

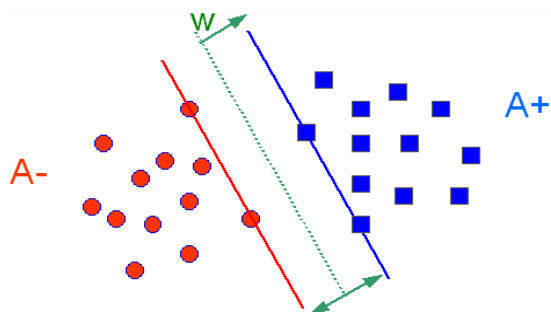


ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

Τμήμα Μηχανικών Παραγωγής & Διοίκησης .

Διπλωματική Εργασία

*«Μηχανές διανύσματος υποστήριξης για την ανάπτυξη υποδειγμάτων
ταξινόμησης: Μια συγκριτική ανάλυση»*



ΔΙΑΜΑΝΤΑΚΗΣ ΕΜΜΑΝΟΥΗΛ

Επιβλέπων : Δούμπος Μιχαήλ

ΧΑΝΙΑ 2003

Ευχαριστίες:

Στο σημείο αυτό θα ήθελα να ευχαριστήσω τους γονείς μου για την υλική και ηθική υποστήριξη τους τόσο κατά τη διάρκεια των σπουδών μου όσο και κατά την παρούσα διπλωματική εργασία. Επίσης θα ήθελα να ευχαριστήσω τον καθηγητή μου Μιχάλη Δούμπο για την πολύτιμη βοήθεια του στην ολοκλήρωση της διπλωματικής εργασίας. Τέλος θα ήθελα να ευχαριστήσω όλους τους κοντινούς μου ανθρώπους για την ηθική υποστήριξη τους.

Περιεχόμενα

Κεφάλαιο 1: Εισαγωγή.....	5
Κεφάλαιο 2: Μηχανές διανύσματος υποστήριξης.....	9
2.1 Εισαγωγή.....	9
2.2.1 Η γραμμική περίπτωση.....	9
2.2.2 Γενίκευση στη μη γραμμική περίπτωση.....	15
2.2 Υπολογιστικές τεχνικές.....	18
2.2.1 Η προσέγγιση Lagrange.....	18
2.2.2 Προσεγγιστικές μηχανές διανύσματος υποστήριξης.....	22
Κεφάλαιο 3: Συγκριτική ανάλυση.....	27
3.1 Σκοπός της συγκριτικής ανάλυσης.....	27
3.2 Εξεταζόμενες μέθοδοι.....	28
3.2.1 Νευρωνικά δίκτυα	28
3.2.2 Αλγόριθμος πλησιέστερου γείτονα.....	32
3.2.3 Διακριτική ανάλυση.....	34
3.2.4 Δέντρα ταξινόμησης (CART).....	35
3.3: Πείραμα και Αποτελέσματα.....	37
3.3.1 Αποτελέσματα μεθόδου L-LSVM.....	39
3.3.2 Αποτελέσματα μεθόδου L-PSVM.....	44
3.3.3 Αποτελέσματα μεθόδου NL-LSVM.....	48
3.3.4 Αποτελέσματα μεθόδου NL-PSVM.....	52
3.3.5 Σύγκριση μεθόδων SVM.....	57
3.3.5.1 Σύγκριση γραμμικών μεθόδων SVM.....	57
3.3.5.2 Σύγκριση μη γραμμικών μεθόδων SVM.....	59
3.3.6 Σύγκριση SVM με εναλλακτικές μεθόδους.....	62

Κεφάλαιο 4: Συμπεράσματα και προοπτικές.....	72
Βιβλιογραφία.....	75

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Η σημερινή κατάσταση της παγκόσμιας αγοράς, υποχρεώνει τους αποφασίζοντες να γίνουν περισσότερο ανταγωνιστικοί μέσω της βελτίωσης των λαμβανομένων αποφάσεων [1]. Η μέθοδος της ταξινόμησης συμβάλλει στην κατεύθυνση αυτή. Κατα τον Mirkin [2], η ταξινόμηση ορίζεται ως εξής:

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

- Τη διαμόρφωση, οργάνωση και διατήρηση της γνώσης.
- Την ανάλυση της δομής του φαινομένου που εξετάζεται .
- Τη συσχέτιση των διαφόρων πλευρών του υπό εξέταση φαινομένου.»

Η ταξινόμηση εκτός από την ερευνητική σημασία που έχει για τον επιστημονικό κόσμο, εφαρμόζεται σε πολλούς τομείς στην καθημερινή ζωή, όπως το περιβάλλον, την ιατρική και στις επιχειρήσεις που πρέπει να λάβουν σημαντικές αποφάσεις. Μερικά χαρακτηριστικά παραδείγματα αναφέρονται παρακάτω.

- Ιατρική: Πραγματοποίηση ιατρικών διαγνώσεων ταξινομώντας τους ασθενείς σε κατηγορίες (παθήσεις) ανάλογα με τα συμπτώματα που παρουσιάζουν [3],[4] .

- Βιολογία: Ταξινόμηση των στοιχείων παρουσίασης των γονιδίων από τις μικρογραμμές του DNA [5].
- Αναγνώριση γραπτών κειμένων. [6]
- Αναγνώριση ανθρώπινων χαρακτηριστικών. [7]
- Πρόβλεψη της πτώχευσης επιχειρήσεων, εκτίμηση του πιστωτικού κινδύνου επιχειρήσεων, οργανισμών και φυσικών προσώπων, επιλογή και διαχείριση χαρτοφυλακίων επενδύσεων, αξιολόγηση των οικονομικών επιδόσεων και της δανειοληπτικής ικανότητας κρατών. [8],[9].

Λόγω της μεγάλης ερευνητικής σημασίας της ταξινόμησης, υλοποιήθηκαν πολλές μέθοδοι προς την κατεύθυνση αυτή. Από τις μεθόδους αυτές, οι μηχανές διανύσματος υποστήριξης (support vector machines, [16]) είναι 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 στο πλαίσιο αυτό, 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 πρόσφατα έχουν προταθεί ειδικές διαδικασίες και αλγόριθμοι για την αποτελεσματική ανάπτυξη μοντέλων ταξινόμησης μέσω SVM.

Σκοπός της παρούσας εργασίας είναι η αξιολόγηση της αποτελεσματικότητας αυτών των νέων τεχνικών, όσο αφορά την ακρίβεια των μοντέλων που

αναπτύσσονται. Ειδικότερα εξετάζονται δύο τεχνικές, τα Lagrangian SVM [13] και τα; Proximal SVM [14]. Η διερεύνηση της αποτελεσματικότητας των τεχνικών αυτών πραγματοποιείται μέσα από μια εκτενή σύγκριση με άλλες διαδεδομένες μεθοδολογίες για την ανάπτυξη μοντέλων ταξινόμησης, όπως τα νευρωνικά δίκτυα (neural network), αλγόριθμοι πλησιέστερου γείτονα (nearest neighbors), γραμμική διακριτική ανάλυση (linear discriminant analysis) και δέντρα ταξινόμησης και παλινδρόμησης (CART). Η σύγκριση των μεθόδων, έγινε σε δεδομένα διαφόρων διαστάσεων τα οποία ακολουθούν συγκεκριμένες στατιστικές ιδιότητες. Για τον έλεγχο της αποτελεσματικότητας των μεθόδων χρησιμοποιήθηκαν δεδομένα με ιδιότητες ίδιες με αυτών της εκπαίδευσης. Θα γίνει λεπτομερής αναφορά των ιδιοτήτων αυτών στα παρακάτω κεφάλαια.

Το υπόλοιπο της διπλωματικής εργασίας οργανώνεται σε 4 κεφάλαια και γίνεται αναφορά στις μεθόδους, και στην παρουσίαση των αποτελεσμάτων. Πιο συγκεκριμένα:

Στο δεύτερο κεφάλαιο γίνεται αναφορά στις μηχανές διανύσματος υποστήριξης. Αρχικά γίνεται μια εισαγωγή στη νέα αυτή μεθοδολογία για την ανάπτυξη μοντέλων ταξινόμησης και στη συνέχεια παρουσιάζονται κάποιες υπολογιστικές διαδικασίες εκπαίδευσης των μηχανών διανύσματος υποστήριξης.

Στο τρίτο κεφάλαιο γίνεται αναφορά στη συγκριτική ανάλυση. Συγκεκριμένα γίνεται αναφορά στο σκοπό και στις μεθόδους που χρησιμοποιήθηκαν για τη σύγκριση. Επίσης δίνεται ο σχεδιασμός του πειράματος, η παραγωγή των

δεδομένων, η διαδικασία υλοποίησης και τέλος η παρουσίαση των αποτελεσμάτων.

Στο τέταρτο κεφάλαιο παρουσιάζονται τα συμπεράσματα και οι προοπτικές της θεωρίας ταξινόμησης.

ΚΕΦΑΛΑΙΟ 2

Μηχανές διανύσματος υποστήριξης

2.1 Εισαγωγή

2.1.1. Η γραμμική περίπτωση

Οι μηχανές διανύσματος υποστήριξης είναι συστήματα μάθησης τα οποία διαχωρίζουν τα δεδομένα εκπαίδευσης μέσω γραμμικών συναρτήσεων. Στις περισσότερες των περιπτώσεων το σύνολο των δεδομένων δεν μπορούν να διαχωριστούν, και απεικονίζονται σε ένα πολυδιαστατο χώρο και έπειτα υπολογίζεται ένα υπερεπίπεδο το οποίο διαχωρίζει τα δεδομένα. Οι μηχανές διανύσματος υποστήριξης εκπαιδεύονται σύμφωνα με αλγορίθμους εκπαίδευσης από την στατιστική θεωρία μάθησης (statistical learning theory).

Για ένα σύνολο σημείων (x_1, x_2, \dots, x_n) τα οποία πρέπει να διαχωριστούν σε δυο κατηγορίες $+1$ και -1 , η ταξινόμηση πραγματοποιείται από μία γραμμική συνάρτηση της μορφής:

$$f(x) = (w \cdot x) + \gamma = \sum_{i=1}^n w_i x_i + \gamma \quad (1)$$

Όπου τα w και γ , είναι παράμετροι του υπερεπιπέδου $w \cdot x = \gamma$. Ο κανόνας απόφασης για κάποιο σημείο είναι: το διάνυσμα $x = (x_1, x_2, \dots, x_n)$ ταξινομείται στη κατηγορία $+1$ εάν $f(x) \geq 0$, ενώ στην αντίθετη περίπτωση ταξινομείται στην κατηγορία -1 . Αν υπάρχει υπερεπίπεδο f το οποίο ταξινομεί με απόλυτη ακρίβεια τα δεδομένα τότε θεωρείται ότι οι κατηγορίες είναι γραμμικά

διαχωρίσιμες. Στην αντίθετη περίπτωση τα σημεία δεν είναι γραμμικά διαχωρίσιμα. Στο σχήμα 2.1 παρουσιάζεται ο παραπάνω διαχωρισμός σε σημεία τα οποία ανήκουν σε επίπεδο δυο διαστάσεων.

Σχήμα 2.1: Γραμμικός διαχωρισμός σημείων σε χώρο δυο διαστάσεων

Ο παραπάνω διαχωρισμός γίνεται έτσι ώστε η απόσταση των κοντινότερων σημείων από την ευθεία να είναι μέγιστη. Όταν υπάρχει υπερεπίπεδο το οποίο διαχωρίζει όλα τα σημεία χωρίς σφάλματα και τα κοντινότερα σημεία ως προς το υπερεπίπεδο, με τη μέγιστη απόσταση (περιθώριο) d τότε το υπερεπίπεδο ονομάζεται βέλτιστο υπερεπίπεδο (ή υπερεπίπεδο μέγιστου περιθωρίου). Στο σχήμα 2.2 απεικονίζεται ένα τέτοιο υπερεπίπεδο. Το περιθώριο d αποδεικνύεται ότι υπολογίζεται από την σχέση $d = \frac{2}{\|w\|_2}$. Η μεγιστοποίηση του περιθωρίου d επιτυγχάνεται με την ελαχιστοποίηση της νόρμας του w .

Σχήμα 2.2: Διαχωρισμός σημείων με υπερεπίπεδο μέγιστου περιθωρίου.

Στα πλαίσια αυτής της θεώρησης, η ανάπτυξη του βέλτιστου υπερεπιπέδου επιτυγχάνεται με την επίλυση του ακόλουθου προβλήματος τετραγωνικού προγραμματισμού

$$\min_{w,b} \frac{1}{2} (w \cdot w)$$

υ.π.

$$y_i((w * x) + \gamma) \geq 1, \quad i = 1, \dots, l$$

όπου x_1, x_2, \dots, x_l είναι τα σημεία ενός δείγματος εκπαίδευσης και y_1, y_2, \dots, y_l είναι η ταξινόμηση των σημείων αυτών. Από υπολογιστικής πλευράς η

επίλυση του παραπάνω προβλήματος γίνεται ευκολότερη μέσω του δυϊκού προγράμματος, από την επίλυση του οποίου προκύπτει ότι (ως u συμβολίζονται οι πολλαπλασιαστές Lagrange του δυϊκού, βλ Vapnik [10]):

$$w = \sum_{i=1}^l y_i u_i x_i \quad (2)$$

$$0 = \sum_{i=1}^l y_i u_i \quad (3)$$

Τα όρια της κάθε κατηγορίας δίνονται από τις σχέσεις

$$x'w = \gamma + 1$$

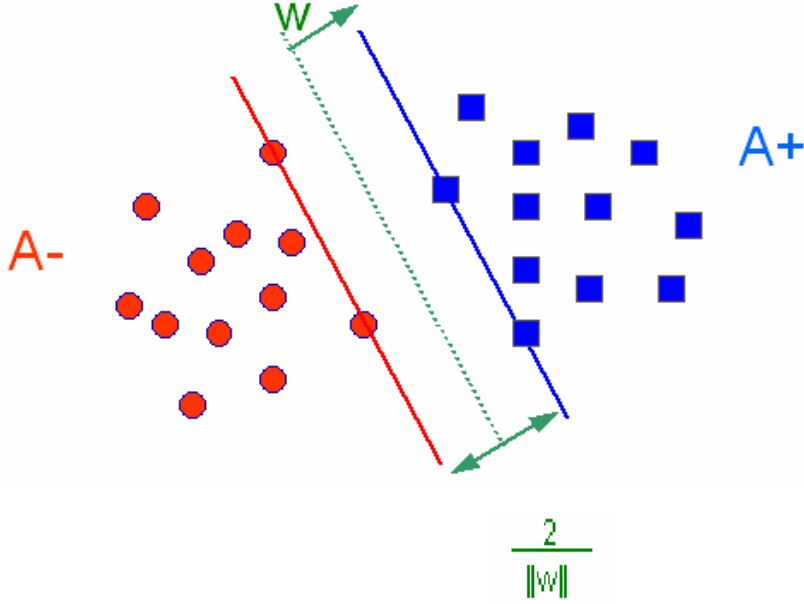
$$x'w = \gamma - 1$$

Η σχέση που χαρακτηρίζει αν κάποιο σημείο ανήκει στην κατηγορία +1 ή στην κατηγορία -1 παρουσιάζεται παρακάτω

$$x'w \geq \gamma + 1 \text{ για } y_i = 1 \quad (4)$$

$$x'w \leq \gamma - 1 \text{ για } y_i = -1 \quad (5)$$

Στο σχήμα 2.3 παρουσιάζονται τα όρια των κατηγοριών και το περιθώριο d .



Σχήμα 2.3: Όρια κατηγοριών και περιθώριο.

Αν πρέπει να ταξινομηθούν l σημεία σε ένα n διαστάσεων πραγματικό χώρο. Τότε απεικονίζονται τα σημεία, σε έναν $l \times n$ πίνακα A έτσι ώστε κάθε στοιχείο του πίνακα A να ανήκει στο χώρο A^+ ή A^- όπως καθορίζεται από ένα διαγώνιο πίνακα D του οποίου τα στοιχεία είναι $+1$ και -1 , τα οποία αντιστοιχούν στο χώρο A^+ και A^- αντίστοιχα. Άρα οι σχέσεις (4) και (5) γίνονται

$$A_i w \geq \gamma + 1 \text{ για } D_{ii} = 1 \quad (6)$$

$$A_i w \leq \gamma - 1 \text{ για } D_{ii} = -1 \quad (7)$$

Αν τα σημεία που πρόκειται να διαχωριστούν δε διαχωρίζονται γραμμικά, τότε στις σχέσεις (6) και (7) προστίθεται ή αφαιρείται αντίστοιχα μια μη αρνητική ποσότητα γ , η οποία προσδιορίζει το σφάλμα της ταξινόμησης

$$A_i w + y_i \geq \gamma + 1 \text{ για } D_{ii} = 1$$

$$A_i w - y_i \leq \gamma - 1 \text{ για } D_{ii} = -1$$

Έτσι το προηγούμενο τετραγωνικό πρόγραμμα με παράμετρο $\nu > 0$ μετατρέπεται όπως παρακάτω

$$\min_{w, \gamma, y \in R^{n+m+1}} \nu e' y + \frac{1}{2} w' w$$

$$\text{υ.π.} \tag{\pi 1}$$

$$D(Aw - e\gamma) + y \geq e, \quad y \geq 0$$

Από υπολογιστικής πλευράς η επίλυση του προγράμματος (π1) μπορεί να πραγματοποιηθεί πιο απλά μέσω του δυϊκού του το οποίο έχει την εξής μορφή:

$$\min_{u \in R^m} \frac{1}{2} u' D D A A' D u - e' u$$

$$\text{υ.π.} \tag{\pi 2}$$

$$e' D u = 0, \quad 0 \leq u \leq \nu e$$

Σύμφωνα με τους Vapnik και Chervonenkis [10] αν $u_0 = (u_1^0, \dots, u_l^0)$ είναι η λύση του προγράμματος (π2) τότε η νόρμα του w είναι

$$|w|^2 = \sum_{\substack{\text{sup port} \\ \text{vectors}}} u_i^0 u_j^0 (x_i x_j) y_i y_j$$

Ο κανόνας διαχωρισμού, βασισμένος στο βέλτιστο υπερεπίπεδο είναι:

$$f(x) = \text{sign}(\sum_{\substack{\text{sup port} \\ \text{vectors}}} y_i u_i^0(x_i, x) - \gamma),$$

όπου x_i είναι τα διανύσματα υποστήριξης, u_i^0 είναι οι πολλαπλασιαστές Lagrange και η ποσότητα γ δίνεται από την παρακάτω σχέση

$$\gamma = \frac{1}{2} [(w_0 x^*(1)) + (w_0 x^*(-1))]$$

όπου $x^*(1)$ και $x^*(-1)$ είναι διανύσματα τα οποία ανήκουν στην κατηγορία 1 και -1 αντίστοιχα.

2.1.2. Γενίκευση στη μη γραμμική περίπτωση

Στην πραγματικότητα τα προβλήματα είναι πολύ πιο πολύπλοκα με αποτέλεσμα η ανάπτυξη απλών μοντέλων να μην επαρκεί. Από την σχέση (2) αποδεικνύεται ότι η γραμμική συνάρτηση (1) μπορεί να γραφτεί σαν γραμμικός συνδιασμός όπως παρακάτω:

$$f(x) = \sum_{i=1}^l u_i y_i (x_i \cdot x) + \gamma \quad (8)$$

Εισάγοντας τη μη γραμμική αναπαράσταση των δεδομένων σε έναν χώρο F μπορεί να αναπτυχθεί πλέον μια γραμμική συνάρτηση f στο μη γραμμικό

χώρο ως εξής: $f(x) = \sum_{i=1}^l u_i y_i (\Phi(x_i) \Phi(x)) + \gamma$. Όπου u είναι οι πολλαπλασιαστές

Lagrange που υπολογίζονται από την επίλυση του προβλήματος (π2). Στο σχήμα 2.4 παρουσιάζεται ένα παράδειγμα απεικόνισης ενός συνόλου σημείων που δεν μπορούν να διαχωριστούν από μία γραμμική συνάρτηση, σε ένα δυο διαστάσεων επίπεδο F τέτοιο ώστε $F = \{\Phi(x) : x \in X\}$.



