Τμήμα Μηχανικών Παραγωγής

& Διοίκησης

Πολυτεχνείο Κρήτης

Μη Παραμετρικές Διαδικασίες

Μονότονης Ταξινόμησης

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

Υπό

Βαλυράκη Αλέξανδρο

2010

Η διατριβή του Βαλυράκη Αλέξανδρου εγκρίνεται από τους κ.κ.

Δούμπο Μιχαήλ – Επίκουρο Καθηγητή

Ζοπουνίδη Κωνσταντίνο – Καθηγητή

Γρηγορούδη Ευάγγελο – Επίκουρο Καθηγητή

Περιεχόμενα

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου στον επιβλέποντα καθηγητή μου κ. Μιχάλη Δούμπο ο οποίος με παρότρυνε να ασχοληθώ με το συγκεκριμένο θέμα και παρείχε πολύτιμη καθοδήγηση κατά την εκπόνηση της μεταπτυχιακής μου διατριβής. Επίσης παρείχε υλικό (δεδομένα και κώδικα ορισμένων αλγορίθμων) σημαντικό για την πρόοδο της εν λόγω εργασίας. Αναμφίβολα, η αρωγή του συνέβαλε τα μέγιστα στην ολοκλήρωση του ΜΠΣ. Ακόμα θα ήθελα να εκφράσω τις ευχαριστίες μου σε όλους τους διδάσκοντες των μαθημάτων του μεταπτυχιακού κύκλου σπουδών μου στο Πολυτεχνείο Κρήτης και στο διοικητικό προσωπικό του ιδρύματος.

Επίσης ευχαριστώ πολύ τους καθηγητές κύριο Ζοπουνίδη Κωνσταντίνο και Γρηγορούδη Ευάγγελο, για την τιμή που μου έκαναν να διαβάσουν την διπλωματική μου εργασία και να συμμετάσχουν στην επιτροπή εξέτασης αυτής.

Τέλος, θεωρώ υποχρέωση μου να ευχαριστήσω τους γονείς μου, που όλα αυτά τα χρόνια μου έχουν προσφέρει τα πάντα, αλλά πάνω από όλα με στηρίζουν ηθικά και ψυχολογικά και με υπομένουν, και για αυτό το σκοπό να τους αφιερώσω την παρακάτω εργασία…

Σύντομο Βιογραφικό Σημείωμα

Ο Αλέξανδρος Βαλυράκης γεννήθηκε το 1981 στα Χανιά. Το 1999 εισήχθη στο Τμήμα Μηχανικών της Σχολής Ικάρων (Σχολή Μηχανικών Αεροπορίας), από όπου και αποφοίτησε το 2003, με εξειδίκευση στις Τηλεπικοινωνίες-Ηλεκτρονικά. Από το 2003 είναι τοποθετημένος στην 115 Πτέρυγα Μάχης (Σούδα) όπου έχει αναλάβει τεχνικά, εκπαιδευτικά, επιτελικά και διοικητικά καθήκοντα. Ταυτόχρονα, το 2004 ξεκίνησε μεταπτυχιακές σπουδές στο Πολυτεχνείο Κρήτης, στον τομέα Τηλεπικοινωνιών του Τμήματος Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών, τις οποίες ολοκλήρωσε το 2007. Από το ίδιο έτος ακολούθησε το μεταπτυχιακό κύκλο σπουδών του Τμήματος Μηχανικών Παραγωγής και Διοίκησης Πολυτεχνείου Κρήτης, στον τομέα Οργάνωσης και Διοίκησης.

ΠΡΟΛΟΓΟΣ

Η εισαγωγή εκ των προτέρων πληροφορίας (prior information) σε τεχνικές μάθησης, αποτελεί συνήθης τακτική βελτίωσης της αποτελεσματικότητάς τους. Ένα τέτοιο είδος πληροφορίας, που εμφανίζεται σε πολλές εφαρμογές, είναι ο περιορισμός της μονοτονίας, ο οποίος συνεπάγεται ότι μια αύξηση σε μια συγκεκριμένη είσοδο δεν μπορεί να οδηγήσει σε μια μείωση στην έξοδο. Η ανάπτυξη μοντέλων που λαμβάνουν υπόψη τη μονοτονία των χαρακτηριστικών που περιγράφουν ένα πρόβλημα βοηθά στη βελτίωση της ικανότητας γενίκευσης των μοντέλων, ενώ ταυτόχρονα συμβάλει στην καλύτερη κατανόηση της λειτουργίας τους από τους αναλυτές και τους αποφασίζοντες. Χαρακτηριστικά παραδείγματα όπου η ανάπτυξη μονότονων μοντέλων λήψης αποφάσεων παίζει σημαντικό ρόλο υπάρχουν σε πεδία όπως τα χρηματοοικονομικά, το μάρκετινγκ, η ιατρική, κά.

Σε ερευνητικό επίπεδο, η ενσωμάτωση της μονοτονίας σε τεχνικές μάθησης για την αναγνώριση και ταξινόμηση προτύπων (pattern classification), έχει ελκύσει το ενδιαφέρον των ερευνητών τα τελευταία έτη. Στόχος της παρούσας ερευνητικής εργασίας είναι η αξιολόγηση της αποτελεσματικότητας διαφόρων μεθοδολογιών ανάπτυξης μοντέλων μονότονης ταξινόμησης. Θα εξεταστούν τεχνικές που βασίζονται σε νευρωνικά δίκτυα, τις μηχανές διανυσμάτων υποστήριξης και την πολυκριτήρια ανάλυση. Η συγκριτική ανάλυση θα βασιστεί τόσο σε πειραματικά δεδομένα (μέσω προσομοίωσης Monte Carlo), όσο και σε ένα ευρύ σύνολο πραγματικών δεδομένων.

ΚΕΦΑΛΑΙΟ 1^{ο} ΕΙΣΑΓΩΓΗ

Στο πρώτο αυτό κεφάλαιο, ο στόχος είναι διπλός. Αφενός, να καθοριστεί με μεγαλύτερη σαφήνεια το αντικείμενο στο οποίο επιδιώκεται να γίνει μελέτη στην παρούσα εργασία, δηλαδή αυτό της ταξινόμησης. Αφετέρου, να γίνει αναφορά σε βασικές κατηγορίες μεθόδων που εξυπηρετούν την αντιμετώπιση προβλημάτων ταξινόμησης. Από τα παραπάνω θα δημιουργηθεί κατάλληλο υπόβαθρο σχετικά με το αντικείμενο της εργασίας. Στο κατά πόσο δηλαδή οι καινοτόμες μέθοδοι που – βασιζόμενες σε προηγούμενες - κάνουν χρήση της ιδιότητας της μονοτονίας, είναι αποδοτικότερες στην επίλυση προβλημάτων ταξινόμησης.

1.1 Η Έννοια της Αναγνώρισης Προτύπων (Pattern Recognition)

Το πρότυπο (ή και μοτίβο, στα αγγλικά pattern, προερχόμενο από τη γνωστή σε όλους γαλλική λέξη patron), αποτελεί ένα σχήμα από επαναλαμβανόμενα γεγονότα ή αντικείμενα, που ορισμένες φορές αναφέρονται ως στοιχεία ενός συνόλου, και τα οποία επαναλαμβάνονται με έναν προβλέψιμο τρόπο. Τα πρότυπα μπορούν να χρησιμοποιηθούν στη δημιουργία αντικειμένων, όταν είναι γνωστό το μοντέλο που ακολουθείται, μπορεί όμως και αντίστροφα, σε περίπτωση που δεν είναι αυτά γνωστά, να γίνει ανίχνευση των υπαρχόντων προτύπων, με στόχο αφενός μεν την εξήγηση του τρόπου – λογικής με την οποία συμβαίνει κάτι (π.χ. να λαμβάνεται κάποια απόφαση) και αφετέρου την κατάλληλη κατάταξη των υπό μελέτη αντικειμένων σε κατηγορίες. Αυτό αποκτά μεγαλύτερη αξία, όπως θα φανεί παρακάτω, ειδικά από τη στιγμή που υπάρχει σε αρκετά από τα υπό μελέτη συστήματα (χρηματοοικονομικά, συστήματα λήψεως αποφάσεων) μια εγγενής – εσωτερική αιτιότητα, η οποία είναι σημαντικό να ερμηνευθεί, κάνοντάς μας να κατανοήσουμε τον τρόπο λειτουργίας κάθε συστήματος. Για παράδειγμα, η επόμενη κίνηση σε μια παρτίδα σκάκι βασίζεται στην παρούσα εικόνα στη σκακιέρα, και η αγορά ή η πώληση μετοχών αποφασίζεται από ένα σύνθετο πρότυπο πληροφοριών.

Η προηγούμενη περίπτωση της ανίχνευσης προτύπων, είναι ουσιαστικά η έννοια της αναγνώρισης προτύπων, η οποία μπορεί να οριστεί ως "η διαδικασία της λήψης ανεπεξέργαστων – καθαρών δεδομένων και έπειτα από επεξεργασία τους, η λήψη κάποιας απόφασης - δράσης με βάση την κατηγορία στην οποία τοποθετούνται ότι ανήκουν αντικείμενα που περιγράφονται από τα δεδομένα αυτά" [13].

Βέβαια αξίζει να σημειωθεί ότι καθημερινά εμφανίζονται πλήθος από προβλήματα αναγνώρισης προτύπων: μυρωδιές, εικόνες, ήχοι, πρόσωπα, καταστάσεις και πολλά άλλα, τα οποία αντιμετωπίζονται με ένα διαισθητικό τρόπο, χωρίς να γίνεται χρήση κάποιου ξεκάθαρου και σαφώς ορισμένου αλγορίθμου. Μόλις όμως καταστεί δυνατός ο προσδιορισμός ενός τέτοιου αλγορίθμου, ταυτόχρονα γίνεται εφικτή και η επίλυση του προβλήματος από έναν απλό υπολογιστή, και έτσι στην πραγματικότητα επιτυγχάνεται η υποκατάσταση του ρόλου του ίδιου του ανθρώπου. Σε αυτό ακριβώς βασίζονται και αρκετές από τις εφαρμογές αναγνώρισης προτύπων που θα παρατεθούν παρακάτω.

1.2 Διαχωρισμός Μεθοδολογιών Αναγνώρισης Προτύπων

Υπάρχουν δυο κύριοι τύποι προβλημάτων αναγνώρισης προτύπων οι εποπτευόμενες και οι μη εποπτευόμενες μέθοδοι.

Στην κατηγορία των μη εποπτευόμενων μεθόδων, το πρόβλημα είναι να ανακαλυφθεί η δομή του συνόλου δεδομένων, εφόσον αυτή υπάρχει. Αυτό συνήθως σημαίνει ότι είναι επιθυμητό να γίνει γνωστό εάν υπάρχουν ομάδες στα δεδομένα, και ποια είναι αυτά τα χαρακτηριστικά τα οποία κάνουν τα αντικείμενα όμοια μέσα στην ομάδα, και τα διαφοροποιούν από τις υπόλοιπες. Σε αυτή την περίπτωση, έχει αναπτυχθεί πλήθος από αλγόριθμους συσταδοποίησης (clustering). Το θετικό και ταυτόχρονα αρνητικό των μεθόδων αυτών είναι ότι δεν υπάρχει κάποιο σημείο αναφοράς, βάσει του οποίου να συγκρίνουμε τα αποτελέσματα των

διαφόρων αλγορίθμων. Η μόνη ένδειξη για το πόσο καλό είναι κάποιο αποτέλεσμα είναι η υποκειμενική εκτίμηση που μας δίνει ο αποφασίζοντας.

Στην περίπτωση των εποπτευόμενων μεθόδων, το κάθε αντικείμενο στο σύνολο δεδομένων έρχεται με μια προκαθορισμένη ετικέτα κλάσης. Στόχος είναι η εκπαίδευση ενός ταξινομητή για να κάνει με σωστό τρόπο την απόδοση ετικετών σε επόμενα αντικείμενα με τα οποία θα τροφοδοτείται. Τις περισσότερες φορές, η διαδικασία την οποία ακολουθεί ο ταξινομητής δε μπορεί να περιγραφεί και να γίνει εύκολα αντιληπτή. Αντικείμενο είναι απλά να δίδονται δυνατότητες μάθησης στον ταξινομητή, και αφού γίνει η εκπαίδευσή του, με τα δεδομένα που ήδη ανήκουν σε κάποια γνωστή κατηγορία, ακολούθως να εκτιμηθεί η ακρίβειά του.

1.3 Βασικές Εφαρμογές Αναγνώρισης Προτύπων

Η αναγνώριση προτύπων έχει μελετηθεί σε πολλούς τομείς, μεταξύ των οποίων είναι η ψυχολογία και η επιστήμη υπολογιστών. Στην ιατρική επιστήμη, η αναγνώριση προτύπων είναι η βάση συστημάτων διάγνωσης βοηθούμενης από υπολογιστές computer-aided diagnosis (CAD). Τα συστήματα περιγράφουν μια διαδικασία η οποία υποστηρίζει τα ευρήματα και τις ερμηνείες του γιατρού. Άλλες τυπικές εφαρμογές είναι η αυτόματη αναγνώριση ομιλίας, η ταξινόμηση του κειμένου σε διάφορες κατηγορίες (π.χ. spam / μη-spam μηνύματα ηλεκτρονικού ταχυδρομείου), η αυτόματη αναγνώριση των χειρόγραφων ταχυδρομικών κωδικών για στους ταχυδρομικούς φακέλους, ή η αυτόματη αναγνώριση των εικόνων ανθρωπίνων προσώπων. Τα τελευταία δύο παραδείγματα αποτελούν υποενότητα της ανάλυσης εικόνας της αναγνώρισης προτύπων που ασχολείται με ψηφιακές εικόνες ως είσοδο για επεξεργασία στα συστήματα αναγνώρισης προτύπων.

1.4 Η έννοια της Ταξινόμησης: Από την παραμετρική στην μη παραμετρική ταξινόμηση.

Η στατιστική ταξινόμηση είναι μια διαδικασία στην οποία ανεξάρτητα αντικείμενα (οι οποίες μπορεί να είναι εναλλακτικές λύσεις ενός προβλήματος) τοποθετούνται σε ομάδες – κατηγορίες (classes) με βάση την ποσοτική πληροφορία που προκύπτει από τη επεξεργασία των χαρακτηριστικών των συγκεκριμένων αντικειμένων. Τυπικά, το πρόβλημα μπορεί να διατυπωθεί ως εξής: εφόσον δοθεί ένα σύνολο εκπαίδευσης $\{(\mathrm{x}_1, y_1),\ldots,(\mathrm{x}_n, y_n)\}$, να παραχθεί ένας ταξινομητής $h: X \to Y$, ο οποίος να αντιστοιχεί κάθε αντικείμενο $x \in X$ στην πραγματική κλάση – κατηγορία που αυτό ανήκει.

Οι αλγόριθμοι ταξινόμησης είναι αυτοί που συνήθως χρησιμοποιούνται στα συστήματα αναγνώρισης προτύπων. Για αυτό και η αναγνώριση προτύπων, ή σε ευρύτερη έννοια η λήψη αποφάσεων, μπορεί να θεωρηθεί ως ένα πρόβλημα εκτίμησης συναρτήσεων πυκνότητας πιθανότητας σε ένα πολυδιάστατο χώρο, και το χωρισμό του έπειτα σε περιοχές κατηγοριών απόφασης ή κλάσεις. Εξαιτίας αυτής της θεώρησης, η στατιστική και θεωρία πιθανοτήτων παίζει σημαντικό ρόλο στον καθορισμό του αντικειμένου, όπως βέβαια και η χρήση πινάκων και διανυσμάτων, για την αναπαράσταση γραμμικών τελεστών και δειγμάτων αντίστοιχα, δηλαδή βασικά εργαλεία γραμμικής άλγεβρας.

Το πρώτο ερώτημα που είναι ενδιαφέρον να απαντηθεί, σχετίζεται με το ποιος είναι ο θεωρητικά καλύτερος ταξινομητής, θεωρώντας όμως ότι είναι γνωστές οι κατανομές των τυχαίων διανυσμάτων. Το συγκεκριμένο πρόβλημα είναι γνωστό ως στατιστικός έλεγχος υποθέσεων (*statistical hypothesis testing),* και ο ταξινομητής *Bayes* είναι αυτός ο οποίος έχει βρεθεί ότι ελαχιστοποιεί το σφάλμα ταξινόμησης. Διάφορες τέτοιες τεχνικές έχουν μελετηθεί στα [1,2,3].

Η παράμετρος κλειδί στην αναγνώριση προτύπων είναι η πιθανότητα σφάλματος. Το σφάλμα που εισάγει ο ταξινομητής Bayes (σφάλμα *Bayes)* δίνει το μικρότερο θεωρητικό σφάλμα που μπορούμε να επιτύχουμε εφόσον είναι γνωστές

οι κατανομές. Υπάρχει επίσης δυνατότητα υπολογισμού άνω ορίου για το συγκεκριμένο σφάλμα.

