

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ & ΔΙΟΙΚΗΣΗΣ



Τεχνικές Προγραμματισμού Στόχων στην Ανάπτυξη Μηχανών  
Διανύσματος Υποστήριξης

Διατριβή που υπεβλήθη για την μερική ικανοποίηση των απαιτήσεων για την  
απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης

Υπό  
ΛΑΡΙΣΑ ΗΛΙΑ

Χανιά 2005

© Copyright υπό Λάρισα Ηλία  
Χανιά 2005

Η διατριβή του Λάρισα Ηλία εγκρίνεται:

Ζοπουνίδης Κωνσταντίνος:

Δούμπος Μιχάλης:

Γρηγορούδης Ευάγγελος:

Περιεχόμενα

<b>ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ.....</b>	<b>6</b>
<b>ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ .....</b>	<b>6</b>
<b>ΕΥΧΑΡΙΣΤΙΕΣ.....</b>	<b>7</b>
<b>ΒΙΟΓΡΑΦΙΚΟ .....</b>	<b>9</b>
<b>ΠΕΡΙΛΗΨΗ .....</b>	<b>10</b>
<b>ΕΙΣΑΓΩΓΗ.....</b>	<b>11</b>

Κεφάλαιο 1<sup>ο</sup>

<b>1. ΟΡΙΣΜΟΣ ΤΑΞΙΝΟΜΗΣΗΣ .....</b>	<b>15</b>
<b>2. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ .....</b>	<b>17</b>
<b>3. ΟΡΙΣΜΟΣ ΤΩΝ ΚΑΤΗΓΟΡΙΩΝ .....</b>	<b>17</b>
<b>4. ΑΚΡΙΒΕΙΑ ΤΑΞΙΝΟΜΗΣΗΣ .....</b>	<b>18</b>
<b>5. ΟΡΙΣΜΟΙ ΒΑΣΙΚΩΝ ΕΝΝΟΙΩΝ ΚΑΙ ΜΑΘΗΜΑΤΙΚΗ ΔΙΑΤΥΠΩΣΗ .....</b>	<b>19</b>
<b>6. ΠΕΔΙΑ ΕΦΑΡΜΟΓΗΣ .....</b>	<b>19</b>

Κεφάλαιο 2<sup>ο</sup>

<b>1. ΕΙΣΑΓΩΓΗ.....</b>	<b>22</b>
<b>2. ΘΕΩΡΙΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΣΤΟΧΩΝ.....</b>	<b>23</b>
<b>3. ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΣΤΟΧΩΝ ΚΑΙ ΤΑΞΙΝΟΜΗΣΗ.....</b>	<b>25</b>
3.1 ΜΟΝΤΕΛΟ ΜΕΓΙΣΤΟΠΟΙΗΣΗΣ ΤΗΣ ΕΛΑΧΙΣΤΗΣ ΑΠΟΣΤΑΣΗΣ (MMD).....	26
3.2 ΜΟΝΤΕΛΟ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ΤΩΝ ΑΠΟΚΛΙΣΕΩΝ (OSD) .....	28
3.3 ΥΒΡΙΔΙΚΟ ΜΟΝΤΕΛΟ .....	30

Κεφάλαιο 3<sup>ο</sup>

<b>1. ΕΙΣΑΓΩΓΗ.....</b>	<b>33</b>
<b>2. Η ΓΡΑΜΜΙΚΗ ΠΕΡΙΠΤΩΣΗ .....</b>	<b>34</b>
2.1 ΔΙΑΧΩΡΙΣΙΜΑ ΣΤΟΙΧΕΙΑ.....	34
2.2 ΜΗ ΔΙΑΧΩΡΙΣΙΜΑ ΣΤΟΙΧΕΙΑ.....	36
2.3 ΕΝΔΕΙΚΤΙΚΑ ΠΑΡΑΔΕΙΓΜΑΤΑ .....	38
<b>3. ΜΗ ΓΡΑΜΜΙΚΗ ΠΕΡΙΠΤΩΣΗ.....</b>	<b>39</b>
<b>4. ΜΟΝΤΕΛΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΣΤΟΧΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΣΤΗΝ ΑΝΑΠΤΥΞΗ ΜΗΧΑΝΩΝ ΔΙΑΝΥΣΜΑΤΟΣ ΥΠΟΣΤΗΡΙΞΗΣ.....</b>	<b>40</b>
4.1 1 <sup>ο</sup> ΜΟΝΤΕΛΟ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΣΤΟΧΩΝ .....	40
4.1.1 Περίπτωση διαχωρίσιμων στοιχείων .....	41
4.1.2 Περίπτωση μη διαχωρίσιμων στοιχείων .....	42
4.2 2 <sup>ο</sup> ΜΟΝΤΕΛΟ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΣΤΟΧΩΝ .....	43
4.2.1 Περίπτωση διαχωρίσιμων στοιχείων .....	43
4.2.2 Περίπτωση μη διαχωρίσιμων στοιχείων .....	44
4.3 3 <sup>ο</sup> ΜΟΝΤΕΛΟ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΣΤΟΧΩΝ .....	45
4.4 4 <sup>ο</sup> ΜΟΝΤΕΛΟ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΣΤΟΧΩΝ .....	46
4.4.1 Γραμμική περίπτωση .....	46
4.4.2 Μη γραμμική περίπτωση.....	48

Κεφάλαιο 4<sup>ο</sup>

<b>1. ΕΙΣΑΓΩΓΗ.....</b>	<b>51</b>
-------------------------	-----------

<b>2. ΔΕΔΟΜΕΝΑ .....</b>	<b>52</b>
<b>3. Η ΜΕΘΟΔΟΛΟΓΙΑ CROSS-VALIDATION .....</b>	<b>53</b>
<b>4. ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ .....</b>	<b>54</b>
4.1 ΚΡΙΤΗΡΙΑ ΑΞΙΟΛΟΓΗΣΗΣ .....	54
4.2 ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΓΡΑΜΜΙΚΟ ΠΥΡΗΝΑ .....	57
4.3 ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΠΟΛΥΩΝΥΜΙΚΟ ΠΥΡΗΝΑ 2 <sup>ΟΥ</sup> ΒΑΘΜΟΥ .....	60
4.4 ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΠΟΛΥΩΝΥΜΙΚΟ ΠΥΡΗΝΑ 3 <sup>ΟΥ</sup> ΒΑΘΜΟΥ .....	61
4.5 ΣΥΝΟΨΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ .....	63
<i>Κεφάλαιο 5<sup>ο</sup></i>	
<b>ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ.....</b>	<b>69</b>
 <b>ΠΑΡΑΡΤΗΜΑ .....</b>	 <b>72</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ .....</b>	<b>85</b>

## Ευρετήριο Σχημάτων

ΣΧΗΜΑ 2. 1: ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΤΗΣ ΕΛΑΧΙΣΤΗΣ ΑΠΟΣΤΑΣΗΣ ΣΤΟ ΜΟΝΤΕΛΟ MMD (ΠΗΓΗ: ΖΟΠΟΥΝΙΔΗΣ, 1998) .....	27
ΣΧΗΜΑ 2.2: ΑΝΑΠΑΡΑΣΤΑΣΗ ΤΩΝ ΣΦΑΛΜΑΤΩΝ ΚΑΙ ΤΩΝ ΑΠΟΣΤΑΣΕΩΝ ΣΤΟ ΜΟΝΤΕΛΟ OSD (ΠΗΓΗ: ΖΟΠΟΥΝΙΔΗΣ, 1998) .....	29
ΣΧΗΜΑ 2.3: ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΥΒΡΙΔΙΚΟΥ ΜΟΝΤΕΛΟΥ (ΠΗΓΗ: ΖΟΠΟΥΝΙΔΗΣ, 1998).....	31
ΣΧΗΜΑ 3.1: ΓΡΑΜΜΙΚΗ ΠΕΡΙΠΤΩΣΗ ΥΠΕΡΕΠΙΠΕΔΩΝ ΔΙΑΧΩΡΙΣΜΟΥ ΓΙΑ ΔΙΑΧΩΡΙΣΙΜΑ ΔΕΔΟΜΕΝΑ. ΤΑ ΔΙΑΝΥΣΜΑΤΑ ΥΠΟΣΤΗΡΙΞΗΣ ΒΡΙΣΚΟΝΤΑΙ ΠΑΝΩ ΣΤΙΣ ΔΙΑΚΕΚΟΜΜΕΝΕΣ ΓΡΑΜΜΕΣ. (ΠΗΓΗ: BURGESS, 1998) .....	35
ΣΧΗΜΑ 3.2: ΓΡΑΜΜΙΚΗ ΠΕΡΙΠΤΩΣΗ ΥΠΕΡΕΠΙΠΕΔΩΝ ΓΙΑ ΜΗ ΔΙΑΧΩΡΙΣΙΜΑ ΣΤΟΙΧΕΙΑ. (ΠΗΓΗ: BURGESS, 1998) .....	38
ΣΧΗΜΑ 3.3: ΓΡΑΜΜΙΚΗ ΠΕΡΙΠΤΩΣΗ ΥΠΕΡΕΠΙΠΕΔΩΝ ΓΙΑ ΜΗ ΔΙΑΧΩΡΙΣΙΜΑ ΣΤΟΙΧΕΙΑ. (ΠΗΓΗ: BURGESS, 1998) .....	38
ΣΧΗΜΑ 4.1: ΕΦΑΡΜΟΓΗ ΤΟΥ 10-FOLD CROSS-VALIDATION ΣΤΟ ΣΥΝΟΛΟ ΑΝΤΙΚΕΙΜΕΝΩΝ A (ΠΗΓΗ: EFRON ΚΑΙ TIBSHIRANI, 1993).....	54
ΣΧΗΜΑ 4.2: ΔΙΑΓΡΑΜΜΑ ROC.....	56

## Ευρετήριο Πινάκων

ΠΙΝΑΚΑΣ 4.1: ΤΑ ΕΞΕΤΑΖΟΜΕΝΑ ΔΕΔΟΜΕΝΑ.....	52
ΠΙΝΑΚΑΣ 4.2: ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΙΚΤΗ ΑΚΡΙΒΕΙΑΣ ΚΑΙ ΠΟΣΟΣΤΟΥ ΑΚΡΙΒΕΙΑΣ ΓΙΑ ΓΡΑΜΜΙΚΟ ΠΥΡΗΝΑ .	58
ΠΙΝΑΚΑΣ 4.3: ΠΙΝΑΚΑΣ ΥΠΕΡΟΧΗΣ ΜΟΝΤΕΛΩΝ ΓΙΑ ΔΕΙΚΤΗ ΑΚΡΙΒΕΙΑΣ ΚΑΙ ΠΟΣΟΣΤΟ ΑΚΡΙΒΕΙΑΣ ΣΤΗΝ ΓΡΑΜΜΙΚΗ ΠΕΡΙΠΤΩΣΗ.....	59
ΠΙΝΑΚΑΣ 4.4: ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΙΚΤΗ ΑΚΡΙΒΕΙΑΣ ΚΑΙ ΠΟΣΟΣΤΟΥ ΑΚΡΙΒΕΙΑΣ ΓΙΑ ΠΟΛΥΩΝΥΜΙΚΟ ΠΥΡΗΝΑ 2 <sup>ΟΥ</sup> ΒΑΘΜΟΥ.....	60
ΠΙΝΑΚΑΣ 4.5: ΠΙΝΑΚΑΣ ΥΠΕΡΟΧΗΣ ΜΟΝΤΕΛΩΝ ΓΙΑ ΔΕΙΚΤΗ ΑΚΡΙΒΕΙΑΣ ΚΑΙ ΠΟΣΟΣΤΟ ΑΚΡΙΒΕΙΑΣ ΣΤΗΝ ΠΕΡΙΠΤΩΣΗ ΠΟΛΥΩΝΥΜΙΚΟΥ ΠΥΡΗΝΑ 2 <sup>ΟΥ</sup> ΒΑΘΜΟΥ.....	61
ΠΙΝΑΚΑΣ 4.6: ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΙΚΤΗ ΑΚΡΙΒΕΙΑΣ ΚΑΙ ΠΟΣΟΣΤΟΥ ΑΚΡΙΒΕΙΑΣ ΓΙΑ ΠΟΛΥΩΝΥΜΙΚΟ ΠΥΡΗΝΑ 3 <sup>ΟΥ</sup> ΒΑΘΜΟΥ.....	62
ΠΙΝΑΚΑΣ 4.7: ΠΙΝΑΚΑΣ ΥΠΕΡΟΧΗΣ ΜΟΝΤΕΛΩΝ ΓΙΑ ΔΕΙΚΤΗ ΑΚΡΙΒΕΙΑΣ ΚΑΙ ΠΟΣΟΣΤΟ ΑΚΡΙΒΕΙΑΣ ΣΤΗΝ ΠΕΡΙΠΤΩΣΗ ΠΟΛΥΩΝΥΜΙΚΟΥ ΠΥΡΗΝΑ 3 <sup>ΟΥ</sup> ΒΑΘΜΟΥ.....	63
ΠΙΝΑΚΑΣ 4.8: ΣΥΓΚΕΝΤΡΩΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΕΣΩΝ ΤΙΜΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΓΙΑ ΤΙΣ ΤΡΕΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΠΥΡΗΝΑ .....	64
ΠΙΝΑΚΑΣ 4.9: ΣΥΓΚΕΝΤΡΩΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΕΣΩΝ ΤΙΜΩΝ ΤΩΝ ΜΟΝΤΕΛΩΝ ΚΑΙ ΓΙΑ ΤΡΕΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΠΥΡΗΝΑ. ....	65
ΠΙΝΑΚΑΣ 4.10: ΣΥΓΚΕΝΤΡΩΤΙΚΟΣ ΠΙΝΑΚΑΣ ΥΠΕΡΟΧΗΣ ΜΟΝΤΕΛΩΝ . ....	66

## Ευχαριστίες

Με την ολοκλήρωση της εργασίας μου, θα ήθελα να ευχαριστήσω το καθηγητή Δούμπο Μιχάλη για την αμέριστη συμβουλή του στην παρούσα εργασία. Η μεταδοτικότητα και η υπευθυνότητα του ήταν πολύτιμη για μένα. Η συνεχής καθοδήγησή του και ο χρόνος που μου αφιέρωσε όποτε και αν του ζητήθηκε συντέλεσαν στο τελικό αποτέλεσμα.

Θα ήταν παράλειψη να μην εκφράσω τις ευχαριστίες μου στα μέλη της επιτροπής για το χρόνο που διέθεσαν και τις παρατηρήσεις τους για την βελτίωση της εργασίας.

Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου για την ηθική και υλική συμπαράστασή τους όλα αυτά τα χρόνια. Χωρίς την βοήθεια και την προτροπή τους τίποτα δεν θα γινόταν. Επίσης ευχαριστώ για την υποστήριξη τους όλους τους φίλους μου.

Στους γονείς μου,  
Νίκο και Παναγιώτα



## ΛΑΡΙΣΑΣ ΗΛΙΑΣ

### Σπουδές

---

**2003- 2005** Πολυτεχνείο Κρήτης, Τμήμα Μηχανικών Παραγωγής & Διοίκησης, Μεταπτυχιακό πρόγραμμα ειδίκευσης στην Διοίκηση Επιχειρήσεων

**1999-2003** Πανεπιστήμιο Πατρών, Οικονομική Σχολή, Τμήμα Διοίκησης Επιχειρήσεων, βαθμός πτυχίου **7.91**, σειρά αποφοίτησης **5<sup>ος</sup>**.

### Εργασιακή Εμπειρία

---

**Ιούλιος -Αύγουστος 2003:** Εθνική Τράπεζα της Ελλάδος (στα πλαίσια του προγράμματος ΕΠΕΑΕΚ του τμήματος Διοίκησης Επιχειρήσεων του Πανεπιστημίου Πατρών) Απασχόληση στο τμήμα δανείων.

### Ξένες Γλώσσες

---

ΑΓΓΛΙΚΑ: Πολύ καλά (Πτυχίο FCE), Ιούνιος 2002  
Βαθμολογία TOEFL: 183, Δεκέμβριος 2002

### Γνώση Η/Υ

---

Δίπλωμα χρήσης Η/Υ ECDL, 8/12/2003

Γνώση και χρήση Η/Υ σε περιβάλλον Windows XP

Προγράμματα: Word, Excel, Access, PowerPoint, Outlook Express, SPSS

### Επιστημονικές Εργασίες

---

- Μοντέλα πρόβλεψης συμπεριφοράς καταναλωτών
- Πληροφοριακά συστήματα ικανοποίησης πελατών
- Πρόβλεψη Εξαγορών & Συγχωνεύσεων μια πολυκριτήρια προσέγγιση
- Μοντελοποίηση Καταναλωτών στο Ηλεκτρονικό Εμπόριο
- Χρηματοοικονομική Ανάλυση Κατασκευαστικού τομέα 2000-2003
- Κατάταξη Ευρωπαϊκών χωρών βάση της επίδοσης τους στην καινοτομία με την μέθοδο Promethee,

**Πτυχιακή** με θέμα « **Μέθοδοι Διαχείρισης και Αξιολόγησης προσωπικού στον Τραπεζικό Τομέα**»

- Marketing plan της εταιρείας, DUR ESCAPE LAND
- Παραοικονομία και Φοροδιαφυγή
- Αναλυτική παρουσίαση του τμήματος πωλήσεων της εταιρείας MOBITELE
- Έρευνα Αγοράς για την εταιρεία ΛΟΥΞ
- Διαδικασία επιλογής ΧΥΤΑ για το νομό Αχαΐας,
- Στρατηγικό σχέδιο της εταιρείας DUR ESCAPE LAND

## Περίληψη

Οι μηχανές διανύσματος υποστήριξης (SVM) είναι μια σχετικά νέα μέθοδος εκμάθησης. Πρόκειται για συστήματα μάθησης για την ανάπτυξη μοντέλων παλινδρόμησης ή ταξινόμησης. Κύρια χαρακτηριστικά των SVM αποτελούν η υψηλή δυνατότητα γενίκευσης, η δυνατότητα ανάπτυξης μη γραμμικών μοντέλων, καθώς και η εισαγωγή της έννοιας της πολυπλοκότητας και της σύνδεσής της με την αποτελεσματικότητα του μοντέλου. Η παρούσα έρευνα επικεντρώνεται στην ανάπτυξη μοντέλων ταξινόμησης μέσω των SVM. Η ανάπτυξη των μοντέλων αυτών βασίζεται σε τεχνικές τετραγωνικού προγραμματισμού, πρόσφατα όμως έχουν προταθεί και διαδικασίες βασιζόμενες στον γραμμικό προγραμματισμό και τον προγραμματισμό στόχων. Το κύριο πλεονέκτημα τέτοιων διαδικασιών έναντι τεχνικών μη γραμμικής βελτιστοποίησης αφορά τον περιορισμένο υπολογιστικό φόρτο που απαιτείται για την ανάπτυξη του βέλτιστου μοντέλου. Το στοιχείο αυτό έχει οδηγήσει σε διάφορα μαθηματικά πρότυπα για την ανάπτυξη SVM μέσω τεχνικών προγραμματισμού στόχων. Σκοπός της προτεινόμενης εργασίας είναι να καταγράψει τις μοντελοποιήσεις που έχουν αναπτυχθεί, να εντοπίσει τις ιδιαιτερότητες και τα χαρακτηριστικά τους και να προχωρήσει σε συγκριτική αξιολόγηση της αποτελεσματικότητάς τους σε ένα ευρύ σύνολο δεδομένων.

## Εισαγωγή

Αναμφισβήτητα οι σύγχρονες επιχειρήσεις δραστηριοποιούνται και ενεργούν μέσα σ' ένα ανταγωνιστικό συνεχώς μεταβαλλόμενο περιβάλλον. Μέσα στο περιβάλλον αυτό καλούνται, λαμβάνοντας υπόψιν τις συνθήκες που επικρατούν, να αντεπεξέρθουν στις έντονες πιέσεις και να πετύχουν τους στόχους τους. Τα στελέχη των επιχειρήσεων μέσα στο καθεστώς αυτό έχουν κατανοήσει πλήρως την ανάγκη να συλλέγουν, να αξιολογούν και να αξιοποιούν τις πληροφορίες λαμβάνοντας κάθε φορά την καταλληλότερη απόφαση που θα τους φέρει πιο κοντά στην επίτευξη των επιχειρηματικών σκοπών.

Το κύριο πρόβλημα που αντιμετωπίζουν οι εταιρείες κινούμενες προς αυτή την κατεύθυνση, είναι η διαρκής αλλαγή των αγορών με ταχύτατους ρυθμούς και η νέα πληροφορία που παράγεται και πρέπει να ληφθεί υπόψιν. Τα τελευταία τριάντα χρόνια η παραγωγή πληροφόρησης είναι συνολικά ανώτερη από την αντίστοιχη των προηγούμενων 5.000 ετών. Σήμερα εκδίδονται στον κόσμο 4.000 βιβλία ημερησίως, έναντι 600 πριν από δέκα χρόνια. Η κυριακάτικη έκδοση των «New York Times» περιέχει περισσότερες πληροφορίες από αυτές που ένα άτομο μπορούσε να αποκτήσει στην διάρκεια της ζωής του τον 17<sup>ο</sup> αιώνα. Αυτή η παραγωγή και ροή πληροφορίας λογικό είναι να δυσκολεύει τις αποφάσεις των επιχειρήσεων αφού πρέπει να αξιολογηθούν όλες για την χρησιμότητα τους και να συνεκτιμηθούν για την τελική απόφαση.

Σημαντική βοήθεια στο πρόβλημα αυτό έρχεται να δώσει η ταξινόμηση. Οι τεχνικές ταξινόμησης παίζουν σημαντικό ρόλο στις διάφορες επιχειρηματικές αποφάσεις, κατατάσσοντας τις διαθέσιμες πληροφορίες σε κατηγορίες, οι οποίες προσδιορίζονται βάσει διαφόρων κριτηρίων. Η ταξινόμηση των πληροφοριών αποτελεί σημαντικό συστατικό για τις διάφορες επιχειρηματικές αποφάσεις. Πολλά θέματα αποφάσεων αποτελούν τμήματα ενός προβλήματος ταξινόμησης ή μπορούν εύκολα να μετατραπούν σε πρόβλημα ταξινόμησης (π.χ. προβλήματα πρόβλεψης, διάγνωσης, αναγνώρισης προτύπων κ.λ.π.). Η ταξινόμηση απέκτησε μεγαλύτερη χρησιμότητα με την έλευση του διαδικτύου. Το διαδίκτυο ως κανάλι επικοινωνίας και συναλλαγής, παρέχει τη δυνατότητα εφαρμογής νέων τεχνολογιών, όπως collaborative filtering και

recommender systems, τα οποία διευκολύνουν το μάρκετινγκ και την εξυπηρέτηση μαζικής πελατείας (Resnick και Varian, 1997). Ένας πρωτεύον σκοπός αυτών των συστημάτων είναι η ταξινόμηση της διαθέσιμης πληροφορίας βάσει κάποιων κριτηρίων.

Λόγω της μεγάλης ερευνητικής σημασίας της ταξινόμησης, υλοποιήθηκαν πολλές μέθοδοι προς την κατεύθυνση αυτή. Από τις μεθόδους αυτές, το ενδιαφέρον της εργασίας αυτής θα εστιαστεί στις μηχανές διανύσματος υποστήριξης (support vector machines), μια νέα μεθοδολογία η οποία έχει γνωρίσει σημαντική διάδοση τα τελευταία χρόνια. Το κύριο μεθοδολογικό πλαίσιο των SVM αφορά την ανάπτυξη μοντέλων ταξινόμησης έχοντας ως κύριο στόχο τον έλεγχο της πολυπλοκότητας των μοντέλων σε σχέση με τη διακριτική τους ικανότητα (ποιότητα των μοντέλων). Μέσα στα πλαίσια αυτά, η ανάπτυξη μοντέλων ταξινόμησης μέσω των SVM ανάγεται στην επίλυση ενός προβλήματος βελτιστοποίησης το οποίο έχει τη μορφή ενός τετραγωνικού προγράμματος. Η επίλυση του προγράμματος αυτού στην περίπτωση που εξετάζεται ένα μεγάλο σύνολο δεδομένων απαιτεί αυξημένο υπολογιστικό φόρτο. Για το λόγο αυτό έχουν προταθεί ειδικές διαδικασίες και αλγόριθμοι για την αποτελεσματική ανάπτυξη μοντέλων ταξινόμησης μέσω SVM.