Σχήμα 2.4: Αναπαράσταση των δεδομένων X στο μη γραμμικό χώρο F .

Για την πραγματοποίηση της μη γραμμικής απεικόνισης των δεδομένων σε ένα χώρο υψηλότερων διαστάσεων χρησιμοποιείται μια συνάρτηση K , τέτοια ώστε $K(x, z) = \Phi(x) \cdot \Phi(z)$ η οποία ονομάζεται συνάρτηση πυρήνα (Kernel function). Ένα χαρακτηριστικό παράδειγμα πυρήνα είναι η χρήση μιας πολυωνυμικής συνάρτησης της ακόλουθης μορφής [7]:

$$K(x, z) = (x \cdot z)^d, \text{ όπου } d \text{ είναι η διάσταση του επίπεδου } F.$$

Η συνάρτηση αυτή αντικαθιστά τα γινόμενα της μορφής $\Phi(x) \cdot \Phi(z)$ ως εξής:

$$(x \cdot z)^2 = ([x_1 \ x_2] \cdot [z_1 \ z_2])^2 = [x_1^2 \ \sqrt{2x_1x_2} \ x_2^2] \cdot [z_1^2 \ \sqrt{2z_1z_2} \ z_2^2] = \Phi(x)\Phi(z).$$

Στην περίπτωση των μηχανών διανύσματος ανάπτυξης, χρησιμοποιούνται εκτός από την πολυωνυμική συνάρτηση και άλλοι πυρήνες όπως ο πυρήνας RBF (radial basis function) που εκφράζεται ως εξής:

$$K(x, y) = \exp\left(\frac{-|x-y|^2}{2\sigma^2}\right)$$

Επίσης μεγάλη χρήση παρουσιάζει και η σιγμοειδής συνάρτηση πυρήνα, η οποία εκφράζεται ως εξής:

$$K(x, y) = \tanh(K(xy) + \Theta)$$

Το πρόβλημα μεγιστοποίησης του περιθωρίου όπως και στη γραμμική του μορφή λύνεται με το παρακάτω πρόγραμμα.

$$\min_{w, \gamma, y \in R^{n+m+1}} v e' y + \frac{1}{2} w' w v e$$

υ.π. (π3)

$$D(K(A, A')Du - e\gamma) + y \geq e, \quad y \geq 0$$

Το πρόγραμμα (π3) προκύπτει από το πρωτεύων πρόγραμμα (π1) θέτοντας $w = A'Du$ σύμφωνα με τη σχέση (2) και $AA' = K(A, A')$.

Το δυϊκό πρόγραμμα προκύπτει από το πρόγραμμα (π2) θέτοντας $AA' = K(A, A')$. Έτσι το νέο δυϊκό πρόγραμμα για τις μη γραμμικές μηχανές διανύσματος υποστήριξης εκφράζεται ως εξής:

$$\min_{u \in R^m} \frac{1}{2} u' DK(A, A') Du - e'u$$

υ.π.

(π4)

$$e'Du = 0, \quad 0 \leq u \leq ve$$

2.2 Υπολογιστικές τεχνικές

Το κύριο πρόβλημα που ανακύπτει κατά τη χρήση των SVM είναι ο αυξημένος υπολογιστικός φόρτος που απαιτεί η εκπαίδευση τους, κυρίως στη περίπτωση που το δείγμα εκμάθησης είναι μεγάλων διαστάσεων. Για την αντιμετώπιση του προβλήματος αυτού έχουν προταθεί διάφορες υπολογιστικές τεχνικές. Στην παρούσα εργασία εξετάζονται δυο τέτοιες τεχνικές, τα Lagrangian SVM (LSVM, [13]) και τα προσεγγιστικά SVM (proximal SVM, [14]). Στις ενότητες που ακολουθούν αναλύονται τα βασικά χαρακτηριστικά αυτών των υπολογιστικών τεχνικών καθώς και η λειτουργία τους.

2.2.1 Η προσέγγιση Lagrange

Στην περίπτωση των SVM έχουν γίνει οι παρακάτω μεταβολές σε σχέση με τις κανονικές γραμμικές μηχανές διανύσματος υποστήριξης (π1).

- Έγινε αλλαγή της πρώτης νόρμας σε δεύτερη τετραγωνική νόρμα του y , με αποτέλεσμα να μην απαιτείται πλέον η εξέταση του περιορισμού $y \geq 0$.

- Προστέθηκε η ποσότητα γ^2 στο $w'w$ έτσι ώστε να μεγαλώνει το περιθώριο μεταξύ των παράλληλων υπερεπιπέδων.

Έτσι το πρωτεύων πρόγραμμα μετατρέπεται όπως παρακάτω [11].

$$\min_{w, \gamma, y \in R^{n+m+1}} v \frac{y'y}{2} + \frac{1}{2} (w'w + \gamma^2)$$

(π5)

$$D(Aw - e\gamma) + y \geq e.$$

Το δυϊκό αυτού του προγράμματος είναι [12]:

$$\min_{0 \leq u \in R^m} \frac{1}{2} u' \left(\frac{I}{v} + D(AA' + ee')D \right) u - e'u$$

(π6)

Η λύση του παραπάνω δυϊκού προβλήματος (π6) δίνει τις τιμές για τα w, γ, y

$$w = A'Du, y = \frac{u}{v}, \gamma = -e'Du.$$

(9)

Το παραπάνω δυϊκό πρόγραμμα (π6) για $H = D[A \ -e]$ και $Q = \frac{I}{v} + HH'$ γίνεται

$$\min_{0 \leq u \in R^m} \frac{1}{2} u' Qu - e'u.$$

(π7)

Η ποσότητα $[A - e]$ προκύπτει αν προστεθεί μία επιπλέον μοναδιαία στήλη με όλα τα στοιχεία της -1 στον πίνακα A . Για το παραπάνω δυϊκό πρόγραμμα (π7) ικανή και αναγκαία συνθήκη ώστε μια λύση u να είναι βέλτιστη είναι [12]:

$$0 \leq u \perp Qu - e \geq 0 \quad (10)$$

Δεδομένου ότι για κάθε δύο πραγματικούς αριθμούς a και b ισχύει ότι :

$$0 \leq a \perp b \geq 0 \Leftrightarrow a = (a - ab)_+, a > 0 \quad (11)$$

Από τη σχέση (11), η σχέση (10) μπορεί να γραφτεί ως εξής:

$$Qu - e = ((Qu - e) - au)_+. \quad (12)$$

Σύμφωνα με το παραπάνω προκύπτει η παρακάτω επαναληπτική σχέση [13] :

$$u^{i+1} = Q^{-1}(e + ((Qu^i - e) - au^i)_+), \quad i = 0, 1, \dots$$

$$\text{με } a \in \left(0, \frac{2}{v}\right)$$

Η παραπάνω ανάλυση η οποία εφαρμόζεται στην περίπτωση των γραμμικών SVM μπορεί εύκολα να επεκταθεί στην μη γραμμική περίπτωση. Στην περίπτωση αυτή το υπερεπίπεδο που διαχωρίζει τις κατηγορίες διαμορφώνεται ως εξής:

$$K([x' \quad -1], \begin{bmatrix} A' \\ -e' \end{bmatrix})Du = 0$$

Όπου u είναι η λύση του δυϊκού προγράμματος (π7) όπου το Q προσδιορίζονται ως εξής:

$$Q = \frac{I}{\nu} + DK(G, G')D, \text{ όπου } G = [A \quad -e] \quad (\pi 3)$$

Το δυϊκό πρόβλημα (π7) για οποιαδήποτε συνάρτηση πυρήνα, διαμορφώνεται ως εξής:

$$\min_{0 \leq u \in R^m} \frac{1}{2} u' \left(\frac{I}{\nu} + DK(G, G')D \right) u - e'u \quad (\pi 8)$$

Η ικανή και αναγκαία συνθήκη για τον προσδιορισμό της βέλτιστης λύσης u στο παραπάνω πρόβλημα είναι [13]:

$$0 \leq u \perp \left(\frac{I}{n} + DK([A \quad -e], \begin{bmatrix} A' \\ -e' \end{bmatrix})D \right) u - e \geq 0 \quad (14)$$

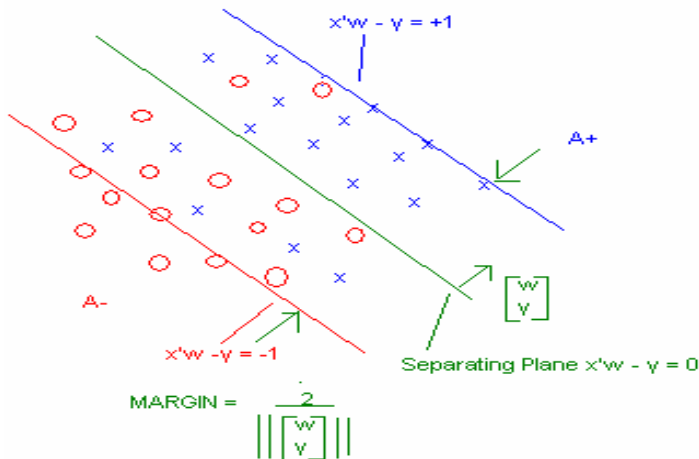
Απο τη σχέση (14) και με χρήση της σχέσης (13) προκύπτει ο επαναληπτικός τύπος του u :

$$u^{i+1} = Q^{-1}(e + ((Qu^i - e) - au^i)_+), \quad i = 0, 1, \dots$$

$$\text{με } Q = \frac{I}{\nu} + DK(G, G')D, \text{ και } G = [A - e]$$

2.2.2 Προσεγγιστικές μηχανές διανύσματος υποστήριξης

Η προσεγγιστική (proximal) μέθοδος βασίζεται στο ίδιο πρόβλημα βελτιστοποίησης που εξετάζεται και στην περίπτωση των LSVM (πρόβλημα (π5)). Στην περίπτωση όμως αυτή ο περιορισμός $D(Aw - e\gamma) + y \geq e$ μετατρέπεται σε ισότητα $D(Aw - e\gamma) + y = e$. Γεωμετρικά η προσεγγιστική μέθοδος παρουσιάζεται στο σχήμα 2.5. Αναλυτικά, τα επίπεδα $x'w = \gamma \pm 1$ δεν διαχωρίζουν πλέον τις δυο κατηγορίες, αλλά μπορεί να θεωρηθεί ότι βρίσκονται κεντρικά των σημείων σε κάθε επίπεδο A^+ και A^- . Τα σημεία χωρίζονται όσο είναι δυνατόν σύμφωνα με την σχέση $(w'w + \gamma^2)$ που βρίσκεται στην αντικειμενική συνάρτηση που δεν είναι τίποτε άλλο από το αντίστροφο του τετραγώνου της απόστασης μεταξύ των δύο επιπέδων στο χώρο (w, γ) του R^{n+1} της δεύτερης νόρμας [14].



Σχήμα 2.5: Γεωμετρική απεικόνιση της προσεγγιστικής μέθοδου

Ικανή και αναγκαία συνθήκη για τον προσδιορισμό της βέλτιστης λύσης στο πρόβλημα είναι να τεθεί ίση με μηδέν, η παράγωγος της παρακάτω συνάρτησης Lagrange ως προς (w, γ, y, u) .

$$L(w, \gamma, y, u) = \frac{v}{2} \|y\|^2 + \frac{1}{2} \left\| \begin{bmatrix} w \\ \gamma \end{bmatrix} \right\|^2 - u'(D(Aw - e\gamma) + y - e).$$

Αν τεθούν ίσες με μηδέν οι παράγωγοι ως προς (w, γ, y, u) της παραπάνω συνάρτησης Lagrange, τότε έχουμε:

$$w - A'Du = 0 \tag{15}$$

$$\gamma + e'Du = 0 \tag{16}$$

$$vy - u = 0 \tag{17}$$

$$D(Aw - e\gamma) + y - u = 0 \tag{18}$$

Από τις εξισώσεις (15), (16), (17) προκύπτουν οι παρακάτω σχέσεις για τα w , y , γ , ενώ από την εξίσωση (18) προκύπτει η σχέση βάση της οποίας προσδιορίζεται το u . [14]:

$$w = A'Du \tag{19}$$

$$\gamma = -e'Du \tag{20}$$

$$y = \frac{u}{v} \tag{21}$$

$$u = v(I - H(\frac{I}{v} + HH')^{-1}H')e, \text{ όπου } H = D[A - e] \tag{22}$$

Για να γίνει ταξινόμηση των σημείων με τη μη γραμμική προσεγγιστική μέθοδο μετατρέπεται το πρωτεύων πρόγραμμα (π5) της μεθόδου Lagrange

κατάλληλα, αντικαθιστώντας στον περιορισμό, όπου $w = A'Du$. Το νέο πρόβλημα διαμορφώνεται ως εξής:

$$\min_{(u, \gamma, y) \in R^{m+1+m}} v \frac{1}{2} \|y\|^2 + \frac{1}{2} (u'u + \gamma^2)$$

υ.π.

$$D(AA'Du - e\gamma) + y = e$$

Ο πίνακας AA' αποτελεί την γραμμική συνάρτηση πυρήνα. Έτσι μπορεί να αντικατασταθεί ο πίνακας AA' με οποιανδήποτε μη γραμμική συνάρτηση πυρήνα $K(A, A')$. Έτσι το πρωτεύων πρόγραμμα γίνεται

$$\min_{(u, \gamma, y) \in R^{m+1+m}} v \frac{1}{2} \|y\|^2 + \frac{1}{2} (u'u + \gamma^2)$$

υ.π.

$$D(K(A, A')Du - e\gamma) + y = e$$

Η λύση αυτού του προβλήματος γίνεται μέσω των πολλαπλασιαστών Lagrange εισάγοντας την ακόλουθη συνάρτηση

$$L(u, \gamma, v, y) = L(u, \gamma, v, y) = \frac{v}{2} \|y\|^2 + \frac{1}{2} \left\| \begin{bmatrix} u \\ \gamma \end{bmatrix} \right\|^2 - v'(D(KDu - e\gamma) + y - e).$$

Παραγωγίζοντας ως προς (u, γ, v, y) και θέτοντας τις παραγώγους ίσες με το μηδέν προκύπτουν οι παρακάτω σχέσεις.

$$u - DK'Dv = 0 \quad (23)$$

$$\gamma + e'Dv = 0 \quad (24)$$

$$vy - v = 0 \quad (25)$$

$$D(KDu - e\gamma) + y = e \quad (26)$$

Από τις εξισώσεις (23), (24), και (25) προκύπτουν οι τιμές των μεταβλητών u , γ και y , ενώ από την εξίσωση (26) προκύπτει η σχέση που προσδιορίζει το u :

$$u = DK'Dv \quad (27)$$

$$\gamma = -e'Dv \quad (28)$$

$$y = \frac{v}{v} \quad (29)$$

$$u = DK'Dv \quad (27)$$

$$\gamma = -e'Dv \quad (28)$$

$$y = \frac{v}{v} \quad (29)$$

$$v = \left(\frac{I}{v} + D(KK' + ee')D\right)^{-1}e = \left(\frac{I}{v} + GG'\right)^{-1}e \quad (30)$$

$$\text{όπου } G = D[K - e]$$

Με αυτή τη λύση η ταξινόμηση κάθε σημείου x μπορεί να πραγματοποιηθεί απλά με τους ακόλουθους κανόνες [14]:

$$(K(x', A')K(A, A')' + e')Dv \begin{cases} > 0, & \tau_o \tau_e & x \in A+ \\ < 0, & \tau_o \tau_e & x \in A- \\ = 0, & \tau_o \tau_e & x \in A+ \end{cases} \quad \dot{\eta} \quad x \in A-$$

ΚΕΦΑΛΑΙΟ 3

Συγκριτική ανάλυση

3.1 Σκοπός της συγκριτικής ανάλυσης

Στη παρούσα διπλωματική εργασία όπως αναφέρθηκε και στην εισαγωγή γίνεται σύγκριση μεθόδων για μηχανές διανύσματος υποστήριξης με άλλες εναλλακτικές μεθόδους, όπως τα νευρωνικά δίκτυα, ο αλγόριθμος πλησιέστερου γείτονα, η γραμμική διακριτική ανάλυση και τα δέντρα ταξινόμησης. Αυτές οι μέθοδοι όπως παρουσιάστηκε και στην εισαγωγή έχουν μεγάλη πρακτική σημασία, τόσο στις σημαντικές αποφάσεις που καλούνται να πάρουν οι επιχειρήσεις ή στην υποστήριξή τους (όπως η χορήγηση δανείων σε καταναλωτές και ο διαχωρισμός τους ανάλογα με την συνέπεια τους), όσο και σε άλλους επιστημονικούς χώρους όπως στη βιολογία στην ιατρική και σε πληθώρα άλλων πρακτικών εφαρμογών.

Μέσω της σύγκρισης των μεθόδων οι αποφασίζοντες έχουν τη δυνατότητα να επιλέξουν ποια μέθοδος είναι καλύτερη για το αντικείμενο τους, αφού η σύγκριση γίνεται σε δεδομένα με διάφορες στατιστικές ιδιότητες

Σημαντικό στοιχείο στη πρακτική εφαρμογή των μεθόδων είναι η ακρίβεια που παρουσιάζουν. Οι απαιτήσεις σε ακρίβεια σε κάθε είδους εφαρμογής είναι αυξημένες, και έτσι η παρούσα σύγκριση εστιάζεται στην παρουσίαση της ακρίβειας των μεθόδων. Η ακρίβεια των μεθόδων υπολογίζεται με το ποσοστό των εσφαλμένων ταξινομήσεων. Ανάλογα με τις στατιστικές ιδιότητες των δεδομένων που θα εξετάζονται, οι μέθοδοι δίνουν κάθε φορά

διαφορετικά αποτελέσματα, με αποτέλεσμα να υπάρχει διακύμανση στη κατάταξη των μεθόδων. Επίσης γίνεται ανάλυση της συμπεριφοράς των μεθόδων σε σχέση με τις ιδιότητες αυτές.

Για τους παραπάνω λόγους, η ανάλυση της αποτελεσματικότητας των μεθόδων κρίνεται σημαντική τόσο για την πρακτική εφαρμογή των μεθόδων όσο και για τον επιστημονικό κόσμο.

3.2 Εξεταζόμενες μέθοδοι

Σε αυτό το σημείο θα γίνει αναφορά στη θεωρία που διέπει τις μεθόδους που εξετάζονται στη παρούσα σύγκριση, τα νευρωνικά δίκτυα, τον αλγόριθμο πλησιέστερου γείτονα, τη γραμμική διακριτική ανάλυση και τα δέντρα ταξινόμησης.

3.2.1 Νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα έχουν πάρει το όνομα τους από την λειτουργία του ανθρώπινου εγκεφάλου. Ο ανθρώπινος εγκέφαλος αποτελείται από μια πληθώρα νευρώνων. Οι νευρώνες αυτοί παίρνουν τα ερεθίσματα από το περιβάλλον μέσω των αισθητήριων οργάνων και μετά από επεξεργασία παράγουν ένα αποτέλεσμα, το οποίο είτε χρησιμοποιείται σαν είσοδος σε άλλους νευρώνες ή χρησιμοποιείται σαν αποτέλεσμα στο αρχικό ερέθισμα.

[15]

Τα νευρωνικά δίκτυα ή αλλιώς τεχνητά νευρωνικά δίκτυα βασίζονται σε αυτήν την διεργασία του ανθρώπινου οργανισμού. Στο σχήμα 3.1 παρουσιάζεται ένα τυπικό νευρωνικό δίκτυο [16].

