους Κύριους Άξονες

Στους πίνακες 6.25 και 6.26 φαίνονται δύο διαφορετικές ομαδοποιήσεις που προέκυψαν με την ομαδοποίηση στους δύο πρώτους κύριους άξονες.

Ομαδοποίηση με αλγόριθμο MD στους δύο πρώτους Κύριους Άξονες									
Αρ. Ατόμων	Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4	Ομάδα 5	Ομάδα 6	Ομάδα 7	Ομάδα 8	Ομάδα 9
1	BMW	ACCENT	ALFA	AUDI	PUNTO	XSARA	SIRION	C180	SAAB
2	VECTRA	LANCER	CIVIC	CORDOBA	PEUGEOT	LANOS			
3	IMPREZA	ALMERA	SEPHIA			ESCORT			
4		BALENO	MAZDA			DELTA			
5			MEGANE						
6			ROVER						

Πίνακας 6.25: Εννέα ομάδες που προκύπτουν με χρήση του αλγορίθμου Min Distance στους δύο πρώτους κύριους άξονες

Ομαδοποίηση με αλγόριθμο MD στους δύο πρώτους Κύριους Άξονες					
Αρ. Ατόμων	Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4	Ομάδα 5
1	ACCENT	ALFA	XSARA	SIRION	C180
2	LANCER	AUDI	LANOS	PUNTO	SAAB
3	ALMERA	BMW	ESCORT	PEUGEOT	
4	BALENO	CIVIC	DELTA		
5		SEPHIA			
6		MAZDA			
7		VECTRA			
8		MEGANE			
9		ROVER			
10		CORDOBA			
11		IMPREZA			

Πίνακας 6.26: Πέντε ομάδες που προκύπτουν με χρήση του αλγορίθμου Min Distance στους δύο πρώτους κύριους άξονες

Από τη σύγκριση των ομαδοποιήσεων για τις εννέα ομάδες στους δύο πρώτους κύριους άξονες με τον αλγόριθμο του Nearest Neighbour (NN) και με τον αλγόριθμο Min Distance (MD) που φαίνονται στους Πίνακες 6.17 και 6.25 αντίστοιχα φαίνεται η εξής μοναδική διαφορά: Το μοντέλο CORDOBA ομαδοποιείται μαζί με το μοντέλο AUDI με τη χρήση του αλγορίθμου NN αντί με τα μοντέλα ALFA, CIVIC,... με τα οποία ομαδοποιείται με χρήση του αλγορίθμου MD.

Από τη σύγκριση των ομαδοποιήσεων για τις πέντε ομάδες στους δύο πρώτους κύριους άξονες με τον αλγόριθμο του Κοντινού Γείτονα (NN) και με τον αλγόριθμο Ελάχιστης Απόστασης (MD) που φαίνονται στους Πίνακες 6.18 και 6.26 αντίστοιχα φαίνεται ότι οι δύο αλγόριθμοι δημιουργούν ακριβώς τις ίδιες ομάδες.

Ομαδοποίηση σε όλους τους Κύριους Άξονες

Στους πίνακες 6.27 και 6.28 φαίνονται δύο διαφορετικές ομαδοποιήσεις που προέκυψαν με την ομαδοποίηση σε όλους τους κύριους άξονες.

Ομαδοποίηση με αλγόριθμο MD σε όλους τους Κύριους Άξονες									
Αρ. Ατόμων	Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4	Ομάδα 5	Ομάδα 6	Ομάδα 7	Ομάδα 8	Ομάδα 9
1	ACCENT	XSARA	SIRION	BMW	ALFA	C180	SAAB	CORDOBA	IMPREZA
2	LANCER	LANOS	PUNTO	VECTRA	AUDI				
3	ALMERA	ESCORT	PEUGEOT		CIVIC				
4	BALENO	DELTA			SEPHIA				
5					MAZDA				
6					MEGANE				
7					ROVER				

Πίνακας 6.27: Εννέα ομάδες που προκύπτουν με χρήση του αλγορίθμου Min Distance σε όλους τους κύριους άξονες