Βέβαια, αν και ο ταξινομητής Bayes είναι βέλτιστος, η υλοποίησή του είναι συχνά δύσκολη εξαιτίας της πολυπλοκότητάς του, ειδικά όταν ο χώρος του προβλήματος είναι πολλών διαστάσεων. Για αυτό το σκοπό συχνά οδηγούμαστε στη θεώρηση ενός απλούστερου, παραμετρικού ταξινομητή. Οι παραμετρικοί ταξινομητές βασίζονται στη θεώρηση μαθηματικών τύπων είτε για τις συναρτήσεις πυκνότητας πιθανότητας, είτε για τις συναρτήσεις διαχωρισμού. Οι γραμμικοί, τετραγωνικοί ή τμηματικοί (piecewise) ταξινομητές είναι οι απλούστερες και πιο συχνές επιλογές, και έχουν μελετηθεί από την επιστημονική κοινότητα ευρέως [3].

Αν και μπορούν να θεωρηθούν οι μαθηματικοί τύποι, οι τιμές των παραμέτρων πρέπει να εκτιμηθούν από τα διαθέσιμα δείγματα, καθώς δεν είναι γνωστές. Όταν ο αριθμός δειγμάτων είναι πεπερασμένος, οι εκτιμήσεις των παραμέτρων και επακόλουθα των ίδιων των ταξινομητών που βασίζονται σε αυτές τις εκτιμήσεις, γίνονται τυχαίες μεταβλητές. Το σφάλμα ταξινόμησης που προκύπτει επίσης είναι τυχαία μεταβλητή, που χαρακτηρίζεται από μέση τιμή και διασπορά. Έτσι, είναι επίσης σημαντικό να καταλάβουμε πως ο αριθμός των δειγμάτων επηρεάζει την σχεδίαση και την απόδοση του ταξινομητή.

Όταν όμως δε μπορεί να θεωρηθεί κάποια παραμετρική δομή για τις συναρτήσεις πυκνότητας πιθανότητας, πρέπει να χρησιμοποιηθούν μη παραμετρικές τεχνικές, όπως οι προσεγγίσεις *k-nearest neighbor,* για εκτίμηση των συναρτήσεων. Ειδικότερα, οι μη παραμετρικές μέθοδοι αναπτύχθηκαν για να χρησιμοποιούνται στις περιπτώσεις όπου οι ερευνητές δε γνωρίζουν τίποτα για τις παραμέτρους των μεταβλητών που μας ενδιαφέρουν. Σε πιο τεχνικούς όρους, οι μη παραμετρικές μέθοδοι δεν βασίζονται στην εκτίμηση παραμέτρων (όπως είναι η μέση τιμή και η τυπική απόκλιση) για την περιγραφή της κατανομής της μεταβλητής που μας ενδιαφέρει. Για αυτό και οι συγκεκριμένες μέθοδοι καλούνται ελεύθερες από παραμέτρους (*parameter-free*) ή ελεύθερες από κατανομές (*distribution-free*).

Γενικά, οι μη παραμετρικές τεχνικές είναι πολύ ευαίσθητες στον αριθμό των παραμέτρων ελέγχου, και έχουν την τάση να δίνουν αποτελέσματα με μεγάλη απόκλιση, εφόσον οι τιμές των παραμέτρων αυτών δεν είναι προσεκτικά επιλεγμένες.

Σε αυτό το σημείο πρέπει να γίνει αντιληπτός και ο λόγος για τον οποίο οδηγούμαστε στη χρήση μη παραμετρικών μεθόδων στην παρούσα εργασία. Η φύση των προβλημάτων που μας απασχολεί, είναι άλλες φορές τόσο πολύπλοκη και άλλες φορές απλά άγνωστη σε εμάς, ώστε να μην επιτρέπει τη σαφή και με αποτελεσματικό τρόπο μοντελοποίηση – εισαγωγή παραμέτρων, για επίλυση μέσω χρήσης μιας παραμετρικής μεθόδου. Για παράδειγμα, στην περίπτωση πρόβλεψης για πιθανή πτώχευση μιας εταιρίας, η οποία είναι άμεσου ενδιαφέροντος για ένα χρηματοπιστωτικό ίδρυμα, είναι τρομερά δύσκολο, έως και αδύνατο να μοντελοποιήσουμε τους παράγοντες, τόσο της ίδιας της εταιρίας, όσο και εξωτερικούς, που μπορεί να επιδράσουν στη βιωσιμότητά της, ώστε να κάνουμε αξιόπιστη πρόβλεψη και ταξινόμησή σε κάποια κατηγορία.

Έτσι, παρόλο που στατιστικές μέθοδοι όπως η διακριτική ανάλυση και η λογιστική παλινδρόμηση είχαν χρησιμοποιηθεί ευρύτατα στο παρελθόν, τα τελευταία χρόνια έχουν γίνει δημοφιλείς μη παραμετρικές τεχνικές μάθησης, όπως τα νευρωνικά δίκτυα, οι αλγόριθμοι εξαγωγής κανόνων, τα ασαφή συστήματα, τα συστήματα πολυκριτήριας λήψης αποφάσεων, οι μηχανές διανυσμάτων υποστήριξης και ο γραμμικός προγραμματισμός. Υπάρχουν πλέον στη βιβλιογραφία πλήθος εργασιών που καταδεικνύουν ότι οι συγκεκριμένες τεχνικές αποδίδουν καλύτερα από τις κλασσικές στατιστικές μεθόδους, σε όρους ακρίβειας πρόβλεψης του ταξινομητή.

Υπάρχουν πολλά θέματα που σχετίζονται με την επιλογή του καλύτερου ταξινομητή. Παρακάτω, παραθέτουμε τα σημαντικότερα από αυτά.

- Ακρίβεια. Εννοούμε την αξιοπιστία του κανόνα ταξινόμησης, που συνήθως αναπαριστάται από το ποσοστό των σωστών ταξινομήσεων, αν και μπορεί

να ισχύει κάποια σφάλματα να είναι πιο σημαντικά από τα υπόλοιπα, και επομένως να είναι σημαντικό να ελεγχθεί το σφάλμα σε κάποια κλάση «κλειδί».

- Ταχύτητα. Σε κάποιες περιπτώσεις, η ταχύτητα του ταξινομητή είναι πολύ σημαντικό θέμα. Ένας ταξινομητής ο οποίος έχει 90% ακρίβεια, μπορεί να προτιμηθεί από κάποιον άλλο που παρουσιάζει ακρίβεια 95%, εφόσον είναι 100 φορές ταχύτερος στον έλεγχο (και τέτοιες διαφορές σε κλίμακες χρόνου δεν είναι ασυνήθιστες σε περιπτώσεις όπως αυτές των νευρωνικών δικτύων, για παράδειγμα). Τέτοιες παραδοχές θα ήταν σημαντικές σε εφαρμογές όπως αυτές της αυτόματης ανάγνωσης ταχυδρομικών κωδικών, ή την αυτόματη ανίχνευση σφάλματος σε αντικείμενα σε μια γραμμή παραγωγής, στις οποίες ο χρόνος απόκρισης του συστήματος παίζει σημαντικό ρόλο.

- Διαφάνεια. Εφόσον πρόκειται ένας χρήστης να εφαρμόσει τη διαδικασία ταξινόμησης, η διαδικασία θα πρέπει να γίνεται εύκολα κατανοητή, διαφορετικά μπορεί να γίνουν σφάλματα κατά την εφαρμογή του κανόνα. Επίσης, είναι σημαντικό οι χρήστες να πιστεύουν στο σύστημα. Ένα συχνά αναφερόμενο παράδειγμα είναι η περίπτωση του ατυχήματος στο νησί Three-Mile [4], όπου οι αυτόματες συσκευές σωστά συνέστησαν κλείσιμο των πυρηνικών αντιδραστήρων, αλλά δεν εισακούστηκαν, καθώς οι χρήστες δεν θεώρησαν ότι οι συστάσεις ήταν σωστά τεκμηριωμένες. Παρόμοια ιστορία συνέβη και στην καταστροφή του Chernobyl.

- Χρόνος Μάθησης. Ειδικά σε ένα ταχέως μεταβαλλόμενο περιβάλλον, μπορεί να είναι απαραίτητη η συχνή επανεκπαίδευση του κανόνα ταξινόμησης, ή ακόμα και η πραγματοποίηση προσαρμογών σε έναν υπάρχοντα κανόνα, σε πραγματικό χρόνο. Η έννοια «σύντομα» μπορεί επίσης να σημαίνει ότι πρέπει να χρειαζόμαστε μονάχα ένα μικρό αριθμό από παρατηρήσεις (δείγματα) για να εκπαιδεύσουμε τον ταξινομητή μας.

Οι μέθοδοι που στοχεύουμε να συγκρίνουμε λοιπόν, ανήκουν στην κατηγορία αυτών της μη παραμετρικής ταξινόμησης. Η διαδικασία που θα ακολουθηθεί σε όλες τις παρακάτω μεθόδους, παρουσιάζεται παρακάτω.

Αρχικά, στόχος είναι να προσδιοριστεί η μορφή των δεδομένων τα οποία πρόκειται να επεξεργαστούμε. Όπως είναι προφανές, όλοι οι υπό μελέτη αλγόριθμοι θα λαμβάνουν τα ίδια δεδομένα εισόδου. Σε όλες τις περιπτώσεις, βασιζόμενοι σε ένα σύνολο εκπαίδευσης, για το οποίο είναι γνωστά σε ποια κλάση ανήκουν τα αντίστοιχα δεδομένα, θα γίνεται εκπαίδευση ενός ταξινομητή. Έπειτα, βάσει ενός συνόλου ελέγχου, θα μελετάται η απόδοση κάθε αλγορίθμου (χωρίς ωστόσο να γίνεται αναφορά σε χρόνο εκτέλεσης, χρόνο εκπαίδευσης, διαφάνεια).

Όσον αφορά την επιλογή των τεχνικών ταξινόμησης, πέρα από ορισμένες τεχνικές που είναι state of the art, θα γίνει επέκταση σε διαδικασίες ταξινόμησης που χρησιμοποιούν μια σημαντική ιδιότητα, αυτή της μονοτονίας. Έτσι, θα μπορέσει να γίνει αντιληπτό αν και σε ποιο βαθμό η εκμετάλλευση της συγκεκριμένης ιδιότητας βελτιώνει την απόδοση ταξινόμησης ή με απλά λόγια θα φανεί η αξία των τεχνικών μονότονης ταξινόμησης.

Είναι επίσης σημαντικό στην παρούσα φάση να αναφέρουμε ότι τόσο στη θεωρητική ανάλυση των αλγορίθμων που θα ακολουθήσει, όσο και στα μετέπειτα πειράματα που θα γίνουν για να μελετηθεί η απόδοσή τους, θα περιοριστούμε στη μελέτη του προβλήματος ταξινόμησης σε δυο τάξεις, χωρίς σφάλμα γενικότητας. Αυτό συμβαίνει διότι υπάρχουν πλήθος ενδιαφερόντων προβλημάτων που εμπίπτουν σε αυτή την κατηγορία αλλά και επειδή στην πραγματικότητα οποιοδήποτε πρόβλημα ταξινόμησης σε περισσότερες κατηγορίες μπορεί να αποδειχθεί ότι επαγωγικά ανάγεται σε μια σειρά από προβλήματα δυαδικής ταξινόμησης [10].

1.5 Δομή Εργασίας

Με βάση τα παραπάνω, περιγράφεται η δομή της εργασίας στα κεφάλαια που ακολουθούν

- $2^{ο}$ Κεφάλαιο. Στόχος του είναι η περιγραφή βασικών μεθόδων ταξινόμησης. Ειδικότερα, θα γίνει αναφορά στα νευρωνικά δίκτυα, τις μηχανές διανυσμάτων υποστήριξης και την πολυκριτήρια μέθοδο UTADIS.

- $3^{ο}$ Κεφάλαιο. Περιγράφεται η ιδιότητα της μονοτονίας καθώς και λόγοι που μας οδηγούν στην εκμετάλλευσή της για τροποποίηση υπαρχόντων αλγορίθμων. Ακολούθως, παρατίθονται οι αλγόριθμοι νευρωνικών δικτύων και μηχανών διανυσμάτων υποστήριξης που κάνουν χρήση της μονοτονίας.

- $4^{ο}$ Κεφάλαιο. Αφού προσδιοριστεί ο τρόπος δημιουργίας των τεχνητών δεδομένων, καθώς και χαρακτηριστικά των πραγματικών δεδομένων, παρουσιάζονται αποτελέσματα των υπό μελέτη αλγορίθμων, και ακόλουθα συμπεράσματα που προκύπτουν τόσο για την απόδοση του κάθε ενός χωριστά, όσο και από τη μεταξύ τους σύγκριση.

Καταλήγοντας, στα συμπεράσματα κάνουμε κάποιες γενικότερες διαπιστώσεις και αναφέρουμε διάφορα θέματα τα οποία χρήζουν μελλοντικής έρευνας.

ΚΕΦΑΛΑΙΟ 2^{ο} ΑΛΓΟΡΙΘΜΟΙ ΤΑΞΙΝΟΜΗΣΗΣ

Τα τελευταία χρόνια έχει γίνει μεγάλη πρόοδος στο αντικείμενο της ταξινόμησης. Αρκετές μέθοδοι έχουν παρουσιαστεί στη βιβλιογραφία και έχουν υλοποιηθεί σε εφαρμογές, με σπουδαίο αντίκρισμα στην καθημερινότητα. Ειδικότερα στο αντικείμενο της μη παραμετρικής ταξινόμησης, για το οποίο εντοπίζονται συγκεκριμένες κατηγορίες προβλημάτων προς επίλυση, αλγόριθμοι όπως k-nearest neighboor, τα νευρωνικά δίκτυα, τα δένδρα αποφάσεων, οι μηχανές διανυσμάτων υποστήριξης και οι πολυκριτήριες μέθοδοι έχουν διατυπωθεί αναλυτικά στη βιβλιογραφία [3, 11, 13]. Ταυτόχρονα, πλήθος παραλλαγών τους τόσο για την επίτευξη καλύτερης προσαρμογής σε συγκεκριμένα μοντέλα, όσο και για βελτίωση της απόδοσής τους σε τομείς (όπως αναφέρθηκαν προηγουμένως: χρόνος εκτέλεσης, διαφάνεια κτλ.) αποδεικνύουν το μεγάλο ενδιαφέρον της ερευνητικής κοινότητας.

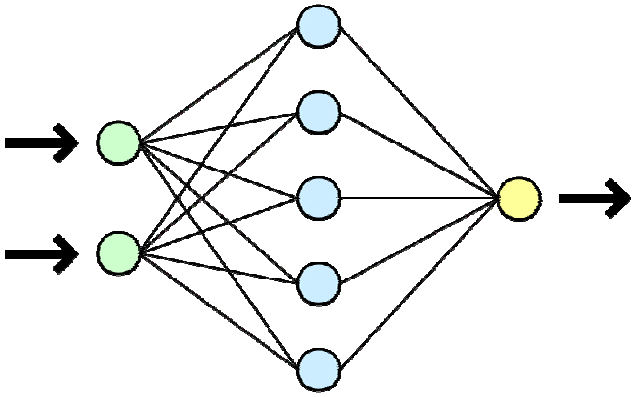
Από τους παραπάνω αλγόριθμους έγινε επιλογή των νευρωνικών δικτύων και των μηχανών διανυσμάτων υποστήριξης για την παρούσα εργασία. Οι συγκεκριμένες τεχνικές πέραν των εγνωσμένων δυνατοτήτων που έχουν, έχουν ένα επιπλέον πλεονέκτημα. Μπορούν να αποτελέσουν τη βάση για την ανάπτυξη αλγορίθμων που θα εκμεταλλεύονται την ιδιότητα της μονοτονίας, και έτσι να εξυπηρετήσουν το αντικείμενο της μελέτης, παρέχοντας δυνατότητα να εξεταστεί το κατά πόσο βελτιώνεται η απόδοση τεχνικών μέσω της χρήσης της εν λόγω ιδιότητας. Πέρα από αυτές, επιλέχθηκε και η πολυκριτήρια μέθοδος UTADIS, η οποία είναι φτιαγμένη ήδη με τρόπο ώστε να θεωρείται εγγενώς μονότονη μέθοδος. Ακολούθως αναλύονται οι τρεις αυτές μέθοδοι, που θα αποτελέσουν βάση για τα περεταίρω πειράματα και σχολιασμό της εργασίας.

2.1 Τεχνητά Νευρωνικά Δίκτυα

2.1.1 Εισαγωγή

Ο όρος τεχνητά νευρωνικά δίκτυα (ΤΝΔ) αναφέρεται σε ένα μαθηματικό μοντέλο αποτελούμενο από ένα μεγάλο αριθμό ανεξάρτητων υπολογιστικών στοιχείων, που ονομάζονται νευρώνες, τα οποία διασυνδέονται μεταξύ τους και είναι οργανωμένα σε στρώματα. Με άλλα λόγια, ένα ΤΝΔ είναι ένας μαζικά παράλληλος κατανεμημένος επεξεργαστής, που έχει την έμφυτη ιδιότητα να αποθηκεύει έμφυτη γνώση και να την έχει διαθέσιμη για χρήση στο μέλλον. Τα ΤΝΔ προσομοιάζουν τον ανθρώπινο εγκέφαλο σε δυο σημεία : (α) η γνώση του ΤΝΔ αποκτάται μέσω μιας διαδικασίας μάθησης, και (β) οι σύνδεσμοι μεταξύ των ΤΝΔ (που ονομάζονται συντελεστές βάρους ή απλά βάρη) χρησιμοποιούνται για να αποθηκευτεί αυτή η γνώση. Η διαδικασία που ακολουθείται για να επιτευχθεί η εκπαίδευση του ΤΝΔ ονομάζεται αλγόριθμος εκπαίδευσης.