Σκοπός της προσπάθειας αυτής είναι να εστιάσει σε κάποιες από τις παραπάνω ειδικές διαδικασίες και πιο συγκεκριμένα στην ανάπτυξη SVM με τη χρήση προγραμματισμού στόχων. Το κύριο πλεονέκτημα αυτών των διαδικασιών είναι ο περιορισμένος υπολογιστικός φόρτος που απαιτείται για την ανάπτυξη του βέλτιστου μοντέλου. Έτσι, λοιπόν, θα καταγραφούν τα εν λόγω μοντέλα, θα παρουσιαστούν οι ιδιαιτερότητες και τα χαρακτηριστικά τους και θα επιχειρηθεί μια συγκριτική αξιολόγηση τους σε ένα ευρύ σύνολο δεδομένων.

Προς τη κατεύθυνση αυτή η διατριβή οργανώνεται με τον ακόλουθο τρόπο:

Στο κεφάλαιο 1 πραγματοποιείται μια εισαγωγή στο πρόβλημα της ταξινόμησης. Παρουσιάζονται οι βασικές έννοιες και οι ορισμοί του προβλήματος της ταξινόμησης, καθώς και πεδία εφαρμογής της.

Το 2<sup>ο</sup> κεφάλαιο αναφέρεται στις τεχνικές προγραμματισμού στόχων στα προβλήματα ταξινόμησης. Αρχικά παρατίθεται η θεωρία του προγραμματισμού στόχων και εν συνεχεία ακολουθούν οι πιο διαδεδομένες τεχνικές προγραμματισμού στόχων που

χρησιμοποιούνται στην ταξινόμηση όπως τα μοντέλα MMD, OSD και το υβριδικό μοντέλο.

Το 3<sup>ο</sup> κεφάλαιο περιλαμβάνει το θεωρητικό υπόβαθρο των SVM για γραμμική και μη γραμμική περίπτωση και κατόπιν καταγράφονται τέσσερις μοντελοποιήσεις των SVM με τη χρήση προγραμματισμού στόχων.

Η εφαρμογή των δεδομένων και η ανάλυση των αποτελεσμάτων για τα 4 μοντέλα γίνεται στο 4<sup>ο</sup> κεφάλαιο. Αξιολογείται η αποτελεσματικότητα του κάθε μοντέλου και γίνεται μια συγκριτική καταγραφή της συμπεριφοράς του καθενός.

Τέλος στο 5<sup>ο</sup> κεφάλαιο παρουσιάζονται τα γενικά συμπεράσματα από την εφαρμογή των νέων μοντέλων και προτείνονται διάφορες κατευθύνσεις για μελλοντικές έρευνες.

# ΚΕΦΑΛΑΙΟ 1<sup>ο</sup>

## Βασικές Έννοιες Ταξινόμησης

## 1. Ορισμός ταξινόμησης

Η ταξινόμηση λαμβάνει χώρα σ' ένα μεγάλο εύρος των ανθρώπινων δραστηριοτήτων. Καθημερινά από τις πιο απλές δραστηριότητες μας (αγορά ενός ρούχου) ως τις πιο σύνθετες (επιλογή εργασίας) όλοι μας ασυνείδητα ή συνειδητά κάποια άλλοι, οδηγούμαστε στην τελική απόφαση με την χρήση ταξινόμησης. Ο όρος λοιπόν αυτός καλύπτει οποιοδήποτε πλαίσιο, μέσα στο οποίο κάποια απόφαση ή πρόβλεψη συντελείται στη βάση μιας διαθέσιμης πληροφορίας. Η διαδικασία ταξινόμησης είναι μια έγκυρη μέθοδος για την επανάληψη τέτοιων κρίσεων σε καινούργιες καταστάσεις.

Η ταξινόμηση έχει δύο διακριτές έννοιες την ομαδοποίηση (Clustering) και την ταξινόμηση (Discrimination). Η πρώτη αφορά ένα σύνολο δεδομένων με το οποίο πρέπει να καθοριστεί η ύπαρξη κατηγοριών ή ομάδων στα δεδομένα. Η δεύτερη αφορά την ύπαρξη συγκεκριμένων κατηγοριών, όπου σκοπός είναι να αναπτυχθεί κάποιος κανόνας με τον οποίο να ταξινομείται οποιαδήποτε νέα παρατήρηση σε κάποια από τις υπάρχουσες κατηγορίες.

Ένας πιο δομημένος ορισμός της ταξινόμησης, πέρα από την γενική περιγραφή που αναφέρθηκε παραπάνω δόθηκε από τον Mirkin το 1996, ο οποίος προσδιορίζει την λειτουργία και την χρησιμότητας της ως εξής:

*«Ταξινόμηση είναι η ρεαλιστική ή ιδεατή τοποθέτηση μαζί παρόμοιων αντικειμένων και ο διαχωρισμός των αντικειμένων τα οποία διαφέρουν, με απότερο σκοπό:*

- 1. τη διαμόρφωση, οργάνωση και διατήρηση της γνώσης,*
- 2. την ανάλυση της δομής του φαινομένου που εξετάζεται,*
- 3. τη συσχέτιση των διαφόρων πλευρών του υπό εξέταση φαινομένου.»*

Στην αγγλική ορολογία χρησιμοποιούνται διάφοροι όροι για την αναφορά στο πρόβλημα της ταξινόμησης, οι πιο διαδεδομένοι είναι οι ακόλουθοι:

- ✓ Discrimination (διάκριση)
- ✓ Classification (ταξινόμηση)
- ✓ Sorting (διατεταγμένη ταξινόμηση)

Οι δύο πρώτοι όροι χρησιμοποιούνται κυρίως από στατιστικούς επιστήμονες καθώς και από ερευνητές που δραστηριοποιούνται στο χώρο της τεχνητής νοημοσύνης (νευρωνικά δίκτυα, μηχανική μάθηση, κ.λ.π.). Αντίθετα ο τρίτος όρος έχει εισαχθεί και χρησιμοποιείται κυρίως από ερευνητές του χώρου της πολυκριτήριας ανάλυσης αποφάσεων.

Όπως αναφέρθηκε παραπάνω, μια διαφορετική από την ταξινόμηση κατηγορία τεχνικών, προσδιοριζόμενες με τον όρο ομαδοποίηση, αναλαμβάνει το διαχωρισμό των δεδομένων σε άγνωστες εκ των προτέρων ομάδες, γνωστού ή άγνωστου αριθμού, τοποθετώντας στην ίδια ομάδα αντικείμενα με όμοια ή παραπλήσια χαρακτηριστικά. Έτσι, στην ομαδοποίηση οι συγκρίσεις είναι σχετικές και η σύνθεση του πληθυσμού των αντικειμένων επηρεάζει τόσο τις εξαγόμενες κατηγορίες όσο και την τοποθέτηση των αντικειμένων σε αυτές. Η κυριότερη διαφορά μεταξύ ταξινόμησης και ομαδοποίησης είναι γνωσιολογική. Στην ταξινόμηση με *a priori* γνωστές τις κατηγορίες, η γνώση που παράγεται τοποθετείται σε μια υπάρχουσα δομή της οποίας η ποιότητα παραμένει αμετάβλητη, αντίθετα η ομαδοποίηση εξάγει την ίδια τη δομή της γνώσης, καθορίζοντας τις κατηγορίες των αντικειμένων βάσει εμπειρίας. Συνεπώς ενώ κατά την ταξινόμηση υπάρχει μια εκ των προτέρων γνώση της μορφής του αποτελέσματος το οποίο θα πρέπει να εξαχθεί, στην ομαδοποίηση πραγματοποιείται ουσιαστικά διαμόρφωση αυτής της γνώσης βάσει των υπαρχόντων δεδομένων, χωρίς όμως να είναι δυνατόν ο αποφασίζων να ελέγξει τον τρόπο με τον οποίο διαμορφώνονται οι ομάδες, τουλάχιστον όσον αφορά τα χαρακτηριστικά των εναλλακτικών δραστηριοτήτων που θα περιλαμβάνουν (Δούμπος, Ζοπουνίδης, 2001).

Είναι αναγκαίο να τονιστεί, ότι παρ' όλες τις διαφορές μεταξύ τους, η ταξινόμηση και η ομαδοποίηση δεν είναι ανταγωνιστικές προσεγγίσεις. Αντιθέτως, μπορούν να χρησιμοποιηθούν διαδοχικά, με την ομαδοποίηση να εξάγει τη δομή της γνώσης που εμπεριέχει το δείγμα και την ταξινόμηση να ακολουθεί παρέχοντας ένα πρότυπο για την τοποθέτηση νέων αντικειμένων στη δομή αυτή.



## 2. Χαρακτηριστικά

Υπάρχουν πολλά ζητήματα που αφορούν τις τεχνικές ταξινόμησης και προσδιορίζουν την εγκυρότητα, την αξιοπιστία και πάνω από όλα την αποτελεσματικότητά τους. Τα τέσσερα πιο σημαντικά χαρακτηριστικά που δίνουν σημαντικές πληροφορίες για την χρηστικότητα μιας τεχνικής ταξινόμησης είναι τα ακόλουθα (Χατζή, 2003):

- ✓ Ακρίβεια: Είναι η αξιοπιστία του κανόνα ταξινόμησης και συνήθως αντιπροσωπεύεται από την αναλογία των σωστών ταξινομήσεων. Παρ' όλα αυτά είναι σημαντικό να ελέγχεται και ο δείκτης σφαλμάτων.
- ✓ Ταχύτητα: Σε πολλές περιπτώσεις η ταχύτητα του υποδείγματος ταξινόμησης αποτελεί κύριο ζήτημα. Ένα υπόδειγμα ταξινόμησης το οποίο δίνει 90% ακρίβεια, είναι προτιμότερο από κάποιο άλλο το οποίο δίνει 95% ακρίβεια, εάν το πρώτο είναι 100 φορές πιο γρήγορο στον έλεγχο.
- ✓ Σαφήνεια: Η διαδικασία ταξινόμησης πρέπει να είναι εύκολα κατανοητή, διότι διαφορετικά είναι πολύ πιθανό να συμβούν σφάλματα κατά την εφαρμογή της.
- ✓ Χρόνος εκμάθησης: Σε ένα γρήγορα μεταβαλλόμενο περιβάλλον, είναι αναγκαία η γρήγορη εκμάθηση των κανόνων ταξινόμησης ή οι τροποποιήσεις σε ένα υπάρχοντα κανόνα, σε πραγματικό χρόνο. Επίσης, γρήγορα μπορεί να υπονοείται ότι ένας μικρός αριθμός παρατηρήσεων είναι αρκετός για την εφαρμογή ενός συγκεκριμένου κανόνα ταξινόμησης.

## 3. Ορισμός των κατηγοριών

Ένα σημαντικό ζήτημα, το οποίο δεν κατανοείται σωστά σε πολλές μελέτες ταξινόμησης, είναι η φύση των κατηγοριών στις οποίες τοποθετούνται τα αντικείμενα και ο τρόπος με τον οποίο αυτές ορίζονται. Υπάρχουν τρεις διαφορετικοί τρόποι προκειμένου να προσδιοριστούν οι κατηγορίες (Χατζή, 2003):

1. Οι κατηγορίες ανταποκρίνονται στα χαρακτηριστικά διαφορετικών πληθυσμών.

2. Οι κατηγορίες προκύπτουν από μια διαδικασία πρόβλεψης. Σε αυτή την περίπτωση, η κατηγορία είναι το αποτέλεσμα το οποίο έχει προβλεφθεί από τη γνώση των κριτηρίων. Σε στατιστικούς όρους, η κατηγορία είναι μια τυχαία μεταβλητή.
3. Οι κατηγορίες έχουν προκαθοριστεί από τη διαίρεση του δείγματος (π.χ. από τα ίδια τα κριτήρια). Η κατηγορία θεωρείται ως μια συνάρτηση των κριτηρίων. Κατά συνέπεια, ένα αντικείμενο μπορεί να ταξινομηθεί εσφαλμένα, εάν κάποια κριτήρια βρίσκονται εκτός καθορισμένων ορίων, και σωστά διαφορετικά.

## 4. Ακρίβεια ταξινόμησης

Για την αξιολόγηση και τον έλεγχο ενός μοντέλου ταξινόμησης, είναι απαραίτητο να υπολογίζεται η ακρίβεια του μοντέλου, ώστε να υπάρχει ένα ποσοτικό στοιχείο που να προσδιορίζει την αποτελεσματικότητά του και να καθιστά δυνατή την σύγκριση με άλλα. Ένας απλός ορισμός της ακρίβειας είναι ο λόγος του πλήθους των σωστών ταξινομήσεων προς το πλήθος των εναλλακτικών (Δούμπος, 2003α). Πρέπει να διευκρινιστεί ότι η ακρίβεια που εκτιμάται στο δείγμα εκμάθησης και η ακρίβεια που εκτιμάται στο δείγμα ελέγχου, είναι συχνά εντελώς διαφορετικές. Δεν είναι ασυνήθιστο σε πολλές εφαρμογές, οι τεχνικές ταξινόμησης να παρουσιάζουν ικανοποιητική ακρίβεια στο δείγμα εκμάθησης, ενώ στο δείγμα ελέγχου τα αποτελέσματα να είναι απογοητευτικά. Συνήθως αυτό που είναι σημαντικό είναι η ακρίβεια που δίνει η ταξινόμηση ενός άγνωστου αντικειμένου. Μια γενικά αποδεκτή μέθοδος για την εκτίμηση του παραπάνω, είναι η χρήση των γνωστών δεδομένων, των οποίων οι κατηγορίες ταξινόμησης είναι γνωστές. Αρχικά χρησιμοποιείται ένα δείγμα εκμάθησης για την εφαρμογή της διαδικασίας. Στη συνέχεια, αυτό ελέγχεται με το δείγμα ελέγχου και τα αποτελέσματα συγκρίνονται με ήδη υπάρχουσες γνωστές ταξινομήσεις. Η αναλογία που τα αποτελέσματα είναι σωστά στο δείγμα ελέγχου, είναι μια αμερόληπτη εκτίμηση της ακρίβειας.

## 5. Ορισμοί βασικών εννοιών και μαθηματική διατύπωση

Στην παράγραφο αυτή θα περιγραφούν οι ορισμοί των βασικών εννοιών της ταξινόμησης (Bauer και Kohavi, 1999).

Παράδειγμα ή labeled instance ονομάζεται ένα ζεύγος  $(x, y)$ , όπου  $x \in X$  και  $X$  ο χώρος των αντικειμένων ή παραδειγμάτων (instance space) ή το γνωστικό πεδίο (domain),  $y \in Y$  και  $Y$  ο διακριτός χώρος των κατηγοριών (space of classes). Το  $x$  έχει  $n$  συνιστώσες και ονομάζεται διάνυσμα χαρακτηριστικών (vector of attributes) ενώ το  $y$  υποδεικνύει την κατηγορία στην οποία ανήκει το αντικείμενο.

Δείγμα (sample) ονομάζεται ένα σύνολο από αντικείμενα.

Υπόδειγμα ταξινόμησης (classifier) ή υπόθεση (hypothesis) είναι μια συνάρτηση  $h : X \rightarrow Y$  που ουσιαστικά αναπαριστά την σύνθεση των χαρακτηριστικών-κριτηρίων με την οποία γίνεται η ταξινόμηση των εναλλακτικών. Επιπλέον αιτιοκρατική διαδικασία επαγωγής (deterministic inducer) ονομάζεται μια αντιστοιχία  $I : 2^X \rightarrow H$ , όπου  $H$  ο χώρος των υποδειγμάτων ταξινόμησης και  $2^X$  τα δυνατά υποσύνολα του χώρου  $X$ .

## 6. Πεδία εφαρμογής

Η ταξινόμηση, εκτός από το ενδιαφέρον που παρουσιάζει σε επιστημονικό επίπεδο λόγω της χρησιμότητάς της, των πληροφοριών και κατευθύνσεων που μπορεί να δίνει στους αποφασίζοντες, έχει εφαρμογή σε πολλούς τομείς της καθημερινότητας. Έτσι, λοιπόν, έχει εφαρμοστεί σε πολλούς κλάδους όλων των επιστημών με εντυπωσιακά αποτελέσματα. Στη συνέχεια θα παρατεθούν οι πιο σημαντικοί από αυτούς (Δούμπος, Ζοπουνίδης, 2001):

- Ιατρική: Πραγματοποίηση ιατρικών διαγνώσεων ταξινομώντας τους ασθενείς σε κατηγορίες ανάλογα με τα συμπτώματα που παρουσιάζουν (Tsumoto, 1998, Belacel, 2000).

- Αναγνώριση προτύπων (pattern recognition): Αναγνώριση των χαρακτηριστικών φυσικών προσώπων ή αντικειμένων και ταξινόμηση τους σε ανάλογες κατηγορίες. Χαρακτηριστικά είναι τα παραδείγματα της αναγνώρισης βασικών ανθρώπινων χαρακτηριστικών όπως η ομιλία, η ίριδα του ματιού και η εφαρμογή τους στην ασφάλεια καίριων συστημάτων (Ripley, 1996, Young και Fu, 1997).
- Διαχείριση ανθρώπινου δυναμικού: Αξιολόγηση προσωπικού βάσει των προσόντων του με σκοπό τον προσδιορισμό της κατάλληλης θέσης εργασίας.
- Διαχείριση παραγωγικών συστημάτων: Παρακολούθηση της λειτουργίας πολύπλοκων συστημάτων παραγωγής και την έγκαιρη διάγνωση πιθανών βλαβών (Catelani και Fort, 2000).
- Μάρκετινγκ: Καθορισμός της κατάλληλης πολιτικής μάρκετινγκ για τη διείσδυση προϊόντων στην αγορά, μελέτη των επιμέρους χαρακτηριστικών διαφορετικών κατηγοριών καταναλωτών, σχεδιασμός των προϊόντων ώστε να ανταποκρίνονται στις απαιτήσεις των καταναλωτών, μέτρηση ικανοποίησης πελατών κ.λ.π. (Siskos et al., 1998).
- Περιβαλλοντική και ενεργειακή διαχείριση: Ανάλυση και έγκαιρη διάγνωση των περιβαλλοντικών επιπτώσεων διάφορων ενεργειακών πολιτικών, διερεύνηση της αποτελεσματικότητας ενεργειακών πολιτικών σε κρατικό επίπεδο (Diakoulaki et. al., 1999).

# ΚΕΦΑΛΑΙΟ 2<sup>ο</sup>

## Θεωρία Προγραμματισμού Στόχων

## 1. Εισαγωγή

Οι τεχνικές που έχουν εφαρμοσθεί σε προβλήματα ταξινόμησης περιλαμβάνουν κυρίως στατιστικές μεθόδους όπως τη διακριτική ανάλυση (γραμμική-LDA και τετραγωνική-QDA), το λογιστικό και το κανονικό υπόδειγμα πιθανότητας. Οι στατιστικές μέθοδοι αποτέλεσαν την πρώτη και την πιο δημοφιλή προσέγγιση στην μελέτη προβλημάτων ταξινόμησης, ωστόσο δέχθηκαν έντονη κριτική κυρίως λόγω των στατιστικών περιορισμών στους οποίους υπόκεινται (Eisenbeis, 1977).

Μια άλλη τεχνική η οποία χρησιμοποιήθηκε για την αντιμετώπιση προβλημάτων ταξινόμησης τις τελευταίες δύο δεκαετίες είναι η τεχνική προγραμματισμού στόχων (goal programming). Τα προβλήματα ταξινόμησης μοντελοποιούνται με την χρήση προγραμματισμού στόχων ως προβλήματα βελτιστοποίησης, με αντικειμενικό σκοπό την ελαχιστοποίηση κάποιου μέτρου του σφάλματος ταξινόμησης. Οι έρευνες αναφορικά με την χρήση προγραμματισμού στόχων για την αντιμετώπιση προβλημάτων ταξινόμησης έχουν δώσει ιδιαίτερο βάρος στα ακόλουθα δύο σημεία (Ζοπουνίδης, 1998):

1. Στην πραγματοποίηση συγκριτικών ερευνών με πολυμεταβλητές στατιστικές μεθόδους (χρησιμοποιώντας κυρίως πειραματικά δεδομένα).
2. Στην αντιμετώπιση των θεωρητικών προβλημάτων που εμφανίζονται και για την επίτευξη καλύτερων αποτελεσμάτων ταξινόμησης.

Στις επόμενες παραγράφους που ακολουθούν θα αναφερθεί η βασική θεωρία του προγραμματισμού στόχων και κατόπιν θα παρουσιαστούν οι 3 πιο διαδεδομένες τεχνικές προγραμματισμού στόχων που χρησιμοποιούνται στα προβλήματα ταξινόμησης.

## 2. Θεωρία προγραμματισμού στόχων

Ο προγραμματισμός στόχων αποτελεί μια από τις πιο διαδεδομένες προσεγγίσεις στο χώρο της πολυκριτήριας ανάλυσης με σημαντικές εφαρμογές σε πλήθος πρακτικών περιπτώσεων. Ο προγραμματισμός στόχων αποτελεί μια ειδική μορφή προβλημάτων γραμμικού προγραμματισμού, τα οποία στοχεύουν στην αντιμετώπιση προβλημάτων όπου οι στόχοι είναι πολλαπλοί. Σε αντίθεση όμως με τις τεχνικές του πολυκριτήριου γραμμικού προγραμματισμού ο προγραμματισμός στόχων δεν αποσκοπεί στον εντοπισμό λύσεων εξετάζοντας ταυτόχρονα όλες τις αντικειμενικές συναρτήσεις, αλλά αντίθετα αποσκοπεί στον εντοπισμό λύσεων που ανταποκρίνονται σε κάποιους προκαθορισμένους στόχους. Οι στόχοι αυτοί αναφέρονται σε κάποια ικανοποιητικά επίπεδα των αντικειμενικών συναρτήσεων τα οποία πρέπει να επιτευχθούν. Επομένως η έννοια του στόχου διαφέρει από την έννοια της αντικειμενικής συνάρτησης, διότι η αντικειμενική συνάρτηση προσδιορίζει μια κατεύθυνση η οποία ακολουθείται στον εντοπισμό μιας λύσης, ενώ η έννοια του στόχου οριοθετεί επιθυμητά επίπεδα που πρέπει να επιτευχθούν.

Η γενική μαθηματική διατύπωση ενός προβλήματος προγραμματισμού στόχων στη γραμμική μορφή του είναι η εξής (Δούμπος, 2003):

$$\min z = h_i(w_i, \sigma_i^-, \sigma_i^+)$$

με τους περιορισμούς

$$f_i(x) + \sigma_i^- - \sigma_i^+ = s_i, \quad \forall i = 1, 2, \dots, k$$

$$x \in A, \quad \sigma_i^-, \sigma_i^+ \geq 0$$

όπου:

- $s_i$  είναι η τιμή στόχος για την αντικειμενική συνάρτηση  $f_i$ ,
- $\sigma_i^-$  και  $\sigma_i^+$  είναι οι αποκλίσεις από το στόχο  $s_i$ ,
- $h_i$  είναι μια συνάρτηση των αποκλίσεων από τον στόχο  $s_i$ ,

- $w_i$  είναι ένας συντελεστής στάθμισης ο οποίος δηλώνει την σημαντικότητα επίτευξης του στόχου  $s_i$ .

Από την διατύπωση αυτή προκύπτει ότι τα προβλήματα προγραμματισμού στόχων αφορούν την βελτιστοποίηση μιας αντικειμενικής συνάρτησης η οποία αφορά ένα συνδυασμό των αποκλίσεων από τους στόχους. Οι αντικειμενικές συναρτήσεις  $f_1, f_2, \dots, f_k$  ενσωματώνονται ως περιορισμοί στο πρόβλημα χρησιμοποιώντας στο δεξιό μέλος των περιορισμών αυτών τις αντίστοιχες τιμές-στόχους  $s_1, s_2, \dots, s_k$ . Οι τιμές στόχοι προσδιορίζονται από τον αποφασίζοντα και ουσιαστικά αποτελούν τα σημεία αναφοράς βάσει των οποίων προσδιορίζεται η καταλληλότητα κάθε λύσης. Ο καθορισμός των στόχων απαιτεί αυξημένη προσοχή, καθώς η εσφαλμένη οριοθέτηση των στόχων είναι πιθανόν να οδηγήσει σε μη αποτελεσματικές λύσεις.