Ομαδοποίηση με αλγόριθμο MD σε όλους τους Κύριους Άξονες					
Αρ. Ατόμων	Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4	Ομάδα 5
1	XSARA	SIRION	ALFA	C180	CORDOBA
2	LANOS	PUNTO	AUDI	SAAB	
3	ESCORT	PEUGEOT	BMW		
4	DELTA		CIVIC		
5			ACCENT		
6			SEPHIA		
7			MAZDA		
8			LANCER		
9			ALMERA		
10			VECTRA		
11			MEGANE		
12			ROVER		
13			IMPREZA		
14			BALENO		

Πίνακας 6.28: Πέντε ομάδες που προκύπτουν με χρήση του αλγορίθμου Min Distance σε όλους τους κύριους άξονες

Οι διαφορές που προκύπτουν μεταξύ των ομαδοποιήσεων στους δύο πρώτους κύριους άξονες και σε όλους τους κύριους άξονες με χρήση του αλγορίθμου Ελάχιστης Απόστασης είναι:

Για την ομαδοποίηση σε εννέα ομάδες (Πίνακες 6.25 και 6.27) παρατηρούμε ότι:

- Τέσσερις από τις εννέα ομάδες που δημιουργήθηκαν αποτελούνται από τα ίδια μοντέλα.
- Με την ομαδοποίηση σε όλους τους κύριους άξονες το μοντέλο SIRION ομαδοποιείται με τα μοντέλα PUNTO και PEUGEOT, το μοντέλο AUDI ομαδοποιείται με τα μοντέλα ALFA, CIVIC κ.τ.λ., ενώ το μοντέλο IMPREZA ομαδοποιείται μόνο του αντί με τα μοντέλα BMW και VECTRA.

Για την ομαδοποίηση σε πέντε ομάδες (Πίνακες 6.26 και 6.28) παρατηρούμε ότι:

- Τρεις από τις πέντε ομάδες που δημιουργήθηκαν αποτελούνται από τα ίδια μοντέλα.
- Τα μοντέλα των ομάδων 1 και 2 του Πίνακα 6.26 κατανέμονται στις ομάδες 3 και 5 του Πίνακα 6.28 με το μοντέλο CORDOBA να είναι το μοναδικό στην ομάδα του.

Οι διαφορές που προκύπτουν μεταξύ των ομαδοποιήσεων σε εννέα ομάδες σε όλους τους κύριους άξονες με χρήση των αλγορίθμων Nearest Neighbour και Min Distance (Πίνακες 6.21 και 6.27) είναι:

- Πέντε από τις εννέα ομάδες που δημιουργήθηκαν αποτελούνται από τα ίδια μοντέλα.
- Με τη χρήση του αλγορίθμου MD το μοντέλο SIRION ομαδοποιείται με τα μοντέλα PUNTO και PEUGEOT, τα μοντέλα ACCENT, LANCER, ALMERA, BALENO αποτελούν μία ομάδα και το μοντέλο AUDI ομαδοποιείται με τα μοντέλα ALFA, CIVIC κ.τ.λ., αντί με τα μοντέλα BMW και VECTRA όπως συνέβη με τη χρήση του αλγορίθμου NN.

Μεταξύ των ομαδοποιήσεων σε πέντε ομάδες σε όλους τους κύριους άξονες με χρήση των αλγορίθμων Nearest Neighbour και Min Distance (Πίνακες 6.22 και 6.28) δεν υπάρχουν διαφορές. Οι δύο αλγόριθμοι δημιουργούν ακριβώς τις ίδιες ομάδες.

Στην περίπτωση της εφαρμογής του αλγορίθμου Min Distance δεν παραθέτουμε τις τιμές συμμετοχής των πέντε ομάδων στους δύο πρώτους κύριους άξονες. Κι αυτό γιατί δεδομένου ότι οι δύο εκδοχές του αλγορίθμου δίνουν την ίδια ομαδοποίηση οι γεωμετρικές συντεταγμένες των κέντρων των ομάδων θα είναι κι αυτές ίδιες. Επομένως οι τιμές συμμετοχής των ομάδων που δίνονται από τη Σχέση 4.2 θα είναι αυτές που φαίνονται στους Πίνακες 6.19, 6.20, 6.23 και 6.24.