Στο ακόλουθο σχήμα φαίνεται η τυπική δομή ενός απλού ΤΝΔ με ένα κρυφό επίπεδο, το οποίο λαμβάνει ένα διάνυσμα 2 εισόδων (δυο κόμβοι εισόδου) και έχει μια έξοδο (ένας κόμβος εξόδου). Το ενδιάμεσο – κρυφό – επίπεδο έχει επιλεγεί να έχει 5 νευρώνες, οι οποίοι συνδέονται με τους κόμβους εισόδου και εξόδου, ωστόσο θα μπορούσε να αποτελούνταν και από περισσότερους ή λιγότερους.



Σχήμα 2.1 Τυπική δομή ΤΝΔ με ένα κρυφό επίπεδο

2.1.2 Κατηγορίες Μεθόδων Εκπαίδευσης

Οι μέθοδοι εκπαίδευσης ΤΝΔ, όπως αναφέραμε και προηγουμένως, μπορούν να χωριστούν σε δυο βασικές κατηγορίες : (α) εποπτευόμενες μεθόδους και (β) μη εποπτευόμενες μεθόδους. Στην πρώτη περίπτωση είναι αναγκαία η παρουσία ενός «δασκάλου», δηλαδή ενός συνόλου εκπαίδευσης, ενώ στην δεύτερη το ΤΝΔ πρέπει να οργανωθεί και να εκπαιδευτεί από μόνο του.

Εκπαίδευση με εποπτεία ονομάζεται η διαδικασία της προσαρμογής ενός συστήματος έτσι ώστε να έχει συγκεκριμένη απόκριση σε συγκεκριμένες εισόδους. Στην περίπτωση των ΤΝΔ η πραγματική έξοδος του ΤΝΔ συγκρίνεται με την επιθυμητή έξοδο και υπολογίζεται η διαφορά τους. Η διαφορά αυτή αποτελεί το σφάλμα εκπαίδευσης του δικτύου. Στη συνέχεια τα βάρη εκπαίδευσης του δικτύου μεταβάλλονται με τέτοιο τρόπο ώστε στην επόμενη επανάληψη η τιμή του σφάλματος να μειωθεί. Για να είναι δυνατή η εκπαίδευση με εποπτεία πρέπει να είναι στην αρχή της εκπαίδευσης διαθέσιμο ένα σύνολο με πρότυπα εκπαίδευσης και για καθένα από αυτά η επιθυμητή απόκριση του δικτύου. Εκεί βρίσκεται και η συμβολή του «δασκάλου» που πρέπει να χαρακτηρίσει όλα τα πρότυπα εισόδου.

Πολλές φορές βέβαια, ο χαρακτηρισμός των προτύπων εισόδου είναι δύσκολος (όταν αυτά είναι πάρα πολλά) ή αδύνατος (όταν προέρχονται από μια άγνωστη διεργασία). Στις περιπτώσεις αυτές είναι δυνατό να χρησιμοποιηθεί η εκπαίδευση χωρίς εποπτεία. Συγκεκριμένα, τα βάρη πλέον του ΤΝΔ μεταβάλλονται μόνο σε σχέση με τις εισόδους. Με τον τρόπο αυτό συνήθων δημιουργούνται ομαδοποιήσεις και το ΤΝΔ μαθαίνει τις συσχετίσεις των δεδομένων εισόδου. Έτσι, μπορεί να επιτευχθεί αναγνώριση προτύπων.

Φυσικά, από όσα έχουμε ως τώρα αναφέρει, στην παρούσα εργασία μας αφορά η περίπτωση των ΤΝΔ με εποπτεία.

2.1.3 Εκπαίδευση ΤΝΔ – Αλγόριθμος Back Propagation

Υπάρχουν διάφοροι τρόποι εκπαίδευσης ενός πολυεπίπεδου εμπρόσθιας τροφοδοσίας νευρωνικού δικτύου. Ο πιο δημοφιλής αλγόριθμος είναι ο error backpropagation [5]. Επειδή ο συγκεκριμένος αλγόριθμος έχει ήδη μελετηθεί διεξοδικά στη βιβλιογραφία, θα δώσουμε μια σύντομη τυπική περιγραφή του τρόπου λειτουργίας του. Ο αλγόριθμος αυτός βασίζεται στην επαναλαμβανόμενη εφαρμογή των ακόλουθων δυο διαδρομών:

1. Εμπρόσθια διαδρομή : Το δίκτυο ενεργοποιείται για μια είσοδο, και υπολογίζεται το σφάλμα ανάμεσα στο δοσμένο στόχο και την πραγματική έξοδο του δικτύου.

2. Διαδρομή προς τα πίσω : Το σφάλμα του δικτύου χρησιμοποιείται για να ανανεωθούν τα βάρη του. Ξεκινώντας από την έξοδο του δικτύου, το σφάλμα διαδίδεται προς τα πίσω, μέσω του δικτύου, από επίπεδο σε επίπεδο. Αυτό γίνεται με τον αναδρομικό υπολογισμό του τοπικού gradient για τον κάθε νευρώνα. Έτσι, εξηγείται η ονομασία του αλγορίθμου.

Αναλόγως του τύπου του προβλήματος πρόβλεψης, το σφάλμα δικτύου υπολογίζεται με διαφορετικούς τρόπους. Ειδικότερα στην ταξινόμηση έχουμε ένα δίκτυο με ένα αριθμό από κόμβους εξόδου που αντιστοιχούν στον αριθμό των κλάσεων – κατηγοριών. Έτσι, όπως έχει αποδείξει ο Bishop [6], η συνάρτηση σφάλματος που επιθυμούμε να ελαχιστοποιήσουμε (γνωστή και ως συνάρτηση cross-entropy) δίδεται από τη σχέση

$$E = -\sum_{n=1}^{N}\sum_{c=1}^{l_{\max}} l_{\mathrm{x}^n}^c \ln\left(\frac{O_{\mathrm{x}^n}^c}{l_{\mathrm{x}^n}^c}\right) \qquad (2.1).$$

Εδώ, ο στόχος $l_{\mathrm{x}^n}^c$ μπορεί να θεωρηθεί ως η πιθανότητα ότι η είσοδος x_n ανήκει στην κλάση l_c. Για να είμαστε πιο ακριβείς, ο στόχος $l_{\mathrm{x}^n}^c$ αναπαρίσταται από ένα δυαδικό διάνυσμα που έχει την τιμή 1 εφόσον $c = l_{\mathrm{x}^n}^c$ και την τιμή μηδέν σε

οποιαδήποτε άλλη περίπτωση. Έτσι, και η έξοδος του δικτύου, $O^c_{\mathrm{x}^n}$, πρέπει να υπολογιστεί ως μια πιθανότητα, η οποία λαμβάνει τιμές στο διάστημα (0, 1). Αυτό επιτυγχάνεται με τη χρήση μιας «softmax» συνάρτησης ενεργοποίησης στο επίπεδο εξόδου, η οποία για την κλάση c ορίζεται ως εξής:

$$O^c_{\mathrm{x}^n} = \frac{e^{\mathrm{w}_c \mathrm{x}^n + \theta_c}}{\sum_{c'=1}^{l_{\max}} e^{\mathrm{w}_{c'} \mathrm{x}^n + \theta_{c'}}} \tag{2.2}$$

Αυτή η συνάρτηση είναι μια παραλλαγή του μοντέλου ενεργοποίησης «ο νικητής τα παίρνει όλα» (winner-takes-all activation model), το οποίο είναι ίσο με τη μονάδα για τη μεγαλύτερη έξοδο, και μηδέν για όλες τις άλλες εξόδους. Έτσι, η συνάρτηση σφάλματος που προκύπτει από τη σχέση (2.1) είναι μη αρνητική και λαμβάνει το ολικό ελάχιστο όταν $O^c_{\mathrm{x}^n} = l^c_{\mathrm{x}^n}$, για όλες τις κατηγορίες c και διανύσματα x^n.

Στο δεύτερο βήμα του αλγορίθμου backpropagation, διαδίδουμε το σφάλμα που υπολογίστηκε προηγουμένως, ώστε να ανανεωθούν τα βάρη και να μειωθεί το σφάλμα. Καθώς και η απόδειξη του κανόνα με βάση τον οποίο τροποποιούνται τα βάρη έχει μελετηθεί διεξοδικά στη βιβλιογραφία, παρουσιάζουμε μόνο τη γενική μορφή του κανόνα, η οποία για το βήμα s είναι

$$w^s = w^{s-1} - \eta \sum_{n=1}^{N} \nabla E \Big|_{w^s} \tag{2.3}$$

όπου $\nabla E \Big|_{w^s}$ είναι οι μερικές παράγωγοι της συνάρτησης σφάλματος ως προς τα βάρη w^s. Ο κανόνας αυτός είναι γνωστός και ως κλαδική μάθηση (batch learning), καθώς τα βάρη τροποποιούνται μετά την ολοκλήρωση κάθε epoch. Σε αντίθεση, στην ακολουθιακή μάθηση, τα βάρη τροποποιούνται μετά την παρουσίαση κάθε διαφορετικής εισόδου.

Η παράμετρος η στον κανόνα τροποποίησης των βαρών ονομάζεται ρυθμός μάθησης (learning rate), και είναι αυτή που καθορίζει το μέγεθος βήματος στη διαδικασία μάθησης. Το να βρεθεί η βέλτιστη τιμή του δεν είναι τετριμμένο. Εάν το η είναι σχετικά μικρό, αναμένουμε να βρούμε το ελάχιστο δυνατό σφάλμα, αλλά με ένα κόστος πολύ αργής μάθησης (μεγάλο υπολογιστικό χρόνο). Εάν το η είναι σχετικά μεγάλο, η διαδικασία μάθησης επιταχύνεται, με το ρίσκο όμως εγκλωβισμού σε ένα τοπικό ελάχιστο. Πλήθος διαδικασιών έχουν αναπτυχθεί για να αντιμετωπιστούν αυτές οι δυσκολίες [6].

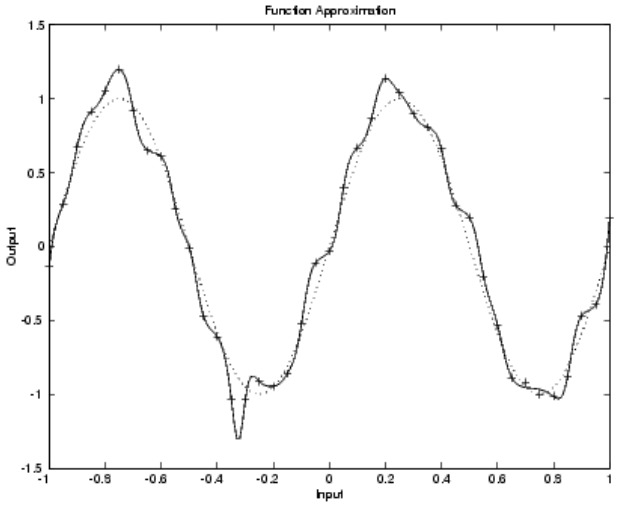
Συνοψίζοντας, ο αλγόριθμος backpropagation στοχεύει στην ελαχιστοποίηση του σφάλματος του δικτύου κάνοντας χρήση της παραγώγου του σφάλματος E, ως προς τα βάρη w^s στο βήμα s.

2.1.4 Τεχνικές Βελτίωσης της Ικανότητας Γενίκευσης ενός ΤΝΔ

Βασικός μας στόχος είναι να φτιάξουμε μοντέλα πρόβλεψης με καλές ικανότητες γενίκευσης (generalization capability), δηλαδή, τα μοντέλα μας να μην είναι αποδοτικά μόνο για τα δεδομένα εκπαίδευσης, αλλά να έχουν αποδεκτή ακρίβεια πρόβλεψης και για νέα δεδομένα. Ένα βασικό πρόβλημα των νευρωνικών δικτύων είναι η τάση τους να προσαρμόζονται υπερβολικά στα δεδομένα, ειδικά για μικρά δείγματα. Αυτό προκαλείται από (i) ένα υπερβολικό αριθμό των παραμέτρων του δικτύου (επίπεδα, κρυμμένοι νευρώνες και συνδέσεις βαρών) και (ii) από πολύ μεγάλα βάρη.

Ουσιαστικά, πρόκειται για υπερβολικό ταίριασμα στα δεδομένα εκπαίδευσης του μοντέλου που δημιουργείται, ώστε ναι μεν το σφάλμα στο σύνολο δεδομένων οδηγείται σε πολύ μικρές τιμές, αλλά όταν παρουσιάζονται νέα δεδομένα στο δίκτυο, το σφάλμα είναι μεγάλο. Το δίκτυο έχει στην πραγματικότητα

προσαρμοστεί με μεγάλη ακρίβεια στα δείγματα εκπαίδευσης, αλλά δεν έχει μάθει να γενικεύει σε νέες καταστάσεις.



Σχήμα 2.2 Προσέγγιση συνάρτησης ημιτόνου με χρήση 1-20-1 ΤΝΔ

Το παραπάνω σχήμα δείχνει την απόκριση ενός νευρωνικού δικτύου 1-20-1 το οποίο έχει εκπαιδευτεί για να προσεγγίζει μια συνάρτηση ημιτόνου, στην οποία έχει προστεθεί θόρυβος. Η καθαρή συνάρτηση ημιτόνου φαίνεται με τη διακεκομμένη γραμμή, ενώ οι πραγματικές μετρήσεις – δείγματα δίδονται από τα σύμβολα `+' , και η απόκριση του δικτύου που εκπαιδεύτηκε δίδεται από τη σκούρα γραμμή. Είναι προφανές ότι το δίκτυο έχει προσαρμοστεί υπερβολικά στα δεδομένα εκπαίδευσης, και δε θα μπορέσει να γενικεύσει σωστά σε επόμενα δείγματα.

Μια μέθοδος για να βελτιωθεί η ικανότητα γενίκευσης του δικτύου είναι να περιοριστεί το μέγεθος του δικτύου. Όσο μεγαλύτερο δίκτυο χρησιμοποιείται τόσο πιο πολύπλοκες συναρτήσεις μπορεί να δημιουργήσει. Αν χρησιμοποιηθεί ένα αρκετά μικρό δίκτυο, δε θα μπορεί να παρουσιαστεί υπερβολική προσαρμογή στα δεδομένα. Δυστυχώς βέβαια, δεν είναι εύκολο να γνωρίζουμε πόσο μεγάλο πρέπει να είναι ένα δίκτυο για μια συγκεκριμένη εφαρμογή.

Μπορούμε ωστόσο να χρησιμοποιήσουμε διάφορες μεθόδους που θα μας επιτρέψουν να οδηγηθούμε στην κατάλληλη επιλογή μοντέλου. Το αντικείμενο της επιλογής μοντέλου είναι να βρεθεί ένα μοντέλο με τις κατάλληλες παραμέτρους δικτύου για το συγκεκριμένο υπό μελέτη πρόβλημα. Μια μέθοδος – παρόμοια με αυτή των δένδρων αποφάσεων – είναι να εφαρμόσουμε μια διαδικασία περιορισμού/κλαδέματος (pruning), δηλαδή να ξεκινήσουμε με ένα μεγάλο δίκτυο και ακολούθως να αφαιρούμε συνδέσεις και νευρώνες κατά τη διαδικασία εκπαίδευσης. Έτσι, το τελικό μοντέλο θα επιλέγεται με βάση το ελάχιστο σφάλμα εκτίμησης που προσδιορίστηκε.

Μια άλλη προσέγγιση, έχει να κάνει απλά με την επιλογή ενός συνόλου δικτύων με διαφορετικό αριθμό παραμέτρων, στα ίδια δεδομένα, και να συγκρίνουμε την απόδοσή τους, οπότε και πάλι επιλέγουμε το μοντέλο που θα έχει το χαμηλότερο σφάλμα πρόβλεψης. Φυσικά, αυτή η διαδικασία μπορεί να είναι υπολογιστικά δύσκολη, και να μην εγγυάται ότι το σύνολο των προεπιλεγμένων δικτύων θα οδηγήσει σε ένα ικανοποιητικό μοντέλο. Μπορεί όμως αυτή η μέθοδος να είναι πιο αποδοτική εφόσον έχουμε κάποια εκ των προτέρων (a priori) γνώση, η οποία να μπορέσει να μας οδηγήσει στο να ρυθμίσουμε τις παραμέτρους των διαφορετικών δικτύων.