Δεδομένης μιας τιμή στόχου  $s_i$  οι αποκλίσεις ορίζονται ως  $\sigma_i^+ = \max\{0, f_i(x) - s_i\}$  και  $\sigma_i^- = \max\{0, s_i - f_i(x)\}$ . Οι αποκλίσεις είναι τέτοιες ώστε  $\sigma_i^+ \sigma_i^- = 0$ . Είναι αναγκαίο να τονιστεί ότι σε πολλές περιπτώσεις δεν απαιτείται η χρησιμοποίηση δύο αποκλίσεων. Όταν για παράδειγμα σε μια αντικειμενική συνάρτηση  $f_i$  ο αποφασίζων καθορίζει ως στόχο το βέλτιστο επίπεδο  $f_i^*$ , τότε η χρήση της απόκλισης  $\sigma_i^+$  που υποδηλώνει στόχο είναι περιττή, δεδομένου ότι δεν είναι δυνατό να βρεθεί λύση  $x$  τέτοια ώστε  $f_i(x) > f_i^*$ .

Βασικό στοιχείο στην χρήση τεχνικών προγραμματισμού στόχου αποτελεί τρόπος με τον οποίο εξετάζεται ο συνδυασμός των αποκλίσεων (συνάρτηση  $h_i$ ). Ο συνδυασμός αυτός μπορεί να υλοποιηθεί με διάφορους τρόπους. Συνήθως χρησιμοποιούνται η πρώτη, η δεύτερη και η άπειρη νόρμα του διανύσματος  $(w_1(\sigma_1^+ + \sigma_1^-), w_2(\sigma_2^+ + \sigma_2^-), \dots, w_k(\sigma_k^+ + \sigma_k^-))$ . Η πρώτη νόρμα ( $L_1$ ) αντιστοιχεί στο άθροισμα όλων των στοιχείων του διανύσματος, η δεύτερη νόρμα ( $L_2$ ) αντιστοιχεί στην τετραγωνική ρίζα του αθροίσματος των τετραγώνων, ενώ η άπειρη νόρμα ( $L_\infty$ ) αντιστοιχεί στο μέγιστο στοιχείο του διανύσματος. Έτσι η αντικειμενική συνάρτηση ενός προβλήματος στόχων στις τρεις αυτές περιπτώσεις διαμορφώνεται ως εξής (Δούμπος, 2003):



$$L_1: h_i(w_i, \sigma_i^-, \sigma_i^+) = \sum_{i=1}^k w_i (\sigma_i^+ + \sigma_i^-)$$

$$L_2: h_i(w_i, \sigma_i^-, \sigma_i^+) = \sqrt{\sum_{i=1}^k w_i^2 (\sigma_i^+ + \sigma_i^-)^2}$$

$$L_\infty: h_i(w_i, \sigma_i^-, \sigma_i^+) = \max_i \{w_i (\sigma_i^+ + \sigma_i^-)\}$$

Στην περίπτωση της νόρμας  $L_\infty$  η μέγιστη απόκλιση  $d$  μπορεί να προσδιοριστεί εισάγοντας επιπλέον περιορισμούς στο πρόβλημα της μορφής  $w_i (\sigma_i^+ + \sigma_i^-) - d \leq 0$ , για κάθε  $i = 1, 2, \dots, k$ .

### 3. Προγραμματισμός στόχων και ταξινόμηση

Στη συνέχεια θα παρουσιαστούν τεχνικές προγραμματισμού στόχων που χρησιμοποιούνται για την αντιμετώπιση προβλημάτων ταξινόμησης. Οι τεχνικές αυτές είναι προσαρμοσμένες για την αντιμετώπιση προβλημάτων ταξινόμησης σε δύο κατηγορίες. Οι κατηγορίες αυτές, θα συμβολίζονται στο εξής ως  $G_0$  και  $G_1$ . Στόχος των τεχνικών είναι η ανάπτυξη ενός υπερεπιπέδου το οποίο θα ταξινομεί κάθε αντικείμενο  $A_i$  ως εξής:

$$\left. \begin{aligned} A_i X &\geq c \Leftrightarrow A_i \in G_1 \\ A_i X &\leq c \Leftrightarrow A_i \in G_0 \end{aligned} \right\}$$

Με βάση τον παραπάνω κανόνα ταξινόμησης είναι πιθανό να προκύψουν προβλήματα όταν το σκορ ορισμένων αντικειμένων  $A_i X$  είναι ακριβώς ίσο με το όριο  $c$ , δηλαδή όταν ισχύει η ισότητα στον παραπάνω κανόνα και το υπό εξέταση αντικείμενο είναι πάνω στο υπερεπίπεδο ταξινόμησης. Στην περίπτωση αυτή είναι αδύνατη η λήψη μιας απόφασης σχετικά με την ταξινόμηση των αντικειμένων. Για την αντιμετώπιση αυτής τη δυσκολίας χρησιμοποιείται ένας μικρός αριθμός  $\varepsilon$  (ο οποίος θα πρέπει να είναι σημαντικά μικρότερος του  $A_i X$ ) στο δεξί μέρος μιας εκ των δύο παραπάνω ανισοτήτων ως εξής (Ζοπουνίδης, 1998):

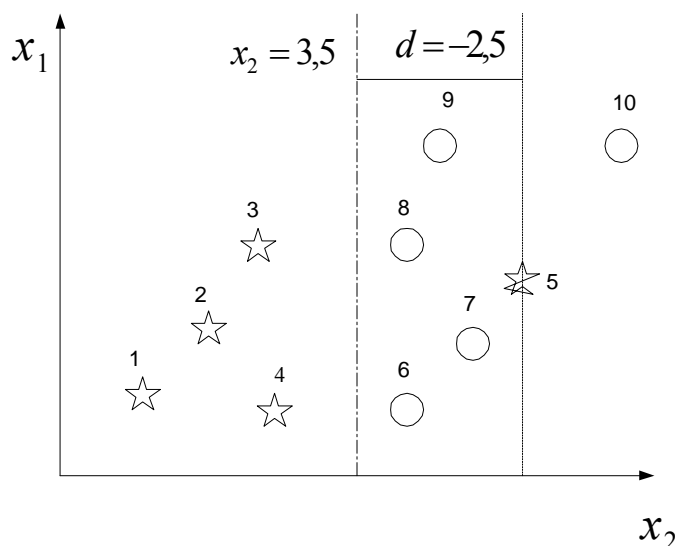
$$\left. \begin{array}{l} A_i X \geq c \Leftrightarrow A_i \in G_1 \\ A_i X \leq c - \varepsilon \Leftrightarrow A_i \in G_0 \end{array} \right\} \quad (1)$$

Με τον τρόπο αυτό ουσιαστικά θεωρείται ότι κάθε αντικείμενο που βρίσκεται πάνω στο υπερεπίπεδο διαχωρισμού ανήκει στην ομάδα  $G_1$ , ενώ τα αντικείμενα της ομάδας  $G_0$  θα πρέπει να έχουν σκορ αυστηρά μικρότερου του ορίου.

Στα πλαίσια αυτά θα αναλυθούν τρεις από τις πλέον διαδεδομένες τεχνικές για την ανάπτυξη του υπερεπιπέδου το οποίο διαχωρίζει τις ομάδες. Η πρώτη αναφέρεται στη μεγιστοποίηση της ελάχιστης απόστασης (Maximize the Minimum Distance-MMD), η δεύτερη αναφέρεται στην βελτιστοποίηση του αθροίσματος των αποκλίσεων (Optimize the Sum of Deviations-OSD) και η τρίτη αποτελεί ένα υβριδικό μοντέλο (Hybrid model).

### 3.1 Μοντέλο μεγιστοποίησης της ελάχιστης απόστασης (MMD)

Το μοντέλο μεγιστοποίησης της ελάχιστης απόστασης παρουσιάστηκε από τους Freed και Glover το 1981. Στο μοντέλο αυτό στόχος είναι η ανάπτυξη του ζητούμενου υπερεπιπέδου να γίνεται με κριτήριο τη μεγιστοποίηση της ελάχιστης απόστασης των αντικειμένων από το όριο που διαχωρίζει τις ομάδες. Η απόσταση αυτή θεωρείται ότι είναι ανεξάρτητη προσήμου (θετική ή αρνητική). Η ελάχιστη απόσταση είναι θετική όταν όλα τα αντικείμενα ταξινομούνται σωστά, ενώ αρνητική είναι όταν ένα ή περισσότερα αντικείμενα ταξινομούνται εσφαλμένα όπως φαίνεται και στο σχήμα 2.1. Στο σχήμα η ελάχιστη απόσταση είναι ίση με την απόσταση από το υπερεπίπεδο του εσφαλμένα ταξινομημένου αντικειμένου 5. Η μεγιστοποίηση αυτή προφανώς θα οδηγήσει στην μετατόπιση του υπερεπιπέδου προς τα δεξιά.



**Σχήμα 2. 1:** Γραφική αναπαράσταση της ελάχιστης απόστασης στο μοντέλο MMD (Πηγή: Ζοπουνίδης, 1998)

Για να δημιουργηθεί το υπερεπίπεδο ταξινόμησης, το μαθηματικό πρόγραμμα που πρέπει να λυθεί είναι το ακόλουθο:

$$\max\{d\}$$

με τους περιορισμούς:

$$A_i X - d \geq c \quad \forall i \in G_1$$

$$A_i X + d \leq c - \varepsilon \quad \forall i \in G_0$$

$$\sum X = M$$

Οι δύο πρώτοι περιορισμοί προέρχονται από τους δύο βασικούς κανόνες ταξινόμησης (1). Με  $d$  συμβολίζεται η ελάχιστη απόσταση η οποία θα πρέπει να μεγιστοποιηθεί, ενώ το  $c$  είναι μια δεδομένη σταθερά η οποία αναπαριστά το όριο που διαχωρίζει τις δύο ομάδες. Ο τελευταίος περιορισμός χρησιμοποιείται ως περιορισμός κανονικοποίησης. Στον περιορισμό αυτό το  $M$  είναι μια σταθερά η οποία συνήθως παίρνει τιμές στο διάστημα  $[1, 10]$ .

### 3.2 Μοντέλο βελτιστοποίησης των αποκλίσεων (OSD)

Η μέτρηση της ακρίβειας ταξινόμησης, με την χρήση της ελάχιστης απόστασης όπως αυτή γίνεται στο μοντέλο MMD περιέχει δύο βασικά μειονεκτήματα (Freed και Glover, 1986):

- Δεν λαμβάνει καθόλου υπόψη το πλήθος των εσφαλμένων ταξινομήσεων καθώς δεν χρησιμοποιείται κάποια μεταβλητή ή οποία δείχνει εάν ένα αντικείμενο είναι σωστά ταξινομημένο ή όχι.
- Έχει παρατηρηθεί ότι μπορεί να επηρεαστεί σημαντικά από την ύπαρξη μη φυσιολογικών στοιχείων στα δεδομένα βάσει των οποίων γίνεται η ανάπτυξη του υπερεπιπέδου.

Τα προβλήματα αυτά αντιμετωπίζονται ενσωματώνοντας στο γραμμικό πρόβλημα που χρησιμοποιείται για την ανάπτυξη του υπερεπιπέδου δύο μεταβλητές για κάθε αντικείμενο: (1) μια μεταβλητή  $a_i$  η οποία αναπαριστά τα σφάλμα ταξινόμησης ενός συγκεκριμένου αντικειμένου  $i$ , δηλαδή το βαθμό στον οποίο το αντικείμενο παραβιάζει το όριο, και (2) μια μεταβλητή  $d_i$  η οποία δείχνει πόσο «σωστά» είναι ταξινομημένο το αντικείμενο, δηλαδή την απόσταση του σωστά ταξινομημένου αντικειμένου από το όριο. Με βάση τις μεταβλητές αυτές όταν ένα αντικείμενο ταξινομείται εσφαλμένα τότε  $a_i > 0$  και  $d_i = 0$ , ενώ σε περίπτωση σωστής ταξινόμησης ενός αντικειμένου θα είναι  $a_i = 0$  και  $d_i \geq 0$ . Η σημασία αυτών των μεταβλητών φαίνεται στο σχήμα 2.2. Η ανάπτυξη του υπερεπιπέδου γίνεται έτσι ώστε να ελαχιστοποιηθεί το άθροισμα των μεταβλητών σφάλματος  $a_i$  και να μεγιστοποιηθεί το άθροισμα των μεταβλητών απόστασης  $d_i$ . Αυτό επιτυγχάνεται με το παρακάτω γραμμικό πρόβλημα:

$$\min \left\{ \sum p_i a_i - \sum q_i d_i \right\}$$

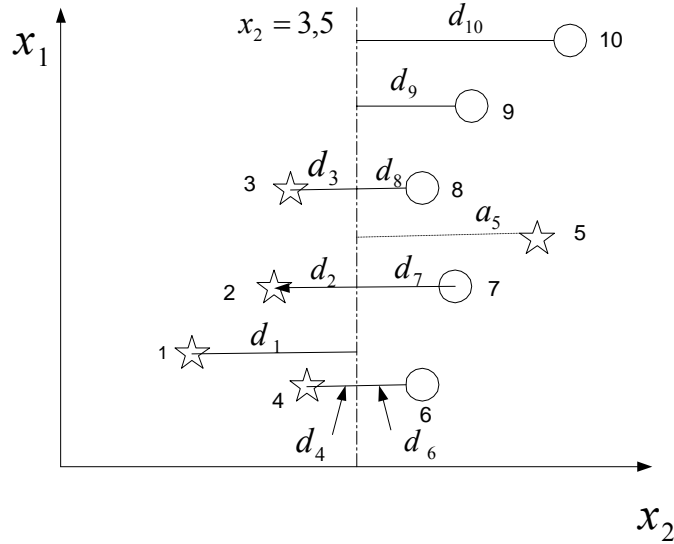
με τους περιορισμούς:

$$A_i X - d = c - a_i \quad \forall i \in G_1$$

$$A_i X + d = c + a_i - \varepsilon \quad \forall i \in G_0$$

$$c + \sum X = M$$

$$d_i, a_i \geq 0$$



**Σχήμα 2.2:** Αναπαράσταση των σφαλμάτων και των αποστάσεων στο μοντέλο OSD (Πηγή: Ζοπουνίδης, 1998)

Στο παραπάνω γραμμικό πρόβλημα το όριο  $c$  είναι μια μεταβλητή η οποία είναι ανεξάρτητη πρόσημου και υπολογίζεται κατά την επίλυση του προβλήματος. Οι παράμετροι  $p_i$  και  $q_i$  καθορίζουν την σχετική σημασία της ελαχιστοποίησης των σφαλμάτων σε σχέση με την μεγιστοποίηση των αποστάσεων. Ο αποφασίζων έχει τη δυνατότητα να καθορίσει τις παραμέτρους αυτές ανεξάρτητα για κάθε αντικείμενο. Λόγου χάρη, αν κριθεί ότι ένα σύνολο αντικειμένων ή ακόμα και ένα μοναδικό αντικείμενο πρέπει να ληφθούν ιδιαίτερα υπόψη στην ανάπτυξη του υπερεπιπέδου, τότε στις αντίστοιχες παραμέτρους  $p_i$  μπορούν να αποδοθούν μεγαλύτερες τιμές σε σχέση με τις παραμέτρους  $q_i$ . Αυτή η δυνατότητα που παρέχει το μοντέλο OSD δίνει ιδιαίτερη ευελιξία στην ανάπτυξη του υπερεπιπέδου ταξινόμησης.

### 3.3 Υβριδικό μοντέλο

Το υβριδικό μοντέλο αναπτύχθηκε από τους Glover et al. (1988). Το μοντέλο χρησιμοποιεί τους παρακάτω τέσσερις διαφορετικούς παράγοντες, οι οποίοι συνδέονται άμεσα με την ακρίβεια της ταξινόμησης και κατά συνέπεια επηρεάζουν την ανάπτυξη του υπερεπιπέδου όπως φαίνεται και στο σχήμα 2.3:

1. τα σφάλματα ταξινόμησης  $a_i$  των αντικειμένων,
2. το μέγιστο σφάλμα  $a_0$ ,
3. τις αποστάσεις  $d_i$  των σωστά ταξινομημένων αντικειμένων από τα όρια, και
4. την ελάχιστη απόσταση  $d_0$

Οι μεταβλητές σφάλματος (τα ανεξάρτητα σφάλματα  $a_i$  και το μέγιστο σφάλμα  $a_0$ ) θα πρέπει να ελαχιστοποιηθούν, ενώ οι αποστάσεις  $d_i$  και η ελάχιστη απόσταση  $d_0$  θα πρέπει να μεγιστοποιηθούν. Η μαθηματική διατύπωση του μοντέλου είναι η εξής:

$$\min \{p_0 a_0 + \sum p_i a_i - q_0 d_0 - \sum q_i d_i\}$$

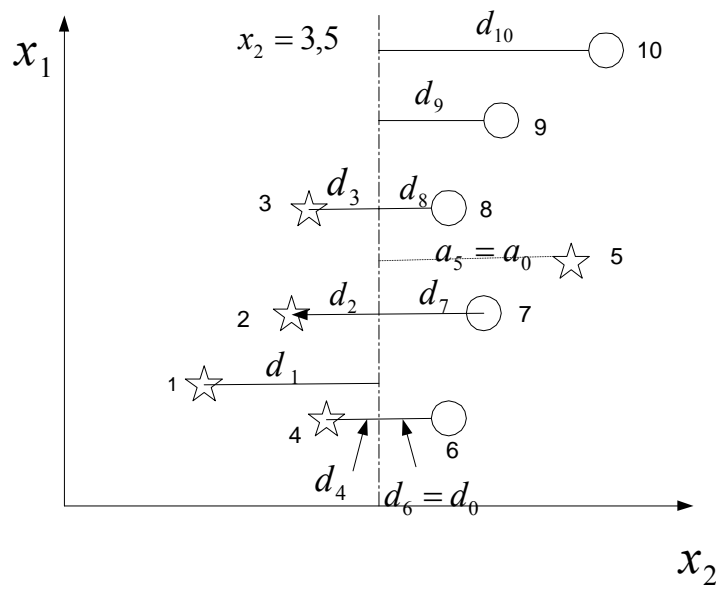
με τους περιορισμούς:

$$A_i X + a_0 + a_i - d_0 - d_i = c \quad \forall i \in G_1$$

$$A_i X - a_0 - a_i + d_0 + d_i = c - \varepsilon \quad \forall i \in G_0$$

$$d_0 + \sum d_i = M$$

$$a_0, a_i, d_0, d_i \geq 0$$



**Σχήμα 2.3:** Γραφική αναπαράσταση υβριδικού μοντέλου (Πηγή: Ζοπουνίδης, 1998)

ΚΕΦΑΛΑΙΟ 3<sup>ο</sup>  
Θεωρία Μηχανών Διανύσματος  
Υποστήριξης



## 1. Εισαγωγή

Η αυξημένη σημαντικότητα του προβλήματος της ταξινόμησης τόσο σε πρακτικό όσο και σε ερευνητικό επίπεδο, έχει ελκύσει το ενδιαφέρον πολλών ερευνητών από διαφορετικούς επιστημονικούς χώρους που στοχεύουν να βελτιστοποιήσουν την αποτελεσματικότητα των μεθόδων ταξινόμησης. Το επίκεντρο της εργασίας αυτής όμως είναι οι μηχανές διανύσματος υποστήριξης και η συμβολή τους στην ταξινόμηση.

Η τεχνική των μηχανών διανύσματος υποστήριξης, παρόλο που δεν είναι ευρέως διαδεδομένη, έχει εφαρμοστεί με θεαματικά αποτελέσματα σε πολλά ζητήματα ταξινόμησης. Έχει εφαρμοστεί, λοιπόν, για ψηφιακή αναγνώριση γραφής (Cortes και Vapnik-1995, Scholkopf, Burges και Vapnik-1995 Scholkopf, Burges και Vapnik-1996, Burges και Scholkopf-1997), αναγνώριση αντικειμένων (Blanz et al-1996), αναγνώριση φωνής (Schmidt-1996), ανίχνευση προσώπων σε εικόνα (Osuna, Freund και Giros- 1997a) και τέλος για κατηγοριοποίηση κειμένων (Joachims-1997).

Όπως έχει ήδη προαναφερθεί οι μηχανές διανύσματος υποστήριξης είναι συστήματα μάθησης για την ανάπτυξη μοντέλων ταξινόμησης, τα οποία κατά κύριο λόγο στηρίζονται σε τεχνικές τετραγωνικού προγραμματισμού. Εν τούτοις έχουν γίνει πολλές προτάσεις για διαδικασίες που εμπεριέχουν γραμμικό προγραμματισμό και προγραμματισμό στόχων με απώτερο σκοπό να περιοριστεί ο υπολογιστικός φόρτος που απαιτείται για την ανάπτυξη βέλτιστων μοντέλων. Οι προσπάθειες αυτές οδήγησαν σε διάφορα μαθηματικά μοντέλα για την ανάπτυξη μηχανών διανύσματος υποστήριξης μέσω προγραμματισμού στόχων. Στα πλαίσια αυτά, λοιπόν, στο κεφάλαιο αυτό εκτός από την θεωρία των SVM θα εξεταστούν και τα μαθηματικά μοντέλα που ενσωματώνουν τεχνικές προγραμματισμού στόχων στις μηχανές διανύσματος υποστήριξης.

## 2. Η γραμμική περίπτωση

### 2.1 Διαχωρίσιμα στοιχεία

Αρχικά παρουσιάζεται η πιο απλή περίπτωση: μηχανές διανύσματος υποστήριξης εκπαιδευόμενες σε διαχωρίσιμα δεδομένα. Έχουμε το δείγμα εκπαίδευσης  $\{x_i, y_i\}$   $i = 1, \dots, l$ ,  $y_i \in \{-1, 1\}$ ,  $x_i \in R^d$ . Υποτίθεται ότι υπάρχει κάποιο υπερεπίπεδο (hyperplane) που διαχωρίζει τα θετικά από τα αρνητικά δεδομένα. Τα σημεία  $x$  που βρίσκονται πάνω στο υπερεπίπεδο ικανοποιούν την σχέση  $w^*x + b = 0$ . Στην γραμμική περίπτωση διαχωρίσιμων στοιχείων ο αλγόριθμος υποστήριξης διανύσματος έχει ως στόχο την ανάπτυξη του υπερεπιπέδου με το μεγαλύτερο περιθώριο. Αυτό μπορεί να αποδοθεί ως εξής (υποθέτουμε ότι όλα τα δεδομένα εκπαίδευσης ικανοποιούν τους ακόλουθους περιορισμούς):

$$x_i^* w + b \geq +1 \quad \text{για } y_i = +1 \quad (1)$$

$$x_i^* w + b \leq -1 \quad \text{για } y_i = -1 \quad (2)$$

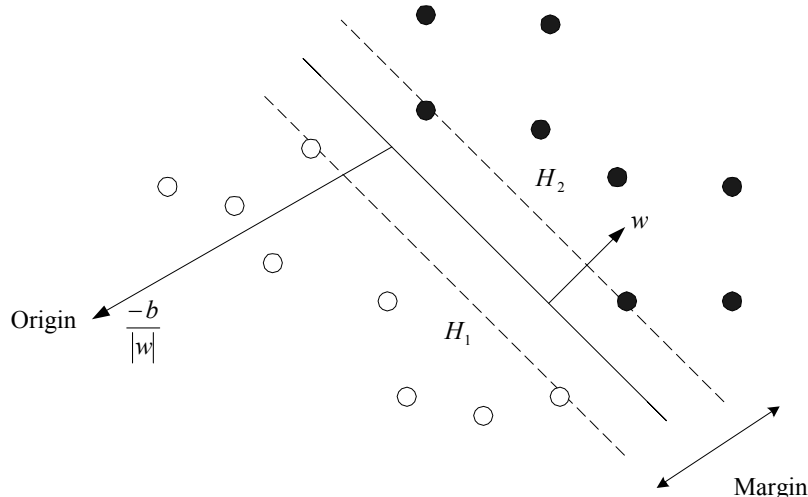
Αυτά μπορούν να εκφραστούν σε μια ανισότητα:

$$y_i(x_i^* w + b) \geq 1 \quad \forall i \quad (3)$$

Τώρα εξετάζονται τα σημεία για τα οποία ισχύει η ισότητα (1). Αυτά τα σημεία βρίσκονται πάνω στο υπερεπίπεδο  $H_1$ :  $x_i^* w + b = 1$ . Αντίστοιχα τα σημεία τα οποία επαληθεύουν την ισότητα (2) βρίσκονται πάνω στο υπερεπίπεδο  $H_2$ :  $x_i^* w + b = -1$ . Το περιθώριο μεταξύ των δύο υπερεπιπέδων είναι  $2/\|w\|$ . Πρέπει να σημειωθεί ότι τα υπερεπίπεδα  $H_1, H_2$  είναι παράλληλα και δεν υπάρχουν δεδομένα εκπαίδευσης ανάμεσα τους. Επομένως μπορούμε να βρούμε το ζευγάρι των υπερεπιπέδων που δίνουν το μεγαλύτερο περιθώριο ελαχιστοποιώντας το  $\|w\|^2$ , υπό τους περιορισμούς (3).