6.5 ΑΝΑΛΥΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΟΜΑΔΟΠΟΙΗΣΗΣ

Με την εντολή Grouping→Grouping Results... από τη φόρμα του γραφήματος της ΑΚΣ λαμβάνουμε τη φόρμα των αποτελεσμάτων της ομαδοποίησης. Από τη φόρμα αυτή είναι δυνατή η περιήγηση σε όλα τα αποτελέσματα. Παρακάτω παραθέτουμε δύο πίνακες με αναλυτικά αποτελέσματα για τις πέντε ομάδες που δημιουργήθηκαν με τη χρήση της εκδοχής Nearest Neighbour του αλγορίθμου της ομαδοποίησης στους δύο πρώτους κύριους άξονες, εφαρμόζοντας τον Normal Συμμετρικό Κανόνα και τον Normal μη Συμμετρικό Κανόνα και ένα πίνακα με λεπτομέρειες της ομαδοποίησης.

Στον Πίνακα 6.29 φαίνονται οι λεπτομέρειες της ομαδοποίησης.

Grouping Details	
Nr. of Individuals	24
Nr. of Groups	5
Average Nr. of Indiv./Group	4.8
Average Radius	0.125 (5.77%)

Πίνακας 6.29: Λεπτομέρειες της ομαδοποίησης

Οι τιμές στον πίνακα αυτό δίνονται από τις σχέσεις του κεφαλαίου 5.8.1.

Στον Πίνακα 6.30 φαίνονται τα αποτελέσματα με την εφαρμογή του Normal Συμμετρικού Κανόνα.

			Άξονας 1				Άξονας 2			
	MONTELO	Radius	High (%)	Medium (%)	Low (%)	Fuzzy (%)	High (%)	Medium (%)	Low (%)	Fuzzy (%)
ΟΜΑΔΑ 1	ACCENT 1.3	0.079	0	0	100	0	0	0	100	0
	LANCER 1.3 GL	0.02	0	49.45	50.55	0	0	0	100	0
	ALMERA 1.4 GX	0.121	0	100	0	0	0	0	100	0
	BALENO 1.3 GL	0.065	0	26.06	73.94	0	0	0	100	0
	ΟΜΑΔΑ 1	0.071	0	43.88	56.12	0	0	0	100	0
ΟΜΑΔΑ 2	ALFA 145 1.4	0.084	0	100	0	0	0	0	100	0
	AUDI A3 1.6	0.14	100	0	0	0	0	0	100	0
	BMW 316 Compact	0.305	100	0	0	0	77.95	22.05	0	0
	CIVIC 1.4	0.266	0	100	0	0	0	0	100	0
	SEPHIA 2 LS	0.194	0	100	0	0	0	0	100	0
	MAZDA 323 1.5	0.132	0	100	0	0	0	0	100	0
	VECTRA 1.6	0.301	100	0	0	0	0	100	0	0
	MEGANE 1.4	0.135	0	100	0	0	0	0	100	0
	ROVER 214	0.25	0	100	0	0	0	0	100	0
	CORDOBA 1.4	0.043	42.79	57.21	0	0	0	0	100	0
	IMPREZA 1.6	0.24	86.44	13.56	0	0	100	0	0	0
	ΟΜΑΔΑ 2	0.190	39.02	60.98	0	0	16.18	11.1	72.73	0
ΟΜΑΔΑ 3	SIRION CL	0.199	0	0	100	0	0	0	100	0
	PUNTO 60	0.11	0	0	100	0	100	0	0	0
	PEUGEOT 206	0.116	0	0	100	0	100	0	0	0
	ΟΜΑΔΑ 3	0.142	0	0	100	0	66.67	0	33.33	0
ΟΜΑΔΑ 4	XSARA 1.4 LX	0.133	0	71.99	28.01	0	100	0	0	0
	LANOS 1.3	0.061	0	0	100	0	100	0	0	0
	ESCORT 1.4	0.115	0	0	100	0	100	0	0	0
	DELTA 1.4	0.059	0	59.33	40.67	0	100	0	0	0
	ΟΜΑΔΑ 4	0.092	0	32.83	67.17	0	100	0	0	0
ΟΜΑΔΑ 5	C180 CLASSIC	0.128	100	0	0	0	100	0	0	0
	SAAB 9-3	0.128	100	0	0	0	100	0	0	0
	ΟΜΑΔΑ 4	0.128	100	0	0	0	100	0	0	0

Πίνακας 6.30: Αποτελέσματα ομαδοποίησης με την εκδοχή Nearest Neighbour σε πέντε ομάδες και εφαρμογή του Normal Συμμετρικού Κανόνα

Στον Πίνακα 6.31 φαίνονται τα αποτελέσματα με την εφαρμογή του Normal μη Συμμετρικού Κανόνα.