Η πιο συνηθισμένη βέβαια μέθοδος για βελτίωση της ικανότητας γενίκευσης του νευρωνικού δικτύου, είναι αυτή της έγκαιρης διακοπής της εκπαίδευσης (γνωστής ως early stopping). Σε αυτή την τεχνική τα διαθέσιμα δεδομένα χωρίζονται σε τρία υποσύνολα. Το πρώτο υποσύνολο είναι το σύνολο εκπαίδευσης και χρησιμοποιείται για τον υπολογισμό της παραγώγου και την τροποποίηση των παραμέτρων του δικτύου. Το δεύτερο σύνολο είναι το σύνολο ελέγχου αξιοπιστίας (validation set). Το σφάλμα στο συγκεκριμένο σύνολο παρατηρείται κατά τη διαδικασία της εκπαίδευσης. Το σφάλμα ελέγχου αξιοπιστίας (validation error) κανονικά μειώνεται κατά την αρχική φάση εκπαίδευσης, όπως άλλωστε συμβαίνει και με το σφάλμα εκπαίδευσης. Ωστόσο, όταν το νευρωνικό δίκτυο αρχίσει να

παρουσιάζει υψηλή προσαρμογή στα δεδομένα εκπαίδευσης, το σφάλμα στο σύνολο ελέγχου αξιοπιστίας συνήθως αυξάνεται. Όταν λοιπόν το σφάλμα αξιοπιστίας συνεχίσει να αυξάνεται για ένα συγκεκριμένο αριθμό από επαναλήψεις, η διαδικασία εκπαίδευσης σταματάει και επιστρέφονται οι τιμές των παραμέτρων για την επανάληψη που είχαμε το ελάχιστο σφάλμα αξιοπιστίας.

Τέλος, το σφάλμα του συνόλου ελέγχου (test set error) δε χρησιμοποιείται κατά την εκπαίδευση, αλλά για να γίνει σύγκριση διαφορετικών μοντέλων. Επίσης είναι χρήσιμο να γίνεται γραφική παράσταση του σφάλματος ελέγχου κατά τη διαδικασία μάθησης. Εφόσον το σφάλμα στο σύνολο ελέγχου φτάσει στην ελάχιστη τιμή του σε σημαντικά διαφορετικό χρονικό σημείο από το σφάλμα συνόλου αξιοπιστίας, τότε υπάρχει ένδειξη για κακό διαχωρισμό του συνόλου δεδομένων.

Οι μέθοδοι Regularization χρησιμοποιούνται για τον περιορισμό των βαρών, ώστε να βελτιώσουν τις ικανότητες γενίκευσης του δικτύου. Όπως είναι γνωστό, μεγάλα βάρη οδηγούν σε απεικονίσεις δικτύων με υψηλές καμπυλότητες, δηλαδή όλες οι παρατηρήσεις στα δεδομένα προσεγγίζονται ακριβώς. Δυο μέθοδοι χρησιμοποιούνται για να αποτρέψουν τα βάρη από τον να μεγαλώσουν τόσο πολύ, ώστε να εξομαλύνουν την έξοδο του δικτύου: stopped training και weight decay.

Η ιδέα του «stopped training» είναι να τερματίσει η διαδικασία εκπαίδευσης πριν επιτευχθεί σύγκλιση. Αυτό επιτυγχάνεται είτε με τη μείωση του αριθμού των epochs είτε με τη χρήση ενός ανεξάρτητου συνόλου ελέγχου για τον υπολογισμό του σφάλματος πρόβλεψης. Η πρώτη προσέγγιση δουλεύει με έναν ad-hoc τρόπο, καθώς δεν είναι τετριμμένο να καθοριστεί ο ακριβής αριθμός των epochs, και για αυτό δεν είναι πολύ εφαρμόσιμη. Η δεύτερη προσέγγιση είναι περισσότερο ρεαλιστική. Σταματάει την εκπαίδευση του δικτύου μόλις το σφάλμα πρόβλεψης στο σύνολο ελέγχου αρχίσει να αυξάνεται.

Στη περίπτωση του «weight decay», η συνάρτηση σφάλματος E που πρέπει να ελαχιστοποιηθεί, τροποποιείται με την πρόσθεση ενός ακόμα όρου για την επιβολή ποινής στα μεγάλα βάρη, δηλαδή

$$\tilde{E} = E + \lambda \sum_{ij} w_{ij}^2$$

όπου λ είναι η παράμετρος regularization. Με άλλα λόγια, η βελτιστοποίηση του δικτύου απαιτεί πλέον την ανάλυση της παραχώρησης ανάμεσα στο καλό ταίριασμα των δεδομένων και στην ομαλή έξοδο του νευρωνικού δικτύου. Ένα πλεονέκτημα της εν λόγω μεθόδου είναι ότι το πρόβλημα βελτιστοποίησης είναι σαφώς ορισμένο. Ένα μειονέκτημά της όμως είναι ότι η επιπλέον παράμετρος λ που εισάγεται πρέπει να καθοριστεί εκ των προτέρων.

Φυσικά, οφείλουμε να σημειώσουμε ότι εάν ο αριθμός των παραμέτρων του δικτύου είναι πολύ μικρότερος συγκριτικά με το συνολικό αριθμό των δεδομένων του συνόλου εκπαίδευσης, τότε περιορίζεται σημαντικά η πιθανότητα υψηλής προσαρμογής στα δεδομένα εκπαίδευσης. Δηλαδή, εάν είναι δυνατό να έχουμε στη διάθεσή μας πολλά δεδομένα, κατά το δυνατόν αντιπροσωπευτικά του συνόλου δεδομένων, τότε δε θα χρειαστεί να εφαρμοστεί κάποια από τις προηγούμενες τεχνικές.

2.2 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)

2.2.1 Εισαγωγή

Μια δεύτερη μεθοδολογία που χρησιμοποιείται ευρέως τα τελευταία χρόνια στην αναγνώριση προτύπων, βασίζεται στη χρήση μηχανών διανυσμάτων υποστήριξης (Support Vector Machines) [6,7]. Η συγκεκριμένη μεθοδολογία έχει ιδιαίτερο ενδιαφέρον καθώς αφενός μεν θεμελιώνεται σε ορισμένες πολύ βασικές και απλές ιδέες, δίδοντας μια πολύ καλή διαίσθηση σχετικά με την έννοια της μάθησης από δείγματα, και αφετέρου, όπως έχει αποδειχθεί, οδηγεί σε πολύ υψηλές αποδόσεις σε αρκετές πρακτικές εφαρμογές. Έτσι, ο συγκεκριμένος

αλγόριθμος μπορεί να θεωρηθεί ότι βρίσκεται –ιδανικά θα έλεγε κανείς – ανάμεσα στη θεωρία μάθησης και την εφαρμογή, συνδυάζοντας και τα δυο επιτυχώς.

Όπως έχει αποδειχθεί στη βιβλιογραφία, οι τεχνικές μηχανών διανυσμάτων υποστήριξης έχουν δυνατότητες να κατασκευάσουν αρκετά πολύπλοκα μοντέλα (όπως νευρωνικά δίκτυα, δίκτυα ακτινικής βάσης, πολυωνυμικούς ταξινομητές) μπορώντας ταυτόχρονα να αναλυθούν μαθηματικά σχετικά εύκολα. Ο λόγος που επιτυγχάνεται αυτό, και αποτελεί το «μυστικό της επιτυχίας» των σχετικών τεχνικών, είναι επειδή γίνεται η χρήση κατάλληλων μετασχηματισμών που επιτρέπουν στα (συνήθως) μη γραμμικά διαχωρίσιμα δεδομένα εισόδου, να μετατρέπονται σε γραμμικά διαχωρίσιμα (εύκολο προς επίλυση πρόβλημα), σε ένα άλλο υψηλότερων διαστάσεων χώρο χαρακτηριστικών (feature space).

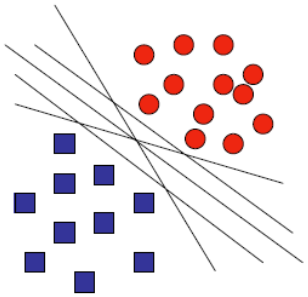
Θα δώσουμε μια συνοπτική περιγραφή του πως δουλεύει ο αλγόριθμος, καθώς αυτό έχει γίνει εκτενώς στη βιβλιογραφία, με βασικό στόχο να επικεντρωθούμε στην εξήγηση των παραμέτρων αυτών που αργότερα παίζουν σημαντικό ρόλο κατά την εκτέλεση των προσομοιώσεων.

2.2.2 Γραμμικές Μηχανές Διανυσμάτων Υποστήριξης

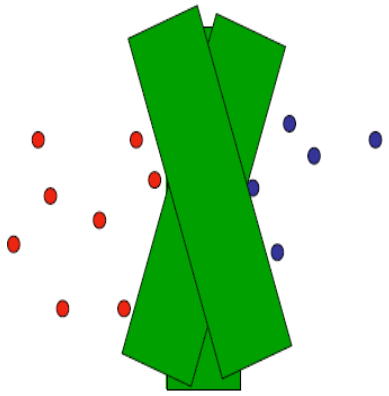
Η απλούστερη εκδοχή των μηχανών διανυσμάτων υποστήριξης προκύπτει όταν στόχος μας είναι η ταξινόμηση γραμμικά διαχωρίσιμων δεδομένων, δεδομένων δηλαδή που μπορούν να διαχωριστούν στις δυο ζητούμενες τάξεις με ένα απλό υπερ-επίπεδο που περιγράφεται από μια εξίσωση της μορφής $\mathrm{x}\cdot\mathrm{w}+b=0$.

Τα δεδομένα εκπαίδευσης είναι πάλι της μορφής $\{\mathrm{x}_i, y_i\},\ i=1,\ldots,n,\ y_i\in\{-1,1\},\ \mathrm{x}_i\in\mathrm{R}^d$. Έστω λοιπόν ότι υπάρχει το υπερεπίπεδο $\mathrm{x}\cdot\mathrm{w}+b=0$ το οποίο μπορεί να τα διαχωρίσει στις δυο τάξεις, και ας θεωρήσουμε $d_+(d_-)$ τις αποστάσεις του κοντινότερου θετικού (αντίστοιχα αρνητικού) δείγματος από το υπερεπίπεδο. Ορίζουμε ως περιθώριο (margin) του υπερεπίπεδου διαχωρισμού το άθροισμα $d_+ + d_-$. Στην απλή περίπτωση των γραμμικά διαχωρίσιμων δεδομένων, στόχος των μηχανών διανυσμάτων υποστήριξης είναι ο

καθορισμός των παραμέτρων w, b για τις οποίες το περιθώριο του σχηματιζόμενου υπερεπίπεδου θα μεγιστοποιείται, και το οποίο υπολογίζεται εύκολα ότι είναι ίσο με $\rho = d_+ + d_- = 2/\|w\|$. Για να γίνει το παραπάνω πιο σαφές, χρήσιμα είναι τα παρακάτω σχήματα.

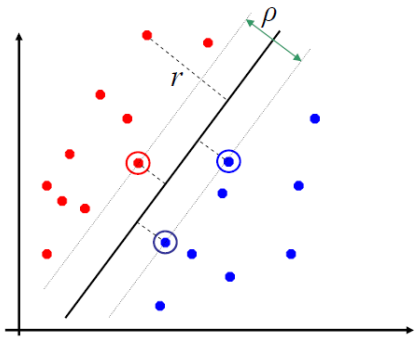


Σχήμα 2.3 Δυνατές επιλογές διαχωρισμού δεδομένων με γραμμές

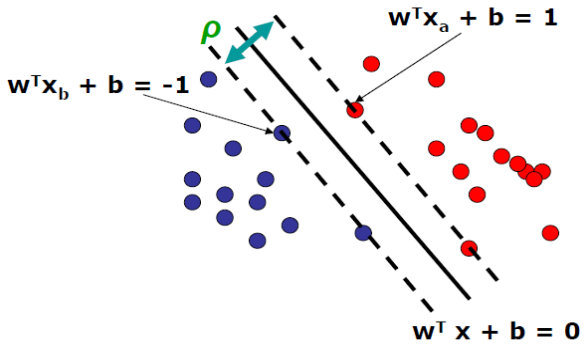


Σχήμα 2.4 Δυνατές επιλογές διαχωρισμού δεδομένων με «παχιές» γραμμές

Στο σχήμα 2.3 φαίνεται ότι για την περίπτωση που θέλουμε να διαχωρίσουμε δεδομένα στο δισδιάστατο χώρο (όπου το υπερεπίπεδο ουσιαστικά είναι γραμμή), οι πιθανές λύσεις είναι πολλές. Αυτό που εμείς επιζητούμε, είναι, όπως φαίνεται στο σχήμα 2.4, να βρούμε τη γραμμή με το μεγαλύτερο περιθώριο (πιο «παχιά») που να τα διαχωρίζει. Διαισθητικά μάλιστα, αυτό εξηγείται και με ένα άλλο τρόπο: εφόσον στοχεύουμε να τοποθετήσουμε ένα «παχύ» διαχωριστή ανάμεσα στις τάξεις, περιορίζονται σημαντικά οι δυνατές επιλογές, και συνακόλουθα η χωρητικότητα του μοντέλου (κάτι που σύμφωνα με τη θεωρία VC dimension είναι επιθυμητό [14]). Επίσης στα σχήματα 2.5 και 2.6 φαίνονται σε δυο παραδείγματα ταξινόμησης τα διανύσματα υποστήριξης καθώς και το πώς αυτά συνεισφέρουν στη διαμόρφωση των προηγουμένων εξισώσεων που αναφέρθηκαν, δικαιολογώντας την ονομασία «μηχανές διανυσμάτων υποστήριξης»



Σχήμα 2.5 Διαχωρισμός δεδομένων και διανύσματα υποστήριξης αποφάσεων



Σχήμα 2.6 Σημασία διανυσμάτων υποστήριξης σε καθορισμό εξισώσεων.

Η επίλυση ενός τέτοιου προβλήματος συνήθως γίνεται με χρησιμοποίηση ενός μοντέλου τετραγωνικού προγραμματισμού [17], αλλά μπορεί να υλοποιηθεί χρησιμοποιώντας διαμορφώσεις γραμμικού προγραμματισμού, μέσω ενός προβλήματος της μορφής που περιγράφεται παρακάτω [15]:

$$
\begin{aligned}
&\min \sum_{j=1}^{d} s_j + C\sum_{i=1}^{n} d_i \\
&\text{υπό περιορισμούς } y_i(\mathrm{x}_i\mathrm{b}\text{-}\gamma) + e_i \geq 1, \quad i = 1,\ldots,n \\
&-s_j \leq b_j \leq s_j, \qquad\qquad j = 1,\ldots,d \\
&\mathrm{b} \in R^d, \gamma \in R \\
&\mathrm{s,e} \geq 0
\end{aligned}
\tag{2.4}
$$

Η αντικειμενική συνάρτηση που χρησιμοποιείται επιτυγχάνει τόσο τη μεγιστοποίηση του περιθωρίου όσο και την ελαχιστοποίηση του σφάλματος εκπαίδευσης. Η σταθερά C χρησιμοποιείται για να ρυθμίσει / αναπαραστήσει την παραχώρηση ανάμεσα σε αυτές τις δυο ανταγωνιστικούς όρους που συνθέτουν την αντικειμενική συνάρτηση.

Αξίζει να επισημάνουμε ότι τα διανύσματα υποστήριξης είναι τα δεδομένα εκείνα του συνόλου εκπαίδευσης τα οποία βρίσκονται κοντύτερα στην επιφάνεια απόφασης. Είναι τα πιο δύσκολα για ταξινόμηση σημεία, και έχουν άμεση επίδραση

στην βέλτιστη τοποθέτηση της επιφάνειας απόφασης. Αν αφαιρεθούν από το σύνολο εκπαίδευσης, τότε θα αλλάξει και η βέλτιστη αυτή τοποθέτηση. Αν μετακινηθεί ένα διάνυσμα υποστήριξης τότε θα αλλάξει και το όριο απόφασης, κάτι το οποίο δεν ισχύει για τα υπόλοιπα δείγματα εκπαίδευσης. Για αυτό και ο αλγόριθμος επίλυσης δουλεύει με τέτοιο τρόπο, ώστε να καθορίζονται μόνο από τα διανύσματα υποστήριξης οι παράμετροι της επιφάνειας απόφασης.