Η λύση για μια τυπική περίπτωση δύο διαστάσεων, αναμένεται να έχει την μορφή που φαίνεται στο σχήμα 3.1. Τα σημεία εκπαίδευσης για τα οποία η ανισότητα (3) επαληθεύεται (αυτά που βρίσκονται πάνω σε ένα από τα δυο υπερεπίπεδα  $H_1, H_2$ ) και

η μετακίνηση τους επηρεάζει την λύση που υπάρχει ονομάζονται διανύσματα υποστήριξης.



**Σχήμα 3.1:** Γραμμική περίπτωση υπερεπίπεδων διαχωρισμού για διαχωρίσιμα δεδομένα. Τα διανύσματα υποστήριξης βρίσκονται πάνω στις διακεκομμένες γραμμές. (Πηγή: Burges, 1998)

Θα εξεταστεί στη συνέχεια η δυϊκή διατύπωση του προβλήματος. Αυτό θα γίνει για δυο λόγους. Πρώτον οι περιορισμοί (3) θα αντικατασταθούν από τους πολλαπλασιαστές Lagrange, οι οποίοι είναι πιο εύκολοι στο χειρισμό και δεύτερον σε αυτή τη διατύπωση του προβλήματος τα δεδομένα εκπαίδευσης θα παρουσιάζονται με τη μορφή εσωτερικού γινόμενου μεταξύ διανυσμάτων.

Εισάγονται, λοιπόν, θετικοί πολλαπλασιαστές Lagrange  $a_i$ ,  $i=1,\dots,l$ , ένας για κάθε περιορισμό της ανίσωσης (3). Με την εισαγωγή των πολλαπλασιαστών Lagrange διαμορφώνεται η συνάρτηση Lagrange ως εξής:

$$L_P \equiv \frac{1}{2} \|w\|^2 - \sum_{i=1}^l a_i y_i (x_i * w + b) + \sum_{i=1}^l a_i \quad (4)$$

Η συνάρτηση αυτή πρέπει να ελαχιστοποιηθεί ως προς τα  $w$  και  $b$  και να μεγιστοποιηθεί ως προς τα  $a_i$ . Για υπολογιστικούς λόγους το πρόβλημα επιλύεται στη δυϊκή του μορφή, η οποία διατυπώνεται ως εξής:

$$\max_a L_D(a) = \max_a \left( \min_{w,b} L_P(w,b,a) \right) \quad (5)$$

Το ελάχιστο της συνάρτησης  $L_P$  ως προς τα  $w$  και  $b$  προσδιορίζεται θέτοντας τις αντίστοιχες παραγώγους της συνάρτησης ίσες με μηδέν:

$$\begin{aligned} \frac{\partial L_P}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^l a_i y_i x_i \\ \frac{\partial L_P}{\partial b} = 0 &\Rightarrow \sum_{i=1}^l a_i y_i = 0 \end{aligned} \quad (6)$$

Βάσει των σχέσεων αυτών το δυϊκό πρόβλημα διατυπώνεται ως εξής:

$$\max \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j x_i \cdot x_j \quad (7)$$

υπό τους περιορισμούς:

$$\begin{aligned} \sum_{i=1}^l a_i y_i &= 0 \\ a_i &\geq 0 \end{aligned} \quad (8)$$

Στη βέλτιστη λύση του προβλήματος αυτού, τα σημεία για τα οποία ισχύει  $a_i > 0$  ονομάζονται διανύσματα υποστήριξης και βρίσκονται πάνω σε ένα από τα υπερεπίπεδα  $H_1, H_2$ . Όλα τα άλλα σημεία εκπαίδευσης έχουν  $a_i = 0$  και βρίσκονται είτε πάνω στα  $H_1, H_2$  (τότε ισχύει η ισότητα (3)) είτε στη σωστή πλευρά του  $H_1$  ή του  $H_2$  όποτε ισχύει η ανισότητα (3). Τα διανύσματα υποστήριξης είναι σημαντικά στοιχεία του συνόλου εκπαίδευσης διότι προσδιορίζουν το μοντέλο ταξινόμησης.

## 2.2 Μη διαχωρίσιμα στοιχεία

Η παραπάνω διαδικασία ανάπτυξης του μοντέλου ταξινόμησης μπορεί να επεκταθεί και στην γενική περίπτωση όπου τα δεδομένα δεν διαχωρίζονται γραμμικά. Αυτό μπορεί να επιτευχθεί και με την εισαγωγή θετικών μεταβλητών απόκλισης  $\xi_i$ ,  $i=1, \dots, l$  στους περιορισμούς (Cortes και Vapnik, 1995) οι οποίες εκφράζονται ως εξής:

$$x_i \cdot w + b \geq +1 - \xi_i \quad \text{για} \quad y_i = +1 \quad (9)$$

$$x_i * w + b \leq -1 + \xi_i \quad \text{για} \quad y_i = -1 \quad (10)$$

$$\xi_i \geq 0 \quad \forall i \quad (11)$$

Έτσι για να συμβεί ένα λάθος, το σφάλμα  $\xi_i$  πρέπει να υπερβεί το όριο, οπότε το  $\sum_i \xi_i$  είναι το ανώτερο όριο στα λάθη του δείγματος εκπαίδευσης. Επομένως ένας φυσικός τρόπος για να οριστεί ένα πρόσθετο κόστος για τα σφάλματα είναι να αλλαχτεί η αντικειμενική συνάρτηση που ελαχιστοποιείται από  $\|w\|^2/2$  σε  $\|w\|^2/2 + C(\sum_i \xi_i)^k$  όπου  $C$  είναι μια παράμετρος που επιλέγεται από τον χρήστη, υψηλό  $C$  αντιστοιχεί σε ανάθεση υψηλής ποινής στα σφάλματα. Όπως παρατηρείται, πρόκειται για ένα κυρτό πρόβλημα για οποιοδήποτε θετικό ακέραιο αριθμό  $k$ . Για  $k=2$  και  $k=1$  είναι επίσης τετραγωνικό πρόβλημα ενώ η επιλογή  $k=1$  έχει το πλεονέκτημα ότι ούτε το  $\xi_i$  ούτε οι πολλαπλασιαστές Lagrange εμφανίζονται στη δυϊκή διατύπωση η οποία έχει τη μορφή:

$$\max L_D \equiv \sum_i a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j x_i * x_j \quad (12)$$

υπό τους περιορισμούς:

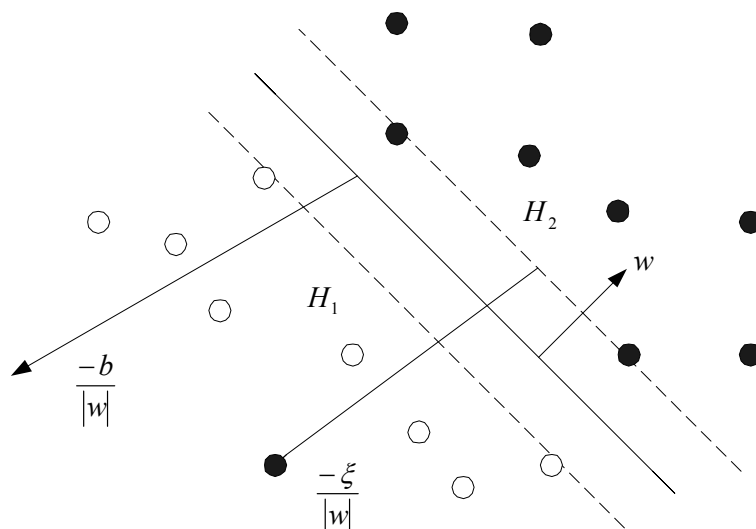
$$0 \leq a_i \leq C, \quad (13)$$

$$\sum_i a_i y_i = 0 \quad (14)$$

Η λύση δίνεται πάλι από:

$$w = \sum_{i=1}^{N_S} a_i y_i x_i \quad (15)$$

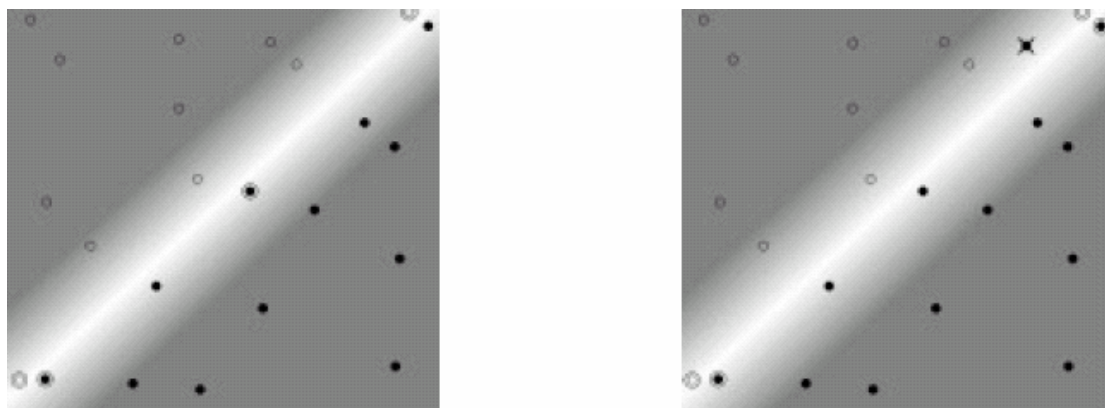
όπου το  $N_S$  ο αριθμός των διανυσμάτων υποστήριξης. Η κατάσταση απεικονίζεται γραφικά στο σχήμα 3.2:



Σχήμα 3.2: Γραμμική περίπτωση υπερεπίπεδων για μη διαχωρίσιμα στοιχεία. (Πηγή: Burges, 1998)

## 2.3 Ενδεικτικά παραδείγματα

Στο σχήμα 3.3 φαίνονται δύο παραδείγματα ενός προβλήματος ταξινόμησης σε δύο ομάδες, ένα με διαχωρίσιμα δεδομένα και ένα με μη διαχωρίσιμα. Η μια ομάδα των δεδομένων αποτελείται από κύκλους και η άλλη από τελείες. Τα διανύσματα υποστήριξης τονίζονται με έναν επιπλέον κύκλο, ενώ το λάθος στα μη διαχωρίσιμα στοιχεία τονίζεται με σταυρό.



Σχήμα 3.3: Γραμμική περίπτωση υπερεπίπεδων για μη διαχωρίσιμα στοιχεία. (Πηγή: Burges, 1998)

### 3. Μη γραμμική περίπτωση

Συχνά παρουσιάζεται η ανάγκη οι παραπάνω μέθοδοι να γενικευτούν, ώστε να έχουν ισχύ όταν η συνάρτηση απόφασης δεν είναι γραμμική. Οι Boser, Guyon και Vapnik (1992) απέδειξαν ότι ένα παλιό τέχνασμα (Aizerman, 1964) μπορεί να χρησιμοποιηθεί, για να επιτευχθεί ο παραπάνω στόχος, μέσω μια απλής διαδικασίας. Καταρχήν πρέπει να σημειωθεί ότι ο τρόπος με τον οποίο εμφανίζονται τα δεδομένα κατά την διαδικασία εκπαίδευσης είναι με τη μορφή εσωτερικού γινομένου  $x_i * x_j$ .

Για την εφαρμογή της διαδικασίας που περιγράφηκε στη γραμμική περίπτωση για την ανάπτυξη μη γραμμικών μοντέλων, τα δεδομένα αναπαριστώνται σε ένα νέο χώρο  $H$  (feature space) μέσω του ορισμού μιας μη γραμμικής αναπαράστασης  $\Phi$ :

$$\Phi : R^d \mapsto H \quad (16)$$

Έτσι ο αλγόριθμος εκπαίδευσης εξαρτάται μόνο από τα δεδομένα με τη μορφή εσωτερικού γινομένου στο  $H$ , δηλαδή από τις συναρτήσεις της μορφής  $\Phi(x_i) * \Phi(x_j)$ .

Αν υπάρχει μια συνάρτηση πυρήνα έτσι ώστε  $K(x_i, x_j) = \Phi(x_i) * \Phi(x_j)$ , τότε μπορεί να χρησιμοποιηθεί η συνάρτηση αυτή στον αλγόριθμο εκπαίδευσης, χωρίς να χρειάζεται να είναι γνωστή η συνάρτηση αναπαράστασης  $\Phi$ . Ένα τυπικό παράδειγμα συνάρτησης πυρήνα είναι ο πυρήνας RBF (radial basis function):

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2} \quad (17)$$

Με την εισαγωγή της συνάρτησης πυρήνα τα εσωτερικά γινόμενα  $x_i * x_j$  που προσδιορίζουν το μοντέλο, αντικαθίστανται από το  $K(x_i, x_j)$  επιτυγχάνοντας έτσι την αναπαράσταση των δεδομένων στο χώρο  $H$  χωρίς να απαιτείται ο προσδιορισμός της συνάρτησης  $\Phi$  και ο υπολογισμός των γινομένων  $\Phi(x_i) * \Phi(x_j)$ . Με τον τρόπο αυτό η εισαγωγή του πυρήνα  $K$  επιτρέπει τη χρήση των διαδικασιών βελτιστοποίησης που περιγράφηκαν προηγούμενα για τη γραμμική περίπτωση στην ανάπτυξη μη γραμμικών μοντέλων.

Ειδικότερα, στη γραμμική περίπτωση το μοντέλο ταξινόμησης έχει την εξής μορφή:

$$f(x) = \sum_{i=1}^{N_s} a_i y_i \Phi(s_i) * \Phi(x) + b = \sum_{i=1}^{N_s} a_i y_i K(s_i, x) + b \quad (18)$$

όπου  $s_i$  είναι τα διανύσματα υποστήριξης

## 4. Μοντέλα προγραμματισμού στόχων που χρησιμοποιούνται στην ανάπτυξη μηχανών διανύσματος υποστήριξης

Στις προηγούμενες ενότητες παρουσιάστηκε το θεωρητικό και μαθηματικό υπόβαθρο των μηχανών διανύσματος υποστήριξης. Στη συνέχεια θα εξεταστούν 4 μοντέλα προγραμματισμού στόχων και ο τρόπος με τον οποίο προσαρμόζονται στις μηχανές διανύσματος υποστήριξης ώστε να βελτιωθούν τα αποτελέσματα ταξινόμησης και να ελαχιστοποιηθούν οι πόροι που χρειάζονται (χρόνος, απαιτήσεις συστήματος κτλ).

### 4.1 1<sup>ο</sup> Μοντέλο προγραμματισμού στόχων

Στην προηγούμενη ενότητα οι μηχανές διανύσματος υποστήριξης χρησιμοποιούσαν την ευκλείδεια νόρμα ( $L_2$ ). Στο συγκεκριμένο μοντέλο χρησιμοποιείται η νόρμα  $L_\infty$  (Pedroso και Murata, 2001). Με τη χρήση αυτής της νόρμας η μεγιστοποίηση του περιθωρίου της ταξινόμησης και η ανάπτυξη του βέλτιστου υπερεπιπέδου διαχωρισμού μπορεί να διαμορφωθεί σαν πρόβλημα γραμμικού προγραμματισμού.

Όταν χρησιμοποιείται η νόρμα  $L_\infty$ , η απόσταση ανάμεσα στα δύο παράλληλα υπερεπίπεδα  $H_1: x_i * w + b_1 = 0$  και  $H_2: x_i * w + b_2 = 0$  δίνεται από τη σχέση :



$$d_{\infty}(H_1, H_2) = \frac{|b_1 - b_2|}{\|w\|_1} \quad (19)$$

#### 4.1.1 Περίπτωση διαχωρίσιμων στοιχείων

Θεωρείται ότι τα διανύσματα υποστήριξης επαληθεύουν την σχέση  $x_i * w + b = \pm 1$ .

Τότε η απόσταση ανάμεσα στα δύο υπερεπίπεδα, που περνούν από τα διανύσματα υποστήριξης της κάθε κατηγορίας, είναι:

$$d_{\infty}(H^+, H^-) = \frac{|(1-b) - (-1-b)|}{\|w\|_1} = \frac{2}{\sum_j |w_j|} \quad (20)$$

Ο στόχος είναι να μεγιστοποιηθεί η (20). Αυτό είναι ισοδύναμο με την ελαχιστοποίηση του:

$$\min \sum_j |w_j| \quad (21)$$

Ο παραπάνω στόχος μπορεί να επιτευχθεί με γραμμικό προγραμματισμό. Αν στην μορφοποίηση προστεθεί για κάθε  $w_j$ , μια βοηθητική μεταβλητή  $a_j$  και οι περιορισμοί  $a_j \geq w_j \quad \forall j$  και  $a_j \geq -w_j \quad \forall j$ , τότε το γραμμικό πρόβλημα είναι η ελαχιστοποίηση του  $\sum_j a_j$  με τους παραπάνω περιορισμούς. Η πλήρης μορφοποίηση του προβλήματος είναι η παρακάτω:

$$\min \sum_{j=1}^d a_j$$

υπό τους περιορισμούς:

$$w * x_i + b \geq +1 \quad \forall i \in P, \text{ (με } P \text{ ορίζεται το σύνολο των θετικών δεδομένων)}$$

$$w * x_i + b \leq -1 \quad \forall i \in N, \text{ (με } N \text{ ορίζεται το σύνολο των αρνητικών δεδομένων)} \quad (22)$$

$$a_j \geq w_j \quad \forall j \in \{1, \dots, d\},$$

$$a_j \geq -w_j \quad \forall j \in \{1, \dots, d\},$$

$$b \in R, a, w \in R^d$$

#### 4.1.2 Περίπτωση μη διαχωρίσιμων στοιχείων

Η περίπτωση αυτή είναι όμοια με την αντίστοιχη για νόρμα  $L_2$ , χρησιμοποιούνται περιορισμοί για την σωστή ταξινόμηση (**soft-margin constraints**) με την ακόλουθη μορφή:

$$\min \left\{ \sum_j |w_j| + C \sum_i \xi_i \right\} \quad (23)$$

Το παραπάνω πρόβλημα λύνεται με γραμμικό προγραμματισμό αν χρησιμοποιηθεί η παρακάτω μορφοποίηση:

$$\min \sum_{j=1}^d a_j + C \left( \sum_{i \in P} \xi_i^+ + \sum_{i \in N} \xi_i^- \right)$$

υπό τους περιορισμούς:

$$w^* x_i + b \geq +1 - \xi_i^+ \quad \forall i \in P,$$

$$w^* x_i + b \leq -1 + \xi_i^- \quad \forall i \in N,$$

$$a_j \geq w_j \quad \forall j \in \{1, \dots, d\},$$

$$a_j \geq -w_j \quad \forall j \in \{1, \dots, d\},$$

$$b \in R, a, w \in R^d, \xi^+ \in R_{0+}^{|P|}, \xi^- \in R_{0+}^{|N|} \quad (24)$$

## 4.2 2<sup>ο</sup> Μοντέλο προγραμματισμού στόχων

Σε αυτό το μοντέλο χρησιμοποιείται η νόρμα  $L_1$  (Pedroso και Murata, 2001), για να υπολογιστεί η απόσταση ανάμεσα στα δεδομένα εκπαίδευσης και τα υπερεπίπεδα διαχωρισμού, και το πρόβλημα μορφοποιείται σαν γραμμικό.

Όταν χρησιμοποιείται η νόρμα  $L_1$  η απόσταση ανάμεσα στα δύο παράλληλα υπερεπίπεδα  $H_1: x_i * w + b_1 = 0$  και  $H_2: x_i * w + b_2 = 0$  δίνεται από τη σχέση:

$$d_1(H_1, H_2) = \frac{|b_1 - b_2|}{\|w\|_\infty} \quad (25)$$

### 4.2.1 Περίπτωση διαχωρίσιμων στοιχείων

Για μια ακόμη φορά τα διανύσματα υποστήριξης επαληθεύουν την σχέση  $x_i * w + b = \pm 1$ , οπότε η απόσταση ανάμεσα στα δύο υπερεπίπεδα που περνούν από τα διανύσματα υποστήριξης της κάθε κατηγορίας είναι:

$$d_1(H^+, H^-) = \frac{|(b+1) - (b-1)|}{\|w\|_\infty} = \frac{2}{\max_j |w_j|} \quad (26)$$

Σκοπός είναι να μεγιστοποιηθεί η (26). Αυτό είναι ισοδύναμο με την ελαχιστοποίηση της παρακάτω σχέσης:

$$\min_{w,b} \max_j |w_j| \quad (27)$$

Αυτό μπορεί να γίνει με γραμμικό προγραμματισμό, εάν προσθέσουμε μια βοηθητική μεταβλητή  $a$  και τους περιορισμούς  $a \geq w_j \quad \forall j$  και  $a \geq -w_j \quad \forall j$ . Στόχος τώρα είναι η ελαχιστοποίηση του  $a$  υπό τους περιορισμούς που αναφέρθηκαν όπως φαίνεται στην παρακάτω μορφοποίηση:

$\min a$

υπό τους περιορισμούς:

$$\begin{aligned}
 w^* x_i + b &\geq +1 \quad \forall i \in P, \\
 w^* x_i + b &\leq -1 \quad \forall i \in N, \\
 a &\geq w_j \quad \forall j \in \{1, \dots, d\}, \\
 a &\geq -w_j \quad \forall j \in \{1, \dots, d\}, \\
 a, b &\in R, w \in R^d
 \end{aligned} \tag{28}$$

#### 4.2.2 Περίπτωση μη διαχωρίσιμων στοιχείων

Όταν τα δεδομένα είναι μη διαχωρίσιμα, χρησιμοποιούνται περιορισμοί σωστής ταξινόμησης (**soft-margin constraints**) συμπεριλαμβανομένου και ενός σφάλματος στην αντικειμενική συνάρτηση όπως στην περίπτωση της  $L_2$  νόρμας. Για να καταστεί δυνατή η μορφοποίηση με γραμμικό προγραμματισμό πρέπει να ισχύσει  $k=1$  στην σχέση  $\|w\|^2/2 + C(\sum_i \xi_i)^k$ . Κατόπιν οι περιορισμοί γίνονται  $w^* x_i + b \geq +1 - \xi_i$  για τα θετικά δεδομένα και  $w^* x_i + b \leq -1 + \xi_i$  για τα αρνητικά δεδομένα, με  $\xi_i \geq 0$ . Η αντικειμενική συνάρτηση τώρα γίνεται :

$$\min_{w, b, \xi} \left\{ \max_j |w_j| + C \sum_i \xi_i \right\} \tag{29}$$

Λύνοντας το πρόβλημα με γραμμικό προγραμματισμό και χρησιμοποιώντας μια βοηθητική μεταβλητή  $\alpha$  που παίρνει την μεγαλύτερη τιμή της  $|w_j|$  η μορφοποίηση του προβλήματος είναι:

$$\min \alpha + C \left( \sum_{i \in P} \xi_i^+ + \sum_{i \in N} \xi_i^- \right)$$

υπό τους περιορισμούς:

$$\begin{aligned}
 w^* x_i + b &\geq +1 - \xi_i^+ \quad \forall i \in P, \\
 w^* x_i + b &\leq -1 + \xi_i^- \quad \forall i \in N, \\
 a &\geq w_j \quad \forall j \in \{1, \dots, d\}, \\
 a &\geq -w_j \quad \forall j \in \{1, \dots, d\}, \\
 a, b &\in R, w \in R^d, \xi^+ \in R_{0+}^{|P|}, \xi^- \in R_{0+}^{|N|}
 \end{aligned} \tag{30}$$

### 4.3 3<sup>ο</sup> Μοντέλο προγραμματισμού στόχων

Μελετώντας την κλασσική θεωρία των μηχανών διανύσματος υποστήριξης κανείς έρχεται αντιμέτωπος με τους παρακάτω περιορισμούς (Mangasarian, 2003):

- 1) Ο υπολογισμός των σφαλμάτων ταξινόμησης γίνεται ως  $\xi_i = \max\{0, 1 - y_i(wx_i + b)\}$ . Ο υπολογισμός αυτός χρησιμοποιείται αντί της απόστασης των στοιχείων που ταξινομούνται εσφαλμένα από τα υπερεπίπεδα ταξινόμησης  $H_1, H_2$ , η οποία είναι

$$\xi'_i = \frac{\max\{0, 1 - y_i(wx_i + b)\}}{\|w\|} \tag{31}$$

- 2) Το περιθώριο λαμβάνεται υπόψη χρησιμοποιώντας τη νόρμα  $L_2$  ( $\|w\|^2$ ) ενώ αντίθετα για τα σφάλματα ταξινόμησης χρησιμοποιείται η νόρμα  $L_1$ .
- 3) Η παράμετρος  $C$  έχει σημαντική επίδραση στα αποτελέσματα, αλλά μπορεί να προσδιοριστεί μόνο μέσω διαδικασιών δοκιμής και λάθους.