Σχήμα 3.1: Αναπαράσταση ενός τυπικού νευρωνικού δικτύου

Το παραπάνω Νευρωνικό δίκτυο αποτελείται από [16]:

(i) $m+1$ επίπεδα: το πρώτο επίπεδο $x(0)$ περιγράφει το διάνυσμα των εισόδων $x = (x^1, \dots, x^n)$, το οποίο δηλώνεται ως $x_i = (x_i^1(0), \dots, x_i^n(0))$ $i = 1, \dots, l$. Για το k -οστό επίπεδο, το διάνυσμα των εισόδων του, δηλώνεται ως $x_i(k) = (x_i^1(k), \dots, x_i^{n_k}(k))$, $i=1, \dots, l$, όπου n_k είναι η διάσταση του διανύσματος $x_i(k)$.

(ii) Το επίπεδο $k-1$ συνδέεται με το επίπεδο k μέσω ενός συνόλου συνδέσεων τα βάρη των οποίων συμβολίζονται ως $w(k)$. Η σχέση που συνδέει τις

εισόδους στους νευρώνες στο επίπεδο k με τις εισόδους στους νευρώνες του επιπέδου $k-1$ έχει την ακόλουθη μορφή:

$$x_i(k) = S\{w(k)x_i(k-1)\}, k=1,2,\dots,m, i=1,\dots,l$$

όπου $S\{w(k)x_i(k-1)\}$ είναι η σιγμοειδής συνάρτηση του διανύσματος $u_i(k) = w(k)x_i(k-1) = (u_i^1(k), \dots, u_i^{n_k}(k))$. Δεν υπάρχει κάποια συγκεκριμένη μεθοδολογία για την εύρεση του αριθμού των επιπέδων ή του αριθμού των κόμβων. Η εύρεση της αρχιτεκτονικής του δικτύου προκύπτει συνήθως μέσω διαδικασιών δοκιμής και λάθους. Σε προβλήματα ταξινόμησης, έρευνες που έχουν πραγματοποιηθεί σχετικές με την εφαρμογή των τεχνητών νευρωνικών δικτύων, έχουν δείξει ότι γενικά ένα ενδιάμεσο επίπεδο αρκεί για την επίτευξη ικανοποιητικών αποτελεσμάτων [17]. Το πλήθος των κόμβων στο ενδιάμεσο επίπεδο μπορεί να κυμαίνεται μεταξύ q και $2n+1$, όπου q είναι το πλήθος των κατηγοριών και n είναι το πλήθος των κριτηρίων αξιολόγησης [15]. Τα βάρη υπολογίζονται κατά τη διάρκεια της εκπαίδευσης του δικτύου και ο υπολογισμός των λαθών υπολογίζεται με το άθροισμα των τετραγώνων των σφαλμάτων.

(iii) Το επίπεδο εξόδου, το οποίο αποτελείται από έναν ή περισσότερους κόμβους. Ανάλογα με το πλήθος των επιπέδων ταξινόμησης ορίζονται και οι κόμβοι της εξόδου. Στην περίπτωση ταξινόμησης σημείων σε δύο κατηγορίες, ο αριθμός των κόμβων μπορεί να είναι ένας και σαν έξοδο να προκύπτουν οι αριθμοί 1 ή 2 ανάλογα με την κατηγορία ταξινόμησης. Ακόμα μπορεί να χρησιμοποιηθούν και 2 κόμβοι από τους οποίους προκύπτουν οι αριθμοί 0 και 1 σε μορφή πίνακα.

Η πλέον διαδεδομένη μέθοδος εκπαίδευσης ενός νευρωνικού δικτύου είναι η back-propagation [18], [19], παραλλαγή της οποίας είναι και η μέθοδος BFGS [20] που χρησιμοποιήθηκε στην παρούσα έρευνα.

Κατά τη διαδικασία του back-propagation στόχος είναι η ελαχιστοποίηση της συνάρτησης

$$I(w(1), \dots, w(m)) = \sum_{i=1}^l (y_i - x_i(m))^2 \quad (31)$$

υ.π.

$$x_i(k) = S\{w(k)x_i(k-1)\}.$$

Ο υπολογισμός του βέλτιστου γίνεται με χρήση των πολλαπλασιαστών $(b_i(k))$ Lagrange. Έτσι η συνάρτηση Lagrange είναι

$$L(w, x, b) = \frac{1}{l} \sum_{i=1}^l (y_i - x_i(m))^2 - \sum_{i=1}^l \sum_{k=1}^m (b_i(k)[x_i(k) - S\{w(k)x_i(k-1)\}])$$

Όπου $(b_i(k)) \geq 0$ είναι οι πολλαπλασιαστές Lagrange. Ικανή και αναγκαία συνθήκη για τοπικό ελάχιστο είναι η σχέση $\nabla L(w, x, b) = 0$. Από τη λύση αυτής της εξίσωσης προκύπτει η επαναληπτική σχέση $x_i(k) = S\{w(k)x_i(k-1)\}$ με $x_i(0) = x_i$. Επίσης προκύπτουν οι πολλαπλασιαστές της εξίσωσης Lagrange. Αναλυτικότερα για το τελευταίο επίπεδο

$$b_i(m) = 2(y_i - x_i(m)), \quad i = 1, \dots, l$$

Για τα υπόλοιπα επίπεδα οι πολλαπλασιαστές είναι [16]

$$b(k) = w^T(k+1)\nabla S\{w(k+1)x_i(k)\}bi(k+1), \quad i = 1, \dots, l \quad \text{και} \quad k = 1, \dots, m-1.$$

Ο υπολογισμός των βαρών προκύπτει από την μέθοδο μάθησης BFGS και διατυπώνεται ως εξής:

$$w(k+1) = w(k) - H_k^{-1}g_k$$

όπου g είναι η παράγωγος της συνάρτησης (31) και H είναι ο Χεσιανός πίνακας των βαρών.

3.2.2 Αλγόριθμος πλησιέστερου γείτονα

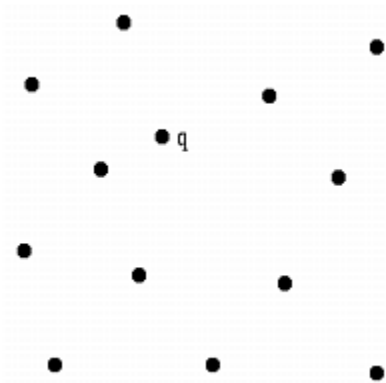
Στον αλγόριθμο πλησιέστερου γείτονα [21], το σύνολο των δεδομένων εκπαίδευσης χρησιμοποιείται για να ταξινομήσει κάθε μέλος ενός συνόλου δεδομένων. Η δομή των στοιχείων είναι τέτοια ώστε να υπάρχει μια μεταβλητή ταξινόμησης, και διάφορες πρόσθετες μεταβλητές πρόβλεψης. Γενικά, ο αλγόριθμος καταγράφεται στα παρακάτω 3 βήματα:

(1) Για κάθε σημείο-αντικείμενο των δεδομένων, εντοπίζονται τα K πλησιέστερα σημεία (οι πλησιέστεροι γείτονες) από τα δεδομένα εκπαίδευσης. Η Ευκλείδεια απόσταση χρησιμοποιείται για να υπολογιστεί πόσο κοντά είναι κάθε μέλος του συνόλου δεδομένων εκπαίδευσης στα δεδομένα που πρόκειται να ταξινομηθούν.

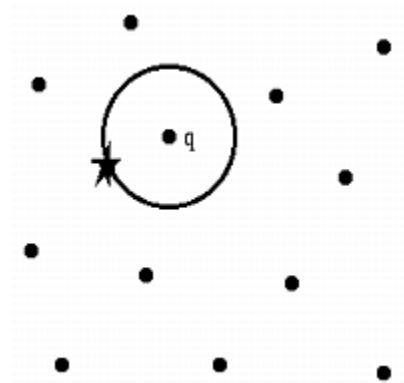
(2) Από τους K πλησιέστερους γείτονες που βρέθηκαν εξετάζεται ποια σημεία έχουν τις περισσότερες κοινές ταξινομήσεις. Η ταξινόμηση αυτή προσδιορίζει την κατηγορία στην οποία εντάσσεται το σημείο-αντικείμενο που εξετάζεται

(3) Επαναλαμβάνεται η διαδικασία για τα υπόλοιπα σημεία-αντικείμενα.

Φυσικά ο χρόνος υπολογισμού αυξάνει καθώς αυξάνεται ο αριθμός K των πλησιέστερων γειτόνων. Το πλεονέκτημα είναι ότι οι υψηλότερες τιμές του K παρέχουν μία ομαλότητα που μειώνει την ευαισθησία των στοιχείων εκπαίδευσης στο θόρυβο. Στις πρακτικές εφαρμογές, επιλέγονται περιττές τιμές για το K όπως 1,3,5....



Σχήμα 3.2: Σημείο-αντικείμενο q από το προς εξέταση σύνολο δεδομένων



Σχήμα 3.3: Εύρεση πλησιέστερου γείτονα.

Στα σχήματα 3.2 και 3.3 παρουσιάζεται ο τρόπος λειτουργίας της μεθόδου πλησιέστερου γείτονα. Στο σχήμα 3.2 με q συμβολίζεται το σημείο-αντικείμενο το οποίο πρέπει να ταξινομηθεί. Στο σχήμα 3.3 με αστεράκι συμβολίζεται ο πλησιέστερος γείτονας στο σημείο-αντικείμενο q . Η

κατηγορία στην οποία ανήκει αυτός ο πλησιέστερος γείτονας προσδιορίζει την ταξινόμηση του σημείου q .

3.2.3 Διακριτική ανάλυση

Η διακριτική ανάλυση είναι μια στατιστική τεχνική που επιτρέπει στους ερευνητές να ερευνήσουν τις διαφορές μεταξύ δυο ή περισσότερων ομάδων από αντικείμενα με εκτίμηση πολλών μεταβλητών ταυτόχρονα [22]. Στο σημείο αυτό θα γίνει παρουσίαση της γραμμικής διακριτικής ανάλυσης [23], [15]. Σκοπός της μεθόδου είναι η ανάπτυξη μιας σειράς διακριτών συναρτήσεων οι οποίες μεγιστοποιούν τη διακύμανση μεταξύ των κατηγοριών σε σχέση με τη διακύμανση εντός των κατηγοριών. Στην περίπτωση που υπάρχουν q κατηγορίες ταξινόμησης, αναπτύσσονται q γραμμικές συναρτήσεις της μορφής [24]:

$$L = b_1x_1 + b_2x_2 + \dots + b_nx_n + c$$

Όπου x_1, x_2, \dots, x_n είναι οι μεταβλητές των εισόδων, και b_1, b_2, \dots, b_n είναι διακριτικοί συντελεστές. Οι τιμές των c και b_i προκύπτουν από τις παρακάτω σχέσεις [15], με την προϋπόθεση ότι οι πίνακες διακύμανσης- συνδιακύμανσης των κατηγοριών είναι ίσοι και οι επιδόσεις των εναλλακτικών δραστηριοτήτων (ή μεταβλητές των εισόδων) ακολουθούν την πολυμεταβλητή κανονική κατανομή.

$$\mathbf{b} = \Sigma^{-1}[\boldsymbol{\mu}_k - \boldsymbol{\mu}_l] \text{ και } c = -[\boldsymbol{\mu}_k + \boldsymbol{\mu}_l]' \mathbf{b}/2$$

όπου μ_k είναι το διάνυσμα των μέσων τιμών των μεταβλητών εισόδων για την κατηγορία C_k και Σ είναι ο πίνακας διακύμανσης-συνδιακύμανσης μεταξύ των κατηγοριών.

Η διακριτική συνάρτηση, που αναπτύσσεται κατά τον παραπάνω τρόπο, χρησιμοποιείται για να υπολογιστεί η ταξινόμηση κάθε νέου σημείου. Για την ταξινόμηση ενός νέου σημείου υπολογίζονται όλες οι διακριτικές συναρτήσεις. Το νέο σημείο ταξινομείται στην κατηγορία i εάν η i -οστή διακριτική συνάρτηση παίρνει την μεγαλύτερη τιμή από όλες τις άλλες [22].

3.2.4 Δέντρα ταξινόμησης (CART)

Κατά τη μεθοδολογία CART αναπτύσσεται ένα δυαδικό δέντρο ταξινόμησης, κάθε κόμβος του οποίου χωρίζεται σε δύο κόμβους “παιδιά” και η διαδικασία επαναλαμβάνεται θεωρώντας κάθε φορά τους κόμβους παιδιά σαν “γονείς” δημιουργώντας έτσι το δέντρο. Τα κύρια σημεία της ανάλυσης CART συνοψίζονται στα ακόλουθα βήματα

- Διαχωρίζει κάθε κόμβο σε δύο επόμενους.
- Αποφασίζει αν ένα δέντρο είναι πλήρες και
- Προσδιορίζει κάθε τερματικό κόμβο σαν αποτέλεσμα ταξινόμησης.

Η διαδικασία αρχίζει με ένα σύνολο δεδομένων των οποίων η ταξινόμηση είναι γνωστή. Σκοπός είναι η δημιουργία ενός δέντρου που διαχωρίζει τα

δεδομένα αυτά. Κάθε κόμβος εξετάζει ένα κριτήριο ταξινόμησης. Για κάθε κόμβο επιλέγεται το καλύτερο κριτήριο, έτσι ώστε αυτό να πετυχαίνει τον καλύτερο διαχωρισμό των σημείων, που περιέχονται στον κόμβο στις προκαθορισμένες κατηγορίες στις οποίες ανήκουν.

Αφού βρεθεί ο καλύτερος διαχωρισμός, η μέθοδος CART επαναλαμβάνει τη διαδικασία εύρεσης για κάθε παιδί κόμβο, συνεχίζοντας μέχρι ο επόμενος διαχωρισμός να είναι δυνατός ή όχι. Έτσι δημιουργείται ένα πλήρες δέντρο. Όταν όλοι οι τερματικοί κόμβοι έχουν εντοπιστεί, σε κάθε έναν από αυτούς καταχωρείται μία κατηγορία ταξινόμησης.

Ένα πρόβλημα που παρουσιάζεται κατά τη δημιουργία των δέντρων είναι ότι συνήθως οδηγούμαστε σε περίπλοκα και εξειδικευμένα δέντρα. Το φαινόμενο αυτό είναι αποτέλεσμα της μεγάλης προσαρμογής του δέντρου στα δεδομένα εκμάθησης (overfitting). Το φαινόμενο αυτό δύναται να αντιμετωπιστεί με την περικοπή του δέντρου, έτσι ώστε να καταστεί πιο γενικό. Διαδικασίες περικοπής ενός δέντρου περιγράφονται στα βιβλία των Breiman [25] και Quinlan [26].

Αφού δημιουργηθεί το μέγιστο δέντρο καθώς και ένα σύνολο από μικρότερα δέντρα, τότε η μέθοδος διαλέγει το καλύτερο δέντρο υπολογίζοντας το λάθος της ταξινόμησης ή το κόστος. Με την ύπαρξη αρκετών δεδομένων χωρίζει τα δεδομένα σε δεδομένα εκπαίδευσης και δεδομένα ελέγχου. Τα δεδομένα εκπαίδευσης χρησιμοποιούνται για την δημιουργία ενός δέντρου μεγάλου μεγέθους. Τα δεδομένα δοκιμής χρησιμοποιούνται για τον υπολογισμό του λάθους εσφαλμένης ταξινόμησης. Το λάθος υπολογίζεται τόσο για το μεγάλο

δέντρο όσο και για τα μικρότερα. Το καλύτερο δέντρο είναι αυτό με το μικρότερο λάθος.

3.3 Πείραμα και αποτελέσματα

Σκοπός της παρούσας εργασίας είναι η σύγκρισή των SVM με άλλες μεθόδους στην ταξινόμηση σημείων σε δύο κατηγορίες. Έτσι πραγματοποιείται έλεγχος κάθε μεθόδου ως προς την αποτελεσματικότητα της σε ένα πλήθος αντικειμένων. Τα δεδομένα αυτά χωρίζονται σε δεδομένα για την εκπαίδευση των μεθόδων και σε δεδομένα για τον έλεγχο των μεθόδων. Η ανάλυση των αποτελεσμάτων κάθε μεθόδου αφορά το ποσοστό εσφαλμένων ταξινομήσεων κατά την εφαρμογή των μεθόδων στα δεδομένα εκπαίδευσης και ελέγχου. Μεγαλύτερη πρακτική σημασία έχει το ποσοστό εσφαλμένων ταξινομήσεων στα δείγματα ελέγχου, γι αυτό και δίνεται περισσότερη σημασία στο σημείο αυτό.

Τα δεδομένα όπως αναφέρθηκε χωρίζονται σε εκείνα που χρησιμοποιούνται για εκπαίδευση και σε εκείνα που χρησιμοποιούνται για τον έλεγχο των μεθόδων. Και στις δυο περιπτώσεις τα δεδομένα ακολουθούν τις ίδιες ακριβώς στατιστικές ιδιότητες οι οποίες αναλύονται ως εξής:

(1) Στατιστική κατανομή. Εξετάζονται τέσσερις συνολικά στατιστικές κατανομές.

A. Κανονική κατανομή με κύρτωση = 0

B. Ομοιόμορφη κατανομή με κύρτωση ≈ -1

Γ. Λογιστική κατανομή με κύρτωση ≈ 1

Δ. Κανονική με απεσπασμένα (outliers) σημεία (contaminated normal). Κατά την κατανομή αυτή το 10% των σημείων έχουν 4 φορές μεγαλύτερη διακύμανση από το υπόλοιπο 90%

(2) Πλήθος αντικειμένων στο δείγμα εκπαίδευσης. Στην παρούσα ανάλυση τα δείγματα εκπαίδευσης αποτελούνται από 500 και 1000 αντικείμενα. Αντίθετα το πλήθος των αντικειμένων στα δεδομένα ελέγχου είναι 200 αντικείμενα.

(3) Πλήθος χαρακτηριστικών/κριτηρίων. Τα χαρακτηριστικά είναι ο αριθμός των διαστάσεων κάθε σημείου (ή αντικειμένου). Τα χαρακτηριστικά πιο πρακτικά είναι ο αριθμός της πληροφορίας που περιγράφει κάθε αντικείμενο. Στην παρούσα ανάλυση εξετάζονται τρεις περιπτώσεις όσον αφορά το πλήθος των χαρακτηριστικών: (α) 5 χαρακτηριστικά, (β) 10 χαρακτηριστικά, (γ) 15 χαρακτηριστικά.

(4) Ομοιογένεια στις διακυμάνσεις των κατηγοριών ταξινόμησης. Τα δεδομένα πρέπει να ταξινομηθούν σε 2 κατηγορίες. Έτσι υπάρχουν δυο περιπτώσεις. Όταν οι διακυμάνσεις των κατηγοριών είναι ίσες και όταν δεν είναι ίσες. Η περίπτωση ίσων διακυμάνσεων μοντελοποιείται θέτοντας για κάθε κατηγορία (1 και 2) διασπορά $\sigma_1^2 = \sigma_2^2 = 1$ και μέση τιμή $\mu_1 = 0$, $\mu_2 = 0.5$. Αντίστοιχα η περίπτωση των άνισων διακυμάνσεων μοντελοποιείται θέτοντας $\sigma_1^2 = 1, \sigma_2^2 = 4$ και $\mu_1 = 0, \mu_2 = 0.8$. Ο συντελεστής συσχέτισης $\rho_{ij} \in [-0.1, 0.1]$ είναι τυχαία μεταβλητή ομοιόμορφα κατανεμημένη.

Η εφαρμογή όλων των μεθόδων καθώς και η παραγωγή των δεδομένων έγινε στη MATLAB. Για την υλοποίηση των μεθόδων χρησιμοποιήθηκαν κώδικες από την εργαλειοθήκη της MATLAB καθώς και κώδικες από το διαδίκτυο.