2.2.3 Επέκταση σε Μη Γραμμικά Διαχωρίσιμα Δεδομένα

Το πρόβλημα που επιλύεται με τη βοήθεια γραμμικού προγραμματισμού, μπορεί να διατυπωθεί και ως δευτέρου βαθμού (quadratic) πρόβλημα προγραμματισμού, το οποίο επιλύεται με τη χρήση της μεθόδου πολλαπλασιαστών Lagrange. Στην περίπτωση που τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα, η κεντρική ιδέα είναι να επιτύχουμε με κατάλληλο μετασχηματισμό των δεδομένων επίλυση του ισοδύναμου προβλήματος σε έναν άλλο χώρο δεδομένων *H*, όπου όμως εκεί είναι γραμμικά διαχωρίσιμα. Για το μετασχηματισμό αυτό γίνεται χρήση κατάλληλων συναρτήσεων πυρήνα. Ο μετασχηματισμός αυτός είναι της μορφής $\mathrm{x}_i\mathrm{x}_j^T = \varphi(\mathrm{x}_i)\varphi^T(\mathrm{x}_j)$, όπου η συνάρτηση πυρήνα $K(\mathrm{x}_i\mathrm{x}_j) = \varphi(\mathrm{x}_i)\varphi^T(\mathrm{x}_j)$. Το μοντέλο απόφασης πλέον εκφράζεται σε όρους των μεταβλητών απόφασης $\mathrm{u} \in \mathrm{R}^n$, ως εξής $f(\mathrm{x}) = K(\mathrm{x},\mathrm{X})\mathrm{u}\text{-}\gamma$, όπου X είναι ένας πίνακας με τα δεδομένα εκπαίδευσης. Δημοφιλείς επιλογές της συναρτήσεως Kernel είναι οι παρακάτω:

- Πολυωνυμική $K(\mathrm{x}_i,\mathrm{x}_j) = (\mathrm{x}_i\mathrm{x}_j^T + 1)^d$, όπου $d \geq 2$ είναι ο βαθμός του πολυωνύμου.
- Ακτινικής βάσης $K(\mathrm{x}_i,\mathrm{x}_j) = \exp\left(-\sigma \left\|\mathrm{x}_i\text{-}\mathrm{x}_j\right\|^2\right)$, με σ>0.
- Σιγμοείδής $K(\mathrm{x}_i,\mathrm{x}_j) = \tanh(\alpha\mathrm{x}_i\mathrm{x}_j^T + \theta)$, όπου $\alpha,\theta \in \mathrm{R}$.

Η αναπαράσταση των δεδομένων χρησιμοποιώντας τις συναρτήσεις αυτές τροποποιεί το προηγούμενο πρόβλημα στο ακόλουθο

$$
\begin{aligned}
&\min \sum_{j=1}^{n} v_j + C\sum_{i=1}^{n} d_i \\
&\text{υπό περιορισμούς } y_i\left[K(x_i, \mathrm{X})\text{u-}\gamma\right] + e_i \geq 1, \quad i = 1,\ldots,n \\
&-v_j \leq u_j \leq v_j, \qquad\qquad\qquad\qquad\qquad\quad j = 1,\ldots,n \qquad (2.5) \\
&\mathrm{u} \in R^n, \gamma \in R \\
&\mathrm{v,e} \geq 0
\end{aligned}
$$

2.3 Πολυκριτήρια Μέθοδος UTADIS

2.3.1 Ιστορικό

Μια μέθοδος πολυκριτήριας ανάλυσης αποφάσεων είναι η UTA (Utilities Addictives), η οποία αναπτύχθηκε από τους Jacquet-Lagreze and Siskos ([9]). Χρησιμοποιώντας μια διαδικασία μονότονης παλινδρόμησης αναπτύσσει ένα σύνολο προσθετικών συναρτήσεων χρησιμότητας με σκοπό την κατάταξη ενός συνόλου εναλλακτικών, βάσει μιας δεδομένης προδιάταξης που καθορίζει ο αποφασίζων.

Η μέθοδος UTADIS (Utilités Addictives DIScriminantes) αποτελεί μια παραλλαγή της UTA, η οποία χρησιμοποιείται όταν σκοπός είναι η ταξινόμηση εναλλακτικών σε προκαθορισμένες κατηγορίες και όχι η κατάταξη των εναλλακτικών. Η UTADIS αναπτύσσει ένα υπόδειγμα σύνθεσης των κριτηρίων αξιολόγησης έτσι ώστε το αποτέλεσμα της σύνθεσης αυτής να αποδίδει υψηλή βαθμολόγηση (σκορ) στις εναλλακτικές δραστηριότητες της καλύτερης κατηγορίας και σταδιακά χαμηλότερη βαθμολόγηση στις δραστηριότητες που ανήκουν στις χειρότερες κατηγορίες. Η μέθοδος UTADIS αναπτύχθηκε με σκεπτικό την όσο το δυνατόν μικρότερη ανάγκη για συμμετοχή του αποφασίζοντα στην ανάλυση, χωρίς ταυτόχρονα να θεωρείται ότι ο αποφασίζων έχει μια ουδέτερη συμπεριφορά. Η μέθοδος στηρίζεται στις αρχές της αναλυτικής-συνθετικής προσέγγισης και

εμφανίζεται αρχικά στα άρθρα των Devaud et al. [16], Jacquet-Lagreze and Siskos [9] και αργότερα στα άρθρα των Zopounidis and Doumpos [18, 19 και 20].

2.3.2 Βασικές αρχές της μεθόδου

Ο σκοπός της μεθόδου, όπως έχει ήδη αναφερθεί, είναι η ταξινόμηση των εναλλακτικών σε κατηγορίες. Έστω ένα πλήθος εναλλακτικών x_1, x_2,…,x_n οι οποίες πρέπει να ταξινομηθούν στις προκαθορισμένες κατηγορίες C_1, C_2,…,C_k με βάση τα κριτήρια αξιολόγησης g_1, g_2,…,g_d. Η κατηγορία C_1 είναι εξ' ορισμού η καλύτερη και η C_k η χειρότερη.

Αναπτύσσεται μια προσθετική συνάρτηση χρησιμότητας της μορφής:

$$U(g) = \sum_{i=1}^{d} u_i(g_i)$$

όπου: U(g) είναι η ολική χρησιμότητα (global utility) μιας εναλλακτικής δραστηριότητας x_j και $u_i(g_i)$ είναι η μερική χρησιμότητα (marginal utility) του κριτηρίου g_i.

Η παραπάνω προσθετική συνάρτηση αποδίδει υψηλές βαθμολογίες στις εναλλακτικές που ανήκουν στην κατηγορία C_1 και σταδιακά χαμηλότερες βαθμολογίες στις εναλλακτικές που ανήκουν σε χαμηλότερες κατηγορίες.

Οι μερικές χρησιμότητες αντιπροσωπεύουν τη σχετική σπουδαιότητα των κριτηρίων αξιολόγησης στο υπόδειγμα ταξινόμησης. Οι συναρτήσεις μερικών χρησιμοτήτων είναι μονότονες συναρτήσεις (γραμμικές ή μη γραμμικές) οριζόμενες στην κλίμακα του κάθε κριτηρίου αξιολόγησης έτσι ώστε να ικανοποιούνται οι ακόλουθες δύο συνθήκες:

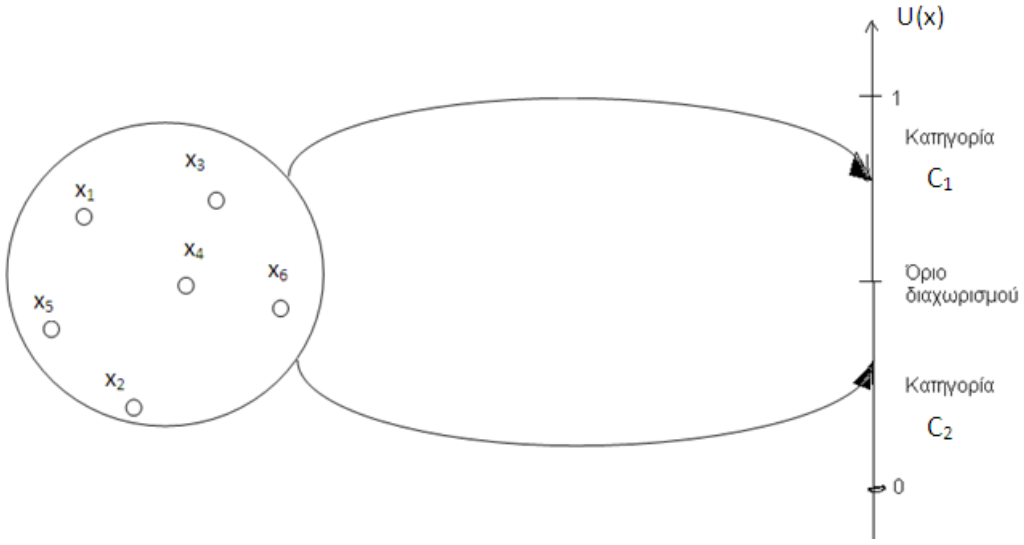
$$\left.\begin{array}{l} u_i(g_{i*}) = 0 \\ u_i(g_i^*) = 1 \end{array}\right\} \forall\, g_i$$

όπου g_{i*} & g_i^* η χείριστη και άριστη επίδοση στο κριτήριο i αντίστοιχα

Η συνάρτηση που δημιουργείται μπορεί να έχει οποιαδήποτε μορφή, χωρίς να ξεφεύγει από το σύνολο [0,1]. Ακόμα πρέπει να είναι μονότονη, δηλαδή να ικανοποιείται ο ακόλουθος περιορισμός:

$$u_i(g_i^{j+1}) - u_i(g_i^{j}) \geq 0, \forall i$$

Στο σχήμα 2.7 απεικονίζεται η ταξινόμηση των εναλλακτικών δραστηριοτήτων στην περίπτωση δύο κατηγοριών, η οποία πραγματοποιείται συγκρίνοντας τις ολικές τους χρησιμότητες με ένα όριο το οποίο διαχωρίζει τις προκαθορισμένες κατηγορίες. Συγκεκριμένα στο παράδειγμα αυτό, δραστηριότητες με ολική χρησιμότητα μεγαλύτερη του ορίου αυτού τοποθετούνται στην κατηγορία C_1, ενώ αντίθετα δραστηριότητες των οποίων η ολική χρησιμότητα είναι μικρότερη από το όριο αυτό εντάσσονται στην κατηγορία C_2.



Σχήμα 2.7 Ταξινόμηση των εναλλακτικών δραστηριοτήτων (πηγή: Doumpos and Zopounidis, 2002 [20])

Στο επόμενο βήμα ορίζονται τα όρια χρησιμότητας, δηλαδή τα όρια τα οποία διαχωρίζουν τις προκαθορισμένες κατηγορίες, u_i ($u_1>u_2>\ldots>u_{k-1}$). Οι κατηγορίες διαχωρίζονται με βάση τη βαθμολογία τους από την προσθετική συνάρτηση χρησιμότητας και τα όρια χρησιμότητας:

$$\begin{cases} U(x_j) \geq u_1 & \Rightarrow x_j \in C_1 \\ u_2 \leq U(x_j) < u_1 & \Rightarrow x_j \in C_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\ U(x_j) < u_{k-1} & \Rightarrow x_j \in C_k \end{cases}$$

όπου u_1, u_2,…,u_{k-1} ορίζονται τα όρια τα οποία διαχωρίζουν τις προκαθορισμένες κατηγορίες (όρια χρησιμότητας). Η ταξινόμηση των εναλλακτικών δραστηριοτήτων επιτυγχάνεται μέσω της σύγκρισης των ολικών χρησιμοτήτων με τα αντίστοιχα όρια χρησιμοτήτων.

Κατά την ταξινόμηση των εναλλακτικών είναι πιθανό να εμφανιστούν δύο είδη σφαλμάτων:

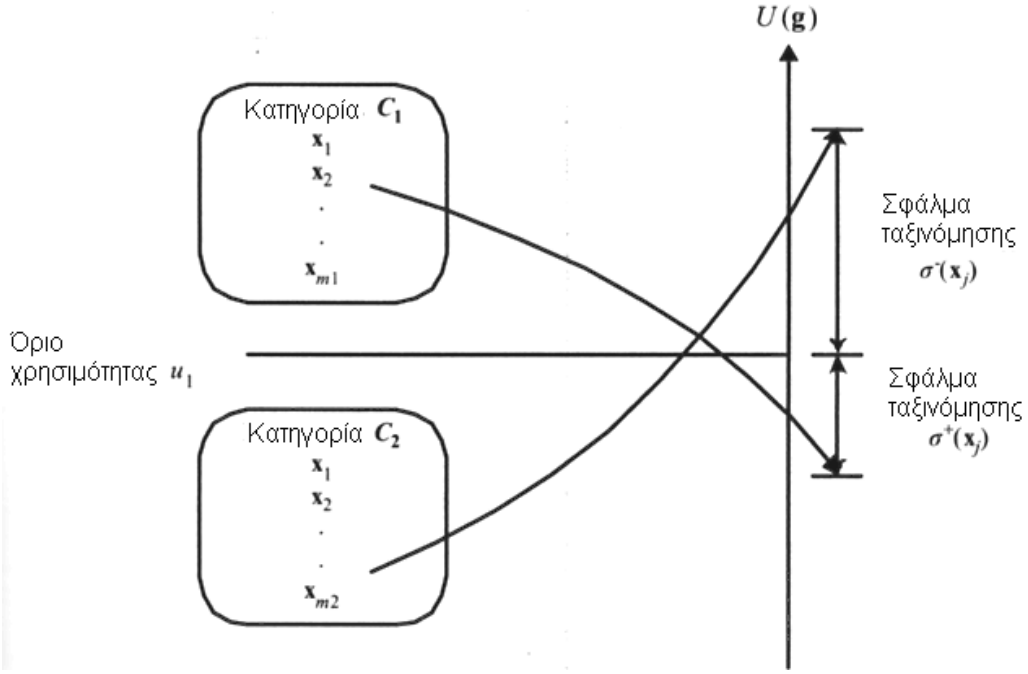
α. Το σφάλμα υπερεκτίμησης $\sigma^+(x)$. Στην περίπτωση αυτή η εναλλακτική με βάση τη βαθμολογία της κατατάχθηκε σε κατηγορία χαμηλότερη από ότι θα έπρεπε. Τότε πρέπει να προστεθεί μια ποσότητα ίση με $\sigma^+(x)$ στη βαθμολογία της, ώστε να ενταχθεί στην κατηγορία που πρέπει. Παρουσιάζεται επομένως, παραβίαση του κάτω ορίου μιας κατηγορίας.

β. Το σφάλμα υποεκτίμησης $\sigma^-(x)$. Ακριβώς αντίθετα από το προηγούμενο σφάλμα υπάρχει παραβίαση του άνω ορίου της κατηγορίας, δηλαδή η εναλλακτική έχει ταξινομηθεί σε υψηλότερη κατηγορία από ότι θα έπρεπε. Στην περίπτωση αυτή η ποσότητα $\sigma^-(x)$ πρέπει να αφαιρεθεί από τη βαθμολογία της εναλλακτικής ώστε να ενταχθεί στην σωστή κατηγορία.

Τα δύο αυτά σφάλματα ταξινόμησης απεικονίζονται στο σχήμα 2.8 για την περίπτωση των δύο κατηγοριών.

Είναι προφανές ότι δεν μπορούμε να έχουμε ταυτόχρονα και τα δύο είδη σφαλμάτων για την ίδια εναλλακτική δραστηριότητα. Τουλάχιστον το ένα από τα δύο είναι πάντα μηδέν, ή μαθηματικά $\sigma^+\sigma^- = 0$. Η αντιμετώπιση των σφαλμάτων

ταξινόμησης με αυτόν τον τρόπο στην πραγματικότητα αποτελεί μια πολύ καλή προσέγγιση του πραγματικού σφάλματος της ταξινόμησης.



Σχήμα 2.8 Σφάλματα ταξινόμησης στην περίπτωση δύο κατηγοριών (πηγή: Δούμπος, Ζοπουνίδης, 2001 [20])

Αν συνυπολογιστούν τα σφάλματα λαμβάνονται οι παρακάτω σχέσεις:

$$U(x_j) + \sigma_j^+ \geq u_1, \qquad \forall x_j \in C_1$$

$$\left.\begin{array}{l} U(x_j) + \sigma_j^+ \geq u_l \\ U(x_j) - \sigma_j^- \geq u_{l-1} \end{array}\right\}, \qquad \forall x_j \in C_l (l = 2,3,...,k-1)$$

$$U(x_j) - \sigma_j^- \geq u_{k-1}, \qquad \forall x_j \in C_k$$

Με βάση τους παραπάνω περιορισμούς, η ελαχιστοποίηση του σφάλματος της ταξινόμησης μπορεί να πραγματοποιηθεί μέσω της επίλυσης ενός προβλήματος μαθηματικού προγραμματισμού, το οποίο έχει την ακόλουθη μορφή:

$$\min \gamma' = \frac{1}{k}\sum_{l=1}^{k}\left(\frac{\sum_{\forall x_j \in C_l}\sigma_j}{n_l}\right) \Leftrightarrow \min\left\{\sum_{l=1}^{k}\left[\frac{\sum_{\forall x_j \in C_l}(\sigma_j^{+}+\sigma_j^{-})}{n_l}\right]\right\}$$

Υπό τους περιορισμούς:

$$U(x_j)-u_1+\sigma_j^{+}\geq\delta, \qquad \forall x_j \in C_1$$

$$\begin{matrix} U(x_j)-u_l+\sigma_j^{+}\geq\delta \\ U(x_j)-u_{l-1}-\sigma_j^{-}\leq-\delta \end{matrix}, \qquad \forall x_j \in C_l (l=2,3,...,k-1)$$

$$U(x_j)-u_{k-1}-\sigma_j^{-}\leq-\delta, \qquad \forall x_j \in C_k$$

$$U(x^{*})=1$$

$$U(x_{*})=0$$

$$u_k-u_{k+1}\geq s, \qquad \forall k=1,2,...,k-2$$

$$u_i(x_i) \qquad \text{αύξουσες συναρτήσεις}$$

$$\sigma_j^{+}\geq 0, \sigma_j^{-}\geq 0 \qquad \forall j=1,2,...,n$$

Στους παραπάνω περιορισμούς η σταθερά δ είναι θετική (δ≥0) και χρησιμοποιείται για την αποφυγή περιπτώσεων $U(x_j)=u_l$, όταν $x_j \in C_l$. Η σταθερά s ορίζεται έτσι ώστε s > δ (δηλώνει την αυστηρή προτίμηση μεταξύ των ορίων χρησιμότητας που διακρίνουν τις κατηγορίες). Ως x^{*} και x_{*}, συμβολίζονται αντίστοιχα τα διανύσματα με τις περισσότερες και τις λιγότερες προτιμητέες τιμές των κριτηρίων αξιολόγησης.