			Άξονας 1				Άξονας 2			
MONTEΛΟ		Radius	High (%)	Medium (%)	Low (%)	Fuzzy (%)	High (%)	Medium (%)	Low (%)	Fuzzy (%)
ΟΜΑΔΑ 1	ACCENT 1.3	0.079	0.0	0.0	71.3	28.7	0.0	0.0	100.0	0.0
	LANCER 1.3 GL	0.020	0.0	33.0	33.7	33.3	0.0	0.0	100.0	0.0
	ALMERA 1.4 GX	0.121	0.0	80.0	0.0	20.0	0.0	0.0	82.9	17.1
	BALENO 1.3 GL	0.065	0.0	17.4	49.3	33.3	0.0	0.0	100.0	0.0
	ΟΜΑΔΑ 1	0.071	0.0	32.6	38.6	28.8	0.0	0.0	95.7	4.3
ΟΜΑΔΑ 2	ALFA 145 1.4	0.084	0.0	77.0	0.0	23.0	0.0	0.0	100.0	0.0
	AUDI A3 1.6	0.140	82.1	0.0	0.0	17.9	0.0	0.0	100.0	0.0
	BMW 316 Compact	0.305	100.0	0.0	0.0	0.0	52.0	14.7	0.0	33.3
	CIVIC 1.4	0.266	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0
	SEPHIA 2 LS	0.194	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0
	MAZDA 323 1.5	0.132	0.0	100.0	0.0	0.0	0.0	0.0	84.7	15.3
	VECTRA 1.6	0.301	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	MEGANE 1.4	0.135	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0
	ROVER 214	0.250	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0
	CORDOBA 1.4	0.043	28.5	38.1	0.0	33.3	0.0	0.0	100.0	0.0
	IMPREZA 1.6	0.240	57.6	9.0	0.0	33.3	100.0	0.0	0.0	0.0
	ΟΜΑΔΑ 2	0.190	33.5	56.7	0.0	9.8	13.8	10.4	71.3	4.4
ΟΜΑΔΑ 3	SIRION CL	0.199	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	PUNTO 60	0.110	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0
	PEUGEOT 206	0.116	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0
	ΟΜΑΔΑ 3	0.142	0.0	0.0	100.0	0.0	0.0	91.9	0.0	8.1
ΟΜΑΔΑ 4	XSARA 1.4 LX	0.133	0.0	48.0	18.7	33.3	100.0	0.0	0.0	0.0
	LANOS 1.3	0.061	0.0	0.0	82.8	17.2	100.0	0.0	0.0	0.0
	ESCORT 1.4	0.115	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0
	DELTA 1.4	0.059	0.0	39.6	27.1	33.3	100.0	0.0	0.0	0.0
	ΟΜΑΔΑ 4	0.092	0.0	21.9	57.1	21.0	100.0	0.0	0.0	0.0
ΟΜΑΔΑ 5	C180 CLASSIC	0.128	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	SAAB 9-3	0.128	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	ΟΜΑΔΑ 5	0.128	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0

Πίνακας 6.31: Αποτελέσματα ομαδοποίησης με την εκδοχή Nearest Neighbour σε πέντε ομάδες και εφαρμογή του Normal μη Συμμετρικού Κανόνα

ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΟΝΤΙΚΕΣ ΠΡΟΟΠΤΙΚΕΣ

Στην παρούσα διπλωματική εργασία παρουσιάστηκε μία μέθοδος λεκτικού χαρακτηρισμού με τη χρήση Ασαφούς Λογικής στις ομάδες που προέκυψαν από τη γεωμετρική ομαδοποίηση των ατόμων σε ένα γράφημα Ανάλυσης σε Κύριες Συνιστώσες.

Σ' αυτό το κεφάλαιο θα προσπαθήσουμε να εξάγουμε γενικά συμπεράσματα για την αποτελεσματικότητα της μεθοδολογίας που προτάθηκε και της εφαρμογής που αναπτύχθηκε.