Η παρουσίαση γίνεται με την ακόλουθη σειρά: L-LSVM (Linear Lagrangian SVM), L-PSVM (Linear Proximal SVM). Ακολουθεί σύγκριση των γραμμικών αυτών μεθόδων. Στη συνέχεια παρουσιάζονται τα αποτελέσματα των μη γραμμικών μεθόδων NL-LSVM (Non-Linear Lagrangian SVM), NL-PSVM (Non-Linear Proximal SVM) και έπειτα η σύγκριση μεταξύ τους. Στη συνέχεια γίνεται σύγκριση των παραπάνω μεθόδων με τις παρακάτω μεθόδους ANN (artificial neural network), NN (Nearest Neighbors), LDA (Linear discriminant analysis), CART (classification and regression trees).

Για λόγους ευκολίας της παρουσίασης των αποτελεσμάτων, τόσο για την εκπαίδευση όσο και για τον έλεγχο οι τέσσερις στατιστικές κατανομές αναγράφονται ως $w = 1,2,3,4$ αντίστοιχα, το πλήθος των αντικειμένων για 500 και 1000 αναγράφεται ως $x = 1,2$ αντίστοιχα, το πλήθος των χαρακτηριστικών για 5, 10, 15 ως $y = 1,2,3$ και η ομοιογένεια των διακυμάνσεων (ίσες, άνισες) ως $z = 1,2$ αντίστοιχα.

3.3.1 Αποτελέσματα μεθόδου L-LSVM

Ξεκινώντας με τα L-LSVM για κάθε παράγοντα που εξετάζεται στην παρούσα ανάλυση το μέσο ποσοστό των εσφαλμένων ταξινομήσεων στο δείγμα ελέγχου παρουσιάζεται στον πίνακα 3.1

Από το πίνακα 3.1 παρατηρούμε ότι όταν τα δεδομένα ακολουθούν την κανονική και την ομοιόμορφη κατανομή ($w=1$ και $w=2$) το ποσοστό των λαθών είναι περίπου το ίδιο 23,43%. Στην περίπτωση της λογιστικής κατανομής το ποσοστό των εσφαλμένων ταξινομήσεων μειώνεται στο 22,70%, ενώ στην κανονική κατανομή με απεσπασμένα δεδομένα το ποσοστό εσφαλμένων ταξινομήσεων είναι υψηλότερο από όλες τις άλλες περιπτώσεις και ανέρχεται στο 24,154%. Χαρακτηριστικό είναι, ότι όσο αυξάνεται το πλήθος των χαρακτηριστικών (y) το ποσοστό των σφαλμάτων μειώνεται, κάτι που είναι αναμενόμενο καθώς όσο περισσότερη πληροφορία υπάρχει για κάποιο αντικείμενο τόσο πιο εύκολα και σωστά ταξινομείται. Άξιο αναφοράς είναι το γεγονός ότι όταν οι διακυμάνσεις (z) των κατηγοριών είναι άνισες η μέθοδος καταφέρει να ταξινομήσει τα αντικείμενα καλύτερα.

Παράγοντες	Ποσοστό σφάλματος
$w=1$	23.42%
$w=2$	23.43%
$w=3$	22.70%
$w=4$	24.15%
$x=1$	23.95%
$x=2$	22.90%
$y=1$	29.31%
$y=2$	22.84%
$y=3$	18.13%
$z=1$	23.89%
$z=2$	22.96%

Πίνακας 3.1: Ποσοστό σφαλμάτων σε σχέση με τους παράγοντες των δεδομένων

Τα αναλυτικά αποτελέσματα (μέσο ποσοστό εσφαλμένων ταξινομήσεων στο δείγμα ελέγχου) για όλους τους συνδυασμούς των εξεταζομένων παραγόντων παρουσιάζονται στον πίνακα 3.2

Συνδυασμός παραγόντων	Ποσοστό σφάλματος	Συνδυασμός παραγόντων	Ποσοστό σφάλματος
w=1,x=1,y=1,z=1	29.65%	w=2,x=1,y=1,z=1	30.47%
w=1,x=1,y=1,z=2	30.72%	w=2,x=1,y=1,z=2	31.07%
w=1,x=1,y=2,z=1	22.87%	w=2,x=1,y=2,z=1	22.67%
w=1,x=1,y=2,z=2	23.02%	w=2,x=1,y=2,z=2	25.12%
w=1,x=1,y=3,z=1	18.17%	w=2,x=1,y=3,z=1	20.00%
w=1,x=1,y=3,z=2	19.15%	w=2,x=1,y=3,z=2	13.70%
w=1,x=2,y=1,z=1	30.57%	w=2,x=2,y=1,z=1	27.15%
w=1,x=2,y=1,z=2	26.12%	w=2,x=2,y=1,z=2	30.17%
w=1,x=2,y=2,z=1	23.05%	w=2,x=2,y=2,z=1	22.40%
w=1,x=2,y=2,z=2	20.20%	w=2,x=2,y=2,z=2	21.77%
w=1,x=2,y=3,z=1	19.32%	w=2,x=2,y=3,z=1	20.52%
w=1,x=2,y=3,z=2	18.27%	w=2,x=2,y=3,z=2	16.12%
w=3,x=1,y=1,z=1	30.22%	w=4,x=1,y=1,z=1	31.25%
w=3,x=1,y=1,z=2	29.95%	w=4,x=1,y=1,z=2	27.35%
w=3,x=1,y=2,z=1	24.22%	w=4,x=1,y=2,z=1	25.22%
w=3,x=1,y=2,z=2	22.00%	w=4,x=1,y=2,z=2	23.47%
w=3,x=1,y=3,z=1	19.30%	w=4,x=1,y=3,z=1	17.57%
w=3,x=1,y=3,z=2	17.35%	w=4,x=1,y=3,z=2	20.32%
w=3,x=2,y=1,z=1	29.22%	w=4,x=2,y=1,z=1	28.92%
w=3,x=2,y=1,z=2	25.62%	w=4,x=2,y=1,z=2	30.47%
w=3,x=2,y=2,z=1	21.22%	w=4,x=2,y=2,z=1	23.25%
w=3,x=2,y=2,z=2	22.20%	w=4,x=2,y=2,z=2	22.75%
w=3,x=2,y=3,z=1	16.07%	w=4,x=2,y=3,z=1	20.05%
w=3,x=2,y=3,z=2	15.05%	w=4,x=2,y=3,z=2	19.20%

Πίνακας 3.2: Σφάλματα σε σχέση με τους συνδυασμούς των παραγόντων

Παρατηρείται ότι το ποσοστό των λαθών αν και βελτιώνεται σε σχέση με το πλήθος των χαρακτηριστικών, ωστόσο παραμένει μεγάλο και αυτό διότι τα

προβλήματα που συνήθως παρουσιάζονται στη πράξη είναι πολύπλοκα και οι απλές γραμμικές μέθοδοι αδυνατούν να δώσουν ικανοποιητικά αποτελέσματα. Παρατηρείται ότι όταν σε κάποιον από τους συνδυασμούς υπάρχουν παράγοντες που δίνουν καλύτερα αποτελέσματα ($y=3, z=2, x=2, w=3$) όπως παρουσιάστηκε στον πίνακα 3.1, τότε το ποσοστό των σφαλμάτων πάντα βελτιώνεται. Το αντίστροφο γίνεται όταν οι παράγοντες δίνουν χειρότερα αποτελέσματα, σύμφωνα με τον πίνακα 3.1. Σε αντίθεση με τον πίνακα 3.1 που δείχνει ότι τα αποτελέσματα είναι καλύτερα, όταν οι διακυμάνσεις των κατηγοριών είναι άνισες ($z=2$), στον πίνακα 3.2 σε ορισμένες περιπτώσεις συνδυασμών, όταν οι διακυμάνσεις των κατηγοριών είναι άνισες ($z=2$) τα αποτελέσματα είναι χειρότερα. Για παράδειγμα στους συνδυασμούς $w=1, x=1, y=1, z=1$ και $w=1, x=1, y=1, z=2$ το ποσοστό των σφαλμάτων είναι 29,650% και 30,725% αντίστοιχα. Στους συνδυασμούς που το πλήθος των χαρακτηριστικών αυξάνεται το ποσοστό των σφαλμάτων πάντα βελτιώνεται. Αξιοσημείωτο είναι ότι το χαμηλότερο ποσοστό εσφαλμένης ταξινόμησης παρατηρείται στους συνδυασμούς της ομοιόμορφης κατανομής ($w=2$) και όχι της λογιστικής όπως προκύπτει από τον πίνακα 3.1. Επίσης αν γίνει σύγκριση των αποτελεσμάτων με βάση την κατανομή, σε αντίθεση με τον πίνακα 3.1 παρατηρείται ότι τα χαμηλότερα αποτελέσματα δεν παρουσιάζονται στην ομοιόμορφη κατανομή.

Ο πίνακας 3.3 παρουσιάζει μια σύγκριση των σφαλμάτων μεταξύ των δειγμάτων εκπαίδευσης και ελέγχου. Παρατηρείται ότι το λάθος είναι μικρότερο κατά την εκπαίδευση, όπως αναμενόταν, αλλά παραμένει μεγάλο. Παρόλο που η μέθοδος κατά την εκπαίδευση παρουσιάζει μεγάλο ποσοστό σφαλμάτων, κατά τον έλεγχο παρουσιάζει ποσοστό εσφαλμένων ταξινομήσεων με μικρές διαφορές από εκείνο της εκπαίδευσης. Οι διαφορές

αυτές κυμαίνονται γύρω στις 0,62 ποσοστιαίες μονάδες. Επίσης παρατηρείται το φαινόμενο ότι όλα τα προαναφερθέντα μη αναμενόμενα σφάλματα του πίνακα 3.2 παρουσιάζονται και για τα δεδομένα εκπαίδευσης στον πίνακα 3.3.

	Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου		Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου
w=1,x=1,y=1,z=1	29.02%	29.65%	w=2,x=1,y=1,z=1	29.64%	30.47%
w=1,x=1,y=1,z=2	29.85%	30.72%	w=2,x=1,y=1,z=2	30.11%	31.07%
w=1,x=1,y=2,z=1	22.39%	22.87%	w=2,x=1,y=2,z=1	22.43%	22.67%
w=1,x=1,y=2,z=2	22.18%	23.02%	w=2,x=1,y=2,z=2	24.02%	25.12%
w=1,x=1,y=3,z=1	16.27%	18.17%	w=2,x=1,y=3,z=1	19.00%	20.00%
w=1,x=1,y=3,z=2	17.55%	19.15%	w=2,x=1,y=3,z=2	12.57%	13.70%
w=1,x=2,y=1,z=1	29.45%	30.57%	w=2,x=2,y=1,z=1	27.30%	27.15%
w=1,x=2,y=1,z=2	27.01%	26.12%	w=2,x=2,y=1,z=2	29.13%	30.17%
w=1,x=2,y=2,z=1	22.78%	23.05%	w=2,x=2,y=2,z=1	22.17%	22.40%
w=1,x=2,y=2,z=2	20.53%	20.20%	w=2,x=2,y=2,z=2	19.66%	21.77%
w=1,x=2,y=3,z=1	18.29%	19.32%	w=2,x=2,y=3,z=1	20.45%	20.52%
w=1,x=2,y=3,z=2	17.05%	18.27%	w=2,x=2,y=3,z=2	15.69%	16.12%
w=3,x=1,y=1,z=1	29.54%	30.22%	w=4,x=1,y=1,z=1	32.58%	31.25%
w=3,x=1,y=1,z=2	28.64%	29.95%	w=4,x=1,y=1,z=2	26.44%	27.35%
w=3,x=1,y=2,z=1	24.53%	24.22%	w=4,x=1,y=2,z=1	24.09%	25.22%
w=3,x=1,y=2,z=2	21.33%	22.00%	w=4,x=1,y=2,z=2	22.69%	23.47%
w=3,x=1,y=3,z=1	18.41%	19.30%	w=4,x=1,y=3,z=1	16.44%	17.57%
w=3,x=1,y=3,z=2	16.35%	17.35%	w=4,x=1,y=3,z=2	18.56%	20.32%
w=3,x=2,y=1,z=1	30.02%	29.22%	w=4,x=2,y=1,z=1	29.44%	28.92%
w=3,x=2,y=1,z=2	26.13%	25.62%	w=4,x=2,y=1,z=2	29.87%	30.47%
w=3,x=2,y=2,z=1	21.59%	21.22%	w=4,x=2,y=2,z=1	22.79%	23.25%
w=3,x=2,y=2,z=2	21.04%	22.20%	w=4,x=2,y=2,z=2	22.57%	22.75%
w=3,x=2,y=3,z=1	15.04%	16.07%	w=4,x=2,y=3,z=1	19.41%	20.05%
w=3,x=2,y=3,z=2	14.38%	15.05%	w=4,x=2,y=3,z=2	18.22%	19.20%

Πίνακας 3.3: Σύγκριση αποτελεσμάτων δείγματος ελέγχου και εκπαίδευσης.

3.3.2. Αποτελέσματα μεθόδου L-PSVM

Για την μέθοδο L-PSVM ο πίνακας 3.4 παρουσιάζει το ποσοστό των εσφαλμένων ταξινομήσεων για κάθε ιδιότητα.

Παράγοντες	Ποσοστό σφάλματος
w=1	23.21%
w=2	23.30%
w=3	22.51%
w=4	24.01%
x=1	23.79%
x=2	22.72%
y=1	29.27%
y=2	22.67%
y=3	17.83%
z=1	23.88%
z=2	22.64%

Πίνακας 3.4: Ποσοστό σφαλμάτων ανά ιδιότητα

Και εδώ όπως και στην μέθοδο linear Lagrange τα δεδομένα με κανονική και ομοιόμορφη κατανομή έχουν περίπου το ίδιο ποσοστό λαθών και το μικρότερο ποσοστό παρουσιάζεται όταν η κατανομή είναι η λογιστική. Η διαφορά μεταξύ ομοιόμορφης και λογιστικής είναι στις 0,70 ποσοστιαίες μονάδες και η διαφορά μεταξύ ομοιόμορφης και κανονικής με απεσπασμένα σημεία παρουσιάζει μικρή διαφορά και είναι ίση με 0.78 ποσοστιαίες μονάδες. Επίσης τα αποτελέσματα βελτιώνονται όσο το πλήθος των αντικειμένων αυξάνονται. Όπως και στην μέθοδο L-LSVM όταν οι διακυμάνσεις είναι άνισες τα αποτελέσματα βελτιώνονται. Η βελτίωση όμως είναι μικρή και ίση με 1,2 ποσοστιαίες μονάδες.

Συνδυασμός παραγόντων	Ποσοστό σφάλματος	Συνδυασμός παραγόντων	Ποσοστό σφάλματος
-----------------------	-------------------	-----------------------	-------------------

w=1,x=1,y=1,z=1	29.65%	w=2,x=1,y=1,z=1	30.47%
w=1,x=1,y=1,z=2	30.62%	w=2,x=1,y=1,z=2	31.17%
w=1,x=1,y=2,z=1	22.85%	w=2,x=1,y=2,z=1	22.67%
w=1,x=1,y=2,z=2	22.60%	w=2,x=1,y=2,z=2	24.97%
w=1,x=1,y=3,z=1	18.12%	w=2,x=1,y=3,z=1	20.00%
w=1,x=1,y=3,z=2	18.40%	w=2,x=1,y=3,z=2	13.15%
w=1,x=2,y=1,z=1	30.57%	w=2,x=2,y=1,z=1	27.15%
w=1,x=2,y=1,z=2	26.05%	w=2,x=2,y=1,z=2	30.02%
w=1,x=2,y=2,z=1	23.05%	w=2,x=2,y=2,z=1	22.40%
w=1,x=2,y=2,z=2	19.75%	w=2,x=2,y=2,z=2	21.40%
w=1,x=2,y=3,z=1	23.32%	w=2,x=2,y=3,z=1	20.45%
w=1,x=2,y=3,z=2	17.62%	w=2,x=2,y=3,z=2	15.77%
w=3,x=1,y=1,z=1	30.20%	w=4,x=1,y=1,z=1	31.30%
w=3,x=1,y=1,z=2	29.87%	w=4,x=1,y=1,z=2	27.30%
w=3,x=1,y=2,z=1	24.27%	w=4,x=1,y=2,z=1	25.17%
w=3,x=1,y=2,z=2	21.67%	w=4,x=1,y=2,z=2	23.30%
w=3,x=1,y=3,z=1	19.27%	w=4,x=1,y=3,z=1	17.52%
w=3,x=1,y=3,z=2	16.77%	w=4,x=1,y=3,z=2	19.75%
w=3,x=2,y=1,z=1	29.22%	w=4,x=2,y=1,z=1	28.97%
w=3,x=2,y=1,z=2	25.55%	w=4,x=2,y=1,z=2	30.17%
w=3,x=2,y=2,z=1	21.15%	w=4,x=2,y=2,z=1	23.22%
w=3,x=2,y=2,z=2	21.85%	w=4,x=2,y=2,z=2	22.50%
w=3,x=2,y=3,z=1	16.07%	w=4,x=2,y=3,z=1	20.05%
w=3,x=2,y=3,z=2	14.27%	w=4,x=2,y=3,z=2	18.85%

Πίνακας 3.5: Ποσοστό σφαλμάτων σε σχέση με τους συνδυασμούς των παραγόντων

Στον πίνακα 3.5 παρουσιάζονται τα αναλυτικά αποτελέσματα (μέσο ποσοστό εσφαλμένων ταξινομήσεων στο δείγμα ελέγχου) για όλους τους συνδυασμούς των εξεταζομένων παραγόντων. Παρατηρείται ότι η μέθοδος δεν διαφέρει ουσιαστικά από την προηγούμενη γραμμική μέθοδο. Από τον πίνακα 3.5 παρατηρείται το ίδιο φαινόμενο όπως και στη μέθοδο L-LPSVM σε σχέση με

τη διακύμανση. Παρατηρείται δηλαδή ότι σε ορισμένους συνδυασμούς όταν οι διακυμάνσεις είναι άνισες, το ποσοστό των σφαλμάτων είναι μεγαλύτερο σε σχέση με την περίπτωση ίσων διακυμάνσεων. Για παράδειγμα για τους συνδυασμούς $w=1, x=1, y=1, z=1$ και $w=1, x=1, y=1, z=2$ το ποσοστό των λαθών είναι 29,65% και 30,6% αντίστοιχα. Επίσης αν και θα περίμενε κανείς στους συνδυασμούς με την λογιστική κατανομή να παρουσιάζονται καλύτερα αποτελέσματα (βλ πίνακας 3.4), σε μόνο 4 περιπτώσεις έχει καλύτερα αποτελέσματα. Συγκεκριμένα στους συνδυασμούς $w=3, x=1, y=2, z=2$, $w=3, x=2, y=1, z=2$, $w=3, x=2, y=3, z=1$, $w=3, x=2, y=3, z=2$. Επίσης μπορεί να παρατηρήσει κανείς ότι τα καλύτερα αποτελέσματα σε σχέση με τις κατανομές παρατηρούνται στην ομοιόμορφη, την λογιστική και την κανονική με τα αποσπασμένα σημεία. Τέλος όσο αυξάνεται το πλήθος των χαρακτηριστικών το ποσοστό των σφαλμάτων πάντα βελτιώνεται.