ΚΕΦΑΛΑΙΟ 3^{ο} Η ΙΔΙΟΤΗΤΑ ΤΗΣ ΜΟΝΟΤΟΝΙΑΣ ΣΕ ΠΡΟΒΛΗΜΑΤΑ ΤΑΞΙΝΟΜΗΣΗΣ

3.1 Εισαγωγή

Αρκετές μεθοδολογίες που έχουν χρησιμοποιηθεί για τη βελτίωση της απόδοσης πλήθους αλγορίθμων (ανάμεσα στους οποίους είναι τα Νευρωνικά Δίκτυα και οι Μηχανές Διανυσμάτων Υποστήριξης), στηρίζονται στη χρήση εκ των προτέρων πληροφορίας που υπάρχει για τα προς εκτίμηση μοντέλα. Το πρόβλημα είναι η εν λόγω πληροφορία μπορεί να διατυπωθεί με διαφορετικούς τρόπους, και επομένως η κύρια δυσκολία στη χρήση της έγκειται στο ότι δεν υπάρχει ένας συστηματικός τρόπος ενσωμάτωσης ετερογενών κομματιών πληροφορίας σε μια εύχρηστη διαδικασία μάθησης, ώστε να είναι δυνατή η εκμετάλλευσή τους. Στην παρούσα εργασία θα μας απασχολήσει την ιδιότητα μονοτονίας, και για αυτό το λόγο θα εξηγήσουμε αμέσως παρακάτω κάποιους βασικούς λόγους που μας οδήγησαν στη χρήση της.

3.2 Ο περιορισμός μονοτονίας στην ταξινόμηση, ως εκ των προτέρων γνώση.

Το κίνητρο για τη χρήση του περιορισμού μονοτονίας βασίζεται στις ακόλουθες παρατηρήσεις:

1. Η μονοτονία ως ιδιότητα παρατηρείται σε πολλά επιστημονικά πεδία. Πρόκειται για μια απλή και διαισθητική ιδιότητα, η οποία δηλώνει ότι όσο μεγαλύτερη είναι κάποια είσοδος, τόσο μεγαλύτερη είναι η έξοδος (με την προϋπόθεση όλες οι άλλες είσοδοι να είναι ίδιες). Σε αυτή την περίπτωση μιλάμε για αύξουσα μονοτονία. Παρόμοια, φθίνουσα μονοτονία ορίζεται η περίπτωση όπου αύξηση της εισόδου προκαλεί μείωση της εξόδου.

Οι ιδιότητα της μονοτονίας είναι γνωστή σε διάφορα επιστημονικά πεδία:

- Διοίκηση και οικονομικά: Η οικονομική θεωρία δηλώνει ότι η ζήτηση ενός προϊόντος μειώνεται εάν αυξάνεται η τιμή του, επομένως υπάρχει μια αρνητική σχέση ανάμεσα σε τιμή και ζήτηση. Ένα άλλο ευρέως γνωστό παράδειγμα είναι στην αποδοχή αιτήσεων δανείων, όπου ο κανόνας απόφασης θα πρέπει να είναι μονότονος σε σχέση με το εισόδημα του αιτούμενου το δάνειο, δηλαδή δε θα γινόταν δεκτή μια πολιτική η οποία θα απέρριπτε έναν αιτούμενο με υψηλό εισόδημα, ενώ την ίδια στιγμή θα ενέκρινε έναν αιτούμενο με χαμηλότερο εισόδημα και όμοια τα υπόλοιπα χαρακτηριστικά.

- Αναμονητικά συστήματα: Είναι ευρέως γνωστό ότι όσο μεγαλύτερο είναι το κυκλοφοριακό πρόβλημα στους δρόμους ή ότι όσο περισσότεροι είναι οι πελάτες σε ένα super-market, τότε οδηγούμαστε σε μεγαλύτερο χρόνο αναμονής. Επομένως υπάρχει μια μονότονη σχέση ανάμεσα σε χρόνο αναμονής και πλήθος πελατών που εξυπηρετεί ένα σύστημα.

- Επιστήμη Υπολογιστών: Μονότονες σχέσεις εμφανίζονται στη διάγνωση προβλημάτων απόδοσης σε συστήματα υπολογιστών, π.χ. στην σελιδοποίηση αύξησης καθυστερήσεων με τον αριθμό των συνδεδεμένων χρηστών [22].

- Νομικές Επιστήμες: Ένα παράδειγμα μιας εφαρμογής στα νομικά όπου οι συντελεστές (χαρακτηριστικά) έχουν μονότονη επιρροή στο αποτέλεσμα μιας διαδικασίας κρίσης είναι ένα σύστημα ημερομισθίου [23]. Το αντικείμενο ενός τέτοιου συστήματος είναι να κατηγοριοποιήσει έναν εργαζόμενο είτε ως αμειβόμενο με ημερομίσθιο είτε όχι (μερικής απασχόλησης, ανεξάρτητο συνεργάτη κτλ), βασιζόμενο σε ένα πλήθος παραγόντων. Αυτό συμβαίνει για σκοπούς εργασιακού δικαίου καθώς οι συγκεκριμένοι υπάλληλοι δικαιούνται σημαντικά μεγαλύτερη άδεια και πλήθος άλλων

προνομίων. Σε ένα τέτοιο σύστημα για παράδειγμα παράγοντες όπως το να εργάζεται ο μισθωτός κάτω από την ευθύνη του μισθωτή και αν ο μισθωτής έχει την εξουσιοδότηση να δίδει εντολές στον μισθωτό, έχουν μονότονη σχέση στην αξιολόγηση του αν ο τελευταίος ανήκει στην κατηγορία του ημερομισθίου ή όχι. Δηλαδή, μια αλλαγή στους παράγοντες αυτούς επηρεάζει την τάση που έχει το σύστημα για να τοποθετήσει στη μια ή την άλλη κατηγορία τον εργαζόμενο.

- Φυσικές Επιστήμες: Πλήθος παραδειγμάτων υπάρχουν στο συγκεκριμένο τομέα. Για παράδειγμα, το μέγεθος του σώματος ενός ζώου βρίσκεται σε μονότονη σχέση με τις ανάγκες του σε υποστήριξη, πχ όσο μεγαλύτερο είναι το ζώο, τόσο μεγαλύτερη είναι η ποσότητα της ενέργειας που απαιτείται για να το κρατήσει ζωντανό για κίνηση, η παραγωγή θερμότητας κτλ, χωρίς να γίνεται μεταβολή του βάρους του σώματος. Επιπλέον, ζώα με το ίδιο μέγεθος, χρειάζονται αναλογικά περισσότερη τροφή για υποστήριξη και μάλιστα τροφή καλύτερης ποιότητας από μεγαλύτερα σε ηλικία ζώα. Ένα άλλο παράδειγμα είναι η επίδραση της αύξησης του σωματικού βάρους του ανθρώπου στην αύξηση της πιθανότητας για καρδιακά επεισόδια, καρκίνους ή άλλες χρόνιες παθήσεις.

2. <u>Η μονοτονία βελτιώνει τη διαδικασία λήψης αποφάσεων.</u> Η εφαρμογή της αρχής της μονοτονίας μειώνει σημαντικά το πλήθος των δεδομένων που απαιτούνται είτε από τους αποφασίζοντες είτε από επαγωγικά συστήματα για τη διενέργεια κρίσεων με υψηλή ακρίβεια [23, 24, 25]. Έτσι, επιταχύνεται η διαδικασία χωρίς όμως να μειώνεται η απόδοσή της. Επιπλέον, το να λαμβάνονται υπόψη μονότονες σχέσεις ανάμεσα σε εξαρτημένα και ανεξάρτητα χαρακτηριστικά, μας επιτρέπει να συμπληρώνουμε τιμές χαρακτηριστικών που λείπουν από τα δεδομένα, καθώς και να κάνουμε ακριβείς προβλέψεις για αντικείμενα που δεν είναι

διαθέσιμα άμεσα με τα δεδομένα που έχουμε [24, 26]. Ουσιαστικά, αυτό το γεγονός βελτιώνει την ποιότητα των δεδομένων και την ανάλυσή τους.

3. Τα μονότονα μοντέλα αποφάσεων έχει παρατηρηθεί ότι αποδίδουν καλύτερα από τα μη μονότονα μοντέλα. Για προβλήματα με ιδιότητες μονοτονίας, έχει παρατηρηθεί ότι τα μοντέλα αυτά αποδίδουν καλύτερα από τα μη μονότονα ισοδύναμά τους.

4. Τα μονότονα μοντέλα είναι ευκολότερα στην κατανόηση καθώς συμφωνούν με την εμπειρία των αποφασιζόντων. Με άλλα λόγια τα μη μονότονα μοντέλα είναι περισσότερο δύσκολα στην αναπαράσταση καθώς παρουσιάζουν αντιφατικές και λιγότερο διαισθητικές εξαρτήσεις [35].

5. Η επιβολή της μονοτονίας σε μοντέλα αναιρεί το θόρυβο και επιλύει ασυνέπειες, και εμποδίζει την υπερβολική προσαρμογή στα δεδομένα. Αποτέλεσμα αυτού είναι τα μονότονα μοντέλα να δίνουν καλύτερες προβλέψεις, δηλαδή να έχουν μικρότερη πιθανότητα σφάλματος σε νέα δεδομένα [32].

6. Τα μονότονα μοντέλα έχουν μικρότερη διακύμανση στη συμπεριφορά τους σε επαναλαμβανόμενες δειγματοληψίες (αναφέρονται και ως «ευσταθή» στην εξόρυξη δεδομένων). Η μονοτονία οδηγεί σε μείωση της διακύμανσης και επομένως τα μοντέλα που προκύπτουν είναι περισσότερο ευσταθή [32].

3.3 Γενικά για τη Χρήση Περιορισμών-Εκ των προτέρων πληροφορίας

Οι ερευνητές στην αναγνώριση προτύπων, τη στατιστική και την εκπαίδευση μηχανών συχνά επιλέγουν ανάμεσα σε γραμμικά και μη γραμμικά μοντέλα, όπως τα νευρωνικά δίκτυα. Τα γραμμικά μοντέλα κάνουν πολύ ισχυρές υποθέσεις για τη συνάρτηση που πρέπει να μοντελοποιηθεί, ενώ τα νευρωνικά δίκτυα δεν κάνουν τέτοιες υποθέσεις, και μπορούν, θεωρητικά, να προσεγγίσουν κάθε ομαλή

συνάρτηση, εφόσον τους δοθούν αρκετές κρυμμένες μονάδες. Ανάμεσα σε αυτά τα δυο άκρα, υπάρχει ένα συχνά παραμελημένο μέσο πεδίο από μη γραμμικά μοντέλα, το οποίο ενσωματώνει ισχυρή εκ των πρότερων (prior) πληροφορία και υπακούει σε ισχυρούς περιορισμούς.

Ένα μοντέλο μονοτονίας είναι ένα παράδειγμα το οποίο μπορεί να ανήκει σε αυτή την περιοχή. Τα μοντέλα μονοτονίας είναι πιο ευέλικτα από τα γραμμικά μοντέλα, αλλά ακόμα υπακούουν σε ισχυρούς περιορισμούς. Πολλές εφαρμογές υπάρχουν στις οποίες για σημαντικούς λόγους θεωρούμε ότι η συνάρτηση που θέλουμε να μάθουμε είναι μονότονη σε ορισμένες ή και σε όλες τις εισόδους μεταβλητών. Για παράδειγμα, παρακολουθώντας τους αιτούντες για πιστωτικές κάρτες, θα μπορούσε κάποιος να αναμένει ότι η πιθανότητα ασυνέπειας μειώνεται μονότονα με το εισόδημα του αιτούντα. Ήδη σε προηγούμενη ενότητα αναφέρθηκε πλήθος πεδίων στα οποία εντοπίζονται μονότονες σχέσεις, αλλά και αρκετά πλεονεκτήματα κατά τη χρήση της ιδιότητας της μονοτονίας. Για αυτό θα ήταν χρήσιμη η ανάπτυξη μη γραμμικών μοντέλων που να διαθέτουν αυτή την ιδιότητα.

Στη βιβλιογραφία υπάρχουν δυο βασικές στρατηγικές ενσωμάτωσης της ιδιότητας της μονοτονίας. Η πρώτη στηρίζεται στη μάθηση από τεχνητά δείγματα μονοτονίας (γνωστά ως hints), τα οποία χρησιμοποιούνται για την εκπαίδευση του αλγορίθμου. Τα hints μπορούν να βελτιώσουν την απόδοση των μοντέλων μάθησης (μειώνοντας τη δυνατότητα αναπαράστασης των μοντέλων), χωρίς όμως να θυσιάζουν δυνατότητα προσέγγισης. Η εναλλακτική προσέγγιση στηρίζεται στην εισαγωγή της ιδιότητας της μονοτονίας με κατάλληλη τροποποίηση των αλγορίθμων μάθησης, χωρίς την παρέμβαση στα δείγματα εκπαίδευσης, ώστε να εξασφαλίζεται η δημιουργία μοντέλων με μονότονες ιδιότητες.

3.4 Χρήση Περιορισμού Μονοτονίας σε Αλγόριθμους Ταξινόμησης

3.4.1 Νευρωνικά Δίκτυα

Παρακάτω αναλύονται δύο τέτοιες διαφορετικοί μέθοδοι που εισάγουν και εκμεταλλεύονται την ιδιότητα της μονοτονίας στα νευρωνικά δίκτυα, οι οποίες έχουν παρουσιαστεί από τον Sill στα [28, 29, 32]. Η πρώτη από αυτές αφού κάνει κατάλληλη τροποποίηση της αντικειμενικής συνάρτησης την οποία το νευρωνικό δίκτυο βελτιστοποιεί, το εκπαιδεύει με χρήση ζευγών τεχνητών δειγμάτων τα οποία έχει δημιουργήσει, εισάγοντας έτσι με έμμεσο τρόπο την ιδιότητα της μονοτονίας. Έτσι, ενώ η δομή του δικτύου είναι η συνήθης, καθώς τα δεδομένα έχουν έμφυτη την ιδιότητα της μονοτονίας, τα μοντέλο που προκύπτει είναι μονότονο. Η δεύτερη χρησιμοποιεί τέτοια δομή του νευρωνικού δικτύου, η οποία το αναγκάζει να είναι μονότονο κατά την κατασκευή του, χωρίς να απαιτείται χρήση μονότονων δεδομένων.

3.4.1.1 Παραδείγματα Μονοτονίας για Εκπαίδευση ΤΝΔ

Bayesian Αναπαράσταση της Αντικειμενικής Συνάρτησης

Στα πλαίσια χρήσης των παραδειγμάτων μονοτονίας στη διαδικασία εκπαίδευσης του νευρωνικού δικτύου απαιτείται να γίνει μια τροποποίηση της αντικειμενικής συνάρτησης που επιθυμούμε να ελαχιστοποιήσουμε, ώστε να εξασφαλίζει ταυτόχρονα την μονοτονία αλλά και στην προσαρμογή στα δεδομένα εκπαίδευσης.