Για την αντιμετώπιση των θεμάτων αυτών ο Mangasarian (2003) πρότεινε μια εναλλακτικής διατύπωση του προβλήματος χρησιμοποιώντας τεχνικές γραμμικού προγραμματισμού. Στόχος στη νέα αυτή διατύπωση του προβλήματος είναι η ελαχιστοποίηση της απόστασης των εσφαλμένως ταξινομημένων στοιχείων από τα

υπερεπίπεδα ταξινόμησης σε σχέση με το περιθώριο μεταξύ των υπερεπιπέδων (περιθώριο ταξινόμησης). Έτσι, ο νέος στόχος της ανάλυσης διατυπώνεται ως εξής:

$$\min_{w,b} \frac{\sum_{i=1}^l \max\{0, 1 - y_i(wx_i + b)\}}{\frac{\|w\|}{2\|w\|}} \quad (32)$$

Η μεγιστοποίηση αυτής της συνάρτησης επιτυγχάνεται με την επίλυση του ακόλουθου γραμμικού προγράμματος:

$$\begin{aligned} \min \quad & \sum_{i=1}^l \xi_i \\ \text{Υπό τους περιορισμούς:} \\ & y_i(wx_i + b) + \xi_i \geq 1, \quad \forall i = 1, 2, \dots, l \\ & \xi_i \geq 0 \end{aligned} \quad (33)$$

Αξιοσημείωτο είναι το γεγονός, ότι αυτό το μαθηματικό πρόγραμμα είναι αμετάβλητο όσον αφορά την νόρμα που υιοθετείται για τον υπολογισμό του περιθωρίου. Δηλαδή όλες οι νόρμες οδηγούν στο ίδιο γραμμικό πρόγραμμα.

#### 4.4 4<sup>ο</sup> Μοντέλο προγραμματισμού στόχων

##### 4.4.1 Γραμμική περίπτωση

Έστω τα δεδομένα  $((x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l))$ ,  $(x, y) \in (R^n, R)$  και οι γραμμικές συναρτήσεις απόφασης (**target functions**)  $f(x, w, b) = w^T x + b$ . Εισάγοντας το δομικό κίνδυνο (**structural risk**) σύμφωνα με το θεώρημα 1<sup>1</sup> προκύπτει η παρακάτω μορφοποίηση για τη περίπτωση των διαχωρίσιμων δεδομένων (Zhou et. al., 2002):

$$\max L = \frac{r}{c\|w\|_\beta} \quad (34)$$

<sup>1</sup> Βλέπε παράρτημα.

υπό τους περιορισμούς:

$$y_i(w^T x_i + b) \geq r \quad i = 1, \dots, l$$

$$r \geq 0,$$

όπου  $\|w\|_\beta$  είναι οποιαδήποτε νόρμα διανύσματος και  $c$  μια σταθερά. Εάν  $\|w\|_\beta$  είναι η  $L_2$  νόρμα τότε  $c=1$ .

Τώρα εξετάζεται η νόρμα  $L_\infty$ . Βεβαίως υπάρχει μια θετική σταθερά  $c$  και οι παράμετροι των συναρτήσεων απόφασης θέτονται έτσι ώστε να ισχύει:

$$h \leq \min \left( \left\lceil \frac{c^2 * R^2 * \|w\|_\infty^2}{\Delta^2} \right\rceil, n \right) + 1.$$

Επομένως η σχέση (34) μπορεί να γραφτεί

$$L = \frac{r}{c\|w\|_\infty}. \quad (35)$$

Ο παρανομαστής της (34) μεταφράζεται σε έναν περιορισμό. Καθώς η σταθερά  $c$  δεν επηρεάζει την λύση του γραμμικού προβλήματος μπορεί να αγνοηθεί. Εάν συμβεί αυτό προκύπτει η εξής μορφοποίηση:

$$\max L = r \quad (36)$$

υπό τους περιορισμούς:

$$y_i(w^T x_i + b) \geq r \quad i = 1, \dots, l \quad (37)$$

$$r \geq 0,$$

$$-1 \leq w_j \leq 1 \quad j=1, \dots, n.$$

Η συνάρτηση απόφασης γίνεται:

$$f(x, w, b) = w^T x + b. \quad (38)$$

και το αποτέλεσμα της ταξινόμησης προέρχεται από τη σχέση

$$y = \text{sgn}(f(x, w, b)), \quad (39)$$

Για την περίπτωση των μη διαχωρίσιμων στοιχείων προκύπτει η παρακάτω μορφοποίηση προσθέτοντας την θετική μεταβλητή απόκλισης  $\xi_i$ :

$$\min \quad L = -r + C \sum_{i=1}^l \xi_i \quad (40)$$

υπό τους περιορισμούς:

$$y_i(w^T x_i + b) \geq r - \xi_i \quad i = 1, \dots, l \quad (41)$$

$$r \geq 0,$$

$$-1 \leq w_j \leq 1 \quad j=1, \dots, n.$$

$$\xi_i \geq 0 \quad i = 1, \dots, l$$

όπου  $C > 0$  μια σταθερά. Η συνάρτηση απόφασης για την περίπτωση των μη διαχωρίσιμων στοιχείων είναι ίδια με αυτή για τα διαχωρίσιμα στοιχεία.

#### 4.4.2 Μη γραμμική περίπτωση

Χρησιμοποιώντας το θεώρημα 2<sup>2</sup> η μορφοποίηση του προβλήματος για την περίπτωση διαχωρίσιμων δεδομένων είναι η εξής (Zhou et. al., 2002):

$$\max \quad L = r \quad (42)$$

υπό τους περιορισμούς:

$$y_i \left( \sum_{j=1}^l a_j y_j K(x_j, x_i) + b \right) \geq r \quad i = 1, \dots, l, \quad (43)$$

$$r \geq 0$$

$$-1 \leq a_k \leq 1 \quad k=1, \dots, l$$

---

<sup>2</sup> Βλέπε παράρτημα.



Η συνάρτηση απόφασης παίρνει την μορφή

$$f(x) = \sum_{i=1}^l a_i y_i K(x_j, x_i) + b \quad (44)$$

και το αποτέλεσμα της ταξινόμησης είναι

$$y = \text{sgn}(f(x)) \quad (45)$$

Για τη περίπτωση των μη διαχωρίσιμων στοιχείων, εισάγεται μια μεταβλητή απόκλισης  $\xi_i$  και οδηγεί στην ακόλουθη μορφοποίηση:

$$\min L = -r + C \sum_{i=1}^l \xi_i \quad (46)$$

υπό τους περιορισμούς

$$y_i f(x_i) = y_i \left( \sum_{j=1}^l a_j y_j K(x_j, x_i) + b \right) \geq r - \xi_i \quad i = 1, \dots, l, \quad (47)$$

$$r \geq 0$$

$$-1 \leq a_k \leq 1 \quad k=1, \dots, l,$$

$$\xi_i \geq 0 \quad i = 1, \dots, l,$$

όπου  $C > 0$  μια σταθερά. Η συνάρτηση απόφασης για την περίπτωση των μη διαχωρίσιμων στοιχείων είναι ίδια με αυτή για τα διαχωρίσιμα στοιχεία

ΚΕΦΑΛΑΙΟ 4<sup>ο</sup>  
Εφαρμογή Μοντέλων -  
Ανάλυση Αποτελεσμάτων

## 1. Εισαγωγή

Στην παρούσα εργασία όπως έχει ήδη αναφερθεί γίνεται καταγραφή των μηχανών διανύσματος υποστήριξης που βασίζονται σε διαδικασίες γραμμικού προγραμματισμού και προγραμματισμού στόχων. Οι εν λόγω μέθοδοι έχουν μεγάλη πρακτική σημασία όχι μόνο στον τομέα των επιχειρήσεων (όπως είναι η χορήγηση δανείων σε καταναλωτές και ο διαχωρισμός τους ανάλογα με την αξιοπιστία τους, η ταξινόμηση των επιχειρήσεων ανάλογα με την απόδοση τους σε επικερδείς και μη κτλ), αλλά και σε άλλους επιστημονικούς χώρους όπως στην βιολογία και την ιατρική (ταξινόμηση των ασθενών με βάση τα συμπτώματα) και σε πληθώρα ακόμη πρακτικών εφαρμογών.

Το κύριο πλεονέκτημα των διαδικασιών γραμμικού προγραμματισμού και προγραμματισμού στόχων έναντι τεχνικών μη γραμμικής βελτιστοποίησης εστιάζεται στον περιορισμένο υπολογιστικό φόρτο που απαιτείται για την ανάπτυξη του βέλτιστου μοντέλου. Το γεγονός αυτό ήταν που οδήγησε σε μαθηματικά πρότυπα για την ανάπτυξη μηχανών διανύσματος υποστήριξης μέσω προγραμματισμού στόχων που παρουσιάστηκαν στο προηγούμενο κεφάλαιο. Στην συγκεκριμένη ενότητα στόχος αποτελεί ο εντοπισμός των ιδιαιτεροτήτων και των χαρακτηριστικών κάθε προτύπου και η συγκριτική αξιολόγηση της αποτελεσματικότητας τους σε ένα ευρύ σύνολο δεδομένων. Σημαντικό στοιχείο στην πρακτική εφαρμογή των προτύπων είναι η ακρίβεια που παρουσιάζουν. Οι απαιτήσεις σε ακρίβεια για κάθε είδους εφαρμογή είναι αυξημένη και έτσι η παρούσα σύγκριση δίνει μεγάλη έμφαση στην ακρίβεια των προτύπων.

Στα πλαίσια αυτά θα αναφερθεί η ταυτότητα των δεδομένων που χρησιμοποιήθηκαν στην εφαρμογή, η μεθοδολογία ανάλυσης και τέλος θα παρουσιαστούν και θα αξιολογηθούν τα αποτελέσματα για κάθε πρότυπο, ενώ θα επιχειρηθεί και μια σύγκριση μεταξύ των προτύπων.

## 2. Δεδομένα

Η εφαρμογή της θεωρίας των μηχανών διανύσματος υποστήριξης σε συνδυασμό με την μεθοδολογία του cross-validation (CV) έγινε σε μια ομάδα διαφορετικών δεδομένων. Αυτά τα δεδομένα προέρχονται από μια πολύ γνωστή βάση δεδομένων για μηχανική μάθηση την UCI repository of machine learning του πανεπιστημίου της Καλιφόρνια (Blake και Merz-1998). Από την βάση αυτή επιλέχθηκαν δεκατρία σύνολα δεδομένων, τα οποία παρουσιάζονται στον παρακάτω πίνακα με τα ιδιαίτερα χαρακτηριστικά τους.

	Σύνολα δεδομένων	Αντικείμενα	Υπό συνθήκη χαρακτηριστικά
1	Bupa liver disorders	350	6
2	Hepatitis	160	19
3	Credit screening	690	14
4	Ionosphere	350	33
5	Pima Indians Diabetes	770	8
6	Tic Tac Toe	960	9
7	Breast cancer Wisconsin	570	30
8	Voting	440	16
9	German credit	1000	20
10	Heart disease	270	13
11	Monks-1	550	6
12	Monks-2	600	6
13	Monks-3	550	6

**Πίνακας 4.1:** Τα εξεταζόμενα δεδομένα

Τα παραπάνω δεδομένα επεξεργάστηκαν με τη χρήση λογισμικού Matlab, με το οποίο εφαρμόστηκαν η μεθοδολογία cross-validation και η τεχνική ταξινόμησης των μηχανών διανύσματος υποστήριξης. Για την επεξεργασία τους χρησιμοποιήθηκε ένας αλγόριθμος που περιέχει τα τέσσερα μαθηματικά μοντέλα που αναφέρθηκαν στο προηγούμενο κεφάλαιο. Ο αλγόριθμος εφαρμόστηκε τρεις φορές για γραμμικό πυρήνα, πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού και πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού από όπου προήρθαν και τα τελικά αποτελέσματα.

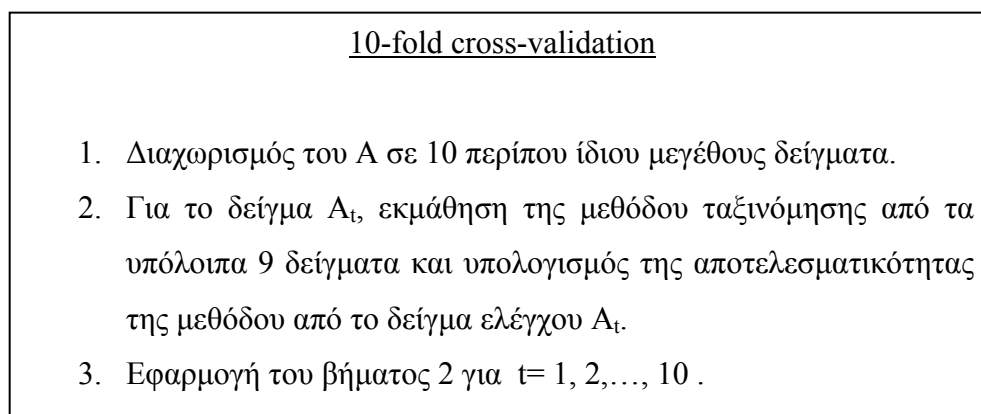
### 3. Η μεθοδολογία cross-validation

Με τη χρήση της μεθοδολογίας cross-validation αξιολογείται ο βαθμός γενίκευσης των αποτελεσμάτων των μεθοδολογιών ταξινόμησης. Στόχος της διαδικασίας είναι η πραγματοποίηση αξιόπιστων εκτιμήσεων για την αποτελεσματικότητα των εξεταζόμενων υποδειγμάτων χρησιμοποιώντας ένα κοινό δείγμα τόσο για την ανάπτυξη υποδειγμάτων όσο και για τον έλεγχο τους. Υπάρχουν δύο σημαντικοί λόγοι για να μετρηθεί ο βαθμός γενίκευσης των αποτελεσμάτων: α) για να εκτιμηθεί η αποτελεσματικότητα μιας μεθόδου ταξινόμησης, β) για να συγκριθούν διάφορες μεθοδολογίες ταξινόμησης μεταξύ τους και να επιλεγεί η πλέον κατάλληλη ανάλογα με τις περιστάσεις.

Στην εργασία αυτή, έγινε εφαρμογή της μεθοδολογίας cross-validation με σκοπό να αξιολογηθεί η αποτελεσματικότητα των μηχανών διανύσματος υποστήριξης με τη χρήση προγραμματισμού στόχων. Κατά την εφαρμογή της θεωρίας των SVM δημιουργήθηκαν δείγματα εκμάθησης και δείγματα ελέγχου. Στην περίπτωση ενός δεδομένου δείγματος εκμάθησης και ενός δείγματος ελέγχου, το δείγμα εκμάθησης χρησιμοποιείται για την ανάπτυξη ενός κατάλληλου μοντέλου ταξινόμησης σύμφωνα με τις αρχές της εκάστοτε μεθόδου, ενώ το δείγμα ελέγχου χρησιμοποιείται για τον έλεγχο της αποτελεσματικότητας του μοντέλου. Όμως, το γεγονός ότι με αυτό τον τρόπο πραγματοποιείται μόνο ένας έλεγχος της αποτελεσματικότητας, αφήνει το ενδεχόμενο τα αποτελέσματα που λαμβάνονται να ήταν τυχαία (μη αντιπροσωπευτικά για την μεθοδολογία). Για την αντιμετώπιση του προβλήματος αυτού, χρησιμοποιήθηκε η επαναληπτική διαδικασία ελέγχου cross-validation.

Κατά την εφαρμογή της διαδικασίας cross-validation, το σύνολο των αντικειμένων  $A$ , που αποτελείται από  $m$  αντικείμενα, χωρίζεται σε  $K$  αμοιβαίως αποκλειόμενα μικρότερου μεγέθους δείγματα  $A_1, A_2, \dots, A_K$  περίπου ίδιου μεγέθους  $m/k$ . Σε κάθε επανάληψη  $t$  αναπτύσσεται ένα μοντέλο, έχοντας ως δείγμα εκμάθησης το δείγμα  $A$  εκτός του  $A_t$ , και ως δείγμα ελέγχου το αποκλειόμενο δείγμα  $A_t$  (Δούμπος, 2003). Συνήθως ο αριθμός των επαναλήψεων  $K$  κυμαίνεται μεταξύ του 1 και του 20. Όμως μπορεί να τεθεί ακόμη και ίσος με  $m$  (leave-one-out cross-validation). Μελέτες

έδειξαν ότι μια τέτοια επιλογή μπορεί να οδηγήσει σε αποτυχία υπολογισμού της πραγματικής αποτελεσματικότητας του μοντέλου ταξινόμησης, ενώ αυξάνει και η διακύμανση των υπολογισμών (Kohavi, 1995). Αντίθετα, αν το  $K$  είναι μικρό, είναι πιθανό ο υπολογισμός του σφάλματος να είναι υπερβολικά απαισιόδοξος, λόγω της διαφοράς στο μέγεθος των δειγμάτων εκμάθησης και ελέγχου που διαμορφώνονται σε κάθε επανάληψη της διαδικασίας cross-validation. Το πρόβλημα γίνεται ακόμη πιο σημαντικό όταν το σύνολο του δείγματος είναι μικρό. Στην περίπτωση αυτή, η επιλογή ενός μικρού αριθμού επαναλήψεων, οδηγεί στην χρήση ανεπαρκών δειγμάτων για την ανάπτυξη του μοντέλου, αφού ο αριθμός των παρατηρήσεων στο σύνολο αναφοράς είναι αρκετά περιορισμένος. Βάσει των παραπάνω παρατηρήσεων, η πιο ευρέως χρησιμοποιούμενη τιμή για τον αριθμό των επαναλήψεων είναι 10 (10-fold cross-validation, Σχήμα 4.1).



**Σχήμα 4.1:** Εφαρμογή του 10-fold cross-validation στο σύνολο αντικειμένων  $A$  (Πηγή: Efron και Tibshirani, 1993)

## 4. Ανάλυση αποτελεσμάτων

### 4.1 Κριτήρια αξιολόγησης

Η παρουσίαση των αποτελεσμάτων αφορά την ακρίβεια ταξινόμησης του κάθε μοντέλου που εφαρμόζεται. Η ακρίβεια δίνεται από δύο μεγέθη, τα οποία θα αποτελέσουν και τα κριτήρια αξιολόγησης των υπό εξέταση μοντέλων, τον δείκτη

ακρίβειας-ΔΑ (accuracy ratio) και το ποσοστό ακρίβειας- ΠΑ (accuracy rate). Με επίκεντρο τα δύο αυτά κριτήρια θα γίνει η προσέγγιση της αποτελεσματικότητας των μοντέλων.

Ο δείκτης ακρίβειας-ΔΑ κυμαίνεται μεταξύ -1 και 1 και υπολογίζεται ως  $\Delta A = 2AUC - 1$ , όπου ως AUC συμβολίζεται το εμβαδόν κάτω από την καμπύλη ROC η οποία αποτυπώνει την αποτελεσματικότητα ενός μοντέλου ταξινόμησης. Ο υπολογισμός του ΔΑ γίνεται εξετάζοντας τα αποτελέσματα των 10 επαναλήψεων της διαδικασίας του cross-validation για καθένα από τα 10 δείγματα ελέγχου.

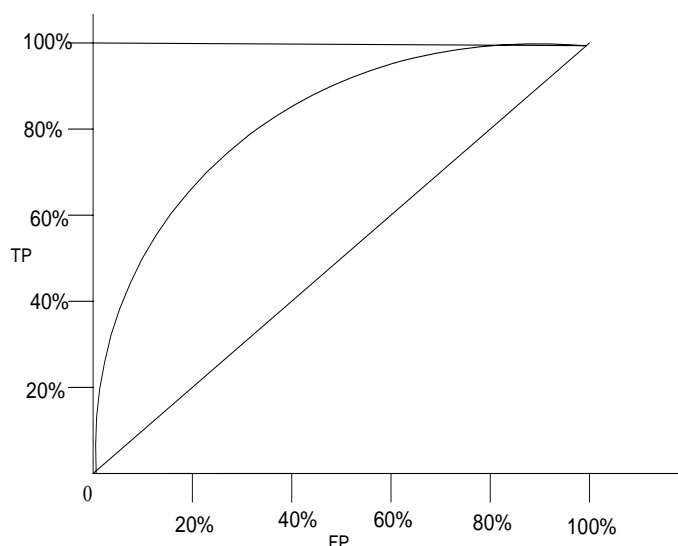
Για να γίνει πιο κατανοητός ο παραπάνω ορισμός του ΔΑ θα πρέπει να προσδιοριστεί η καμπύλη ROC και η χρησιμότητα της. Η καμπύλη ROC (Receiver Operating Characteristic) είναι ένας ιδιαίτερα διαδεδομένος τρόπος αξιολόγησης της αποτελεσματικότητας ενός υποδείγματος ταξινόμησης. Η κατασκευή του διαγράμματος ROC υλοποιείται στην περίπτωση δύο κατηγοριών, αλλά είναι δυνατή η επέκταση της ανάλυσης και στην περίπτωση περισσότερων κατηγοριών. Δεδομένων λοιπόν δύο κατηγοριών,  $C_1$  και  $C_2$ , το διάγραμμα ROC είναι ένα διάγραμμα δύο αξόνων (Σχήμα 4.2). Ο οριζόντιος άξονας δείχνει το ποσοστό των εναλλακτικών που ταξινομήθηκαν από τον αποφασίζοντα στην κατηγορία  $C_2$  και το υπόδειγμα τις ταξινομεί στην κατηγορία  $C_1$  (false positives, FP), ενώ ο κάθετος άξονας δείχνει το ποσοστό των εναλλακτικών που ταξινομήθηκαν από τον αποφασίζοντα στην κατηγορία  $C_1$  και το υπόδειγμα τις ταξινομεί επίσης στην κατηγορία  $C_1$  (true positives, TP).

Για την κατασκευή του διαγράμματος θεωρείται ότι το εξεταζόμενο υπόδειγμα  $U$  αποδίδει μια βαθμολογία  $U(x_i)$  σε κάθε εναλλακτική  $x_i$ . Βάσει αυτής της βαθμολογίας η ταξινόμηση των εναλλακτικών πραγματοποιείται εισάγοντας ένα σημείο διαχωρισμού  $t$ , έτσι ώστε (Δούμπος, 2003):

$$\left. \begin{array}{l} U(x_i) > t \Rightarrow x_i \in C_1 \\ U(x_i) < t \Rightarrow x_i \in C_2 \end{array} \right\}$$

Το διάγραμμα ROC καταγράφει τα αποτελέσματα του υποδείγματος συναρτήσει του σημείου διαχωρισμού  $t$ . Για διαφορετικές τιμές του σημείου διαχωρισμού

καταγράφεται η συμπεριφορά του υποδείγματος στα μεγέθη TP και FP, τα οποία στη συνέχεια τοποθετούνται σε ένα γράφημα δύο αξόνων διαμορφώνοντας έτσι το διάγραμμα ROC του υποδείγματος. Σε ένα εξαιρετικό (άριστο) υπόδειγμα οι βαθμολογίες των εναλλακτικών της κατηγορίας  $C_1$  είναι πάντα υψηλότερες από τις βαθμολογίες των εναλλακτικών της κατηγορίας  $C_2$ . Αντίθετα σε ένα ανεπαρκές υπόδειγμα δεν θα παρατηρείται καμία σχέση μεταξύ των βαθμολογιών των εναλλακτικών και της ταξινόμησης τους οπότε η καμπύλη ROC στην περίπτωση αυτή είναι μια ευθεία γραμμή που ξεκινά από την αρχή των αξόνων. Ένα ικανοποιητικό υπόδειγμα βρίσκεται στο ενδιάμεσο των δύο περιπτώσεων.