Παρατηρείται στον πίνακα 3.6 (σύγκριση των ποσοστών των λαθών κατά την εκπαίδευση και κατά τον έλεγχο) ότι, ενώ όπως αναμενόταν το ποσοστό των λαθών κατά την εκπαίδευση να είναι μικρότερο, η διαφορά κυμαίνεται από 0 μέχρι 1 ποσοστιαία μονάδα και σε 3 μόνο περιπτώσεις ξεπερνάει τη 1 ποσοστιαία μονάδα. Ο μέσος όρος των διαφορών είναι στις 0.6 ποσοστιαίες μονάδες. Παρατηρείται όμως το γεγονός ότι στις περιπτώσεις που το πλήθος των αντικειμένων είναι 1000 ($x=2$) και το πλήθος των χαρακτηριστικών είναι 5, τότε η μέθοδος παρουσιάζει καλύτερα αποτελέσματα στο δείγμα ελέγχου παρά στο δείγμα εκπαίδευσης. Στη περίπτωση που η κατανομή είναι λογιστική ($w=3$), το πλήθος των χαρακτηριστικών είναι 10 ($y=2$) και οι διακυμάνσεις είναι ίσες, τότε και σε αυτήν την περίπτωση η μέθοδος ταξινομεί καλύτερα το δείγμα ελέγχου από το δείγμα εκπαίδευσης. Σε ορισμένες περιπτώσεις του δείγματος ελέγχου όταν οι διακυμάνσεις είναι ίσες

η μέθοδος δίνει καλύτερα αποτελέσματα σε σχέση με εκείνα των άνισων διακυμάνσεων. Στις ίδιες περιπτώσεις το ίδιο φαινόμενο παρατηρείται και στο δείγμα εκπαίδευσης. Επίσης η μείωση του σφάλματος κατά την αύξηση των χαρακτηριστικών συμβαίνει πάντα και στα δυο δείγματα. Η διαφορά στη μείωση αυτή, μεταξύ δείγματος ελέγχου και δείγματος εκπαίδευσης είναι επίσης μικρή και ίση με 0.65 ποσοστιαίες μονάδες. Δηλαδή η μέθοδος δεν παρουσιάζει γενικώς μεγάλη απόκλιση από το ποσοστό σφάλματος κατά την εκπαίδευση.

	Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου		Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου
w=1,x=1,y=1,z=1	29.02%	29.65%	w=2,x=1,y=1,z=1	29.64%	30.48%
w=1,x=1,y=1,z=2	29.78%	30.63%	w=2,x=1,y=1,z=2	30.03%	31.18%
w=1,x=1,y=2,z=1	22.39%	22.85%	w=2,x=1,y=2,z=1	22.42%	22.68%
w=1,x=1,y=2,z=2	21.84%	22.60%	w=2,x=1,y=2,z=2	23.81%	24.98%
w=1,x=1,y=3,z=1	16.23%	18.13%	w=2,x=1,y=3,z=1	18.98%	20.00%
w=1,x=1,y=3,z=2	17.00%	18.40%	w=2,x=1,y=3,z=2	12.14%	13.15%
w=1,x=2,y=1,z=1	29.45%	30.58%	w=2,x=2,y=1,z=1	27.30%	27.15%
w=1,x=2,y=1,z=2	26.93%	26.05%	w=2,x=2,y=1,z=2	29.07%	30.03%
w=1,x=2,y=2,z=1	22.80%	23.05%	w=2,x=2,y=2,z=1	22.17%	22.40%
w=1,x=2,y=2,z=2	20.38%	19.75%	w=2,x=2,y=2,z=2	19.31%	21.40%
w=1,x=2,y=3,z=1	17.41%	18.33%	w=2,x=2,y=3,z=1	20.45%	20.45%
w=1,x=2,y=3,z=2	16.54%	17.63%	w=2,x=2,y=3,z=2	15.07%	15.78%

Πίνακας 3.6: Ποσοστά σφαλμάτων δείγματος εκπαίδευσης και ελέγχου

	Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου		Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου
w=3,x=1,y=1,z=1	29.54%	30.20%	w=4,x=1,y=1,z=1	32.59%	31.30%
w=3,x=1,y=1,z=2	28.59%	29.88%	w=4,x=1,y=1,z=2	26.33%	27.30%
w=3,x=1,y=2,z=1	24.56%	24.28%	w=4,x=1,y=2,z=1	24.15%	25.18%
w=3,x=1,y=2,z=2	21.01%	21.68%	w=4,x=1,y=2,z=2	22.55%	23.30%
w=3,x=1,y=3,z=1	18.39%	19.28%	w=4,x=1,y=3,z=1	16.39%	17.53%
w=3,x=1,y=3,z=2	15.86%	16.78%	w=4,x=1,y=3,z=2	18.18%	19.75%
w=3,x=2,y=1,z=1	30.03%	29.23%	w=4,x=2,y=1,z=1	29.43%	28.98%
w=3,x=2,y=1,z=2	26.04%	25.55%	w=4,x=2,y=1,z=2	29.76%	30.18%
w=3,x=2,y=2,z=1	21.58%	21.15%	w=4,x=2,y=2,z=1	22.79%	23.23%
w=3,x=2,y=2,z=2	20.80%	21.85%	w=4,x=2,y=2,z=2	22.38%	22.50%
w=3,x=2,y=3,z=1	15.02%	16.08%	w=4,x=2,y=3,z=1	19.37%	20.05%
w=3,x=2,y=3,z=2	13.92%	14.28%	w=4,x=2,y=3,z=2	17.80%	18.85%

Πίνακας 3.6: Ποσοστά σφαλμάτων δείγματος εκπαίδευσης και ελέγχου (συνέχεια)

3.3.3. Αποτελέσματα μεθόδου NL-LSVM

Για τη μέθοδο Non linear Lagrange για κάθε ιδιότητα το ποσοστό των λαθών στα δεδομένα δοκιμής παρουσιάζεται στον πίνακα 3.7

Σε αυτή την περίπτωση η ομοιόμορφη κατανομή παρουσιάζει το χαμηλότερο ποσοστό εσφαλμένων ταξινομήσεων, ενώ στην περίπτωση των απεσπασμένων στοιχείων παρουσιάζονται τα χειρότερα αποτελέσματα. Χαρακτηριστικό είναι ότι και σε αυτή την μέθοδο όπως και στις δύο προηγούμενες (L-LSVM και L-PSVM) μεθόδους υπάρχουν 2 κατανομές (κανονική και λογιστική) που θα παρουσιάζουν περίπου ίδια αποτελέσματα με διαφορά 0,1 ποσοστιαίες μονάδες. Όσον αφορά το πλήθος των αντικειμένων και των χαρακτηριστικών όπως και στη γραμμική περίπτωση, η αποτελεσματικότητα της μεθόδου βελτιώνεται όσο το πλήθος αυξάνεται. Ωστόσο η βελτίωση από τα 10 χαρακτηριστικά (y=2) σε 15 χαρακτηριστικά

($y=3$) δεν είναι σημαντική. Σε αντίθεση με τις προηγούμενες περιπτώσεις στις οποίες η βελτίωση των σφαλμάτων μεταξύ ίσων και άνισων διακυμάνσεων ήταν γύρω στο 1%, στη συγκεκριμένη περίπτωση η βελτίωση ανέρχεται στις 4,209 ποσοστιαίες μονάδες.

Παράγοντες	Ποσοστό σφάλματος
w=1	21.35%
w=2	20.96%
w=3	21.17%
w=4	22.51%
x=1	22.54%
x=2	20.46%
y=1	24.38%
y=2	20.45%
y=3	19.67%
z=1	23.60%
z=2	19.39%

Πίνακας 3.7: ποσοστό εσφαλμένων ταξινομήσεων ανά ιδιότητα

Τα αναλυτικά αποτελέσματα (μέσο ποσοστό εσφαλμένων ταξινομήσεων στο δείγμα ελέγχου) για όλους τους συνδυασμούς των εξεταζομένων παραγόντων παρουσιάζονται στον πίνακα 3.8

Παρόλο που σύμφωνα με τον πίνακα 3.7 η ομοιόμορφη κατανομή δίνει καλύτερα αποτελέσματα, στον πίνακα 3.8 το χαμηλότερο ποσοστό εσφαλμένων ταξινομήσεων παρουσιάζεται στη λογιστική κατανομή. Αξιοσημείωτο είναι ότι ενώ στο πίνακα 3.7 η μέθοδος βελτιώνει το σφάλμα όσο το πλήθος των χαρακτηριστικών αυξάνει, στον πίνακα 3.8, και συγκεκριμένα στις περιπτώσεις που η διακύμανση είναι άνιση, δεν συμβαίνει αυτό. Συγκεκριμένα όταν το πλήθος των χαρακτηριστικών είναι 15, η μέθοδος δίνει χειρότερα αποτελέσματα σε σχέση με την περίπτωση 5

χαρακτηριστικών. Το φαινόμενο αυτό δεν παρατηρείται όταν η κατανομή είναι η κανονική με αποσπώμενα στοιχεία. Επίσης παρατηρείται ότι στις περιπτώσεις κανονικής ($w=1$) και ομοιόμορφης κατανομής ($w=3$) και όταν το πλήθος των χαρακτηριστικών είναι 15 ($y=3$) η ιδιότητα των άνισων διακυμάνσεων δίνει χειρότερα αποτελέσματα σε σχέση με τις ίσες. Επίσης παρατηρείται ότι το ποσοστό του σφάλματος για άνισες διακυμάνσεις κυμαίνεται από 14 μέχρι 23%, ενώ για ίσες διακυμάνσεις κυμαίνεται από 16 μέχρι 30%, γεγονός που δικαιολογεί το αποτέλεσμα του πίνακα 3.7.

Συνδυασμός παραγόντων	Ποσοστό σφάλματος	Συνδυασμός παραγόντων	Ποσοστό σφάλματος
$w=1, x=1, y=1, z=1$	30.57%	$w=2, x=1, y=1, z=1$	30.42%
$w=1, x=1, y=1, z=2$	20.20%	$w=2, x=1, y=1, z=2$	18.35%
$w=1, x=1, y=2, z=1$	23.07%	$w=2, x=1, y=2, z=1$	23.15%
$w=1, x=1, y=2, z=2$	19.07%	$w=2, x=1, y=2, z=2$	18.82%
$w=1, x=1, y=3, z=1$	16.07%	$w=2, x=1, y=3, z=1$	21.30%
$w=1, x=1, y=3, z=2$	23.00%	$w=2, x=1, y=3, z=2$	20.77%
$w=1, x=2, y=1, z=1$	29.97%	$w=2, x=2, y=1, z=1$	26.15%
$w=1, x=2, y=1, z=2$	18.02%	$w=2, x=2, y=1, z=2$	16.22%
$w=1, x=2, y=2, z=1$	21.90%	$w=2, x=2, y=2, z=1$	21.25%
$w=1, x=2, y=2, z=2$	16.20%	$w=2, x=2, y=2, z=2$	15.80%
$w=1, x=2, y=3, z=1$	18.00%	$w=2, x=2, y=3, z=1$	20.47%
$w=1, x=2, y=3, z=2$	20.15%	$w=2, x=2, y=3, z=2$	18.87%

Πίνακας 3.8: Ποσοστό σφαλμάτων για όλους τους συνδυασμούς των εξεταζομένων παραγόντων

Συνδυασμός παραγόντων	Ποσοστό σφάλματος	Συνδυασμός παραγόντων	Ποσοστό σφάλματος
w=3,x=1,y=1,z=1	30.22%	w=4,x=1,y=1,z=1	32.15%
w=3,x=1,y=1,z=2	21.20%	w=4,x=1,y=1,z=2	20.57%
w=3,x=1,y=2,z=1	24.80%	w=4,x=1,y=2,z=1	25.07%
w=3,x=1,y=2,z=2	18.15%	w=4,x=1,y=2,z=2	20.55%
w=3,x=1,y=3,z=1	19.82%	w=4,x=1,y=3,z=1	18.62%
w=3,x=1,y=3,z=2	22.20%	w=4,x=1,y=3,z=2	22.90%
w=3,x=2,y=1,z=1	28.85%	w=4,x=2,y=1,z=1	28.52%
w=3,x=2,y=1,z=2	17.97%	w=4,x=2,y=1,z=2	20.70%
w=3,x=2,y=2,z=1	20.00%	w=4,x=2,y=2,z=1	22.35%
w=3,x=2,y=2,z=2	17.52%	w=4,x=2,y=2,z=2	19.55%
w=3,x=2,y=3,z=1	14.12%	w=4,x=2,y=3,z=1	19.70%
w=3,x=2,y=3,z=2	19.25%	w=4,x=2,y=3,z=2	19.50%

Πίνακας 3.8: Ποσοστό σφαλμάτων για όλους τους συνδυασμούς των εξεταζομένων παραγόντων (συνέχεια)

Ο πίνακας 3.9 παρουσιάζει μια σύγκριση των λαθών κατά την εκπαίδευση και κατά την δοκιμή. Παρατηρείται ότι το λάθος είναι μικρότερο κατά την εκπαίδευση, όπως αναμενόταν, αλλά η διαφορά με τα ποσοστά λαθών του δείγματος ελέγχου είναι πολύ μεγάλη. Γεγονός που δείχνει ότι η μέθοδος προσαρμόστηκε υπερβολικά στα δεδομένα εκπαίδευσης (overfitting). Μάλιστα στις περιπτώσεις που το πλήθος των χαρακτηριστικών αυξάνει το ποσοστό λάθους κατά την εκπαίδευση είναι παρά πολύ μικρό πλησιάζοντας το μηδέν. Επίσης παρατηρείται στα δεδομένα εκπαίδευσης ότι η μείωση του ποσοστού των λαθών στις άνισες διακυμάνσεις είναι σημαντική.

	Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου		Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου
w=3,x=1,y=1,z=1	26.44%	30.57%	w=4,x=1,y=1,z=1	26.58%	30.42%
w=3,x=1,y=1,z=2	14.97%	20.20%	w=4,x=1,y=1,z=2	13.50%	18.35%
w=3,x=1,y=2,z=1	7.45%	23.07%	w=4,x=1,y=2,z=1	7.05%	23.15%
w=3,x=1,y=2,z=2	0.59%	19.07%	w=4,x=1,y=2,z=2	0.39%	18.82%
w=3,x=1,y=3,z=1	0.08%	16.07%	w=4,x=1,y=3,z=1	0.01%	21.30%
w=3,x=1,y=3,z=2	0.00%	23.00%	w=4,x=1,y=3,z=2	0.00%	20.77%
w=3,x=2,y=1,z=1	27.89%	29.97%	w=4,x=2,y=1,z=1	25.11%	26.15%
w=3,x=2,y=1,z=2	15.40%	18.02%	w=4,x=2,y=1,z=2	14.60%	16.22%
w=3,x=2,y=2,z=1	12.54%	21.90%	w=4,x=2,y=2,z=1	11.30%	21.25%
w=3,x=2,y=2,z=2	1.55%	16.20%	w=4,x=2,y=2,z=2	0.79%	15.80%
w=3,x=2,y=3,z=1	0.73%	18.00%	w=4,x=2,y=3,z=1	0.29%	20.47%
w=3,x=2,y=3,z=2	0.01%	20.15%	w=4,x=2,y=3,z=2	0.01%	18.87%
w=3,x=1,y=1,z=1	26.20%	30.22%	w=4,x=1,y=1,z=1	27.08%	32.15%
w=3,x=1,y=1,z=2	15.32%	21.20%	w=4,x=1,y=1,z=2	12.82%	20.57%
w=3,x=1,y=2,z=1	9.61%	24.80%	w=4,x=1,y=2,z=1	7.01%	25.07%
w=3,x=1,y=2,z=2	1.22%	18.15%	w=4,x=1,y=2,z=2	0.53%	20.55%
w=3,x=1,y=3,z=1	0.29%	19.82%	w=4,x=1,y=3,z=1	0.04%	18.62%
w=3,x=1,y=3,z=2	0.00%	22.20%	w=4,x=1,y=3,z=2	0.00%	22.90%
w=3,x=2,y=1,z=1	28.16%	28.85%	w=4,x=2,y=1,z=1	25.57%	28.52%
w=3,x=2,y=1,z=2	15.80%	17.97%	w=4,x=2,y=1,z=2	16.24%	20.70%
w=3,x=2,y=2,z=1	11.12%	20.00%	w=4,x=2,y=2,z=1	9.73%	22.35%
w=3,x=2,y=2,z=2	2.16%	17.52%	w=4,x=2,y=2,z=2	1.37%	19.55%
w=3,x=2,y=3,z=1	0.63%	14.12%	w=4,x=2,y=3,z=1	0.52%	19.70%
w=3,x=2,y=3,z=2	0.01%	19.25%	w=4,x=2,y=3,z=2	0.01%	19.50%

Πίνακας 3.9: Ποσοστά σφαλμάτων δείγματος εκπαίδευσης και ελέγχου

3.3.4. Αποτελέσματα μεθόδου NL-PSVM

Όπως στις παραπάνω μεθόδους, έτσι και στη μέθοδο Non Linear Proximal SVM τα ποσοστά των εσφαλμένων ταξινομήσεων σε σχέση με τις ιδιότητες του δείγματος ελέγχου παρουσιάζονται στον πίνακα 3.10.

Παράγοντες	Ποσοστό σφαλμάτων
w=1	15.19%
w=2	14.38%
w=3	15.64%
w=4	17.56%
x=1	16.25%
x=2	15.14%
y=1	22.06%
y=2	14.39%
y=3	10.64%
z=1	23.00%
z=2	8.39%

Πίνακας 3.10: ποσοστό εσφαλμένων ταξινομήσεων ανά παράγοντα

Τα αποτελέσματα δείχνουν ότι η μέθοδος καταφέρνει να ταξινομήσει τα δεδομένα πολύ καλύτερα από τις προηγούμενες μεθόδους αφού τα ποσοστά των εσφαλμένων ταξινομήσεων είναι φανερά πολύ μικρότερα. Μάλιστα όταν το πλήθος των χαρακτηριστικών είναι 5 ($y=1$) και οι διακυμάνσεις είναι ίσες ($z=1$) τότε το ποσοστό εσφαλμένων ταξινομήσεων είναι μεγαλύτερο του 20%. Σε όλες τις υπόλοιπες περιπτώσεις το ποσοστό είναι μικρότερο του 20%. Παρατηρώντας τα ποσοστά των σφαλμάτων των κατανομών, προκύπτει ότι η ομοιόμορφη κατανομή παρουσιάζει το μικρότερο ποσοστό. Το μεγαλύτερο ποσοστό το παρουσιάζει η κανονική κατανομή με απεσπασμένα σημεία. Οι υπόλοιπες κατανομές έχουν ποσοστά κοντά σε αυτό της ομοιόμορφης. Κατά την αύξηση του πλήθους των χαρακτηριστικών το ποσοστό των σφαλμάτων μειώνεται σημαντικά (από 22 σε 14% και από 14 σε 10%). Άξιο αναφοράς είναι το γεγονός ότι όταν οι διακυμάνσεις είναι άνισες το ποσοστό των σφαλμάτων είναι μικρότερο του 10%. Σημαντική είναι η μείωση που επιτυγχάνεται μεταξύ άνισης στην ίσης διακύμανσης

Τα αναλυτικά αποτελέσματα (μέσο ποσοστό εσφαλμένων ταξινομήσεων στο δείγμα ελέγχου) για όλους τους συνδυασμούς των εξεταζομένων παραγόντων παρουσιάζονται στον πίνακα 3.11