Για αυτό το σκοπό, σύμφωνα και με την εργασία του Sill στα [28, 29], καθορίζεται ένα βαθμωτό μέγεθος του βαθμού στον οποίο μια συγκεκριμένη υποψήφια συνάρτηση y υπακούει στη μονοτονία για ένα σύνολο μεταβλητών εισόδου. Ένα τέτοιο μέγεθος που χρησιμοποιείται παρακάτω, καθορίζεται με τον ακόλουθο τρόπο: έστω ότι x είναι το διάνυσμα εισόδου που λαμβάνεται από την κατανομή εισόδου. i είναι ο δείκτης μιας μεταβλητής εισόδου που λαμβάνεται τυχαία μέσω μιας ομοιόμορφης κατανομής ανάμεσα στις μεταβλητές για τις οποίες

ισχύει η μονοτονία. Ορίζουμε μια κατανομή διαταραχής (perturbation distribution), πχ $\mathrm{U}[0,1]$ και λαμβάνονται δx_i δείγματα από αυτή την κατανομή. Ορίζουμε διάνυσμα x' τέτοιο ώστε

$$\begin{aligned} &\forall j \neq i,\ x_j' = x_j \\ &x_i' = x_i + \mathrm{sgn}(i)\delta x_i \end{aligned} \tag{3.1}$$

όπου $\mathrm{sgn}(i)$ =1 ή -1, ανάλογα εάν η η συνάρτηση που θέλουμε να εκτιμήσουμε, έστω f, είναι μονότονα αύξουσα ή φθίνουσα στην μεταβλητή i. Ονομάζουμε σφάλμα μονοτονίας (monotonicity error) της y για το ζεύγος εισόδου (x,x'), το παρακάτω μέγεθος:

$$E_h = \begin{cases} 0 & y(\mathrm{x}') \geq y(\mathrm{x}) \\ (y(\mathrm{x})\text{-}y(\mathrm{x}'))^2 & y(\mathrm{x}') < y(\mathrm{x}) \end{cases} \tag{3.2}$$

Το μέτρο παραβίασης της μονοτονίας της y είναι το $E(E_h)$, όπου η μέση τιμή λαμβάνεται ως προς τις τυχαίες μεταβλητές x, i και δx_i.

Η καλύτερη δυνατή προσέγγιση της f, με τη χρήση της παραπάνω αρχιτεκτονικής, είναι πιθανώς προσεγγιστικά μονότονη. Αυτό, μπορεί να ποσοτικοποιηθεί σε μια εκ των προτέρων κατανομή (prior distribution) για όλες τις υλοποιήσιμες με αυτή την αρχιτεκτονική υποψήφιες συναρτήσεις:

$$\Pr(y) \propto e^{-\lambda E[E_h]}.$$

Αυτή η κατανομή αναπαριστά την a priori πυκνότητα πιθανότητας (density), ή πιθανοφάνεια (likelihood), που ισχύει για μια υποψήφια συνάρτηση, δεδομένου ενός συγκεκριμένου επιπέδου σφάλματος μονοτονίας. Η πιθανότητα ότι μια συνάρτηση είναι η καλύτερη δυνατή προσέγγιση στην f, μειώνεται εκθετικά, όσο αυξάνεται το σφάλμα μονοτονίας. Το λ είναι μια θετική σταθερά που δείχνει πόσο ισχυρή είναι η προκατάληψή μας ανάμεσα στις μονότονες συναρτήσεις.

Εκτός όμως του να υπακούει στην εκ των προτέρων πληροφορία, το μοντέλο θα πρέπει να προσαρμόζεται καλά στα δεδομένα. Για προβλήματα κατάταξης λαμβάνουμε την έξοδο y του δικτύου να αναπαριστά την πιθανότητα κατάταξης c=1, υπό τη συνθήκη της παρατήρησης του διανύσματος εισόδου (οι δυο δυνατές κατατάξεις είναι 0 και 1). Επιθυμούμε να διαλέξουμε το πιο πιθανό μοντέλο με τα δοσμένα δεδομένα ή ισοδύναμα μπορούμε να επιλέξουμε να μεγιστοποιήσουμε το $\log\left(P(\text{model/data})\right)$. Χρησιμοποιώντας το θεώρημα του Bayes, προκύπτει ότι

$$\begin{aligned}\log\left(P(\text{model/data})\right) &\propto \log\left(P(\text{data / model})\right) + \log\left(P(\text{model})\right) \\ &= \sum_{i=1}^{n} c_i \log(y_i) + (1-c_i)\log(1-y_i) - \lambda E[E_h]\end{aligned} \tag{3.3}$$

Για προβλήματα παλινδρόμησης συνεχής εξόδου, αναπαριστούμε το διάνυσμα y ως τον υπό συνθήκη μέσο όπως t εξόδου παρατήρησης, δεδομένης της παρατήρησης x. Αν θεωρήσουμε σταθερής διακύμανσης (constant variance) Γκαουσσιανό θόρυβο, τότε με την ίδια αιτιολόγηση όπως και στη διακριτή περίπτωση κατάταξης, η αντικειμενική συνάρτηση που πρέπει να μεγιστοποιηθεί είναι

$$-\sum_{i=1}^{n} (y_i - t_i)^2 - \lambda E[E_h] \tag{3.4}$$

Ο Bayesian prior οδηγεί σε μια διαισθητικά απλή μορφή τις αντικειμενικής συνάρτησης, με τον πρώτο όρο να αντανακλά τις επιθυμία να γίνει προσαρμογή στα δεδομένα και τον δεύτερο όρο να «τιμωρεί» την απόκλιση από την μονοτονία.

Έχοντας καθορίσει επομένως τις αντικειμενικές συναρτήσεις που στοχεύουμε να μεγιστοποιήσουμε, περιγράφουμε τη διαδικασία εκπαίδευσης των δικτύων με τα παραδείγματα μονοτονίας, η οποία χωρίζεται σε 2 στάδια.

- Εφόσον η σημασία των χαρακτηριστικών δεν είναι γνωστή σε εμάς, οι κατευθύνσεις της μονοτονίας χρειάζεται να προσδιοριστούν. Αυτό, μπορεί να επιτευχθεί μέσω της εκπαίδευσης ενός γραμμικού

perceptron στα δεδομένα εκπαίδευσης για ικανό πλήθος υλοποιήσεων (π.χ. 300), και παρατηρώντας τα βάρη που προκύπτουν. Όταν θα λαμβάνεται θετικό βάρος πρόκειται για αύξουσα μονοτονία, ενώ αρνητικό για τη φθίνουσα.

- Αφού καθορίστηκαν οι κατευθύνσεις της μονοτονίας, τα δίκτυα εκπαιδεύτηκαν, μέσω της μεγιστοποίησης της αντικειμενικής συνάρτησης, που δίδεται από τις σχέσεις (3.3) ή (3.4).

Στην παρούσα φάση κρίνεται απαραίτητο να αναλυθεί ο τρόπος με τον οποίο επιτυγχάνεται η δημιουργία των μονότονων δεδομένων (hints) που εκπαιδεύουν το δίκτυο.

Περιγραφή δεδομένων εκπαίδευσης

Το $E_{h,n}$ αναπαριστά το σφάλμα μονοτονίας του δικτύου, σε ένα συγκεκριμένο ζευγάρι διανυσμάτων εισόδου x, x'. Έχοντας ήδη εντοπίσει τις κατευθύνσεις μονοτονίας για κάθε μεταβλητή εισόδου (αν δηλαδή είναι αύξουσα ή φθίνουσα), στόχος για τη ρύθμιση των παραμέτρων του δικτύου είναι να δημιουργούμε για κάθε μεταβλητή χωριστά ζευγάρια διανυσμάτων τα οποία να υπακούουν στη μονοτονία αυτή. Κάθε ζεύγος δημιουργείται σύμφωνα με τη μέθοδο που περιγράφτηκε προηγουμένως, δηλαδή βασιζόμενοι σε ένα υπάρχον διάνυσμα, δημιουργούμε άλλο ένα στο οποίο μεταβάλλεται ελάχιστα μια μεταβλητή εισόδου, αφήνοντας τις υπόλοιπες ίδιες.

Έτσι, για κάθε μεταβλητή εισόδου, μπορεί να δημιουργηθεί πλήθος (π.χ. 500) ζευγαριών από διανύσματα εισόδου που αναπαριστούν τη μονοτονία σε αυτή τη μεταβλητή ως ακολούθως: Ένα απλό διάνυσμα εισόδου x δημιουργείται από ένα πολυμεταβλητό Gaussian μοντέλο της κατανομής εισόδου. Το δεύτερο διάνυσμα x' καθορίζεται έτσι ώστε

$$\forall j \neq i,\ x_j' = x_j$$
$$x_i' = x_i + \mathrm{sgn}(i)\delta x_i$$

όπου $\mathrm{sgn}(i)\delta x_i = 1$ ή -1, αναλόγως του αν η συνάρτηση f είναι μονότονα αύξουσα ή μονότονα φθίνουσα στη μεταβλητή i, και το δx_i δίδεται από μια κατανομή $U[0,1]$. Έτσι, για το l ζευγάρι, ισχύει ότι

$$e_1^l = \begin{cases} \left(g(\mathrm{x}) - g(\mathrm{x}')\right)^2 & \text{εαν } g(\mathrm{x}) > g(\mathrm{x}') \\ 0 & \text{εαν } g(\mathrm{x}) \le g(\mathrm{x}') \end{cases}$$

Η συγκεκριμένη μέθοδος, εφαρμοζόμενη για κάθε μεταβλητή χωριστά, οδηγεί σε ένα σύνολο ζευγάρια δειγμάτων - hints (π.χ. για 5 μεταβλητές εισόδου και 500 δείγματα ανά μεταβλητή, προκύπτουν 2500 ζεύγη). Η τιμή της παραμέτρου λ μπορεί να πάρει αυθαίρετα μεγάλη τιμή, με στόχο την επιβολή αυστηρής τιμής σε περίπτωση που δεν επιτυγχάνεται η μονοτονία.

Το E_1 επιλέγεται ως το άθροισμα πάνω σε όλα τα ζευγάρια, του σφάλματος μονοτονίας για από το κάθε ζευγάρι, δηλαδή

$$E_1 = \frac{1}{L}\sum_{l=1}^{L} e_1^l \ .$$

Η ικανότητα γενίκευσης του δικτύου, η οποία προσδιορίζεται από το σφάλμα ελέγχου μονοτονίας, μπορεί να γίνει με τη χρήση μικρότερου πλήθους (π.χ. 100) ζευγαριών διανυσμάτων από κάθε μεταβλητή, με τα οποία δεν είχε γίνει εκπαίδευση, αλλά τα χρησιμοποιούμε μόνο για τον υπολογισμό του σφάλματος μονοτονίας.

3.4.1.2 Δίκτυα Μονοτονίας

Μια ειδική περίπτωση μονότονων νευρωνικών δικτύων προτάθηκε από τον Sill [32]. Τα εν λόγω δίκτυα παρέχουν προσεγγιστικές δυνατότητες για όλες τις συνεχείς μονότονες συναρτήσεις.

Ένα δίκτυο σύμφωνα με την εργασία του Sill έχει μια αρχιτεκτονική τριών επιπέδων (δηλαδή δυο κρυφά επίπεδα). Το επίπεδο εισόδου συνδέεται με το

πρώτο κρυμμένο επίπεδο που αποτελείται από ένα σύνολο από γραμμικές μονάδες (υπερεπίπεδα), τα οποία είναι συνδυασμένα σε διάφορες ομάδες (ο αριθμός των μονάδων σε κάθε ομάδα δεν είναι απαραίτητα ο ίδιος). Σε αντιστοιχία με κάθε ομάδα υπάρχει μια μονάδα δεύτερου κρυφού επιπέδου, η οποία υπολογίζει το μέγιστο ανάμεσα σε όλες τις μονάδες του πρώτου κρυφού επιπέδου. Η τελική μονάδα, που είναι η μονάδα εξόδου, υπολογίζει το ελάχιστο ανάμεσα σε όλες τις ομάδες του δευτέρου επιπέδου.

Θέλοντας να δώσουμε μια πιο τυπική περιγραφή, θεωρούμε ότι έχουμε K ομάδες στο πρώτο κρυμμένο επίπεδο, με εξόδους $g_1, g_2, \ldots g_K$. Κάθε ομάδα k αποτελείται από h_k υπερεπίπεδα. Τότε, οι παράμετροι (βάρη) των υπερεπιπέδων στην ομάδα k είναι h_k-διαστάσεων διανύσματα $\mathrm{w}^{(\mathrm{k},1)}, \mathrm{w}^{(\mathrm{k},2)}, \ldots \mathrm{w}^{(\mathrm{k},\mathrm{h}_k)}$ και ο πίνακας όλων των βαρών και των αποκλίσεων (biases) συμβολίζεται με W. Τότε, η έξοδος στην ομάδα k είναι

$g_k(\mathrm{x}) = \max_j \left(\mathrm{w}^{(\mathrm{k},\mathrm{j})}\mathrm{x} - \theta^{(k,j)} \right), 1 \le j \le h_k$, όπου θ είναι ο όρος της απόκλισης.

Τότε η τελική έξοδος του δικτύου είναι:

$$y = \min_k g_k(\mathrm{x})$$

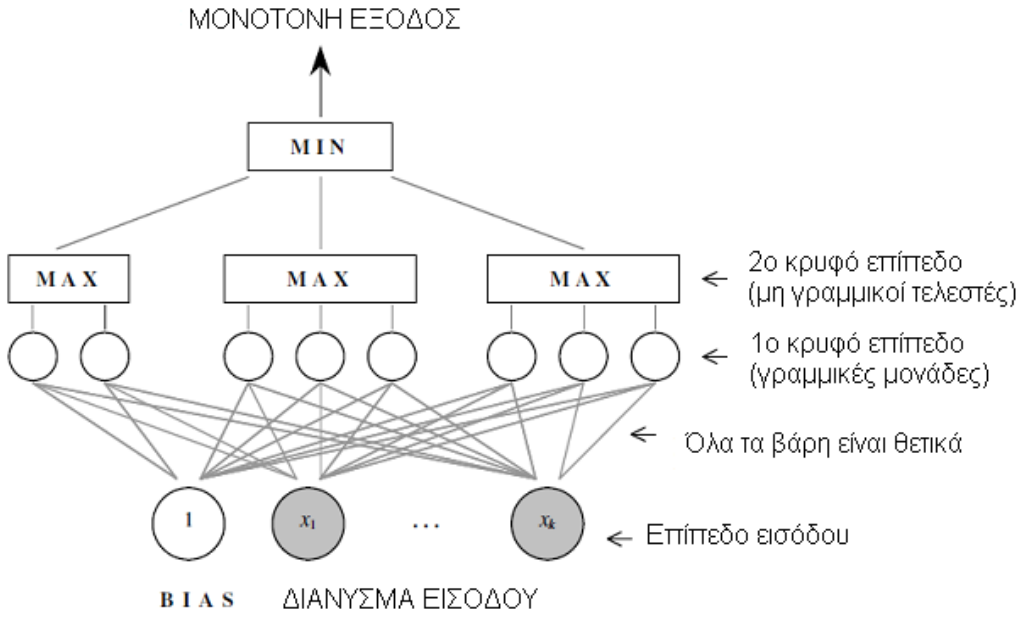
ή, για προβλήματα ταξινόμησης

$$y = \sigma\left(\min_k g_k(\mathrm{x}) \right),$$

όπου $\sigma(u)$ είναι μια συνάρτηση μετασχηματισμού (π.χ. η λογιστική συνάρτηση $\sigma(u) = \frac{1}{1+e^{-u}}$).

Ένα δίκτυο μονοτονίας έχει επομένως μια εμπρόσθιας τροφοδοσίας (feedforward), τριών επιπέδων αρχιτεκτονική. Το πρώτο επίπεδο των μονάδων υπολογίζει διαφορετικούς γραμμικούς συνδυασμούς από το διάνυσμα εισόδου.

Εφόσον επιθυμείται αύξουσα μονοτονία για μια συγκεκριμένη είσοδο, τότε όλα τα βάρη που συνδέονται σε αυτή την είσοδο περιορίζονται να είναι θετικά. Ομοίως, τα βάρη που συνδέονται σε μια είσοδο όπου απαιτείται φθίνουσα μονοτονία, πρέπει να είναι αρνητικά. Οι μονάδες του πρώτου επιπέδου τμηματοποιούνται σε διαφορετικές ομάδες (ο αριθμός των μονάδων σε κάθε ομάδα δεν χρειάζεται να είναι ο ίδιος). Αντίστοιχα σε κάθε ομάδα, βρίσκεται μια μονάδα του δευτέρου επιπέδου, η οποία υπολογίζει το μέγιστο ανάμεσα σε όλες τις μονάδες του πρώτου επιπέδου. Η τελική μονάδα εξόδου υπολογίζει το ελάχιστο ανάμεσα σε όλες τις μονάδες του $2^{ου}$ επιπέδου.

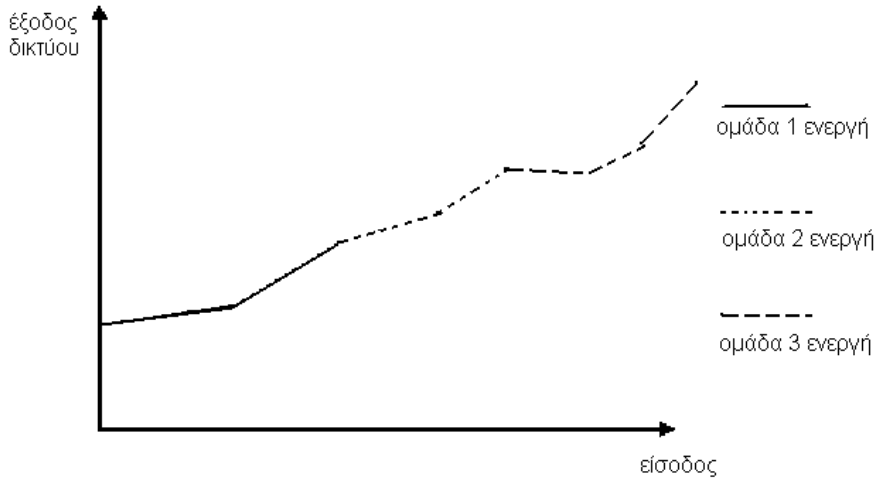


Για την παρακάτω μελέτη, είναι χρήσιμο να καθορίσουμε τον όρο ενεργός (active). Θα ονομάζουμε μια ομάδα l ενεργή στο x εφόσον

$$g_l(\mathrm{x}) = \min_k g_k(\mathrm{x}),$$

δηλαδή εφόσον η ομάδα καθορίζει την έξοδο του δικτύου σε εκείνο το σημείο (x). Αντίστοιχα, θα λέμε ότι ένα υπερεπίπεδο είναι ενεργό στο x αν η ομάδα του είναι ενεργή και το υπερεπίπεδο είναι μέγιστο από όλα τα άλλα υπερεπίπεδα στην ομάδα.