Σχήμα 4.2: Διάγραμμα ROC

Η αποτελεσματικότητα του υποδείγματος βάσει του διαγράμματος ROC εξετάζεται από το εμβαδόν της περιοχής κάτω από την καμπύλη ROC (Area Under Curve, AUC). Προφανώς το εμβαδόν αυτό για ένα άριστο υπόδειγμα είναι  $AUC = 1$ , ενώ για ένα ανεπαρκές υπόδειγμα είναι  $AUC = 0,5$ . Στην γενική περίπτωση το εμβαδόν της περιοχής κάτω από την καμπύλη ROC κυμαίνεται μεταξύ 0,5 και 1 με υψηλότερες τιμές να υποδεικνύουν υψηλότερη αποτελεσματικότητα. Στην περίπτωση όπου  $AUC < 0,5$  το υπόδειγμα είναι απολύτως ανεπαρκές.



Για το υπολογισμό του AUC, έστω ότι εξετάζεται ένα υπόδειγμα το οποίο παρέχει  $k$  διακριτές βαθμολογίες  $s_1 < s_2 < \dots < s_k$  στις υπό εξέταση εναλλακτικές. Η πιθανότητα μια εναλλακτική της κατηγορίας  $C_1$  να λάβει το πολύ τη βαθμολογία  $s_i$  συμβολίζεται ως  $p_1^i$  (προφανώς  $p_1^k = 1$ ). Η πιθανότητα αυτή ουσιαστικά αντιστοιχεί στο ποσοστό των εναλλακτικών που ταξινομήθηκαν από τον αποφασίζοντα στην κατηγορία  $C_1$  και λαμβάνουν από το υπόδειγμα βαθμολογία μικρότερη ή ίση με  $s_i$ . Αντίστοιχα, ως  $p_2^i$  συμβολίζεται η αναλογία (πιθανότητα) των εναλλακτικών της κατηγορίας  $C_2$  που λαμβάνουν βαθμολογία μικρότερη ή ίση με  $s_i$  (προφανώς  $p_2^k = 1$ ). Βάσει αυτών των στοιχείων μπορεί να υπολογιστεί το εμβαδόν κάτω από την καμπύλη ROC του υποδείγματος ως εξής (Δούμπος, 2003):

$$AUC = \sum_{i=1}^k \frac{1}{2} (p_1^i - p_1^{i-1}) (p_2^i - p_2^{i-1})$$

Έχοντας έτσι ορίσει την προέλευση και την σημασία του AUC είναι εφικτό να υπολογιστεί ο  $\Delta A$  μέσω της σχέσης  $\Delta A = 2AUC - 1$ .

Αναφορικά με το ποσοστό ακρίβειας-ΠΑ δείχνει το πλήθος των σωστών ταξινομήσεων σε σχέση με το σύνολο των δεδομένων που αξιολογήθηκαν και παίρνει τιμές μεταξύ 0 και 1. Τα αποτελέσματα και σε αυτή την περίπτωση δίνονται για 10 επιμέρους πειράματα σε κάθε σύνολο δεδομένων. Για πρακτικούς λόγους στους πίνακες που αναφέρονται στο ποσοστό ακρίβειας, κάθε στοιχείο είναι ο μέσος όρος των 10 επιμέρους πειραμάτων που πραγματοποιήθηκαν για το αντίστοιχο σύνολο δεδομένων. Αναλυτικά οι τιμές για το ποσοστό ακρίβειας δίνονται στο παράρτημα.

## 4.2 Αποτελέσματα για γραμμικό πυρήνα

Τα αποτελέσματα που προέκυψαν για γραμμικό πυρήνα όπως φαίνεται παρακάτω, είναι σε γενικές γραμμές ικανοποιητικά, ενώ για κάποια δεδομένα οι τιμές είναι πολύ υψηλές. Στον αμέσως επόμενο πίνακα παρουσιάζονται συνοπτικά τα αποτελέσματα του δείκτη ακρίβειας και του ποσοστού ακρίβειας για τον γραμμικό πυρήνα.

	Δείκτης ακρίβειας				Ποσοστό ακρίβειας			
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.34107	0.34752	0.34752	-0.04061	0.6375	0.6375	0.6375	0.475
2	0.71303	0.71127	0.71083	0.69897	0.775	0.7937	0.7812	0.8562
3	0.82842	0.82822	0.82817	0.77432	0.8550	0.8550	0.8550	0.8579
4	0.70614	0.68219	0.68048	0.71478	0.8828	0.8628	0.8628	0.8571
5	0.65953	0.65941	0.65967	<b>0.34122</b>	0.7753	0.7727	0.7727	0.7194
6	0.19416	0.16775	0.20034	0.10758	0.6520	0.6520	0.6520	0.6385
7	0.97705	0.961	0.96761	0.95515	0.9614	0.9526	0.9456	0.9473
8	0.97303	0.97238	0.97251	0.91094	0.9454	0.9477	0.9477	0.9386
9	0.55017	0.54999	0.54998	<b>0.1869</b>	0.754	0.755	0.754	0.669
10	0.80967	0.81067	0.81011	0.69489	0.837	0.837	0.837	0.8037
11	0.50934	0.49817	0.50254	<b>0.26997</b>	0.669	0.6672	0.669	0.5854
12	0.056195	0.03074	0.07199	<b>-0.17118</b>	0.6566	0.6566	0.6566	0.5533
13	0.78472	0.78379	0.77998	0.6836	0.8	0.8	0.8	0.7527

**Πίνακας 4.2:** Αποτελέσματα δείκτη ακρίβειας και ποσοστού ακρίβειας για γραμμικό πυρήνα

Οι τιμές του δείκτη ακρίβειας είναι ικανοποιητικές για όλα τα δεδομένα με εξαίρεση τα δεδομένα 1, 6, 11, 12. Οι τιμές του ποσοστού ακρίβειας για όλα τα σύνολα δεδομένων είναι πάνω από 0.6 με ελάχιστες εξαιρέσεις.

Αξιοσημείωτο είναι το γεγονός ότι το 4<sup>ο</sup> μοντέλο δίνει χαμηλότερες τιμές από όλα τα υπόλοιπα και για το δείκτη ακρίβειας (εκτός από το 4<sup>ο</sup> σύνολο δεδομένων) και για το ποσοστό ακρίβειας (με εξαίρεση τα δεδομένα 2, 3, 7). Σε δύο μάλιστα περιπτώσεις (για τα δεδομένα 1 και 12) υπάρχουν αρνητικές τιμές για τον δείκτη ακρίβειας. Πρέπει να τονιστεί ότι ενώ στα περισσότερα δεδομένα η τιμή του δείκτη ακρίβειας για το 4<sup>ο</sup> μοντέλο είναι χαμηλότερη από αυτή των υπολοίπων μοντέλων, στα δεδομένα 5, 9, 11, 12 (οι τιμές με έντονους χαρακτήρες στον πίνακα 4.2) όμως είναι εντυπωσιακά μικρότερη. Για τα ίδια δεδομένα οι τιμές του ποσοστού ακρίβειας του 4<sup>ου</sup> μοντέλου είναι μικρότερη από αυτές των άλλων μοντέλων αλλά δεν παρουσιάζουν τόσο μεγάλη διαφορά όσο παρατηρείται για το δείκτη ακρίβειας.

Το 1ο μοντέλο δίνει καλύτερο δείκτη ακρίβειας σε όλα τα δεδομένα εκτός, από τα 1 και 10, συγκριτικά με το 2<sup>ο</sup>

Όσον αφορά το ποσοστό ακρίβειας, οι τιμές του για τα μοντέλα 1 και 3 είναι ακριβώς ίδιες για όλα τα δεδομένα εκτός από τα 2, 4, 5, 7, 8 στα οποία οι διαφορές είναι οριακές. Επιπλέον οι τιμές του 4<sup>ου</sup> μοντέλου για τον συγκεκριμένο δείκτη σε 10 από τα 13 (1, 4, 5, 6, 8, 9, 10, 11, 12, 13) σύνολα δεδομένων είναι χαμηλότερη από τα υπόλοιπα μοντέλα.

Για γραμμικό πυρήνα, λοιπόν, είναι φανερό ότι ο δείκτης ακρίβειας και το ποσοστό ακρίβειας είναι ικανοποιητικά στα περισσότερα δεδομένα, μάλιστα για τα 2, 3, 7, 8, 10, 13 φτάνει σε πολύ υψηλά επίπεδα για όλα τα μοντέλα. Αυτό που πρέπει να αναφερθεί είναι η χαμηλότερη ακρίβεια ταξινόμησης που πετυχαίνει το 4<sup>ο</sup> μοντέλο σε σχέση με τα άλλα τρία.

	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
ΜΠΣ1		11/4	8/3	12/11
ΜΠΣ2	2/3		6/3	12/11
ΜΠΣ3	5/2	6/1		12/10
ΜΠΣ4	1/2	1/2	1/3	

**Πίνακας 4.3:** Πίνακας υπεροχής μοντέλων για δείκτη ακρίβειας και ποσοστό ακρίβειας στην γραμμική περίπτωση.

Στον πίνακα υπεροχής 4.3 αναπαριστάται μια συγκεντρωτική εικόνα σχετικά με την υπεροχή του κάθε μοντέλου έναντι των υπόλοιπων για τον δείκτη ακρίβειας και το ποσοστό ακρίβειας. Σε κάθε κελί υπάρχουν δύο τιμές που δείχνουν την υπεροχή του μοντέλου που βρίσκεται στον κάθετο άξονα έναντι του μοντέλου που βρίσκεται στον οριζόντιο άξονα. Η πρώτη τιμή αναφέρεται στον δείκτη ακρίβειας και η δεύτερη στο ποσοστό ακρίβειας. Για παράδειγμα το 11/4 σημαίνει ότι το 1<sup>ο</sup> μοντέλο υπερέχει έναντι του 2<sup>ου</sup> για 11 δεδομένα στο δείκτη ακρίβειας και για 4 δεδομένα στο ποσοστό ακρίβειας. Αυτό που πρέπει να υπογραμμιστεί είναι η απόλυτη υπεροχή όλων των μοντέλων συγκριτικά με το 4<sup>ο</sup> για την πλειοψηφία των δεδομένων και για το δείκτη ακρίβειας και για το ποσοστό ακρίβειας. Υψηλότερη είναι επίσης και η απόδοση του 1<sup>ου</sup> μοντέλου από το 2<sup>ο</sup> όσον αφορά τον δείκτη ακρίβειας για 11 δεδομένα, ενώ αυτό δεν ισχύει για το ποσοστό ακρίβειας όπου υπερέχει μόνο για τα 4 από τα 13 σύνολα δεδομένων. Τέλος το 1<sup>ο</sup> μοντέλο είναι καλύτερο του 3<sup>ου</sup> για 8 από τα 13 δεδομένα για το δείκτη ακρίβειας.

### 4.3 Αποτελέσματα για πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού

Στην συνέχεια με τα ίδια δεδομένα εφαρμόστηκε ο αλγόριθμος για πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού. Η ακρίβεια ταξινόμησης για την δεύτερη περίπτωση σε σχέση με την περίπτωση του γραμμικού πυρήνα για άλλα δεδομένα έχει βελτιωθεί αισθητά, ενώ για κάποια άλλα έχει μειωθεί. Ακολουθεί πίνακας με τις τιμές του δείκτη ακρίβειας και του ποσοστού ακρίβειας (αναλυτικά οι τιμές του ποσοστού ακρίβειας για πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού βρίσκονται στο παράρτημα).

	Δείκτης Ακρίβειας				Ποσοστό ακρίβειας			
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.05996	0.01611	0.01611	-0.02256	0.5125	0.5375	0.5375	0.45
2	0.66469	0.72358	0.28148	0.72709	0.825	0.8437	0.7625	0.8437
3	0.73246	0.74171	0.73844	<b>0.13702</b>	0.8101	0.8101	0.8101	0.5666
4	0.79804	0.83454	0.65483	0.83206	0.88	0.8971	0.8571	0.8971
5	0.63999	0.6401	0.64013	<b>0.085806</b>	0.7597	0.7597	0.7597	0.5519
6	1	1	1	1	0.9979	1	1	1
7	0.91351	0.92363	0.95433	0.92297	0.9245	0.9333	0.9508	0.9315
8	0.88894	0.9128	0.86491	0.91141	0.8954	0.9151	0.9113	0.9151
9	0.36108	0.3635	0.34248	<b>0.14551</b>	0.6840	0.6771	0.6840	0.6044
10	0.56433	0.55989	0.46789	0.51978	0.7185	0.7259	0.7111	0.7259
11	0.94922	0.94922	0.94922	0.87701	0.9709	0.9709	0.9709	0.86
12	0.69361	0.69356	0.69356	<b>0.14233</b>	0.7816	0.7816	0.7816	0.6566
13	0.98587	0.98781	0.99011	0.82064	0.9763	0.9763	0.9763	0.7381

Πίνακας 4.4: Αποτελέσματα δείκτη ακρίβειας και ποσοστού ακρίβειας για πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού

Παρατηρείται, ότι για το 1<sup>ο</sup> σύνολο δεδομένων ο δείκτης ακρίβειας είναι σημαντικά μικρότερος από τον αντίστοιχο για γραμμικό πυρήνα, για όλα τα μοντέλα που εφαρμόστηκαν. Αυτό διαπιστώνεται και για τα υπόλοιπα δεδομένα άλλα σε μικρότερη έκταση. Αντίθετα παρά την μείωση, συγκριτικά με τα αποτελέσματα για γραμμικό πυρήνα, ο δείκτης ακρίβειας για τα δεδομένα 7 και 8 παραμένει σε υψηλά επίπεδα. Αξίζει να αναφερθεί όμως η εντυπωσιακή βελτίωση του δείκτη ακρίβειας σε όλα τα μοντέλα για τα δεδομένα 11 και 12, καθώς και η απόλυτη ακρίβεια ταξινόμησης που πετυχαίνεται και από τα τέσσερα μοντέλα για το 6<sup>ο</sup> σύνολο

δεδομένων, το οποίο παρουσιάζει τις χαμηλότερες τιμές για γραμμικό πυρήνα. Τα μοντέλα 1, 2 και 3 έχουν πετύχει ίδιο η σχεδόν ίδιο δείκτη ακρίβειας για τα δεδομένα 5, 6, 7, 9, 11, 12 και 13.

Και για πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού το 4<sup>ο</sup> μοντέλο δίνει πολύ χαμηλότερο δείκτη ακρίβειας, σε σχέση με τα υπόλοιπα μοντέλα για τα δεδομένα 3, 5, 9, 12 (οι τιμές με έντονους χαρακτήρες στον πίνακα 4.4), έτσι για το 3<sup>ο</sup> σύνολο δεδομένων από το 0,74 που είναι ο δείκτης ακρίβειας περίπου στα μοντέλα 1, 2 και 3, για το 4<sup>ο</sup> μοντέλο πέφτει σε 0,13. Ακόμη το 4<sup>ο</sup> μοντέλο δίνει και αρνητική τιμή δείκτη ακρίβειας για το σύνολο δεδομένων 1.

	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
ΜΠΣ1		3/1	7/3	8/7
ΜΠΣ2	8/7		6/4	10/8
ΜΠΣ3	4/4	3/2		7/8
ΜΠΣ4	4/6	2/0	4/4	

**Πίνακας 4.5:** Πίνακας υπεροχής μοντέλων για δείκτη ακρίβειας και ποσοστό ακρίβειας στην περίπτωση πολυωνυμικού πυρήνα 2<sup>ου</sup> βαθμού.

Από τον παραπάνω πίνακα προκύπτει η υπεροχή όλων των μοντέλων έναντι του 4<sup>ου</sup> μοντέλου τόσο για δείκτη ακρίβειας όσο και για ποσοστό ακρίβειας. Χαρακτηριστική είναι ακόμη και η υπεροχή του 2<sup>ου</sup> έναντι του 1<sup>ου</sup> μοντέλου και για δείκτη ακρίβειας αλλά και για ποσοστό ακρίβειας. Το 1<sup>ο</sup> μοντέλο είναι καλύτερο από το 3<sup>ο</sup> σε 7 από τα 13 σύνολα δεδομένων για το δείκτη ακρίβειας. Αναφορικά με τα μοντέλα 2 και 3 μπορεί να ειπωθεί ότι σε γενικές γραμμές είναι ισορροπημένα.

#### 4.4 Αποτελέσματα για πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού

Τα αποτελέσματα για πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού δεν παρουσιάζουν μια κοινή συμπεριφορά ως προς αυτά των άλλων δύο περιπτώσεων, δεν είναι δυνατόν να δηλωθεί απόλυτα αν τα αποτελέσματα του είναι καλύτερα ή χειρότερα. Έτσι σε κάποια δεδομένα βλέπουμε εντυπωσιακή άνοδο και σε κάποια άλλα μια σταθερότητα

σε σχέση με τα αποτελέσματα του γραμμικού και του πολυωνυμικού πυρήνα 2<sup>ου</sup> βαθμού, όπως φαίνεται και στον πίνακα 4.6.

	Δείκτης Ακρίβειας				Ποσοστό Ακρίβειας			
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.36944	0.30496	0.099936	0.28562	0.675	0.6125	0.5625	0.6125
2	0.65634	0.47264	0.51835	0.46122	0.8375	0.775	0.8	0.7687
3	0.62205	0.66025	0.55865	0.67624	0.7753	0.7884	0.7478	0.7811
4	0.72123	0.74681	0.56342	0.7388	0.8742	0.8714	0.8542	0.8685
5	0.45423	0.07290	0.32815	<b>-0.09087</b>	0.7103	0.5454	0.6961	0.5051
6	0.99804	0.98625	0.99857	0.99467	0.9875	0.9791	0.9864	0.9843
7	0.95381	0.95467	0.87636	0.95433	0.9473	0.9438	0.9192	0.9438
8	0.94081	0.94068	0.84742	0.94085	0.9363	0.9227	0.9136	0.9249
9	0.40609	0.4335	0.21475	<b>0.3899</b>	0.705	0.716	0.638	0.705
10	0.69256	0.71789	0.64611	0.71467	0.7740	0.7851	0.7592	0.7851
11	1	1	1	1	1	1	1	1
12	0.96254	0.95698	0.96713	<b>0.30215</b>	0.9816	0.9683	0.9633	0.6366
13	0.974	0.97167	0.98998	0.98919	0.9563	0.9545	0.9727	0.9672

**Πίνακας 4.6:** Αποτελέσματα δείκτη ακρίβειας και ποσοστού ακρίβειας για πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού.

Ο δείκτης ακρίβειας για πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού δίνει πολύ υψηλές τιμές για τα δεδομένα 6, 7, 8, 10, 11, 12, 13 ενώ για τα υπόλοιπα οι τιμές είναι χαμηλές. Πρέπει να σημειωθεί ότι οι τιμές του δείκτη ακρίβειας για τα δεδομένα 2, 3, 5 και 9 είναι χαμηλότερες από αυτές για γραμμικό και πολυωνυμικό 2<sup>ου</sup> βαθμού πυρήνα. Για το σύνολο δεδομένων 11 πετυχαίνεται η απόλυτη ταξινόμηση και από τα τέσσερα μοντέλα.

Το 4<sup>ο</sup> μοντέλο για γραμμικό και πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού δίνει χαμηλότερες τιμές από τα υπόλοιπα μοντέλα αναφορικά με τον δείκτη ακρίβειας, στην περίπτωση όμως του πολυωνυμικού πυρήνα 3<sup>ου</sup> βαθμού δεν συμβαίνει αυτό, για αρκετά δεδομένα οι τιμές που δίνει είναι υψηλότερες από αυτές των άλλων μοντέλων. Αξιοσημείωτο είναι το γεγονός ότι για τα δεδομένα 5, 9 και 12 οι τιμές του 4<sup>ου</sup> μοντέλου (οι τιμές με έντονους χαρακτήρες στον πίνακα 4.6) είναι πολύ μικρότερες από τις υπόλοιπες. Για τα ίδια δεδομένα, αυτό έχει παρατηρηθεί και για την περίπτωση του γραμμικού και του πολυωνυμικού 2<sup>ου</sup> βαθμού πυρήνα. Αυτή η διαφορά στην απόδοση του 4<sup>ου</sup> μοντέλου δεν αντικατροπτίζεται στο ποσοστό ακρίβειας, όπου εκεί η απόδοση του μοντέλου είναι μεν μικρότερη άλλα είναι κοντά

στις τιμές των υπόλοιπων μοντέλων, με εξαίρεση το σύνολο δεδομένων 9 όπου δεν έχει την χειρότερη τιμή από τα άλλα μοντέλα.

Για 8 από τα 13 σύνολα δεδομένων το 1<sup>ο</sup> μοντέλο δίνει καλύτερο ποσοστό ακρίβειας συγκριτικά με τα υπόλοιπα. Επιπλέον για τα δεδομένα 6, 7, 8, 12, 13 το ποσοστό ακρίβειας είναι πολύ υψηλό, ενώ για το 11 είναι το ιδανικό (ΠΑ=1 για όλα τα μοντέλα).

	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
ΜΠΣ1		7/9	9/11	6/8
ΜΠΣ2	5/3		7/8	7/5
ΜΠΣ3	3/1	5/4		5/5
ΜΠΣ4	6/3	5/4	7/8	

**Πίνακας 4.7:** Πίνακας υπεροχής μοντέλων για δείκτη ακρίβειας και ποσοστό ακρίβειας στην περίπτωση πολυωνυμικού πυρήνα 3<sup>ου</sup> βαθμού.

Στον πίνακα 4.7 αξίζει να υπογραμμιστεί η υπεροχή του 1<sup>ου</sup> μοντέλου έναντι του 3<sup>ου</sup> και για το δείκτη ακρίβειας και για το ποσοστό ακρίβειας. Το 2<sup>ο</sup> μοντέλο έχει καλύτερες επιδόσεις από το 3<sup>ο</sup> για 7 δεδομένα για το δείκτη ακρίβειας και για 8 για το ποσοστό ταξινόμησης. Το 4<sup>ο</sup> μοντέλο είναι καλύτερο από το 3<sup>ο</sup> για 7 και 8 δεδομένα για τον δείκτη ακρίβειας και το ποσοστό ακρίβειας αντίστοιχα. Γενικά για την περίπτωση του πολυωνυμικού πυρήνα 3<sup>ου</sup> βαθμού, είναι φανερό ότι με εξαίρεση τα δεδομένα 1, 5, 9 που έχουν πολύ χαμηλούς δείκτες, οι τιμές για τα άλλα δεδομένα είναι από ικανοποιητικές ως πολύ καλές.

## 4.5 Σύνοψη αποτελεσμάτων

Συνοψίζοντας τα παραπάνω αποτελέσματα είναι δυνατή η εξαγωγή ορισμένων γενικών παρατηρήσεων αναφορικά με την αποτελεσματικότητα των εξεταζόμενων μεθόδων. Για το λόγο αυτό είναι χρήσιμο να συγκεντρωθούν σε πίνακες, οι μέσες τιμές όλων των μοντέλων του δείκτη ακρίβειας και του ποσοστού ακρίβειας κάθε συνόλου δεδομένων για τις τρεις περιπτώσεις πυρήνα (γραμμικό πυρήνα, πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού και πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού):

	Γραμμικός πυρήνας		Πολυωνυμικός πυρήνας 2 <sup>ου</sup> βαθμού		Πολυωνυμικός πυρήνας 3 <sup>ου</sup> βαθμού	
	ΔΑ	ΠΑ	ΔΑ	ΠΑ	ΔΑ	ΠΑ
1	0.2488	0.5968	0.0174	0.5093	0.2650	0.6156
2	0.7085	0.8015	0.5992	0.8187	0.5271	0.7953
3	0.8147	0.8557	0.5874	0.7492	0.6293	0.7732
4	0.6958	0.8663	0.7798	0.8828	0.6926	0.8671
5	0.5799	0.76	0.5015	0.7077	0.1911	0.6142
6	0.1674	0.6486	1	0.9994	0.9944	0.9843
7	0.9652	0.9517	0.9286	0.9350	0.9348	0.9385
8	0.9572	0.9448	0.8945	0.9092	0.9174	0.9244
9	0.4592	0.733	0.3031	0.6623	0.3611	0.6910
10	0.7813	0.8286	0.5279	0.7203	0.6928	0.7759
11	0.445	0.6476	0.9311	0.9431	1	1
12	-0.003	0.6307	0.5557	0.7503	0.7972	0.8875
13	0.758	0.7881	0.9461	0.9167	0.9812	0.9627
M.O.	0.5829	0.7733	0.6594	0.8080	0.6911	0.8331

**Πίνακας 4.8:** Συγκεντρωτικός πίνακας μέσων τιμών των δεδομένων και για τις τρεις περιπτώσεις πυρήνα

Ο δείκτης ακρίβειας για γραμμικό πυρήνα είναι μεγαλύτερος για τα περισσότερα σύνολα δεδομένων με εξαίρεση τα σύνολα 4, 6, 11, 12, 13 τα οποία έχουν καλύτερο δείκτη ακρίβειας για πολυωνυμικό πυρήνα 2<sup>ου</sup> ή 3<sup>ου</sup> βαθμού. Ο δείκτης ακρίβειας του 6<sup>ου</sup> συνόλου δεδομένων, ενώ για γραμμικό πυρήνα παρουσιάζει πολύ χαμηλές τιμές (την δεύτερη χαμηλότερη από όλα τα δεδομένα), για πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού πετυχαίνει την μέγιστη τιμή, δηλαδή την ιδεατή ταξινόμηση και για πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού αγγίζει το ιδανικό. Αξιοσημείωτο είναι το γεγονός ότι στο 6<sup>ο</sup> σύνολο δεδομένων για πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού και στο 11<sup>ο</sup> για πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού πετυχαίνεται ο ιδανικός δείκτης ακρίβειας δηλαδή ΔΑ=1. Αναφορικά με το ποσοστό ακρίβειας, για το 6<sup>ο</sup> και 11<sup>ο</sup> σύνολο δεδομένων, είναι εντυπωσιακή η βελτίωση του από την περίπτωση του γραμμικού πυρήνα στις άλλες δυο περιπτώσεις πυρήνων. Πρέπει να σημειωθούν και οι υψηλές τιμές του ποσοστού ακρίβειας για τα δεδομένα 7 και 8 σε όλες τις περιπτώσεις πυρήνων.