Συνδυασμός παραγόντων	Ποσοστό σφάλματος	Συνδυασμός παραγόντων	Ποσοστό σφάλματος
w=1,x=1,y=1,z=1	30.05%	w=2,x=1,y=1,z=1	30.52%
w=1,x=1,y=1,z=2	16.35%	w=2,x=1,y=1,z=2	14.62%
w=1,x=1,y=2,z=1	21.97%	w=2,x=1,y=2,z=1	21.32%
w=1,x=1,y=2,z=2	6.97%	w=2,x=1,y=2,z=2	4.25%
w=1,x=1,y=3,z=1	15.45%	w=2,x=1,y=3,z=1	18.72%
w=1,x=1,y=3,z=2	2.47%	w=2,x=1,y=3,z=2	0.55%
w=1,x=2,y=1,z=1	29.62%	w=2,x=2,y=1,z=1	26.37%
w=1,x=2,y=1,z=2	13.12%	w=2,x=2,y=1,z=2	11.25%
w=1,x=2,y=2,z=1	21.45%	w=2,x=2,y=2,z=1	20.85%
w=1,x=2,y=2,z=2	5.52%	w=2,x=2,y=2,z=2	4.30%
w=1,x=2,y=3,z=1	17.22%	w=2,x=2,y=3,z=1	19.05%
w=1,x=2,y=3,z=2	2.15%	w=2,x=2,y=3,z=2	0.82%
w=3,x=1,y=1,z=1	29.67%	w=4,x=1,y=1,z=1	31.22%
w=3,x=1,y=1,z=2	16.77%	w=4,x=1,y=1,z=2	15.62%
w=3,x=1,y=2,z=1	24.17%	w=4,x=1,y=2,z=1	23.70%
w=3,x=1,y=2,z=2	7.30%	w=4,x=1,y=2,z=2	10.40%
w=3,x=1,y=3,z=1	18.30%	w=4,x=1,y=3,z=1	18.97%
w=3,x=1,y=3,z=2	3.92%	w=4,x=1,y=3,z=2	6.67%
w=3,x=2,y=1,z=1	29.10%	w=4,x=2,y=1,z=1	27.85%
w=3,x=2,y=1,z=2	14.02%	w=4,x=2,y=1,z=2	16.77%
w=3,x=2,y=2,z=1	19.70%	w=4,x=2,y=2,z=1	21.80%
w=3,x=2,y=2,z=2	6.52%	w=4,x=2,y=2,z=2	10.05%
w=3,x=2,y=3,z=1	14.15%	w=4,x=2,y=3,z=1	20.75%
w=3,x=2,y=3,z=2	4.12%	w=4,x=2,y=3,z=2	6.92%

Πίνακας 3.11: Ποσοστό σφαλμάτων για όλους τους συνδυασμούς των εξεταζομένων παραγόντων

Ξεκινώντας την ανάλυση του πίνακα 3.11 από τις κατανομές παρατηρούμε ότι τα μικρότερα ποσοστά σφαλμάτων επιτυγχάνονται στους συνδυασμούς της ομοιόμορφης κατανομής με εξαίρεση 4 περιπτώσεις που τα μικρότερα

ποσοστά παρατηρούνται στη λογιστική κατανομή. Η παρατήρηση του πίνακα σε σχέση με το πλήθος των χαρακτηριστικών οδηγεί στα ίδια συμπεράσματα με αυτά του πίνακα 3.10. Επίσης όταν οι διακυμάνσεις είναι άνισες το ποσοστό περιορίζεται σε σχέση με την περίπτωση των ίσων διακυμάνσεων. Ειδικά όταν σε κάποιο συνδυασμό υπάρχουν ιδιότητες που δίνουν χαμηλά αποτελέσματα, τότε η μείωση είναι ακόμα μεγαλύτερη. Άξιο προσοχής είναι οι συνδυασμοί με 15 χαρακτηριστικά ($y=3$) και άνισες διακυμάνσεις. Τα ποσοστά των σφαλμάτων είναι τα μικρότερα. Στους συνδυασμούς αυτούς, όταν η κατανομή είναι η ομοιόμορφη τότε το ποσοστό των σφαλμάτων είναι μικρότερο του 1%. Στην περίπτωση που σε αυτούς τους συνδυασμούς η κατανομή είναι η κανονική με αποσπώμενα σημεία (η οποία δίνει το μεγαλύτερο ποσοστό σφάλματος από όλες τις κατανομές) το ποσοστό των σφαλμάτων είναι γύρω στο 6.5 %. Συνοπτικά η μέθοδος δεν παρουσιάζει κανένα μη αναμενόμενο σφάλμα, όπως εκείνα των προηγούμενων μεθόδων. Χαρακτηριστικό είναι ότι η μεγάλη βελτίωση των ποσοστών των σφαλμάτων που επιτυγχάνει η μέθοδος.

Στον πίνακα 3.12 παρουσιάζεται η σύγκριση των αποτελεσμάτων μεταξύ των δειγμάτων εκπαίδευσης και ελέγχου.

Από τη σύγκριση του ποσοστού των λαθών στα δείγματα εκπαίδευσης και ελέγχου σύμφωνα με τον πίνακα 3.12 παρατηρείται μια μεγάλη διαφορά σε όλες τις περιπτώσεις με εξαίρεση όταν τα δεδομένα έχουν 5 χαρακτηριστικά και οι διακυμάνσεις είναι ίσες. Στην περίπτωση των 5 χαρακτηριστικών η διαφορά είναι περίπου 1% και 3% όταν οι κατανομή είναι κανονική ή ομοιόμορφη. Στις υπόλοιπες περιπτώσεις το ποσοστό κατά την εκπαίδευση βελτιώνεται κατά πολύ και στην περίπτωση των 15 χαρακτηριστικών είναι

μηδενικό. Τέλος από την παρατήρηση των δεδομένων φαίνεται ότι η μέθοδος εκπαιδεύεται σωστά και διατηρεί την αποτελεσματικότητα της στο δείγμα ελέγχου.

	Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου		Σφάλμα εκπαίδευσης	Σφάλμα ελέγχου
w=3,x=1,y=1,z=1	28.23%	30.05%	w=4,x=1,y=1,z=1	27.85%	30.52%
w=3,x=1,y=1,z=2	13.43%	16.35%	w=4,x=1,y=1,z=2	6.55%	14.62%
w=3,x=1,y=2,z=1	5.18%	21.97%	w=4,x=1,y=2,z=1	3.60%	21.32%
w=3,x=1,y=2,z=2	0.08%	6.97%	w=4,x=1,y=2,z=2	0.04%	4.25%
w=3,x=1,y=3,z=1	0.00%	15.45%	w=4,x=1,y=3,z=1	0.00%	18.72%
w=3,x=1,y=3,z=2	0.00%	2.47%	w=4,x=1,y=3,z=2	0.00%	0.55%
w=3,x=2,y=1,z=1	28.62%	29.62%	w=4,x=2,y=1,z=1	25.96%	26.37%
w=3,x=2,y=1,z=2	13.09%	13.12%	w=4,x=2,y=1,z=2	6.98%	11.25%
w=3,x=2,y=2,z=1	10.86%	21.45%	w=4,x=2,y=2,z=1	8.54%	20.85%
w=3,x=2,y=2,z=2	3.41%	5.52%	w=4,x=2,y=2,z=2	0.09%	4.30%
w=3,x=2,y=3,z=1	0.02%	17.22%	w=4,x=2,y=3,z=1	0.00%	19.05%
w=3,x=2,y=3,z=2	0.00%	2.15%	w=4,x=2,y=3,z=2	0.00%	0.82%
w=3,x=1,y=1,z=1	28.04%	29.67%	w=4,x=1,y=1,z=1	30.56%	31.22%
w=3,x=1,y=1,z=2	14.69%	16.77%	w=4,x=1,y=1,z=2	15.64%	15.62%
w=3,x=1,y=2,z=1	7.46%	24.17%	w=4,x=1,y=2,z=1	4.67%	23.70%
w=3,x=1,y=2,z=2	0.32%	7.30%	w=4,x=1,y=2,z=2	0.14%	10.40%
w=3,x=1,y=3,z=1	0.01%	18.30%	w=4,x=1,y=3,z=1	0.00%	18.97%
w=3,x=1,y=3,z=2	0.00%	3.92%	w=4,x=1,y=3,z=2	0.00%	6.67%
w=3,x=2,y=1,z=1	29.15%	29.10%	w=4,x=2,y=1,z=1	28.57%	27.85%
w=3,x=2,y=1,z=2	14.92%	14.02%	w=4,x=2,y=1,z=2	15.93%	16.77%
w=3,x=2,y=2,z=1	10.62%	19.70%	w=4,x=2,y=2,z=1	8.145%	21.80%
w=3,x=2,y=2,z=2	10.59%	6.52%	w=4,x=2,y=2,z=2	2.17%	10.05%
w=3,x=2,y=3,z=1	0.08%	14.15%	w=4,x=2,y=3,z=1	0.01%	20.75%
w=3,x=2,y=3,z=2	0.00%	4.12%	w=4,x=2,y=3,z=2	0.00%	6.92%

Πίνακας 3.12: Ποσοστά σφαλμάτων δείγματος εκπαίδευσης και ελέγχου

3.3.5. Σύγκριση μεθόδων SVM

Μεγάλο ενδιαφέρον παρουσιάζει η σύγκριση των τεσσάρων παραπάνω μεθόδων L-LSVM, L-PSVM, NL-LSVM και NL-PSVM. Οι συγκρίσεις θα πραγματοποιηθούν σε δύο φάσεις. Κατά την πρώτη φάση θα γίνει σύγκριση των γραμμικών μεθόδων και στη συνέχεια θα γίνει σύγκριση των μη γραμμικών μεθόδων.

3.3.5.1. Σύγκριση γραμμικών μεθόδων SVM

Η σύγκριση των γραμμικών μεθόδων L-LSVM και L-PSVM θα γίνει από τους πίνακες των ιδιοτήτων και τους πίνακες των συνδυασμών αυτών. Έτσι ο παρακάτω πίνακας 3.13 παρουσιάζει τα ποσοστά των σφαλμάτων για κάθε παράγοντα.

Παράγοντες	Σφάλμα ελέγχου L-LSVM	Σφάλμα ελέγχου L-PSVM
w=1	23.42%	23.21%
w=2	23.43%	23.30%
w=3	22.70%	22.51%
w=4	24.15%	24.01%
x=1	23.95%	23.79%
x=2	22.90%	22.72%
y=1	29.31%	29.27%
y=2	22.84%	22.67%
y=3	18.13%	17.83%
z=1	23.89%	23.88%
z=2	22.96%	22.64%

Πίνακας 3.13: Συγκριτικός πίνακας μεθόδων L-LSVM και L-PSVM

Από τον πίνακα 3.13 παρατηρείται ότι η μέθοδος L-PSVM παρουσιάζει καλύτερα αποτελέσματα από την μέθοδο L-LSVM. Η διαφορά όμως των αποτελεσμάτων είναι πολύ μικρή και δεν υπερβαίνει το 0,2%. Έτσι μπορούμε να θεωρήσουμε ότι οι μέθοδοι δίνουν τα ίδια περίπου αποτελέσματα.

	Σφάλμα Ελέγχου L-LSVM	Σφάλμα Ελέγχου L-PSVM		Σφάλμα Ελέγχου L-LSVM	Σφάλμα Ελέγχου L-PSVM
w=1,x=1,y=1,z=1	29.65%	29.65%	w=2,x=1,y=1,z=1	30.47%	30.47%
w=1,x=1,y=1,z=2	30.72%	30.62%	w=2,x=1,y=1,z=2	31.07%	31.17%
w=1,x=1,y=2,z=1	22.87%	22.85%	w=2,x=1,y=2,z=1	22.67%	22.67%
w=1,x=1,y=2,z=2	23.02%	22.60%	w=2,x=1,y=2,z=2	25.12%	24.97%
w=1,x=1,y=3,z=1	18.17%	18.12%	w=2,x=1,y=3,z=1	20.00%	20.00%
w=1,x=1,y=3,z=2	19.15%	18.40%	w=2,x=1,y=3,z=2	13.70%	13.15%
w=1,x=2,y=1,z=1	30.57%	30.57%	w=2,x=2,y=1,z=1	27.15%	27.15%
w=1,x=2,y=1,z=2	26.12%	26.05%	w=2,x=2,y=1,z=2	30.17%	30.02%
w=1,x=2,y=2,z=1	23.05%	23.05%	w=2,x=2,y=2,z=1	22.40%	22.40%
w=1,x=2,y=2,z=2	20.20%	19.75%	w=2,x=2,y=2,z=2	21.77%	21.40%
w=1,x=2,y=3,z=1	19.32%	23.32%	w=2,x=2,y=3,z=1	20.52%	20.45%
w=1,x=2,y=3,z=2	18.27%	17.62%	w=2,x=2,y=3,z=2	16.12%	15.77%
w=3,x=1,y=1,z=1	30.22%	30.20%	w=4,x=1,y=1,z=1	31.25%	31.30%
w=3,x=1,y=1,z=2	29.95%	29.87%	w=4,x=1,y=1,z=2	27.35%	27.30%
w=3,x=1,y=2,z=1	24.22%	24.27%	w=4,x=1,y=2,z=1	25.22%	25.17%
w=3,x=1,y=2,z=2	22.00%	21.67%	w=4,x=1,y=2,z=2	23.47%	23.30%
w=3,x=1,y=3,z=1	19.30%	19.27%	w=4,x=1,y=3,z=1	17.57%	17.52%
w=3,x=1,y=3,z=2	17.35%	16.77%	w=4,x=1,y=3,z=2	20.32%	19.75%
w=3,x=2,y=1,z=1	29.22%	29.22%	w=4,x=2,y=1,z=1	28.92%	28.97%
w=3,x=2,y=1,z=2	25.62%	25.55%	w=4,x=2,y=1,z=2	30.47%	30.17%
w=3,x=2,y=2,z=1	21.22%	21.15%	w=4,x=2,y=2,z=1	23.25%	23.22%
w=3,x=2,y=2,z=2	22.20%	21.85%	w=4,x=2,y=2,z=2	22.75%	22.50%
w=3,x=2,y=3,z=1	16.07%	16.07%	w=4,x=2,y=3,z=1	20.05%	20.05%
w=3,x=2,y=3,z=2	15.05%	14.27%	w=4,x=2,y=3,z=2	19.20%	18.85%

Πίνακας 3.14 Σφάλματα L-LSVM και L-PSVM για όλους τους συνδυασμούς των παραγόντων

Η παρατήρηση του πίνακα 3.14 επιβεβαιώνει το παραπάνω συμπέρασμα. Δηλαδή, ότι οι μέθοδοι είναι περίπου εξίσου αποτελεσματικές και έχουν την ίδια συμπεριφορά. Οι διαφορές του πίνακα 3.14 δείχνουν καλύτερη τη μέθοδο L-PSVM, αλλά η διαφορά είναι της τάξης του 0,05% κατά μέσο όρο και σε μερικές μόνο περιπτώσεις η διαφορά να φτάνει το 0,6%. Στην περίπτωση του συνδυασμού κανονικής κατανομής ($w=1$), πλήθος αντικειμένων 200 ($x=2$), πλήθος χαρακτηριστικών 15 ($y=3$) και ίσες διακυμάνσεις ($z=1$) που η διαφορά είναι στο 4%.

3.3.5.2. Σύγκριση μη γραμμικών μεθόδων SVM

Στον πίνακα 3.15 παρουσιάζονται τα ποσοστά των σφαλμάτων των παραγόντων για τις μεθόδους NL-LSVM και NL-PSVM. Από τον πίνακα γίνεται φανερό ότι η μέθοδος NL-PSVM ταξινομεί τα δεδομένα καλύτερα από την NL-LSVM. Η διαφορά στην ομοιόμορφη κατανομή (που δίνει τα καλύτερα αποτελέσματα και στις δύο μεθόδους) είναι στο 6,5%, ενώ στην κανονική κατανομή με απεσπασμένα σημεία (που δίνει τα χειρότερα αποτελέσματα και τις δυο μεθόδους) η διαφορά είναι στο 5%. Στις υπόλοιπες δύο μεθόδους (κανονική και λογιστική) η διαφορά κυμαίνεται στο 6% και 5,5% αντίστοιχα. Στην περιπτώσεις 5 και 10 χαρακτηριστικών οι διαφορές κυμαίνονται στο 2% και 6%. Για ίσες διακυμάνσεις η διαφορά είναι μικρότερη (0.62%). Τέλος για 15 χαρακτηριστικά και άνισες διακυμάνσεις η διαφορά είναι στο 9% και στο 11%. Από τα παραπάνω βλέπουμε ότι η μέθοδος NL-PSVM διαφέρει από την NL-LSVM γύρω στις 6 ποσοστιαίες μονάδες, με εξαίρεση τους παράγοντες ($y=3$ και $z=2$) που δίνουν πολύ καλά

αποτελέσματα όπου η μέθοδος NL-PSVM μικραίνει το λάθος πάρα πολύ. Στο τελευταίο θα μπορούσε να πει κανείς ότι στηρίζεται η ανωτερότητα της NL-PSVM.

Παράγοντες	Σφάλμα ελέγχου NL-LSVM	Σφάλμα ελέγχου NL-PSVM
w=1	21.35%	15.19%
w=2	20.96%	14.38%
w=3	21.17%	15.64%
w=4	22.51%	17.56%
x=1	22.54%	16.25%
x=2	20.46%	15.14%
y=1	24.38%	22.06%
y=2	20.45%	14.39%
y=3	19.67%	10.64%
z=1	23.60%	23.00%
z=2	19.39%	8.39%

Πίνακας 3.15: Συγκριτικός πίνακας μεθόδων NL-LSVM και NL-PSVM

Η παρατήρηση του πίνακα 3.16 επιβεβαιώνει το τελευταίο συμπέρασμα, ότι δηλαδή η ανωτερότητα της μεθόδου NL-PSVM βασίζεται στο γεγονός ότι μειώνει το ποσοστό των εσφαλμένων ταξινομήσεων όσο αυξάνεται το πλήθος των χαρακτηριστικών. Το ίδιο ισχύει και για τις άνισες διακυμάνσεις, που η μείωση από τις ίσες είναι 11%. Οι διαφορές των ποσοστών του πίνακα 3.16 δείχνουν ότι στις περιπτώσεις (y=1, z=2), (y=2, z=2) και (y=3, z=2) η διαφορές είναι στις 4 και 10 και 15 ποσοστιαίες μονάδες περίπου. Στις υπόλοιπες περιπτώσεις οι διαφορές είναι μικρότερες της μιας ποσοστιαίας μονάδας. Επίσης η μέθοδος NL-PSVM δεν παρουσιάζει μη-αναμενόμενα σφάλματα σε σχέση με την μέθοδο NL-LSVM που παρουσιάζει παρά πολλά (βλ παράγραφο 3.3.3). Τέλος σε ορισμένες περιπτώσεις (π.χ.

$w=4, x=1, y=1, z=1$) η μέθοδος NL-LSVM παρουσιάζει καλύτερα αποτελέσματα. Αυτό συμβαίνει πάντα στην περίπτωση που οι διακυμάνσεις είναι ίσες. Υπενθυμίζεται ότι η διαφορά των δύο μεθόδων στις ίσες διακυμάνσεις είναι στις 0,6 ποσοστιαίες μονάδες, γεγονός που δικαιολογεί το παραπάνω φαινόμενο.