7.1 Η ΜΕΘΟΔΟΣ

7.1.1 Ο Αλγόριθμος της Ομαδοποίησης

Ο αλγόριθμος της γεωμετρικής ομαδοποίησης (δύο εκδοχές) αναπτύχθηκε με βασικό στόχο την αντιμετώπιση των ποικίλων αρχικών δεδομένων της ΑΚΣ.

Γενική συμπεριφορά του αλγορίθμου

Από την εφαρμογή του αλγορίθμου της ομαδοποίησης προκύπτουν τα εξής συμπεράσματα:

Σε γενικές γραμμές ο αλγόριθμος ικανοποιεί με τη συμπεριφορά του παρουσιάζοντας ομαδοποιήσεις που ανταποκρίνονται στην πραγματικότητα. Η εφαρμογή του σε πλήθος δεδομένων έδειξε ότι έχει την τάση να ομαδοποιεί ικανοποιητικά ακόμα και σε περιπτώσεις πυκνού νέφους σημείων, αλυσίδων και άλλων ιδιαίτερων περιπτώσεων πλην ελαχίστων εξαιρέσεων. Μία τέτοια περίπτωση είναι όταν ανάμεσα στα δεδομένα υπάρχουν άτομα με εξαιρετικά μεγάλες ή μικρές προβολές οπότε όλα τα υπόλοιπα παρουσιάζονται όμοια σε σύγκριση με αυτά. Το αποτέλεσμα είναι η δημιουργία ομάδων με πάρα πολλά και πάρα πολύ λίγα άτομα. Δηλαδή παρουσιάζεται άνιση κατανομή των ατόμων στις ομάδες. Ωστόσο το πρόβλημα αυτό αντιμετωπίζεται με την αφαίρεση από τα αρχικά δεδομένα αυτών των ατόμων και την εφαρμογή του αλγορίθμου στα εναπομείναντα.

Σε περιπτώσεις που ο αλγόριθμος παρουσιάζει μη επιθυμητές ομαδοποιήσεις είναι δυνατή η καθοδήγησή του με κατάλληλες ρυθμίσεις όπως είναι η επιλογή ομαδοποίησης σε περισσότερους άξονες και ο πειραματισμός με τα δύο κριτήρια τερματισμού.

Διαφορές εκδοχών Nearest Neighbour και Min Distance

Η εκδοχή Nearest Neighbour του αλγορίθμου της ομαδοποίησης μπορούμε να πούμε ότι προσεγγίζει περισσότερο τη μορφή του αλγορίθμου που θα επιζητούσαμε για ενσωμάτωση στην εφαρμογή Fuzzy Data Grouping. Κι αυτό γιατί δεδομένου ότι αναζητούμε ένα αλγόριθμο που θα χειρίζεται ενιαία διάφορους τύπους δεδομένων, αναπτύσσει τεχνικές αποφυγής ιδιαίτερων περιπτώσεων όπως είναι για παράδειγμα αλυσίδες στα δεδομένα, μη σαφείς οπτικοί διαχωρισμοί και πυκνά νέφη σημείων.

Η εκδοχή Min Distance είναι ελαφρώς γρηγορότερη όμως θα πρέπει να προτιμάται μόνο σε περιπτώσεις πυκνού νέφους σημείων οπότε και τα αποτελέσματα δεν διαφέρουν σημαντικά από την εκδοχή Nearest Neighbour.

Ενδιαφέρον παρουσιάζει το γεγονός ότι περιορίζοντας τον αριθμό των ομάδων οι δύο αλγόριθμοι τείνουν προς παρόμοια αποτελέσματα κάτι που σημαίνει ότι είναι δυνατή η χρήση και των δύο εκδοχών σε περιπτώσεις που ο επιθυμητός αριθμός ομάδων είναι μικρός, ακόμα και όταν δεν έχουμε πυκνό νέφος σημείων.

7.1.2 Λεκτικός Χαρακτηρισμός των Ομάδων

Ο λεκτικός χαρακτηρισμός των ομάδων που προκύπτουν κρίνεται επίσης ικανοποιητικός. Η δυνατότητα επιλογής προτεινόμενων κανόνων και η εφαρμογή τους αναλογικά σε όλους τους κύριους άξονες με ρυθμιστή την τυπική απόκλιση των ατόμων σε κάθε κύριο άξονα επιτρέπει τον ορισμό συναρτήσεων συμμετοχής που ανταποκρίνονται σε διάφορους τύπους δεδομένων και κατανομές των ατόμων στο γράφημα της ΑΚΣ.