Στο πίνακα 4.9 που ακολουθεί παραθέτουμε συγκεντρωτικά για κάθε δεδομένο τη μέση τιμή (για δείκτη ακρίβειας και ποσοστό ακρίβειας) του κάθε μοντέλου και για τις τρεις περιπτώσεις πυρήνα.



	Δείκτης Ακρίβειας				Ποσοστό Ακρίβειας			
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.2568	0.2229	0.1545	0.0742	0.6083	0.5958	0.5792	0.5125
2	0.6780	0.6358	0.5036	0.6291	0.8125	0.8041	0.7812	0.8229
3	0.7276	0.7434	0.7084	0.5292	0.8135	0.8178	0.8043	0.7352
4	0.7418	0.7545	0.6329	0.7619	0.8790	0.8771	0.8580	0.8742
5	0.5846	0.4575	0.5427	0.1121	0.7484	0.6926	0.7428	0.5921
6	0.7307	0.7180	0.7330	0.7008	0.8791	0.8770	0.8795	0.8743
7	0.9481	0.9464	0.9328	0.9442	0.9444	0.9432	0.9385	0.9409
8	0.9343	0.9420	0.8949	0.9211	0.9257	0.9285	0.9242	0.9262
9	0.4391	0.4490	0.3691	0.2408	0.7143	0.7160	0.6920	0.6595
10	0.6889	0.6962	0.6414	0.6431	0.7765	0.7827	0.7691	0.7716
11	0.8195	0.8158	0.8173	0.7157	0.8800	0.8794	0.8800	0.8151
12	0.5708	0.5604	0.5776	0.0911	0.8066	0.8022	0.8005	0.6155
13	0.9149	0.9144	0.9200	0.8311	0.9109	0.9103	0.9163	0.8193
M.O.	0.6950	0.6813	0.6483	0.5534	0.8230	0.8174	0.8127	0.7661

**Πίνακας 4.9:** Συγκεντρωτικός πίνακας μέσων τιμών των μοντέλων και για τρεις περιπτώσεις πυρήνα.

Παρατηρείται στον πίνακα 4.9 ότι σε γενικές γραμμές οι μέσες τιμές των μοντέλων είναι πιο κοντά τόσο για τον δείκτη ακρίβειας όσο και για το ποσοστό ακρίβειας. Οι μέσοι όροι αποδεικνύουν την υπεροχή του 1<sup>ου</sup> μοντέλου έναντι των άλλων τριών τόσο για το δείκτη ακρίβειας όσο και για το ποσοστό ακρίβειας. Η διαφορά που είχε εντοπιστεί στην επίδοση του 4<sup>ου</sup> μοντέλου συγκριτικά με τα υπόλοιπα 3 δεν είναι τόσο μεγάλη στις μέσες τιμές. Τούτο συμβαίνει γιατί οι τιμές αυτές έχουν επηρεαστεί πολύ από την απόδοση του 4<sup>ου</sup> μοντέλου για πολωνυμικό πυρήνα 3<sup>ου</sup> βαθμού που είναι στις περισσότερες περιπτώσεις πολύ υψηλή. Μόνο για τα δεδομένα 5 και 12 υπάρχει σημαντική διαφορά ανάμεσα στην επίδοση του 4<sup>ου</sup> μοντέλου και των υπόλοιπων τριών. Για τα δεδομένα 6, 7, 8, 10, 11, 13 οι μέσες τιμές του δείκτη ακρίβειας είναι πάνω από 0.65. Σχετικά με το ποσοστό ακρίβειας για τα δεδομένα 2, 3, 4, 6, 7, 8, 9, 10, 11, 12 και 13 είναι πάνω από 0.65 και σε μερικές περιπτώσεις πάνω και από 0.90.

Με βάση τον πίνακα 4.9 θα κατασκευαστεί ένας νέος πίνακας που θα περιέχει την υπεροχή κάθε μοντέλου ως προς τα υπόλοιπα για το δείκτη ακρίβειας και το ποσοστό ακρίβειας. Στην ανάλυση που προηγήθηκε για κάθε πυρήνα, υπάρχει ένας τέτοιος πίνακας μόνο που αναφέρεται στην περίπτωση του κάθε πυρήνα ξεχωριστά. Ο πίνακας που θα προκύψει από τις μέσες τιμές των μοντέλων, θα δίνει συγκεντρωτικές

πληροφορίες για την υπεροχή των μοντέλων λαμβάνοντας υπόψιν τα αποτελέσματα και των τριών πυρήνων.

	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
ΜΠΣ1		8/9	10/10	12/11
ΜΠΣ2	5/4		8/9	8/12
ΜΠΣ3	3/2	5/4		8/8
ΜΠΣ4	1/2	5/1	5/5	

**Πίνακας 4.10:** Συγκεντρωτικός πίνακας υπεροχής μοντέλων .

Επαναλαμβάνεται ότι ο πρώτος αριθμός κάθε κελιού αντιστοιχεί στο δείκτη ακρίβειας και ο δεύτερος στο ποσοστό ακρίβειας κι εκφράζει την υπεροχή του μοντέλου του κάθετου άξονα έναντι του μοντέλου του οριζόντιου άξονα. Προκύπτει, λοιπόν από τον πίνακα 4.10 μια ξεκάθαρη υπεροχή του 1<sup>ου</sup> μοντέλου, και για τα δύο μεγέθη, έναντι των υπόλοιπων με πιο εντυπωσιακή την ανωτερότητα του ως προς το 4<sup>ο</sup> μοντέλο. Το 2<sup>ο</sup> μοντέλο είναι καλύτερο από το 3<sup>ο</sup> και το 4<sup>ο</sup> και το 3<sup>ο</sup> καλύτερο του 4<sup>ου</sup>.

Συνοψίζοντας, λοιπόν όλα τα προηγούμενα τρία είναι τα σημεία που πρέπει να δοθεί ιδιαίτερη βαρύτητα ως προς την αποτελεσματικότητα των μοντέλων που αναπτύχθηκαν:

- Η υπεροχή του 1<sup>ου</sup> μοντέλου έναντι των υπόλοιπων τριών και για τα δύο εξεταζόμενα μεγέθη και σε κάποιες περιπτώσεις με μεγάλη διαφορά.
- Η αδυναμία του 4<sup>ου</sup> μοντέλου που δίνει χαμηλές τιμές και για το δείκτη ακρίβειας και για το ποσοστό ακρίβειας. Πρέπει να τονιστεί, ότι για πολυωνυμικό πυρήνα 2<sup>ου</sup> και 3<sup>ου</sup> βαθμού, παρόλο που συνεχίζει να δίνει χαμηλότερες τιμές από τα άλλα μοντέλα, η επίδοση του βελτιώνεται σημαντικά. Η συμπεριφορά του 4<sup>ου</sup> μοντέλου οφείλεται στο γεγονός ότι είναι πιθανό να έχει ως βέλτιστη λύση τη μηδενική<sup>3</sup>. Θα πρέπει να σημειωθεί ότι παρ' όλο που το 4<sup>ο</sup> μοντέλο έχει την χειρότερη απόδοση, οι επιδόσεις του δεν είναι τόσο χαμηλές που να το καταστούν αναξιόπιστο, αλλά παραμένουν κοντά σε αυτές των άλλων μοντέλων. Ο προηγούμενος ισχυρισμός

<sup>3</sup> Αναλυτικά βλέπε παράρτημα.

επιβεβαιώνεται από το γεγονός ότι έχει δείκτη ακρίβειας πάνω από 0.60 για 8 από τα 13 δεδομένα και ποσοστό ακρίβειας πάνω από 0.60 για 11 από τα 13 δεδομένα (βλέπετε πίνακα 4.9).

- Ο χρόνος επεξεργασία που απαιτούνταν. Παρατηρήθηκε ότι όλα τα μοντέλα, για δεδομένα που είχαν πολλά χαρακτηριστικά, χρειάζονταν πολύ χρόνο για την επεξεργασία τους. Έτσι τα δεδομένα 5, 6, και 9 ήταν τα πιο χρονοβόρα κατά την επεξεργασία. Πέρα όμως από τα χαρακτηριστικά ο χρόνος επεξεργασίας των δεδομένων ήταν και συνέπεια του πυρήνα που χρησιμοποιήθηκε. Ο χρόνος που απαιτήθηκε για την επεξεργασία των δεδομένων, ήταν περισσότερος στις περιπτώσεις μη γραμμικού πυρήνα.

ΚΕΦΑΛΑΙΟ 5<sup>ο</sup>  
Συμπεράσματα και  
Μελλοντικές Επεκτάσεις

## Συμπεράσματα και Μελλοντικές Επεκτάσεις

Το πρόβλημα της ταξινόμησης παρουσιάζει αυξημένο ερευνητικό και πρακτικό ενδιαφέρον. Η διαπίστωση αυτή επιβεβαιώνεται από την πληθώρα και το εύρος των πρακτικών εφαρμογών που η προβληματική της ταξινόμησης παρουσιάζει.

Ο σκοπός της συγκεκριμένης έρευνας ήταν η διερεύνηση της αποτελεσματικότητας μιας συγκεκριμένης τεχνικής ταξινόμησης, των μηχανών διανύσματος υποστήριξης. Το ενδιαφέρον εστιάστηκε ιδιαίτερα στην ανάπτυξη των μηχανών διανύσματος υποστήριξης με τη χρήση προγραμματισμού στόχων, από όπου προέκυψαν τέσσερα μοντέλα τα οποία αποτέλεσαν το επίκεντρο της ανάλυσης.

Στα πλαίσια αυτά, τα παραπάνω μοντέλα εφαρμόστηκαν για ένα ευρύ σύνολο δεδομένων που προήρθαν από μια πολύ γνωστή βάση δεδομένων για μηχανική μάθηση, την UCI repository of machine learning του πανεπιστημίου της Καλιφόρνια. Τα δεδομένα επεξεργάστηκαν με τη χρήση λογισμικού Matlab, με το οποίο εφαρμόστηκαν η μεθοδολογία cross-validation και η τεχνική ταξινόμησης των μηχανών διανύσματος υποστήριξης. Ο αλγόριθμος που χρησιμοποιήθηκε για την επεξεργασία των δεδομένων περιείχε τα τέσσερα μαθηματικά μοντέλα που παρουσιάστηκαν. Ο αλγόριθμος εφαρμόστηκε τρεις φορές για γραμμικό πυρήνα, πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού και πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού από όπου προήρθαν τα τελικά αποτελέσματα.

Για την αξιολόγηση των υπό εξέταση μοντέλων χρησιμοποιήθηκαν δύο μεγέθη που προσδιορίζουν την ακρίβεια, ο δείκτης ακρίβειας και το ποσοστό ακρίβειας. Με βάση τα δύο αυτά κριτήρια έγινε η προσέγγιση της αποτελεσματικότητας των μοντέλων. Από τα αποτελέσματα φάνηκε μια διακύμανση στην επίδοση των μοντέλων για όλα τα δεδομένα και στις τρεις περιπτώσεις πυρήνα. Αναφορικά με τα δεδομένα εντυπωσιακή ήταν η συμπεριφορά ορισμένων τα οποία παρουσίασαν έντονη διακύμανση από πυρήνα σε πυρήνα και για τα δύο κριτήρια. Η εκτεταμένη ανάλυση των τεσσάρων μοντέλων οδήγησε στο συμπέρασμα ότι το 1<sup>ο</sup> μοντέλο υπερέχει έναντι των υπολοίπων τριών και παράγει τις υψηλότερες τιμές τόσο για δείκτη ακρίβειας όσο και για ποσοστό ακρίβειας. Αντίθετα το 4<sup>ο</sup> μοντέλο παρουσίασε την χειρότερη απόδοση από όλα.

Η χρήση των μοντέλων που παρουσιάστηκαν στη παρούσα εργασία μπορεί να επεκταθεί και σε δεδομένα που περιέχουν και ποιοτικά χαρακτηριστικά προκειμένου να μελετηθεί η αποτελεσματικότητά τους. Μεγάλο ενδιαφέρον παρουσιάζει επίσης η εξέταση της συμπεριφοράς των μοντέλων και σε πραγματικά δεδομένα. Η εμπειρία έχει αποδείξει ότι η πραγματικότητα διαφέρει από το εργαστηριακό περιβάλλον, με αποτέλεσμα να αποτελεί πρόκληση η βελτίωση των μεθόδων ή η δημιουργία μεθόδων που θα συμπεριφέρονται άριστα σε πραγματικό περιβάλλον. Επιπλέον, δεδομένου ότι στην πράξη συχνά απαιτείται η ανάλυση δεδομένων μεγάλων διαστάσεων, είναι χρήσιμο να αναπτυχθούν κατάλληλοι αλγόριθμοι που θα περιορίσουν το χρόνο εκπαίδευσης των μοντέλων.

Βέβαια κρίνεται απαραίτητη εφαρμογή των μοντέλων και σε άλλα προβλήματα αποφάσεων, τα οποία απαιτούν την ταξινόμηση ενός συνόλου εναλλακτικών ενεργειών όπως είναι η αξιολόγηση της δανειοληπτικής ικανότητας επιχειρήσεων, η εκτίμηση πιστωτικού κινδύνου επιχειρήσεων, η επιλογή προσωπικού κ.λ.π. Χρήσιμη θα ήταν η επέκταση των μοντέλων και στην αντιμετώπιση προβλημάτων ταξινόμησης με περισσότερες από δύο ομάδες, καθώς πολλά πλέον προβλήματα αφορούν ταξινόμηση σε περισσότερες από δύο ομάδες και η πραγματοποίηση συγκρίσεων με μεθόδους ταξινόμησης από το χώρο της πολυκριτήριας ανάλυσης αλλά και τεχνικές ταξινόμησης από άλλους επιστημονικούς χώρους.

Η διερεύνηση των παραπάνω μελλοντικών ερευνητικών κατευθύνσεων θα συμβάλει ουσιαστικά στην πληρέστερη αντιμετώπιση των προβλημάτων λήψης αποφάσεων που απαιτούν τη χρησιμοποίηση της προβληματικής της ταξινόμησης, μέσω του πλαισίου των μηχανών διανύσματος υποστήριξης με τη χρήση προγραμματισμού στόχων, το οποίο θα είναι σε θέση να οδηγήσει στην ανάπτυξη υποδειγμάτων υψηλής αποτελεσματικότητας-ακρίβειας, παρέχοντας παράλληλα ένα διευρυμένο σύνολο πληροφοριών για την υποστήριξη των αποφάσεων υπό καθεστώς βεβαιότητας και αβεβαιότητας.

# ΠΑΡΑΡΤΗΜΑ

## Α) Αποδείξεις θεωρημάτων

### Θεώρημα 1°

Δίνεται το σύνολο δεδομένων  $((x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l))$ ,  $(x, y) \in (R^n, R)$  και το διάνυσμα  $x$  που ανήκει σε μια σφαίρα ακτίνας  $R$ . Αν το περιθώριο  $m_\Delta$  των υπερεπιπέδων  $f(x, w, b) = w^T x + b$  ταξινομεί τα διανύσματα  $x$  ως εξής:

$$y = \begin{cases} 1, w^T x + b \geq \Delta, \\ -1, w^T x + b \leq -\Delta, \end{cases} \quad \Delta \geq 0,$$

τότε υπάρχει μια σταθερά  $0 \leq c \leq +\infty$  ανεξάρτητη από το σύνολο των δεδομένων  $(x, y)$  και τις παραμέτρους  $w$  και  $b$  έτσι ώστε:

$$h \leq \min \left( \left\lceil \frac{c^2 * R^2 * \|w\|_\beta^2}{\Delta^2} \right\rceil, n \right) + 1$$

Όπου  $\|w\|_\beta$  είναι οποιαδήποτε νόρμα και  $h$  είναι η διάσταση VC η οποία συνδέεται άμεσα με τη δυνατότητα γενίκευσης του μοντέλου.

### Θεώρημα 2°

Δίνεται το σύνολο δεδομένων  $((\Phi(x_1), y_1), \dots, (\Phi(x_i), y_i), \dots, (\Phi(x_l), y_l))$ ,  $(\Phi(x), y) \in (H, R)$  και το διάνυσμα  $\chi$  που ανήκει σε μια σφαίρα ακτίνας  $R$ . Το μοντέλο ταξινόμησης είναι:

$$f(x) = \sum_{i=1}^l a_i y_i \Phi(x_i) * \Phi(x) + b = \sum_{i=1}^l a_i y_i K(x_i, x) + b$$

Η ταξινόμηση ενός παραδείγματος προκύπτει από την σχέση:

$$y = \begin{cases} 1, f(x) \geq \Delta, \\ -1, f(x) \leq -\Delta, \end{cases} \quad \Delta \geq 0.$$

Τότε υπάρχει μια σταθερά  $0 < c < +\infty$  ανεξάρτητη από το σύνολο των δεδομένων  $(x, y)$  και τις παραμέτρους  $a$  και  $b$  ώστε να ισχύσει η παρακάτω ανισότητα:



$$h \leq \min \left( \left\lceil \frac{c^2 * R^2 * \|a\|_\beta^2}{\Delta^2} \right\rceil, n \right) + 1$$

όπου  $a = (a_1, \dots, a_l)^T$  και  $\|a\|_\beta$  είναι οποιαδήποτε νόρμα.

## Β) Επεξήγηση συμπεριφοράς 4<sup>ου</sup> μοντέλου

Το 4<sup>ο</sup> μοντέλο είναι προβληματικό γιατί είναι πιθανόν να έχει ως βέλτιστη λύση τη μηδενική:

$$(w, b, r, \xi) = 0$$

Αυτό φαίνεται εάν αγνοηθούν οι περιορισμοί  $-1 \leq w \leq 1$ , τότε το πρόβλημα γίνεται:

$$\max L = r - Ce_1^T \xi_1 - Ce_2^T \xi_2$$

υπό τους περιορισμούς:

$$-w^T x_1 - e_1 b + e_1 r - \xi_1 \leq 0$$

$$w^T x_2 + e_2 b + e_2 r - \xi_2 \leq 0$$

$$r, \xi \geq 0,$$

με  $x_1 =$  δεδομένα κατηγορίας 1,  $x_2 =$  δεδομένα κατηγορίας -1, και  $e_1, e_2$  διανύσματα με 1.

Το δυϊκό του προβλήματος αυτού είναι ένα σύστημα ανισο-εξισώσεων με την εξής μορφή:

$$-x_1^T a_1 + x_2^T a_2 = 0$$

$$-e_1^T a_1 + e_2^T a_2 = 0$$

$$e_1^T a_1 + e_2^T a_2 \geq 1$$

$$a_1 \leq Ce_1$$

$$a_2 \leq Ce_2$$

$$a_1, a_2 \geq 0$$

Αν το σύστημα αυτό έχει λύση τότε  $\max L = r - Ce_1^T \xi_1 - Ce_2^T \xi_2 = 0$  (προφανώς για  $w = b = r = \xi = 0$  αυτό ισχύει). Αν το σύστημα του δυϊκού δεν είναι εφικτό, τότε το πρωτεύον πρόβλημα είτε δεν έχει λύση είτε δεν είναι φραγμένο. Λύση όμως υπάρχει ( $w = b = r = \xi = 0$ ) άρα σε μια τέτοια περίπτωση το πρόβλημα δεν είναι φραγμένο. Η εισαγωγή των περιορισμών  $-1 \leq w \leq 1$  φράσει το πρόβλημα αλλά δεν αποφεύγει την περίπτωση της μηδενικής λύσης, καθώς εάν η μηδενική λύση είναι βέλτιστη για το παραπάνω πρόβλημα, τότε είναι βέλτιστη και μετά την εισαγωγή των περιορισμών  $-1 \leq w \leq 1$ .