	Σφάλμα Ελέγχου NL- LSVM	Σφάλμα Ελέγχου NL- PSVM		Σφάλμα Ελέγχου NL- LSVM	Σφάλμα Ελέγχου NL- PSVM
$w=1, x=1, y=1, z=1$	30.57%	30.05%	$w=2, x=1, y=1, z=1$	30.42%	30.52%
$w=1, x=1, y=1, z=2$	20.20%	16.35%	$w=2, x=1, y=1, z=2$	18.35%	14.62%
$w=1, x=1, y=2, z=1$	23.07%	21.97%	$w=2, x=1, y=2, z=1$	23.15%	21.32%
$w=1, x=1, y=2, z=2$	19.07%	6.97%	$w=2, x=1, y=2, z=2$	18.82%	4.25%
$w=1, x=1, y=3, z=1$	16.07%	15.45%	$w=2, x=1, y=3, z=1$	21.30%	18.72%
$w=1, x=1, y=3, z=2$	23.00%	2.47%	$w=2, x=1, y=3, z=2$	20.77%	0.55%
$w=1, x=2, y=1, z=1$	29.97%	29.62%	$w=2, x=2, y=1, z=1$	26.15%	26.37%
$w=1, x=2, y=1, z=2$	18.02%	13.12%	$w=2, x=2, y=1, z=2$	16.22%	11.25%
$w=1, x=2, y=2, z=1$	21.90%	21.45%	$w=2, x=2, y=2, z=1$	21.25%	20.85%
$w=1, x=2, y=2, z=2$	16.20%	5.52%	$w=2, x=2, y=2, z=2$	15.80%	4.30%
$w=1, x=2, y=3, z=1$	18.00%	17.22%	$w=2, x=2, y=3, z=1$	20.47%	19.05%
$w=1, x=2, y=3, z=2$	20.15%	2.15%	$w=2, x=2, y=3, z=2$	18.87%	0.82%
$w=3, x=1, y=1, z=1$	30.22%	29.67%	$w=4, x=1, y=1, z=1$	32.15%	31.22%
$w=3, x=1, y=1, z=2$	21.20%	16.77%	$w=4, x=1, y=1, z=2$	20.57%	15.62%
$w=3, x=1, y=2, z=1$	24.80%	24.17%	$w=4, x=1, y=2, z=1$	25.07%	23.70%
$w=3, x=1, y=2, z=2$	18.15%	7.30%	$w=4, x=1, y=2, z=2$	20.55%	10.40%
$w=3, x=1, y=3, z=1$	19.82%	18.30%	$w=4, x=1, y=3, z=1$	18.62%	18.97%
$w=3, x=1, y=3, z=2$	22.20%	3.92%	$w=4, x=1, y=3, z=2$	22.90%	6.67%
$w=3, x=2, y=1, z=1$	28.85%	29.10%	$w=4, x=2, y=1, z=1$	28.52%	27.85%
$w=3, x=2, y=1, z=2$	17.97%	14.02%	$w=4, x=2, y=1, z=2$	20.70%	16.77%
$w=3, x=2, y=2, z=1$	20.00%	19.70%	$w=4, x=2, y=2, z=1$	22.35%	21.80%
$w=3, x=2, y=2, z=2$	17.52%	6.52%	$w=4, x=2, y=2, z=2$	19.55%	10.05%
$w=3, x=2, y=3, z=1$	14.12%	14.15%	$w=4, x=2, y=3, z=1$	19.70%	20.75%
$w=3, x=2, y=3, z=2$	19.25%	4.12%	$w=4, x=2, y=3, z=2$	19.50%	6.92%

Πίνακας 3.16 Σφάλματα NL-LSVM και NL-PSVM για όλους τους συνδυασμούς των παραγόντων

Μια σύγκριση μεταξύ γραμμικών και μη γραμμικών μεθόδων δείχνει ότι σε σχέση με τις ιδιότητες (βλ πίνακες 3.13 και 3.15) οι μη γραμμικές δίνουν καλύτερα αποτελέσματα. Στην περίπτωση των ίσων διακυμάνσεων οι γραμμικές μέθοδοι παρουσιάζουν περίπου το ίδιο ποσοστό σφάλματος με τις μη γραμμικές. Μάλιστα οι διαφορές είναι μικρότερη του 1%. Τα ποσοστά σφαλμάτων στις γραμμικές μεθόδους είναι μεγαλύτερα του 23% και μόνο στην περίπτωση 15 χαρακτηριστικών δίνουν 18%. Στην περίπτωση των μη γραμμικών μοντέλων τα ποσοστά είναι γύρω στο 21,5% για την μέθοδο NL-LSVM, ενώ για την μέθοδο NL-PSVM τα ποσοστά των σφαλμάτων είναι μικρότερα. Ενδιαφέρον παρουσιάζει το γεγονός ότι η μέθοδος NL-LSVM αν και δίνει καλύτερα αποτελέσματα από τις γραμμικές μεθόδους, η διαφορά όμως δεν είναι πολύ μεγάλη. Χαρακτηριστικό είναι ότι από τους πίνακες 3.14 και 3.16, στις περιπτώσεις ομοιόμορφης κατανομής ($w=2$) και όταν τα χαρακτηριστικά είναι 10 και 15 ($y=2,3$), οι γραμμικές μέθοδοι δίνουν καλύτερα αποτελέσματα από την NL-LSVM. Σε καμία περίπτωση όμως οι γραμμικές μέθοδοι δεν δίνουν καλύτερα αποτελέσματα από εκείνα της μεθόδου NL-PSVM.

3.3.6. Σύγκριση SVM με εναλλακτικές μεθόδους

Η σύγκριση όλων των μεθόδων που παρουσιάζεται παρακάτω γίνεται σε σχέση με τις ιδιότητες ξεχωριστά, το συνολικό ποσοστό λάθους, συνδυασμός μεταξύ κατανομών και χαρακτηριστικών, συνδυασμός μεταξύ κατανομών και διακυμάνσεων και τέλος συνδυασμός μεταξύ κατανομών, χαρακτηριστικών και διακυμάνσεων. Η σύγκριση παρουσιάζεται με αυτόν τον τρόπο για να

γίνει καλύτερη παρουσίαση των αποτελεσμάτων καθώς επίσης και να γίνει όσο το δυνατόν πιο λεπτομερής σύγκριση.

Κατάταξη	Μέθοδοι	Σφάλματα
1	NLP	15.69%
2	ANN	19.35%
3	LDA	20.69%
4	NLL	21.50%
5	CART	22.33%
6	LP	23.26%
7	LL	23.43%
8	NN	25.79%

Πίνακας 3.17: Κατάταξη των μεθόδων

Ο πίνακας 3.17 είναι ο μέσος όρος των ποσοστών των λαθών κάθε μεθόδου και η κατάταξη τους. Από τον πίνακα αυτό παρατηρούμε ότι η μέθοδος NLP είναι η καλύτερη μέθοδος, έπειτα ακολουθεί η μέθοδος των νευρωνικών δικτύων, έπειτα η μέθοδος LDA, έπειτα είναι η μέθοδος NLL, μετά είναι η μέθοδος CART και τελευταίες είναι οι γραμμικές μέθοδοι με πρώτη την γραμμική μορφή της μεθόδου Proximal. Τέλος χαρακτηριστικό είναι, ότι η μέθοδος NLP έχει ένα ποσοστό λάθους 16% ενώ οι υπόλοιπες μέθοδοι ξεπερνάνε το 19%.

	w=1	w=2	w=3	w=4	x=1	x=2	y=1	y=2	y=3	z=1	z=2
NLL	21.4%	21.0%	21.2%	22.5%	22.5%	20.5%	24.4%	20.5%	19.7%	23.6%	19.4%
LL	23.4%	23.4%	22.7%	24.2%	24.0%	22.9%	29.3%	22.7%	18.1%	23.9%	23.0%
NLP	15.2%	14.4%	15.6%	17.6%	16.3%	15.1%	22.1%	14.4%	10.6%	23.0%	8.4%
LP	23.2%	23.3%	22.5%	24.0%	23.8%	22.7%	29.3%	22.7%	17.8%	23.9%	22.6%
ANN	19.1%	17.4%	19.7%	21.2%	21.1%	17.6%	24.0%	18.7%	15.4%	25.9%	12.8%
NN	25.9%	24.7%	25.6%	27.1%	26.8%	24.8%	28.8%	24.9%	23.8%	33.0%	18.6%
LDA	20.4%	20.6%	20.4%	21.4%	21.3%	20.1%	27.4%	20.1%	14.6%	21.9%	19.5%
CART	25.1%	11.0%	26.4%	26.9%	23.3%	21.4%	24.3%	21.8%	20.9%	31.2%	13.5%

Πίνακας 3.18.: Ποσοστό σφαλμάτων για όλες τις μεθόδους ανά παράγοντα

Στον πίνακα 3.18 παρουσιάζονται τα ποσοστά των λαθών σε σχέση με τις ιδιότητες των δεδομένων. Με έντονα παρουσιάζονται τα καλύτερα αποτελέσματα ανα ιδιότητα. Χαρακτηριστικό είναι ότι η μέθοδος CART δίνει πολύ μικρό ποσοστό σφαλμάτων όταν τα δεδομένα ακολουθούν την ομοιόμορφη κατανομή. Στις υπόλοιπες ιδιότητες η μέθοδος NLP δίνει καλύτερα αποτελέσματα με εξαίρεση την ιδιότητα των ίσων διακυμάνσεων που καλύτερη είναι η μέθοδος LDA. Για 5, 10 και 15 χαρακτηριστικά οι καλύτερες μέθοδοι είναι: NLP, ANN, LDA και NLL. Καλύτερη μέθοδος είναι πάντα η NLP, ενώ οι υπόλοιπες μέθοδοι καταλαμβάνουν τις υπόλοιπες θέσεις, χωρίς να παρατηρείται κάποια συγκεκριμένη κατάταξη. Ωστόσο παρουσιάζονται κάποια αξιοσημείωτα φαινόμενα Στην περίπτωση των 5 χαρακτηριστικών η μέθοδος CART έρχεται τρίτη. Ενώ στην περίπτωση των 15 χαρακτηριστικών η μέθοδος LP έρχεται τέταρτη και έπειτα η NLL. Τέλος για ίσες και άνισες διακυμάνσεις οι κατατάξεις είναι LDA, NLP, NLL, LP, LL, ANN, CART, NN και NLP, ANN, CART, NN, NLL, LDA, LP, LL αντίστοιχα. Στις παραπάνω κατατάξεις χαρακτηριστικό είναι ότι οι μέθοδοι NLP, ANN και LDA είναι σχεδόν πάντα οι 3 πρώτες, ενώ οι γραμμικές μέθοδοι (LP και LL) των SVM καθώς και η μέθοδος NN είναι πάντα τελευταίες. Από τις γραμμικές μέθοδοι καλύτερη είναι πάντα η LP. Ωστόσο η διαφορά είναι πολύ μικρή και σε ορισμένες περιπτώσεις μηδενική. Υπάρχουν βέβαια περιπτώσεις που η κατάταξη αυτή αλλάζει. Για παράδειγμα στην ομοιόμορφη κατανομή η μέθοδος CART είναι η καλύτερη. Ακόμα όταν τα χαρακτηριστικά είναι 15 και οι διακυμάνσεις είναι ίσες τότε βλέπουμε ότι οι γραμμικές μέθοδοι βελτιώνονται συγκριτικά με τις υπόλοιπες μεθόδους. Χαρακτηριστικό είναι ότι η μέθοδος CART δεν παρουσιάζει κάποια σταθερότητα, αλλά η κατάταξη της αλλάζει ανάλογα με την ιδιότητα.

	NLL	LL	NLP	LP	ANN	NN	LDA	CART
w=1,y=1	24.69%	29.27%	22.29%	29.23%	24.54%	29.56%	27.51%	26.40%
w=1,y=2	20.06%	22.29%	13.98%	22.06%	18.17%	24.54%	19.31%	24.44%
w=1,y=3	19.31%	18.73%	9.33%	18.36%	14.68%	23.52%	14.44%	24.39%
w=2,y=1	22.79%	29.72%	20.69%	29.11%	20.49%	25.18%	27.33%	15.24%
w=2,y=2	19.76%	22.99%	12.68%	23.02%	16.74%	24.11%	20.08%	10.11%
w=2,y=3	20.36%	17.59%	9.79%	17.59%	15.05%	24.72%	14.24%	7.56%
w=3,y=1	24.56%	28.76%	22.39%	27.73%	25.16%	29.58%	27.29%	27.76%
w=3,y=2	20.12%	22.41%	14.43%	22.56%	18.98%	24.45%	20.02%	25.96%
w=3,y=3	18.85%	16.94%	10.13%	16.91%	14.91%	22.72%	13.89%	25.54%
w=4,y=1	25.49%	29.50%	22.87%	28.19%	25.68%	30.74%	27.50%	27.98%
w=4,y=2	21.88%	23.68%	16.49%	23.84%	20.80%	26.35%	21.05%	26.56%
w=4,y=3	20.18%	19.29%	13.33%	19.31%	17.06%	24.08%	15.66%	26.03%

Πίνακας 3.19: Ποσοστό σφαλμάτων σε συνδυασμό με τις κατανομές και τα χαρακτηριστικά

Στον πίνακα 3.19 παρουσιάζονται τα αποτελέσματα σε σχέση με τον συνδυασμό των κατανομών και του πλήθους των χαρακτηριστικών. Συνολικά τα καλύτερα αποτελέσματα προέρχονται από την μέθοδο NLP, με εξαίρεση όταν η κατανομή είναι ομοιόμορφη που τα καλύτερα αποτελέσματα προέρχονται από την μέθοδο CART. Οι υπόλοιπες μέθοδοι (CART, LP, LL ,NN) καταλαμβάνουν τις υπόλοιπες 4 θέσεις. Η κατάταξη τους αλλάζει ανάλογα τη κατανομή και το πλήθος των χαρακτηριστικών. Έτσι στην περίπτωση κανονικής και ομοιόμορφης κατανομής η μέθοδος CART είναι η καλύτερη από τις 3 μεθόδους LP, LL ,NN. Ενώ στην περίπτωση λογιστικής και κανονικής με απεσπασμένα σημεία καλύτερες είναι οι μέθοδοι LP και LL. Μάλιστα στις περιπτώσεις 15 χαρακτηριστικών, οι μέθοδοι LP και LL έχουν την 4η και 5η κατάταξη. Δηλαδή στις περιπτώσεις 15 χαρακτηριστικών οι μέθοδοι παρουσιάζουν πολύ καλή συμπεριφορά, αν συγκριθεί με το γεγονός ότι σε όλες τις άλλες περιπτώσεις οι μέθοδοι είναι τελευταίες. Στον πίνακα 3.19 τα έντονα νούμερα δείχνουν ποια μέθοδος δίνει τα καλύτερα αποτελέσματα σε σχέση με τους συνδυασμούς.

	NLL	LL	NLP	LP	ANN	NN	LDA	CART
w=1,z=1	23.79%	23.94%	22.63%	23.92%	25.64%	32.75%	21.45%	34.91%
w=1,z=2	18.14%	22.92%	7.77%	22.51%	12.61%	19.00%	19.39%	15.24%
w=2,z=1	23.82%	24.02%	22.71%	24.00%	25.09%	33.21%	21.80%	19.32%
w=2,z=2	18.14%	23.00%	5.97%	22.75%	9.05%	15.68%	19.37%	2.10%
w=3,z=1	22.85%	23.19%	21.92%	23.17%	25.32%	31.65%	21.58%	33.43%
w=3,z=2	19.38%	22.03%	8.78%	21.67%	13.46%	19.07%	18.93%	18.87%
w=4,z=1	24.14%	24.10%	23.23%	24.11%	25.67%	32.78%	22.17%	35.18%
w=4,z=2	20.63%	23.93%	11.08%	23.65%	16.03%	20.58%	20.33%	17.60%

Πίνακας 3.20: Ποσοστό σφαλμάτων σε συνδυασμό με τις κατανομές και τις διακυμάνσεις.

Στον πίνακα 3.20 παρουσιάζονται τα αποτελέσματα της ταξινόμησης σε σχέση με τη κατανομή των δεδομένων και των διακυμάνσεων των κατηγοριών. Από τον πίνακα φαίνεται για τις κατανομές κανονική, λογιστική και κανονική με αποσπώμενα στοιχεία, όταν οι διακυμάνσεις είναι ίσες η κατάταξη είναι LDA, NLP, NLL, LP, LL, ANN, NN, CART. Όταν οι διακυμάνσεις είναι άνισες τότε η κατάταξη είναι NLP, ANN, CART, NLL, LDA, NN, LP, LL. Παρατηρείται ότι οι μέθοδοι NLP, NLL, NN και στις δυο περιπτώσεις δεν έχουν μεγάλες διακυμάνσεις στην κατάταξη τους. Συγκεκριμένα λαμβάνουν τις θέσεις 1-2, 3-4, 7-6 αντίστοιχα. Οι άλλες μέθοδοι μεταβάλλουν την κατάταξη τους δραματικά. Για παράδειγμα η ANN από 6 (όταν οι διακυμάνσεις είναι ίσες) στην 2 θέση (όταν οι διακυμάνσεις είναι άνισες), η CART από 8 (όταν οι διακυμάνσεις είναι ίσες) στην 3 θέση (όταν οι διακυμάνσεις είναι άνισες). Αξιοσημείωτο είναι ότι και εδώ, οι μέθοδοι LP και LL όταν οι διακυμάνσεις είναι άνισες είναι τελευταίες, ενώ όταν είναι ίσες τότε η μέθοδοι έχουν την 4^η και 5^η κατάταξη. Τέλος όταν η κατανομή είναι η ομοιόμορφη τότε η μέθοδος CART είναι η καλύτερη. Αξιοσημείωτο είναι το γεγονός ότι όταν η κατανομή είναι ομοιόμορφη και οι διακυμάνσεις είναι άνισες τότε οι μέθοδοι CART, NLP και ANN έχουν παρά

πολύ μικρό ποσοστό σφαλμάτων σε σχέση με τις άλλες μεθόδους. Με έντονα παρουσιάζονται οι μέθοδοι που παρουσιάζουν τα καλύτερα αποτελέσματα.

	NLL	LL	NLP	LP	ANN	NN	LDA	CART
y=1,z=1	29.61%	29.68%	29.30%	29.69%	30.90%	38.22%	29.38%	33.58%
y=1,z=2	19.16%	28.94%	14.82%	28.85%	17.04%	19.31%	25.43%	15.11%
y=2,z=1	22.56%	22.97%	21.74%	22.96%	25.48%	32.63%	21.11%	30.60%
y=2,z=2	18.21%	22.57%	6.92%	22.26%	11.71%	16.89%	18.99%	12.74%
y=3,z=1	18.40%	18.76%	17.72%	18.73%	21.10%	27.80%	14.93%	29.06%
y=3,z=2	20.83%	17.40%	3.46%	16.83%	9.62%	19.54%	14.10%	12.52%

Πίνακας 3.21: Ποσοστό σφαλμάτων σε συνδυασμό με το πλήθος των χαρακτηριστικών και τις διακυμάνσεις

Ο πίνακας 3.21 παρουσιάζει το συνδυασμό των ιδιοτήτων μεταξύ του πλήθους των χαρακτηριστικών και των διακυμάνσεων των κατηγοριών. Στον πίνακα αυτόν παρατηρείται ότι όταν οι διακυμάνσεις είναι ίσες τότε για κάθε πλήθος χαρακτηριστικών ισχύει η ακόλουθη σειρά: LDA, NLP, NLL, LL, LP, ANN, CART, NN. Χαρακτηριστικό είναι ότι το ποσοστό σφάλματος των ζευγαριών LDA-NLP και LP-LL είναι περίπου ίδιο. Στην περίπτωση που οι διακυμάνσεις είναι άνισες τότε οι κατάταξη αλλάζει για κάθε πλήθος χαρακτηριστικών. Δηλαδή για πλήθος χαρακτηριστικών 5, η σειρά είναι: NLP, CART, ANN, NLL, NN, LDA, LP, LL. Για 10 χαρακτηριστικά: NLP, ANN, CART, NN, NLL, LDA, LP, LL και τέλος για 15 χαρακτηριστικά: NLP, ANN, CART, LDA, LP, LL, NN, NLL. Παρατηρούμε ότι την πρώτη θέση παίρνει η μέθοδος NLP και την δεύτερη και τρίτη θέση συμπληρώνουν οι μέθοδοι CART και ANN. Για τις τελευταίες 5 θέσεις όταν τα χαρακτηριστικά είναι 5 και 10 η σειρά είναι NLL, NN, LDA, LP, LL. Για 15 χαρακτηριστικά όπως έχει παρατηρηθεί και παραπάνω, οι μέθοδοι LP και LL

είναι καλύτερες και παίρνουν τις θέσεις 5^η και 6^η. Με έντονα παρουσιάζονται οι καλύτερες μέθοδος

	NLL	LL	NLP	LP	ANN	NN	LDA	CART
w=1,y=1,z=1	30,27%	31,43%	29,83%	31,43%	31,62%	38,90%	29,88%	30,28%
w=1,y=1,z=2	19,11%	28,62%	14,73%	28,62%	17,45%	20,21%	25,12%	11,51%
w=1,y=2,z=1	22,48%	23,33%	21,71%	23,33%	25,26%	32,57%	20,73%	22,30%
w=1,y=2,z=2	17,63%	21,68%	6,25%	21,68%	11,07%	16,50%	17,88%	4,81%
w=1,y=3,z=1	17,03%	18,73%	16,33%	18,73%	20,03%	26,76%	13,72%	15,46%
w=1,y=3,z=2	21,57%	18,68%	2,31%	18,68%	9,31%	20,27%	15,16%	2,10%
w=2,y=1,z=1	28,28%	30,23%	28,45%	30,23%	28,28%	36,46%	28,08%	28,37%
w=2,y=1,z=2	17,28%	30,47%	12,93%	30,47%	12,70%	13,90%	26,56%	8,03%
w=2,y=2,z=1	22,20%	22,85%	21,08%	22,85%	24,48%	33,21%	20,28%	21,51%
w=2,y=2,z=2	17,31%	23,46%	4,27%	23,46%	9,00%	15,01%	19,87%	2,43%
w=2,y=3,z=1	20,88%	20,36%	18,88%	20,36%	24,65%	31,32%	16,81%	19,15%
w=2,y=3,z=2	19,82%	15,28%	0,68%	15,28%	5,45%	18,11%	11,67%	0,55%
w=3,y=1,z=1	29,53%	31,08%	29,38%	31,08%	31,86%	38,41%	29,71%	29,93%
w=3,y=1,z=2	19,58%	28,11%	15,40%	28,11%	18,46%	20,73%	24,86%	13,75%
w=3,y=2,z=1	22,40%	22,75%	21,93%	22,75%	25,90%	32,03%	21,50%	22,51%
w=3,y=2,z=2	17,83%	22,30%	6,91%	22,30%	12,06%	16,86%	18,53%	5,93%
w=3,y=3,z=1	16,97%	17,70%	16,22%	17,70%	19,97%	25,82%	14,38%	16,12%
w=3,y=3,z=2	20,72%	16,40%	4,02%	16,40%	9,85%	19,61%	13,40%	2,93%
w=4,y=1,z=1	30,33%	31,62%	29,53%	31,62%	31,81%	39,08%	29,83%	30,46%
w=4,y=1,z=2	20,63%	29,21%	16,20%	29,21%	19,55%	22,38%	25,16%	14,80%
w=4,y=2,z=1	23,71%	24,26%	22,75%	24,26%	26,91%	33,50%	22,42%	24,48%
w=4,y=2,z=2	20,05%	23,21%	10,22%	23,21%	14,68%	19,20%	19,67%	9,43%
w=4,y=3,z=1	19,16%	19,13%	19,86%	19,13%	22,36%	27,98%	15,17%	17,41%
w=4,y=3,z=2	21,20%	19,55%	6,80%	19,55%	13,86%	20,16%	16,15%	6,80%

Πίνακας 3.22: Ποσοστό σφαλμάτων με όλους τους συνδυασμούς των παραγόντων.