Στις περιπτώσεις που παρατηρείται χαρακτηρισμός των ομάδων και με τις τρεις λεκτικές μεταβλητές, η εξήγηση είναι η μικρή διασπορά των ατόμων στον άξονα αυτό άρα και η μικρή τυπική απόκλιση που είναι ρυθμιστής των ορίων των συναρτήσεων συμμετοχής. Αυτό το φαινόμενο παρατηρείται επομένως σε άξονες που αντιπροσωπεύουν μικρό

ποσοστό της αρχικής πληροφορίας. Σε αυτούς τους άξονες τα άτομα έχουν λίγα κοινά χαρακτηριστικά. Επομένως δεν θεωρείται σκόπιμη η ανάπτυξη τεχνικών διάσπασης τέτοιων ομάδων αφού έτσι θα αγνοούσαμε το γεγονός ότι τα άτομα που απαρτίζουν τέτοιες ομάδες έχουν πολύ περισσότερα κοινά χαρακτηριστικά σε άξονες με μεγαλύτερη διασπορά, δηλαδή που αντιπροσωπεύουν μεγάλο ποσοστό της πληροφορίας του αρχικού πίνακα δεδομένων.

7.2 Η ΕΦΑΡΜΟΓΗ

Η εφαρμογή Fuzzy Data Grouping αναπτύχθηκε σε περιβάλλον Windows κάτι που την καθιστά φιλική προς τον χρήστη ιδιαίτερα αν αναλογιστεί κανείς ότι η δυνατότητα σχεδίασης της αλληλεπίδρασης χρήστη-μηχανής με κανόνες Windows Standards μπορεί να προσφέρει την εξοικείωση των χρηστών σε σύντομο χρονικό διάστημα. Σημαντικό είναι επίσης το γεγονός ότι τα όρια της εφαρμογής σε όγκο δεδομένων συμπίπτουν με τα όρια του συστήματος στο οποίο είναι εγκατεστημένη η εφαρμογή.

Δεδομένου επίσης ότι η απόκριση της μεθόδου κρίνεται ικανοποιητική, η εφαρμογή Fuzzy Data Grouping μπορεί να αποτελέσει χρήσιμο εργαλείο σε οποιονδήποτε έχει να αντιμετωπίσει προβλήματα ομαδοποίησης και χαρακτηρισμού των ομάδων ξεκινώντας από ένα αρχικό πίνακα δεδομένων σαν αυτού της ΑΚΣ. Η ποικιλία των δεδομένων στα οποία η εφαρμογή μπορεί να δώσει απαντήσεις είναι τόση όση και η ποικιλία των πηγών των δεδομένων.

7.3 ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΟΠΤΙΚΕΣ ΚΑΙ ΠΡΟΤΑΣΕΙΣ

Διάφορες προοπτικές υπάρχουν για τη βελτίωση τόσο της μεθόδου όσο και της εφαρμογής. Μερικές από αυτές είναι:

- Δημιουργία Input Data Wizard για ευκολότερη εισαγωγή των δεδομένων στην εφαρμογή.

- Χρήση τεχνικών αριθμητικής ανάλυσης για την επιτάχυνση του ήδη υπάρχοντος αλγορίθμου γεωμετρικής ομαδοποίησης.
- Ενσωμάτωση ταχύτερων αλγορίθμων γεωμετρικής ομαδοποίησης.
- Ανάπτυξη τεχνικών για πρόταση αριθμού των δημιουργούμενων ομάδων με κριτήριο την ακτίνα των ομάδων σε σχέση με την ακτίνα όλων των αρχικών των δεδομένων.
- Χρήση καμπανοειδών Συναρτήσεων Συμμετοχής αντί για τραπεζοειδείς.
- Ορισμός περισσότερων λεκτικών μεταβλητών για το χαρακτηρισμό των ομάδων που προκύπτουν.
- Προσεκτικότερη μελέτη των πρότυπων κανόνων με στόχο την αντιμετώπιση περισσότερων δυνατών περιπτώσεων και πρόταση κανόνων στο χρήστη.
- Δυνατότητα δημιουργίας πρότυπων κανόνων από τον χρήστη και δυνατότητα αποθήκευσής τους.
- Ανάπτυξη τεχνικών για παραγωγή καθαρών ομάδων (π.χ. 100%-High).
- Ενσωμάτωση του Ms Agent για εκφώνηση των αποτελεσμάτων.