## Γ) Αναλυτικά αποτελέσματα ποσοστού ακρίβειας

### 1. Για γραμμικό πυρήνα

Σύνολο δεδομένων 1					Σύνολο δεδομένων 2				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.5	0.5	0.5	0.375	1	0.625	0.625	0.625	0.8125
2	0.25	0.25	0.25	0.375	2	0.8125	0.875	0.875	0.875
3	0.75	0.75	0.75	0.625	3	0.75	0.8125	0.75	0.8125
4	0.875	0.875	0.875	0.5	4	0.875	0.875	0.875	0.9375
5	0.375	0.375	0.375	0.625	5	0.8125	0.8125	0.875	0.8125
6	0.75	0.75	0.75	0.375	6	0.75	0.8125	0.8125	0.8125
7	0.75	0.75	0.75	0.5	7	0.8125	0.8125	0.8125	0.8125
8	0.75	0.75	0.75	0.5	8	0.625	0.625	0.625	0.8125
9	0.625	0.625	0.625	0.25	9	0.8125	0.8125	0.8125	0.9375
10	0.75	0.75	0.75	0.625	10	0.875	0.875	0.875	0.9375

Σύνολο δεδομένων 3					Σύνολο δεδομένων 4				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.85507	0.85507	0.85507	0.86957	1	0.88571	0.88571	0.88571	0.88571
2	0.91304	0.91304	0.91304	0.88406	2	0.91429	0.94286	0.94286	0.85714
3	0.84058	0.84058	0.84058	0.84058	3	0.8	0.8	0.8	0.91429
4	0.82609	0.82609	0.82609	0.85507	4	0.85714	0.71429	0.71429	0.71429
5	0.88406	0.88406	0.88406	0.86957	5	0.88571	0.88571	0.88571	0.82857
6	0.84058	0.84058	0.84058	0.85507	6	0.85714	0.82857	0.82857	0.8
7	0.82609	0.82609	0.82609	0.81159	7	0.97143	0.91429	0.94286	0.91429
8	0.85507	0.85507	0.85507	0.85507	8	0.85714	0.85714	0.82857	0.85714
9	0.91304	0.91304	0.91304	0.91304	9	0.91429	0.91429	0.91429	0.91429
10	0.7971	0.7971	0.7971	0.82609	10	0.88571	0.88571	0.88571	0.88571

Σύνολο δεδομένων 5					Σύνολο δεδομένων 6				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.71429	0.71429	0.71429	0.64935	1	0.63542	0.63542	0.63542	0.6875
2	0.80519	0.79221	0.79221	0.74026	2	0.57292	0.57292	0.57292	0.59375
3	0.88312	0.88312	0.88312	0.81818	3	0.625	0.625	0.625	0.61458
4	0.7013	0.68831	0.68831	0.7013	4	0.75	0.75	0.75	0.61458
5	0.76623	0.76623	0.76623	0.67532	5	0.64583	0.64583	0.64583	0.72917
6	0.77922	0.77922	0.77922	0.63636	6	0.60417	0.60417	0.60417	0.5625
7	0.79221	0.79221	0.79221	0.75325	7	0.66667	0.66667	0.66667	0.61458
8	0.77922	0.77922	0.77922	0.77922	8	0.65625	0.65625	0.65625	0.64583
9	0.77922	0.77922	0.77922	0.75325	9	0.63542	0.63542	0.63542	0.58333
10	0.75325	0.75325	0.75325	0.68831	10	0.72917	0.72917	0.72917	0.73958

Σύνολο δεδομένων 7					Σύνολο δεδομένων 8				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.91228	0.91228	0.92982	0.91228	1	1	1	1	0.97727
2	1	1	0.92982	0.94737	2	0.95455	0.95455	0.95455	0.95455
3	0.94737	0.94737	0.96491	0.94737	3	0.95455	0.95455	0.95455	0.95455
4	0.96491	0.92982	0.96491	0.92982	4	0.90909	0.90909	0.90909	0.88636
5	0.98246	0.92982	0.92982	0.96491	5	0.95455	0.95455	0.95455	0.97727
6	0.94737	0.94737	0.92982	0.94737	6	0.97727	0.97727	0.97727	0.97727
7	0.92982	0.89474	0.91228	0.89474	7	0.93182	0.93182	0.93182	0.93182
8	0.94737	0.98246	0.94737	0.96491	8	0.95455	0.95455	0.95455	0.90909
9	0.98246	0.98246	0.98246	0.98246	9	0.90909	0.90909	0.90909	0.93182
10	1	1	0.96491	0.94737	10	0.90909	0.93182	0.93182	0.88636
Σύνολο δεδομένων 9					Σύνολο δεδομένων 10				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.76	0.76	0.76	0.66	1	0.66667	0.66667	0.66667	0.74074
2	0.73	0.73	0.73	0.59	2	0.81481	0.81481	0.81481	0.77778
3	0.71	0.71	0.71	0.66	3	0.85185	0.85185	0.85185	0.81481
4	0.83	0.83	0.83	0.72	4	1	1	1	1
5	0.75	0.75	0.75	0.61	5	0.88889	0.88889	0.88889	0.92593
6	0.75	0.76	0.75	0.66	6	0.88889	0.88889	0.88889	0.81481
7	0.69	0.69	0.69	0.68	7	0.81481	0.81481	0.81481	0.77778
8	0.81	0.81	0.81	0.73	8	0.81481	0.81481	0.81481	0.7037
9	0.74	0.74	0.74	0.73	9	0.81481	0.81481	0.81481	0.74074
10	0.77	0.77	0.77	0.65	10	0.81481	0.81481	0.81481	0.74074
Σύνολο δεδομένων 11					Σύνολο δεδομένων 12				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.70909	0.70909	0.70909	0.67273	1	0.63333	0.63333	0.63333	0.6
2	0.70909	0.70909	0.70909	0.58182	2	0.7	0.7	0.7	0.6
3	0.69091	0.67273	0.69091	0.70909	3	0.66667	0.66667	0.66667	0.51667
4	0.56364	0.58182	0.58182	0.56364	4	0.68333	0.68333	0.68333	0.51667
5	0.70909	0.70909	0.70909	0.54545	5	0.68333	0.68333	0.68333	0.56667
6	0.69091	0.69091	0.69091	0.52727	6	0.68333	0.68333	0.68333	0.58333
7	0.70909	0.69091	0.69091	0.50909	7	0.65	0.65	0.65	0.56667
8	0.58182	0.58182	0.58182	0.49091	8	0.63333	0.63333	0.63333	0.6
9	0.67273	0.67273	0.67273	0.63636	9	0.6	0.6	0.6	0.35
10	0.65455	0.65455	0.65455	0.61818	10	0.63333	0.63333	0.63333	0.63333

Σύνολο δεδομένων 13				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
<b>1</b>	0.78182	0.78182	0.78182	0.8
<b>2</b>	0.74545	0.74545	0.74545	0.72727
<b>3</b>	0.78182	0.78182	0.78182	0.74545
<b>4</b>	0.87273	0.87273	0.87273	0.78182
<b>5</b>	0.76364	0.76364	0.76364	0.65455
<b>6</b>	0.8	0.8	0.8	0.76364
<b>7</b>	0.81818	0.81818	0.81818	0.76364
<b>8</b>	0.8	0.8	0.8	0.78182
<b>9</b>	0.90909	0.90909	0.90909	0.83636
<b>10</b>	0.72727	0.72727	0.72727	0.67273

2. Για πολυωνυμικό πυρήνα 2<sup>ου</sup> βαθμού

Σύνολο δεδομένων 1					Σύνολο δεδομένων 2				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.375	0.5	0.5	0.5	1	0.875	0.875	0.6875	0.875
2	0.5	0.5	0.5	0.25	2	0.875	0.875	0.875	0.875
3	0.625	0.5	0.5	0.5	3	0.8125	0.9375	0.6875	0.9375
4	0.75	0.75	0.75	0.25	4	0.9375	0.875	0.875	0.875
5	0.5	0.5	0.5	0.875	5	0.6875	0.875	0.8125	0.875
6	0.625	0.625	0.625	0.5	6	0.875	0.8125	0.9375	0.8125
7	0.125	0.375	0.375	0.625	7	0.75	0.75	0.625	0.75
8	0.5	0.5	0.5	0.125	8	0.6875	0.8125	0.6875	0.8125
9	0.625	0.625	0.625	0.25	9	0.875	0.8125	0.75	0.8125
10	0.5	0.5	0.5	0.625	10	0.875	0.8125	0.6875	0.8125
Σύνολο δεδομένων 3					Σύνολο δεδομένων 4				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.82609	0.7971	0.81159	0.57971	1	0.91429	0.94286	0.85714	0.94286
2	0.76812	0.75362	0.72464	0.44928	2	0.88571	0.88571	0.85714	0.88571
3	0.7971	0.85507	0.84058	0.52174	3	0.91429	0.91429	0.97143	0.91429
4	0.7971	0.7971	0.7971	0.6087	4	0.88571	0.88571	0.82857	0.88571
5	0.82609	0.84058	0.84058	0.46377	5	0.82857	0.85714	0.82857	0.85714
6	0.89855	0.88406	0.89855	0.55072	6	0.82857	0.94286	0.82857	0.94286
7	0.73913	0.75362	0.75362	0.43478	7	0.88571	0.94286	0.82857	0.94286
8	0.84058	0.85507	0.85507	0.72464	8	0.82857	0.82857	0.82857	0.82857
9	0.88406	0.85507	0.88406	0.5942	9	0.91429	0.82857	0.85714	0.82857
10	0.72464	0.71014	0.69565	0.73913	10	0.91429	0.94286	0.88571	0.94286
Σύνολο δεδομένων 5					Σύνολο δεδομένων 6				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.76623	0.76623	0.76623	0.38961	1	1	1	1	1
2	0.81818	0.81818	0.81818	0.58442	2	1	1	1	1
3	0.83117	0.83117	0.83117	0.67532	3	1	1	1	1
4	0.64935	0.64935	0.64935	0.50649	4	1	1	1	1
5	0.72727	0.72727	0.72727	0.62338	5	1	1	1	1
6	0.75325	0.75325	0.75325	0.54545	6	0.97917	1	1	1
7	0.79221	0.79221	0.79221	0.74026	7	1	1	1	1
8	0.75325	0.75325	0.75325	0.49351	8	1	1	1	1
9	0.79221	0.79221	0.79221	0.55844	9	1	1	1	1
10	0.71429	0.71429	0.71429	0.4026	10	1	1	1	1

Σύνολο δεδομένων 7					Σύνολο δεδομένων 8				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.91228	0.91228	0.91228	0.91228	1	0.93182	0.95455	0.93182	0.95455
2	0.85965	0.94737	0.92982	0.94737	2	0.88636	0.93182	0.86364	0.93182
3	0.91228	0.91228	0.89474	0.91228	3	0.90909	0.95455	0.93182	0.95455
4	0.94737	0.92982	0.96491	0.92982	4	0.90909	0.90909	0.84091	0.90909
5	0.92982	0.89474	0.96491	0.89474	5	0.84091	0.88636	0.97727	0.88636
6	0.94737	0.94737	0.94737	0.94737	6	0.90909	0.93182	0.95455	0.93182
7	0.89474	0.84211	0.96491	0.84211	7	0.88636	0.93182	0.88636	0.93182
8	0.94737	0.96491	0.94737	0.96491	8	0.88636	0.88636	0.86364	0.88636
9	0.98246	1	1	1	9	0.90909	0.88636	0.93182	0.88636
10	0.91228	0.98246	0.98246	0.96491	10	0.88636	0.88636	0.93182	0.88636
Σύνολο δεδομένων 9					Σύνολο δεδομένων 10				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.74	0.73	0.74	0.67	1	0.59259	0.55556	0.62963	0.55556
2	0.68	0.68	0.68	0.4	2	0.55556	0.51852	0.62963	0.51852
3	0.65	0.65	0.64	0.48	3	0.81481	0.81481	0.77778	0.81481
4	0.68	0.69	0.7	0.77	4	0.81481	0.81481	0.88889	0.81481
5	0.73	0.73	0.73	0.63	5	0.74074	0.77778	0.81481	0.77778
6	0.72	0.73	0.69	0.36	6	0.77778	0.77778	0.7037	0.77778
7	0.64	0.64	0.64	0.65	7	0.81481	0.88889	0.74074	0.88889
8	0.71	0.71	0.71	0.58	8	0.66667	0.7037	0.62963	0.7037
9	0.62	0.62	0.65	0.63	9	0.66667	0.66667	0.66667	0.66667
10	0.67	0.66	0.66	0.63	10	0.74074	0.74074	0.62963	0.74074
Σύνολο δεδομένων 11					Σύνολο δεδομένων 12				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.96364	0.96364	0.96364	0.85455	1	0.78333	0.78333	0.78333	0.63333
2	0.94545	0.94545	0.94545	0.89091	2	0.75	0.75	0.75	0.7
3	0.92727	0.92727	0.92727	0.81818	3	0.88333	0.88333	0.88333	0.66667
4	0.94545	0.94545	0.94545	0.85455	4	0.66667	0.66667	0.66667	0.68333
5	0.96364	0.96364	0.96364	0.83636	5	0.75	0.75	0.75	0.68333
6	1	1	1	0.85455	6	0.81667	0.81667	0.81667	0.68333
7	1	1	1	0.94545	7	0.83333	0.83333	0.83333	0.65
8	0.98182	0.98182	0.98182	0.87273	8	0.8	0.8	0.8	0.63333
9	0.98182	0.98182	0.98182	0.83636	9	0.81667	0.81667	0.81667	0.6
10	1	1	1	0.83636	10	0.71667	0.71667	0.71667	0.63333

Σύνολο δεδομένων 13				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
<b>1</b>	0.98182	0.98182	0.98182	0.8
<b>2</b>	0.98182	0.98182	0.98182	0.96364
<b>3</b>	0.98182	0.98182	0.98182	0.52727
<b>4</b>	1	1	1	0.76364
<b>5</b>	0.96364	0.96364	0.96364	0.70909
<b>6</b>	0.94545	0.94545	0.94545	0.63636
<b>7</b>	0.94545	0.94545	0.94545	0.67273
<b>8</b>	0.96364	0.96364	0.96364	0.67273
<b>9</b>	1	1	1	0.81818
<b>10</b>	1	1	1	0.81818



3. Για πολυωνυμικό πυρήνα 3<sup>ου</sup> βαθμού

Σύνολο δεδομένων 1					Σύνολο δεδομένων 2				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.625	0.75	0.625	0.75	1	0.875	0.875	0.8125	0.875
2	0.625	0.5	0.5	0.5	2	0.9375	0.75	0.9375	0.75
3	0.875	0.75	0.5	0.75	3	0.75	0.6875	0.75	0.6875
4	0.5	0.375	0.75	0.375	4	0.8125	0.8125	0.8125	0.8125
5	0.75	0.75	0.5	0.75	5	0.75	0.75	0.625	0.75
6	0.625	0.375	0.5	0.375	6	0.9375	0.9375	0.875	0.9375
7	0.75	0.625	0.5	0.625	7	0.8125	0.75	0.75	0.6875
8	0.75	0.875	0.5	0.875	8	0.625	0.625	0.75	0.625
9	0.5	0.5	0.5	0.5	9	0.9375	0.8125	0.875	0.8125
10	0.75	0.625	0.75	0.625	10	0.9375	0.75	0.8125	0.75

Σύνολο δεδομένων 3					Σύνολο δεδομένων 4				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.85507	0.82609	0.78261	0.82609	1	0.88571	0.94286	0.85714	0.94286
2	0.73913	0.78261	0.72464	0.73913	2	0.91429	0.85714	0.85714	0.85714
3	0.73913	0.78261	0.69565	0.78261	3	1	0.91429	0.94286	0.88571
4	0.75362	0.75362	0.78261	0.71014	4	0.8	0.8	0.88571	0.8
5	0.76812	0.76812	0.72464	0.76812	5	0.77143	0.85714	0.88571	0.85714
6	0.81159	0.78261	0.7971	0.81159	6	0.91429	0.85714	0.77143	0.85714
7	0.76812	0.78261	0.71014	0.78261	7	0.94286	0.85714	0.85714	0.85714
8	0.84058	0.82609	0.76812	0.82609	8	0.85714	0.82857	0.8	0.82857
9	0.76812	0.81159	0.82609	0.81159	9	0.77143	0.91429	0.88571	0.91429
10	0.71014	0.76812	0.66667	0.75362	10	0.88571	0.88571	0.8	0.88571

Σύνολο δεδομένων 5					Σύνολο δεδομένων 6				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.71429	0.58442	0.71429	0.38961	1	0.96875	0.96875	0.97917	0.96875
2	0.7013	0.33766	0.72727	0.51948	2	0.95833	0.90625	0.94792	0.94792
3	0.75325	0.76623	0.75325	0.36364	3	1	1	1	1
4	0.64935	0.54545	0.5974	0.54545	4	1	1	1	0.98958
5	0.75325	0.37662	0.75325	0.31169	5	1	1	1	1
6	0.67532	0.24675	0.68831	0.62338	6	0.96875	0.96875	0.96875	0.96875
7	0.66234	0.71429	0.48052	0.74026	7	0.97917	0.96875	0.98958	0.98958
8	0.76623	0.50649	0.77922	0.49351	8	1	0.98958	1	0.98958
9	0.71429	0.75325	0.75325	0.54545	9	1	1	1	1
10	0.71429	0.62338	0.71429	0.51948	10	1	0.98958	0.97917	0.98958

Σύνολο δεδομένων 7					Σύνολο δεδομένων 8				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.94737	0.92982	0.85965	0.92982	1	0.97727	0.97727	0.88636	0.97727
2	0.91228	0.94737	0.92982	0.94737	2	0.95455	0.90909	0.81818	0.90909
3	0.92982	0.91228	0.91228	0.91228	3	0.97727	0.97727	0.95455	0.97727
4	0.94737	0.94737	0.94737	0.94737	4	0.90909	0.90909	0.86364	0.90909
5	0.96491	0.96491	0.89474	0.96491	5	0.90909	0.93182	0.93182	0.93182
6	0.92982	0.92982	0.92982	0.92982	6	0.97727	0.97727	0.95455	0.97727
7	0.91228	0.91228	0.87719	0.91228	7	0.90909	0.88636	0.95455	0.88636
8	0.98246	0.96491	0.94737	0.96491	8	0.88636	0.86364	0.93182	0.88636
9	0.96491	0.98246	0.98246	0.98246	9	0.93182	0.88636	0.90909	0.88636
10	0.98246	0.94737	0.91228	0.94737	10	0.93182	0.90909	0.93182	0.90909
Σύνολο δεδομένων 9					Σύνολο δεδομένων 10				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	0.68	0.69	0.7	0.68	1	0.66667	0.62963	0.66667	0.62963
2	0.7	0.74	0.61	0.69	2	0.7037	0.77778	0.7037	0.77778
3	0.69	0.67	0.6	0.67	3	0.81481	0.74074	0.81481	0.74074
4	0.8	0.8	0.62	0.8	4	0.77778	0.85185	0.85185	0.85185
5	0.73	0.76	0.69	0.73	5	0.85185	0.88889	0.81481	0.88889
6	0.66	0.7	0.65	0.73	6	0.85185	0.81481	0.74074	0.81481
7	0.7	0.67	0.64	0.67	7	0.81481	0.85185	0.85185	0.85185
8	0.69	0.73	0.64	0.68	8	0.77778	0.77778	0.7037	0.77778
9	0.67	0.68	0.61	0.68	9	0.62963	0.74074	0.7037	0.74074
10	0.73	0.72	0.62	0.72	10	0.85185	0.77778	0.74074	0.77778
Σύνολο δεδομένων 11					Σύνολο δεδομένων 12				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4		ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
1	1	1	1	1	1	1	1	0.98333	0.63333
2	1	1	1	1	2	0.98333	0.96667	0.98333	0.7
3	1	1	1	1	3	1	1	1	0.66667
4	1	1	1	1	4	0.98333	1	0.96667	0.68333
5	1	1	1	1	5	0.96667	0.95	0.88333	0.7
6	1	1	1	1	6	0.96667	0.96667	0.91667	0.68333
7	1	1	1	1	7	0.96667	0.98333	0.95	0.65
8	1	1	1	1	8	0.96667	0.83333	0.98333	0.41667
9	1	1	1	1	9	1	1	0.98333	0.6
10	1	1	1	1	10	0.98333	0.98333	0.98333	0.63333

Σύνολο δεδομένων 13				
	ΜΠΣ1	ΜΠΣ2	ΜΠΣ3	ΜΠΣ4
<b>1</b>	0.90909	0.89091	0.94545	0.92727
<b>2</b>	0.98182	0.98182	0.98182	0.98182
<b>3</b>	0.89091	0.89091	0.94545	0.94545
<b>4</b>	1	1	1	1
<b>5</b>	0.94545	0.92727	0.94545	0.94545
<b>6</b>	0.94545	0.94545	0.96364	0.94545
<b>7</b>	0.94545	0.94545	0.96364	0.96364
<b>8</b>	0.96364	0.96364	0.98182	0.98182
<b>9</b>	0.98182	1	1	1
<b>10</b>	1	1	1	0.98182

# *Βιβλιογραφία*

**Ελληνική:**

- Δούμπος, Μ (2003): “Πολυκριτήρια Συστήματα Αποφάσεων”, Σημειώσεις μεταπτυχιακού μαθήματος, Πολυτεχνείο Κρήτης, Χανιά
- Δούμπος, Μ (2003α): “Τεχνικές Ταξινόμησης, Θεωρία και Εφαρμογές στη λήψη χρηματοοικονομικών αποφάσεων”, Σημειώσεις Εργαστηρίου Συστημάτων Χρηματοοικονομικής Διοίκησης, Πολυτεχνείο Κρήτης, Χανιά
- Δούμπος, Μ και Ζοπουνίδης, Κ. (2001): “Πολυκριτήριες τεχνικές ταξινόμησης”, Εκδόσεις Κλειδάριθμος, Αθήνα
- Ζοπουνίδης, Κ. (1998): “Ανάλυση και Διαχείριση Χρηματοοικονομικών Κινδύνων”, Εκδόσεις Κλειδάριθμος, Αθήνα.
- Χατζή, Ε., (2003): “Μεθοδολογικό πλαίσιο εκτίμησης μεθόδων ταξινόμησης”, μεταπτυχιακή διατριβή, Πολυτεχνείο Κρήτης, Χανιά

**Ξένη:**

- Aizerman, E.M., Braverman, and L.I. Rozoner (1964): “Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control” 25, p 821-837.
- Bauer, E. και Kohavi R. (1999): “An empirical comparison of voting classification algorithms: Bagging, boosting and variants”, Machine Learning 36, p 105-139
- Belacel, N. (2000): “Multicriteria assignment method PROAFTN: Methodology and medical applications”, European Journal of Operational Research 125, p 175-183

- Blake, C.L. και Merz, C.J. (1998): UCI Repository of machine learning, [<http://www.ics.uci.edu/~mllearn/MLRepository.html>], Irvine, CA: University of California, Department of Information and Computer Science.
- Blanz, V., Scholkopf, B., Bulthoff, H., Burges, C., Vapnik, V. και Vetter T. (1996): “Comparison of view-based object recognition algorithms using realistic 3d models”. In C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, and B. Sendhoff, editors, Artificial Neural Networks-ICANN'96, p 251-256
- Boser, B. E., Guyon I. M., και Vapnik V. (1992): “A training algorithm for optimal margin classifiers”. In Fifth Annual Workshop on Computational Learning Theory, Pittsburgh. ACM.
- Burges, C. J. C. και Scholkopf, B. (1997): “Improving the accuracy and speed of support vector learning machines”. In M. Mozer, M. Jordan, and T. Petsche, editors, Advances in Neural Information Processing Systems 9, p 375-381
- Burges, C.J.C (1998): “A Tutorial on Support Vector Machines for Pattern Recognition”. Data Mining and Knowledge Discovery 2. Kluwer Academic Publishers, Boston.
- Catelani, M. και Fort, A. (2000): “Fault diagnosis of electronic analog circuits using a radial basis function network classifier”, Measurement, 28/3, p 147-158
- Cortes, C. και Vapnik, V. (1995): “Support vector networks”. Machine Learning, 20, p 273-297.
- Diakoulaki, D., Zopounidis, C., Mavrotas, G. και Doumpos, M. (1999): “The use of a preference disaggregation method in energy analysis and policy making”, Energy-The International Journal, 24/2, p 157-166
- Edgar Osuna, Robert Freund, και Federico Girosi (1997): “An improved training algorithm for support vector machines”. In Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing, p 276-285
- Efron, B. και Tibshirani, R. J. (1993): “An introduction to the bootstrab”, Chapman and Hall, New York

- Eisenbeis, R. (1977): “The pitfalls in the application of discriminant analysis in business, finance and economics”, The Journal of Finance 32, p 723-739
- Freed, N., και Glover, F. (1981): “A linear programming approach to the discriminant problem”, Decision Sciences 12, p 68-74
- Freed, N., και Glover, F. (1986): “Evaluating alternative linear programming models to solve the two-group discriminant problem”, Decision Sciences 17, p 151-162
- Glover, F., Keene, S., και Due, B. (1988): “A new class of models for the discriminant problem”, Decision Sciences 19, p 269-280
- Joachims, T. (1997): “Text categorization with support vector machines”. Technical report, LS VIII Number 23, University of Dortmund
- Kohavi, R. και Frasca, B. (1995): “Useful Feature Subsets and Rough Set Reducts”, in: T.Y. Lin και A. Wildberger (eds.), Soft Computing: Rough Set, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery, Simylation Councils, San Diego.
- Mangasarian, O. L. (2003): “Support Vector Machine Classification via Parameterless Robust Linear Programming”, Data Mining Institute Technical Report
- Mirkin, B., (1996): “Mathematical Classification and Clustering”, Kluwer Academic Publishers, Dordrecht
- Pedroso, J. P. και Murata, N. (2001): “Support Vector Machines with different norms: motivation, formulations and results”, Pattern Recognition Letters 22, p 1263-1272
- Resnick, P. και Varian, H. (1997): “Recommender Systems”, Communication of the ACM 40 (3), p 56-58
- Ripley, B.D. (1996): “Pattern Recognition and Neural Networks”, Cambridge University Press, Cambridge
- Schmidt, M. (1996): “Identifying speaker with support vector networks”. In Interface '96 Proceedings

- Scholkopf, B., Burges C. και Vapnik V. (1995): “Extracting support data for a given task”. In U. M. Fayyad and R. Uthurusamy, editors, Proceedings, First International Conference on Knowledge Discovery & Data Mining. AAAI Press
- Scholkopf, B., Burges, C. και Vapnik V. (1996): “Incorporating invariances in support vector learning machines”. In C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, and B. Sendhoff, editors, Artificial Neural Networks - ICANN'96.
- Siskos, Y., Grigoroudis, E., Zopounidis, C. και Saurais, O. (1998): “Measuring customer satisfaction using a survey based preference disaggregation model”, Journal of Global Optimization, 12/2, p 175-195
- Tsumoto, S. (1998): “Automated extraction of medical expert system rules from clinical databases based on rough set theory”, Information Sciences 112, p 67-84
- W. Zhou, Li Zhang, L Jiao (2002): “Linear programming support vector machines”, Pattern Recognition 3, p 2927-2936
- Young , T.Y και Fu, K.-S. (1997): “Handbook of Pattern Recognition and Image Processing”, Handbooks in Science and Technology, Academic Press