Τέλος στον πίνακα 3.22 παρουσιάζονται οι συνδυασμοί και των τριών παραγόντων (κατανομές, πλήθος χαρακτηριστικών και ομοιομορφία διακυμάνσεων). Παρατηρείται ότι στην κανονική, λογιστική, κανονική με απεσπασμένα σημεία κατανομή η κατάταξη είναι ίδια. Στις περιπτώσεις που οι διακυμάνσεις είναι ίσες και ανάλογα με το πλήθος των χαρακτηριστικών, οι μέθοδοι NLP, LDA, NLL, CART αλλάζουν την κατάταξη τους στις 4

πρώτες θέσεις. Συγκεκριμένα στη περίπτωση των 5 χαρακτηριστικών η κατάταξη είναι NLP, LDA, NLL, CART. Στη περίπτωση των 10 και 15 χαρακτηριστικών οι κατατάξεις είναι LDA, NLP, CART, NLL και LDA, CART, NLP, NLL. Αξιοσημείωτο είναι ότι στις περιπτώσεις 10 και 15 χαρακτηριστικών η LDA είναι η καλύτερη μέθοδος και όχι η NLP. Στις περιπτώσεις άνισων διακυμάνσεων, όταν τα χαρακτηριστικά είναι 5 η κατάταξη είναι CART, NLP, ANN, NLL. Για 10 χαρακτηριστικά η κατάταξη είναι CART, NLP, ANN, NN και για 15 είναι CART, NLP, ANN, LDA. Από τα παραπάνω φαίνεται ότι η μέθοδος NLP ανάλογα τους παράγοντες κυμαίνεται μεταξύ 1, 2 ή 3 θέσης. Η μέθοδος ANN όταν οι διακυμάνσεις είναι ίσες κατατάσσεται στην 7^η θέση πάντα, ενώ όταν είναι άνισες κατατάσσεται στην 3^η θέση. Όμοια για τις μεθόδους LP και LL όταν είναι οι διακυμάνσεις άνισες τότε έχουν την 7^η και 8^η θέση, ενώ όταν είναι ίσες έχουν την 5^η και 6^η θέση. Οι υπόλοιπες μέθοδοι δεν παρουσιάζουν κάποια σταθερότητα, αλλά κατατάσσονται ή στις πρώτες θέσεις ή στις τελευταίες. Με έντονα παρουσιάζονται οι μέθοδοι με τα καλύτερα ποσοστά. Παρατηρούμε ότι η μέθοδος NLP δεν καταλαμβάνει τις πρώτες θέσεις σε πολλές περιπτώσεις. Αντίθετα οι πρώτες θέσεις καταλαμβάνονται από τις μεθόδους LDA και CART. Αυτό μπορεί να εξηγηθεί από τον πίνακα 3.23 στον οποίον παρουσιάζονται οι κατατάξεις των μεθόδων NLP, LDA και CART. Παρατηρούμε ότι η μέθοδος NLP καταλαμβάνει τις θέσεις 2 πρώτες θέσεις και σε λίγες περιπτώσεις την 3^η θέση. Όμως η μέθοδος LDA καταλαμβάνει τις θέσεις 1^η και 6^η και η μέθοδος CART τις θέσεις 1^η και 3^η. Υπάρχουν δηλαδή μεγάλες διακυμάνσεις στην κατάταξη τους.

	NLP	LDA	CART
w=1,y=1,z=1	1	2	3
w=1,y=1,z=2	2	5	1
w=1,y=2,z=1	2	1	3
w=1,y=2,z=2	2	5	1
w=1,y=3,z=1	3	1	2
w=1,y=3,z=2	2	4	1
w=2,y=1,z=1	4	1	3
w=2,y=1,z=2	3	6	1
w=2,y=2,z=1	2	1	3
w=2,y=2,z=2	2	6	1
w=2,y=3,z=1	2	1	3
w=2,y=3,z=2	2	4	1
w=3,y=1,z=1	1	3	4
w=3,y=1,z=2	2	6	1
w=3,y=2,z=1	3	1	2
w=3,y=2,z=2	2	6	1
w=3,y=3,z=1	3	1	2
w=3,y=3,z=2	2	4	1
w=4,y=1,z=1	1	2	4
w=4,y=1,z=2	2	3	1
w=4,y=2,z=1	2	1	5
w=4,y=2,z=2	2	5	1
w=4,y=3,z=1	6	1	2
w=4,y=3,z=2	1	4	2

Πίνακας 3.23: Κατάταξη των μεθόδων NLP, LDA και CART

Συνοπτικά βλέπουμε ότι η μέθοδος NLP στον πίνακα 3.18 παρουσιάζει καλύτερα αποτελέσματα όταν η ανάλυση γίνεται ανα παράγοντα ξεχωριστά. Με μόνες εξαιρέσεις τις μεθόδους CART (παρουσιάζει καλύτερα αποτελέσματα στην ομοιόμορφη κατανομή) και LDA (παρουσιάζει καλύτερα αποτελέσματα στις ίσες διακυμάνσεις των κατηγοριών). Στον συνδυασμό των κατανομών και των χαρακτηριστικών η μέθοδος LDA δίνει καλύτερα αποτελέσματα, με εξαίρεση όταν η κατανομή είναι η ομοιόμορφη. Στην περίπτωση αυτή καλύτερη μέθοδος είναι η μέθοδος CART. Υπενθυμίζεται ότι η μέθοδος CART δίνει καλύτερα αποτελέσματα στην ομοιόμορφη κατανομή

σύμφωνα με τον πίνακα 3.17. Στον συνδυασμό κατανομών και διακυμάνσεων η μέθοδος NLP δίνει καλύτερα αποτελέσματα όταν οι διακυμάνσεις είναι άνισες, άσχετα με την κατανομή. Στην περίπτωση ίσων διακυμάνσεων καλύτερη είναι η μέθοδος LDA. Υπενθυμίζεται ότι η μέθοδος LDA δίνει καλύτερα αποτελέσματα στις ίσες διακυμάνσεις των κατηγοριών σύμφωνα με τον πίνακα 3.17. Στον συνδυασμό χαρακτηριστικών και διακυμάνσεων, η μέθοδος NLP δίνει καλύτερα αποτελέσματα όταν οι διακυμάνσεις των κατηγοριών είναι άνισες ($z=2$). Όταν οι διακυμάνσεις είναι ίσες ($z=1$) καλύτερη είναι η μέθοδος LDA. Τέλος στον συνδυασμό και των τριών παραγόντων (κατανομή, χαρακτηριστικά και διακυμάνσεις) η μέθοδος NLP φαίνεται να μην δίνει πάντα τα καλύτερα αποτελέσματα. Τα αποτελέσματα στις περισσότερες των περιπτώσεων κατατάσσονται μεταξύ των μεθόδων LDA και CART. Συγκεκριμένα όταν οι διακυμάνσεις είναι ίσες να δίνει καλύτερα αποτελέσματα η μέθοδος LDA και όταν οι διακυμάνσεις είναι άνισες η μέθοδος CART να δίνει τα καλύτερα. Εδώ αξίζει να σημειώσουμε ότι οι διαφορές της μεθόδου NLP (όταν δεν είναι η καλύτερη) από την καλύτερη κάθε φορά μέθοδο, κυμαίνεται από 0,2% μέχρι 2%. Επίσης όπως αναφέρθηκε παραπάνω η μέθοδος NLP σύμφωνα με τον πίνακα 3.23 παρουσιάζει μεγαλύτερη σταθερότητα.

ΚΕΦΑΛΑΙΟ 4

Συμπεράσματα και προοπτικές

Απο την παραπάνω ανάλυση είναι φανερό ότι, όταν τα δεδομένα που πρέπει να ταξινομηθούν ακολουθούν συγκεκριμένες ιδιότητες, τότε ανάλογα με το συνδυασμό των ιδιοτήτων θα πρέπει να χρησιμοποιηθούν οι μέθοδοι, όπως φαίνεται στον παρακάτω πίνακα.

Συνδυασμοί ιδιοτήτων	Καλύτερη μέθοδος	Εξαιρέσεις
Μια ιδιότητα ($w=1$ ή $w=2$ ή $w=3$ ή $w=4$ ή $x=1$ κ.τ.λ)	NLP	Όταν $w=2$ CART. Όταν $z=1$ LDA
Συνδιασμός κατανομών (w) και χαρακτηριστικών (y)	LDA	Όταν $w=2$ CART
Συνδιασμός κατανομών (w) και διακυμάνσεων (z)	NLP (όταν $z=2$) LDA (όταν $z=1$)	
Συνδιασμός χαρακτηριστικών (x) και διακυμάνσεων (z)	NLP (όταν $z=2$) LDA (όταν $z=1$)	
Συνδιασμός κατανομών (w) χαρακτηριστικών (x) και διακυμάνσεων (z)	CART (όταν $z=2$) LDA (όταν $z=1$)	Όταν $y=1$ NLP

Πίνακας 4.1: Πίνακας καλύτερων μεθόδων ανά συνδυασμό ιδιοτήτων

Σε κάθε περίπτωση η μέθοδος NLP δίνει ιδιαίτερα ικανοποιητικά αποτελέσματα, στοιχείο το οποίο υποδεικνύει ότι η χρήση μιας υπολογιστικά απλής διαδικασίας εκπαίδευσης των μηχανών διανύσματος υποστήριξης (στη μη γραμμική τους μορφή) είναι σε θέση να αντιμετωπίσει το πρόβλημα της ταξινόμησης με επιτυχία.

Η ανάλυση που πραγματοποιήθηκε στη παρούσα εργασία μπορεί να επεκταθεί σε δεδομένα στα οποία εξετάζονται τόσο ποσοτικά όσο και ποιοτικά δεδομένα. Μεγάλο ενδιαφέρον παρουσιάζει επίσης η επέκταση της

συγκριτικής ανάλυσης των SVM 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 στην πραγματικότητα. Επίσης δεδομένου ότι στη πράξη συχνά απαιτείται η ανάλυση δεδομένων μεγάλων διαστάσεων, είναι χρήσιμο να αναπτυχθούν κατάλληλοι αλγόριθμοι που θα περιορίσουν το χρόνο εκπαίδευσης κατάλληλων μοντέλων ταξινόμησης [27]. Περιοχές με δεδομένα μεγάλου μεγέθους είναι η ανάκτηση πληροφοριών από γραπτά κείμενα, αναγνώριση ομιλίας, αντικειμένων κ.α.

Παραδοσιακά οι ερευνητές σε μηχανές εκμάθησης ψάχνουν να βρουν γενικούς αλγορίθμους (π.χ δέντρα αποφάσεων, νευρωνικά δίκτυα και αλγόριθμοι πλησιέστερου γείτονα) οι οποίοι έχουν πολλά μειονεκτήματα. Στα πρακτικά προβλήματα στα οποία περιέχεται εκτεταμένη γνώση είναι δύσκολο να ενσωματωθεί αυτή η γνώση σε ένα γενικό αλγόριθμο. Είναι συχνό το φαινόμενο στο οποίο μηχανές εκμάθησης με γενικούς αλγορίθμους δεν μπορούν να ερμηνεύσουν την πραγματικότητα. Ένα ακόμα μοντέλο το οποίο καταφέρνει να περιγράψει την πραγματικότητα είναι το στοχαστικό [27].

Τέλος η παραπάνω συγκριτική ανάλυση των SVM μπορεί να επεκταθεί χρησιμοποιώντας και άλλες μεθόδους εκμάθησης εκτός των Lagrangian και Proximal. Παραδείγματα τέτοιων αλγορίθμων είναι τα παρακάτω:

α) Bayesian Trigonometric Support Vector Classifier [28]. Ο αλγόριθμος αυτός χρησιμοποιεί μια τριγωνομετρική συνάρτηση η οποία έχει το επιθυμητό χαρακτηριστικό της φυσικής κανονικοποίησης στη συνάρτηση πιθανότητας

β) Hidden Markov Support Vector Machines [29]. Η αρχιτεκτονική του βασίζεται στις εξαρτήσεις μεταξύ των γειτονικών πεδίων χρησιμοποιώντας την αποκωδικοποίηση Viterbi.

γ) Semismooth Support Vector Machines [30]. Ο αλγόριθμος αυτός μετατρέπει τους περιορισμούς των τετραγωνικών προβλημάτων σε ημιλεία (semismooth) συστήματα εξισώσεων χρησιμοποιώντας τη συνάρτηση Fischer-Burmeister, και εφαρμόζει τη μέθοδο Newton για την επίλυση του προβλήματος.

Βιβλιογραφία

- [1]. P. Kotler, (1994), Marketing management: Analysis, planning, implementation and control , 8th ed., Prentice-Hall, London.
- [2]. B.Mirkin, (1996), Mathematical Classification and Clustering, Kluwer Academic Publishers, Dordrecht.
- [3]. S. Tsumoto (1998), Automated extraction of medical expert system rules from clinical databases based on rough set theory, Information Sciences, 112, 67-84.
- [4]. N.Belacel, (2000), Multicriteria assignment method PROAFTN: Methodology and medical applications, European Journal of Operational Research, 125, 175-183
- [5]. M. Brown, W. Grundy, D. Lin, N. Cristianini, C. Sugnet, T. Furey, M. Ares, and D. Haussler (1999). Knowledge-based analysis of microarray gene expression data using support vector machines. Technical report, University of California in Santa Cruz.
- [6].N. Cristianini, J. Shawe-Taylor (2000), An Introduction to Support Vector Machines and other Kernel-Based learning methods, Cambridge University Press.
- [7].M. Hearst, Support Vector Machines, IEEE Intelligent Systems, July/August 18-19.
- [8]. C. Zopounidis, (1998), Operational Tools in the Management of Financial Risks, Kluwer Academic Publishers, Dordrecht.
- [9]. M. Doumpos. and C. Zopounidis, (1998), “The use of the preference disaggregation analysis in the assessment of financial risks”, Fuzzy Economic Review, 3/1, 39-57.
- [10]. V. N. Vapnik and A. Ja. Chervonenkis (1974), Theory of Pattern Recognition (in Russian), Nauka, Moscow. (German translation: W. N. Wapnik, A. Ja. Tschervonenkis (1974), Theorie der Zeichenerkennung, Akademie, berlin).

- [11]. O. L. Mangasarian and D. R. Musicant (1999). Successive overrelaxation for support vector machines. *IEEE Transactions on Neural Networks*, 10:1032-1037.
- [12]. O. L. Mangasarian, (1994) *Nonlinear Programming*. SIAM, Philadelphia, PA.
- [13]. O. L. Mangasarian and D. R. Musicant (2000), *Lagrangian Support Vector Machine Classification*, Technical Report 00-06, Data Mining Institute, Computer Sciences Dept., University of Wisconsin, Madison.
- [14]. G. Fung and O. L. Mangasarian, (2001), *Proximal Support Vector Machine Classifiers*, in: F. Provost and R. Srikant, *Proceedings KDD-2001: Knowledge Discovery and Data Mining*, Association for Computing Machinery, 77-86
- [15]. Μ. Δούμπος, Κ. Ζοπουνίδης (2001), *Πολυκριτηριες τεχνικές ταξινόμησης: Θεωρία και εφαρμογές*, Εκδόσεις Κλειδάριθμος.
- [16]. V. N. Vapnik, (2000) *The Nature of Statistical Learning Theory*, Second Edition, Springer-Verlag New-York.
- [17]. E. Patuwo, M.Y. Hu, and M. S. Hung, (1993), “Two-group classification using neural networks”, *Decision Sciences*, 24, 825-845.
- [18]. D.E. Rumelhart, G.E. Hinton, and R. J. Williams (1986), *Learning internal representations by error propagation. Parallel distributed processing: Explorations in macrostructure of cognition*, Vol. I, Badford Books, Cambridge, MA., 318-362.
- [19]. Y. LeCun (1986), *Learning processes in an asymmetric threshold network*, *Disordered systems and biological organizations*, les Houches, France, Springer, 233-240.
- [20]. J. E. Dennis, and R. B. Schnabel, (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Englewood Cliffs, NJ: Prentice-Hall.
- [21]. http://www.resample.com/xlminer/help/k-NN/knn_intro

- [22]. W. R. Klecka, (1980), Discriminant Analysis, Series: Quantitative Applications in the Social Sciences, Sage Publications, 7
- [23]. Fisher, R.A. (1936), The use of multiple measurements in taxonomic problems, Annals of Eugenics, 7, 179-188
- [24]. http://www.resample.com/xlminer/help/DA/da_intro.htm
- [25]. L. Breiman, J. H. Friedman, R. A. Olsen, and C.J. Stone, (1984), Classification and regression Trees, Pacific Grove, California.
- [26]. J. R. Quinlan, (1993), C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, Los Altos, California.
- [27]. T.G. Dietterich, Machine Learning Research: Four Current Directions, Department of Computer Science, Oregon State University, Corvallis, OR 97331
- [28]. <http://guppy.mpe.nus.edu.sg/~mpessk/svm.shtml>
- [29]. <http://www.cs.brown.edu/people/th/papers/AltTsoHof-ICML2003.pdf>
- [30]. http://www.optimization-online.org/DB_HTML/2000/11/240.html