BIBΛΙΟΓΡΑΦΙΑ

1. **Auto Motor and Sport**, No 14 (22 Jul. 1999), Greek Edition.
2. Bandemer Hans (1993), **Fuzzy Analysis of Fuzzy Data**, in *"Fuzzy Logic"*, R. Lowen and M. Roubens (eds.), Kluwer Academic Publishers, 385-394, The Netherlands.
3. Campos L. M. and Moral S. (1990), **Fuzzy Measures with Different Levels of Granularity**, in *"Progress in Fuzzy Sets and Systems"*, Wolfgang H. Janko, M. Roubens and H.-J. Zimmermann (eds.), Kluwer Academic Publishers, 134-146, The Netherlands.
4. Dubois D. and Prade H. (1984), **Criteria Aggregation and Ranking of Alternatives in the Framework of Fuzzy Set Theory**, in *"Fuzzy Sets and Decision Analysis"*, H.J. Zimmermann, L.A. Zadeh, B.R. Gaines (eds.), TIMS Studies in the Management Sciences, Vol 20, 209-240, North Holland.
5. Everitt, B.S. (1993), **Cluster Analysis (third edition)**, Wiley, New York.
6. Grabisch M., Nguyen H. T. and Walker E. A. (1995), **Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference**, Kluwer Academic Publishers, The Netherlands.
7. Ildiko E. Frank and Todeschini R., (1994), **The Data Analysis Handbook**, Elsevier Science Publishers B.V., Vol. 14, Amsterdam.
8. Jackson J.E. (1991), **A User's Guide to Principal Components Analysis**, Wiley, New York.
9. Jamshidi M. (1993), **Fuzzy Logic Software and Hardware**, in *"Fuzzy Logic and Control Software and Hardware Applications"*, M. Jamshidi at al. (eds.), 112-148, PTR Prentice Hall, New Jersey.
10. Jolliffe I.T. (1986), **Principal Component Analysis**, Springer-Verlag, New York.
11. Lebart L., Morineu A. and Warwick K.M. (1984), **Multivariate Descriptive Statistical Analysis**, Wiley, New York.
12. Mandic N.J. and Mamdani E.H. (1984), **A Multi-Attribute Decision-Making Model with Fuzzy Rule Based Modification of Priorities**, in *"Fuzzy Sets and Decision*

- Analysis*", H.J. Zimmermann, L.A. Zadeh, B.R. Gaines (eds.), TIMS Studies in the Management Sciences, Vol 20, 285-306, North Holland.
13. McNeil D. and Freiberger P. (1993), **Fuzzy Logic**, Daniel McNeil and Paul Freiberger, New York.
 14. Perry Greg (1998), **Εγχειρίδιο της Visual Basic 6**, Εκδόσεις Γκιούρδας, Αθήνα.
 15. Simpson Alan (1997), **Πλήρες Εγχειρίδιο της Access 97**, Εκδόσεις Γκιούρδας, Αθήνα.
 16. Tong R.M. and Bonissone P.P. (1984), **Linguistic Solutions to Fuzzy Decision Problems**, in *"Fuzzy Sets and Decision Analysis"*, H.J. Zimmermann, L.A. Zadeh, B.R. Gaines (eds.), TIMS Studies in the Management Sciences, Vol. 20, 323-334, North Holland.
 17. Zimmermann H.-J. (1996), **Fuzzy Set Theory and its Applications (third edition)**, Kluwer Academic Publishers, USA.
 18. Παναγιωτόπουλος X. (1999), **Διακριτά Μαθηματικά**, Α. Σταμούλης, Αθήνα.
 19. Σκιαδάς Χρήστος (1996), **Ανάλυση Δεδομένων**, Σημειώσεις Παραδόσεων, Χανιά.
 20. Φλώρου Γ. (1999), **Χρήση κατάλληλων μετρικών και μεθόδων για την επεξεργασία και Ανάλυση Δεδομένων με μεθόδους Ιεραρχικής Ταξινόμησης**, Πρακτικά 10ου Πανελληνίου Συνεδρίου Στατιστικής, Πειραιάς.