Όπως θα φανεί παρακάτω, η αρχιτεκτονική τριών επιπέδων επιτρέπει σε ένα δίκτυο μονοτονίας να προσεγγίσει κάθε συνεχή, παραγωγίσιμη μονότονη συνάρτηση αυθαίρετα καλά, έχοντας δώσει επαρκώς πολλές ομάδες και επαρκώς πολλά υπερεπίπεδα μέσα στην κάθε ομάδα. Η λειτουργία μεγιστοποίησης μέσα στην κάθε ομάδα επιτρέπει στο δίκτυο να προσεγγίσει κυρτές επιφάνειες (δεύτερη παράγωγος θετική), ενώ η λειτουργία της ελαχιστοποίησης ανάμεσα στις ομάδες ενεργοποιεί το δίκτυο για να υλοποιήσει κοίλες περιοχές (αρνητική δεύτερη παράγωγος) της συνάρτησης στόχου.



Σχήμα 4.1 Η παραπάνω επιφάνεια σχηματίζεται από ένα δίκτυο μονοτονίας αποτελούμενο από 3 ομάδες. Η πρώτη και η δεύτερη αποτελούνται από δυο υπερεπίπεδα, ενώ η τρίτη έχει τρία.

Τα δίκτυα μονοτονίας μπορούν να εκπαιδευτούν με τη χρήση πολλών τεχνικών βελτιστοποίησης βασισμένων σε αναλυτικές μεθόδους, που χρησιμοποιούνται κοινά στην εκπαίδευση μηχανών. Η κλίση για κάθε υπερεπίπεδο βρίσκεται με τον υπολογισμό του σφάλματος ανάμεσα σε όλα τα δείγματα για τα οποία το υπερεπίπεδο είναι ενεργό. Αφού γίνει η ανανέωση των παραμέτρων σύμφωνα με τον κανόνα της τεχνικής βελτιστοποίησης, το κάθε δείγμα εκπαίδευσης επαναπροσδιορίζεται για το υπερεπίπεδο που είναι ενεργό στο σημείο εκείνο. Η

ομάδα δειγμάτων για την οποία ένα υπερεπίπεδο είναι ενεργό, μπορεί για αυτό το λόγο να αλλάξει κατά τη διάρκεια της διαδικασίας εκπαίδευσης.

Οι περιορισμοί για τα πρόσημα των βαρών, επιβάλλονται με τη χρήση ενός εκθετικού μετασχηματισμού. Εφόσον επιθυμούμε αύξουσα μονοτονία σε μια μεταβλητή εισόδου i, τότε $\forall j,k$ τα βάρη που αντιστοιχούν στην είσοδο αναπαρίστανται ως $w_i^{(j,k)} = e^{z_i^{(j,k)}}$. Ο αλγόριθμος βελτιστοποίησης μπορεί να τροποποιεί ελεύθερα τα $z_i^{(j,k)}$ κατά τη διαδικασία εκπαίδευσης, διατηρώντας τον περιορισμό. Αν απαιτείται φθίνουσα μονοτονία, τότε $\forall j,k$ λαμβάνουμε $w_i^{(j,k)} = -e^{z_i^{(j,k)}}$.

3.4.2 Μηχανές Διανυσμάτων Υποστήριξης

Για την απλή περίπτωση των γραμμικά διαχωρίσιμων μηχανών διανυσμάτων υποστήριξης η συνθήκη μονοτονίας είναι εύκολο να εισαχθεί, προσθέτοντας απλά τον μη αρνητικό περιορισμό $\mathrm{b} \geq 0$ στο πρόβλημα που περιγράφεται από τη σχέση (2.4) (για την περίπτωση όπου θεωρούμε όλες οι μεταβλητές να έχουν αύξουσα μονοτονία). Έτσι, με τον τρόπο αυτό έχουμε άμεσα εισάγει στη διαδικασία ανάπτυξης του μοντέλου την εκ των προτέρων γνώση που είχαμε για τη μονοτονία αυτού.

Ωστόσο, στην περίπτωση των μη γραμμικών μοντέλων δεν είναι δυνατό να γίνει η ίδια άμεση διαδικασία. Σε αντιστοιχία με τη μεθοδολογία που περιγράφθηκε στην παράγραφο 3.4.1.1 για χρήση εικονικών δειγμάτων στη διαδικασία της εκπαίδευσης (hints), προτείνεται στην εργασία (αναφορά στην εργασία κου Δούμπου) η παρακάτω μεθοδολογία:

- Δημιουργία ζευγών μονότονων δεδομένων
- Εισαγωγή περιορισμού μονοτονίας σε πρόβλημα προγραμματισμού (2.5) με τη μορφή

$$f(\mathrm{x}_i) \geq f(\mathrm{x}_j) \Rightarrow \left[K(\mathrm{x}_i, \mathrm{X}) \text{-} K(\mathrm{x}_j, \mathrm{X}) \right] \mathrm{u} \geq 0, \; \forall \mathrm{x}_i \succ \mathrm{x}_j$$

κάνοντας χρήση των δειγμάτων μονοτονίας (hints).

Η κατασκευή του κάθε εικονικού μονότονου ζεύγους $\left\{ \mathrm{x}_i^+, \mathrm{x}_i^- \right\}$ ξεκινά από τη δημιουργία του κυριαρχούμενου δείγματος. Ειδικότερα το x_i^- δημιουργείται από μια τυχαία διαταραχή στα δεδομένα εκπαίδευσης. Για κάθε χαρακτηριστικό απόφασης k επιλέγεται μια τυχαία τιμή $x_{lk} < x_k^*$ από τις παρατηρήσεις εκπαίδευσης που αποδίδεται στην x_{ik}^-, όπου x_k^* είναι η μέγιστη τιμή της μεταβλητής k από τα δεδομένα εκπαίδευσης. Έχοντας δημιουργήσει το x_{ik}^-, το δείγμα x_{ik}^+ που κυριαρχεί, δημιουργείται ως ακολούθως: αρχικά επιλέγεται τυχαία ένα σύνολο από δείκτες μεταβλητών $J \subset \{1, 2, \ldots, M\}$. Για το κάθε χαρακτηριστικό με δείκτη $k \in J$ το δείγμα που κυριαρχεί ορίζεται ως $x_{ik}^+ = x_{ik}^-$. Για κάθε χαρακτηριστικό με δείκτη $k \notin J$, επιλέγεται από τα δεδομένα εκπαίδευσης μια τυχαία τιμή $x_{lk} > x_{ik}^-$ για να χρησιμοποιηθεί ως x_{ik}^+, δηλαδή $x_{ik}^+ = x_{lk}$. Έτσι, είναι προφανές από τον τρόπο κατασκευής των δειγμάτων ότι ισχύει $\mathrm{x}_i^+ \succ \mathrm{x}_i^-$.

3.5 Εφαρμογή Μονοτονίας σε άλλες τεχνικές ταξινόμησης

Στο σημείο αυτό, έχοντας ολοκληρώσει την ανάλυση των αλγορίθμων που αποτέλεσαν αντικείμενο μελέτης της παρούσης εργασίας, αξίζει να επισημάνουμε ότι σε πλήθος άλλων αλγορίθμων μάθησης εφαρμόζεται η ιδιότητα της μονοτονίας. Ενδεικτικά αναφέρεται οι εργασίες των Ben-David et al (1989) [24], Potharst and Feelders (2002) [35], Dembczynski et al. (2008) [36], Slowinski et al. (2009)[37, 38], Van de Kamp et al. (2009) σε μοντέλα κανόνων αποφάσεων και δένδρων αποφάσεων. Επίσης έχουν γίνει σημαντικές δημοσιεύσεις σχετικά με χρήση της μονοτονίας σε rough sets [40, 41, 42, 43].

ΚΕΦΑΛΑΙΟ 4^{ο} ΣΥΓΚΡΙΣΗ ΜΕΘΟΔΩΝ ΤΑΞΙΝΟΜΗΣΗΣ

4.1 Εισαγωγή -Πειραματική σύγκριση διαφόρων μεθόδων ταξινόμησης

Δεν είναι λίγοι οι επιστήμονες που διατείνονται ότι δεν υπάρχει μια μοναδική βέλτιστη μέθοδος ταξινόμησης. Οι ταξινομητές, όταν χρησιμοποιούνται σε διαφορετικά προβλήματα και εκπαιδεύονται με διαφορετικά σύνολα δεδομένων, αποδίδουν με διαφορετικό τρόπο. Ο Diettetich [34] αναφέρει τέσσερεις σημαντικές πηγές διαφοροποίησης της απόδοσης, οι οποίες πρέπει να ληφθούν υπόψη κατά τη σύγκριση μοντέλων ταξινόμησης.

1. Η επιλογή του συνόλου ελέγχου. Διαφορετικά σύνολα ελέγχου μπορεί να κατατάξουν τους ταξινομητές με διαφορετικό τρόπο χωρίς στην πραγματικότητα να υπάρχουν ουσιαστικές διαφορές. Για αυτό το σκοπό είναι επικίνδυνο να προκύπτουν συμπεράσματα από ένα απλό πείραμα ελέγχου, ειδικά όταν το μέγεθος του συνόλου δεδομένων είναι μικρό.

2. Η επιλογή του συνόλου εκπαίδευσης. Ορισμένα μοντέλα ταξινομητών είναι ασταθή [11] επειδή μικρές αλλαγές στο σύνολο εκπαίδευσης μπορεί να δημιουργήσουν ουσιαστικές αλλαγές στον υπό εκπαίδευση ταξινομητή. Παραδείγματα ασταθών ταξινομητών είναι τα δένδρα αποφάσεων και κάποια νευρωνικά δίκτυα.

3. Η εσωτερική τυχαιότητα του αλγορίθμου εκπαίδευσης. Ορισμένοι αλγόριθμοι εκπαίδευσης έχουν ένα στοιχείο που εισάγει τυχαιότητα. Αυτό μπορεί να είναι ο τρόπος αρχικοποίησης των παραμέτρων του ταξινομητή, οι οποίες έπειτα ρυθμίζονται με ακρίβεια (πχ αλγόριθμος backpropagation για νευρωνικά δίκτυα), ή μια τυχαία διαδικασία ρύθμισης του ταξινομητή (πχ γενετικός αλγόριθμος). Έτσι μπορεί να παρατηρηθεί ότι ο ίδιος αλγόριθμος ταξινόμησης μπορεί να δίδει διαφορετικό μοντέλο για το ίδιο σύνολο εκπαίδευσης, ίσως ακόμα και για την ίδια αρχικοποίηση παραμέτρων.

4. Το σφάλμα τυχαίας ταξινόμησης. Ο Diettetich [11] θεωρεί ότι υπάρχει πιθανότητα να υπάρχουν λάθος ταξινομημένα δεδομένα στο σύνολο ελέγχου, αλλάζοντας την απόδοση των διαφόρων αλγορίθμων.

Από όλα τα παραπάνω γίνεται προφανές ότι πρέπει να χρησιμοποιηθούν πολλαπλά σύνολα εκπαίδευσης και ελέγχου, τα οποία να τα «τρέξουμε» για πολλές φορές, ώστε να προκύψει μια πιο αξιόπιστη, στοχαστική, εκτίμηση της απόδοσης των διαφόρων αλγορίθμων.

Όταν αναπτύσσουμε μια νέα ή τροποποιούμε μια ήδη υπάρχουσα μέθοδο επίλυσης ενός προβλήματος μοντελοποίησης δεδομένων, κάποιος συνήθως ελέγχει τη μέθοδο σε μεγαλύτερο βαθμό όχι μόνο σε πραγματικά δεδομένα, αλλά και σε τεχνητά. Ένας λόγος για τον συνυπολογισμό των τεχνητών δεδομένων σε τέτοια πειράματα είναι ότι οι παράμετροι των μοντέλων που βρίσκονται κάτω από τα δεδομένα είναι πολύ καλύτερα ελεγχόμενες σε σύγκριση με την περίπτωση των πραγματικών δεδομένων. Για παράδειγμα, όταν θέλουμε να ελέγξουμε την ισχύ μιας νέας μεθόδου μοντελοποίησης δεδομένων για διαφορετικούς βαθμούς θορύβου στα δεδομένα, είναι γενικά πολύ ευκολότερο να δημιουργήσουμε τεχνητά δεδομένα, παρά να ερευνήσουμε για πραγματικά δεδομένα που να έχουν τέτοια επίπεδα θορύβου. Σε πολλές περιπτώσεις τα σύνολα τεχνητών δεδομένων μπορούν να δημιουργηθούν χωρίς πολύ προσπάθεια, χρησιμοποιώντας τεχνικές προσομοίωσης. Για παράδειγμα πρώτα δημιουργώντας τυχαίες τιμές για τις παραμέτρους του μοντέλου μας, έπειτα δημιουργώντας τυχαίες μεταβλητές εισόδου, υπολογίζοντας τις αντίστοιχες εξόδους και προσθέτοντας τυχαίο θόρυβο στις τιμές εξόδου. Ωστόσο, η δημιουργία τεχνητών δεδομένων δεν είναι πάντοτε εύκολη.

Στο πλαίσιο αυτό κρίθηκε σκόπιμο να δημιουργήσουμε τεχνητά δεδομένα τα οποία θα χρησιμοποιηθούν για τον έλεγχο της απόδοσης των υπό μελέτη μεθόδων, με μια μέθοδο που θα αναλύσουμε στην επόμενη παράγραφο.

Οι αλγόριθμοι που συγκρίθηκαν είναι οι παρακάτω, που μπορούν να ομαδοποιηθούν σε μη μονότονες και σε μονότονες τεχνικές:

- Μη μονότονες:
    - Απλό εμπρόσθιας τροφοδοσίας (feed-forward) νευρωνικό δίκτυο με ένα κρυφό επίπεδο
    - Μηχανή διανυσμάτων υποστήριξης, στην οποία επιλέχτηκε συνάρτηση πυρήνα ακτινικής βάσης (radial basis Kenel function). Χρησιμοποιήθηκε η libsvm υλοποίηση των SVM [39]
- Μονότονες:
    - Νευρωνικό δίκτυο μονοτονίας που πρότεινε ο Sill [32] και υλοποίησε η Velikova [33]
    - Μηχανή διανυσμάτων υποστήριξης με χρήση παραδειγμάτων μονοτονίας, στην οποία χρησιμοποιήθηκε η συνάρτηση πυρήνα ακτινικής βάσης
    - Η πολυκριτήρια μέθοδος UTADIS

Οι συγκεκριμένοι αλγόριθμοι κατά την εκτέλεσή τους υποβλήθηκαν σε κατάλληλες ρυθμίσεις παραμέτρων, ούτως ώστε να επιτευχθεί η καλύτερη δυνατή απόδοση. Στη συνέχεια του κεφαλαίου δίδονται σε πίνακες και σχεδιαγράμματα τα αποτελέσματα των διαφόρων προσομοιώσεων που εκτελέστηκαν, μαζί με ανάλυση και συμπεράσματα που προκύπτουν, για τον κάθε αλγόριθμο ξεχωριστά, τόσο για πραγματικά όσο και για τα τεχνητά δεδομένα. Με βάση αυτά, και κρατώντας τις τιμές εκείνες των παραμέτρων που δίνουν βέλτιστη απόδοση, θα καταλήξουμε σε σύγκριση όλων των αλγορίθμων μεταξύ τους.

Οφείλουμε να τονίσουμε η απόδοση σε όλες τις περιπτώσεις εκτιμήθηκε μέσω της χρήσης του σφάλματος ταξινόμησης. Δηλαδή, δεν μας ενδιαφέρει, λόγω της φύσης του προβλήματος, η «απόσταση» (σφάλμα) του εκτιμώμενου μεγέθους

από το πραγματικό, αλλά το αν με βάση την εκτιμώμενη τιμή του μεγέθους, γίνει ταξινόμηση στην σωστή κατηγορία. Έτσι, σε αρκετούς αλγορίθμους, η έξοδός τους χρησιμοποιήθηκε ως κριτήριο για την κατηγοριοποίηση, και ο έλεγχος σφάλματος έγινε σχετικά με τα εάν έγινε ορθή πρόβλεψη για την κατηγορία ταξινόμησης.

4.2 Τεχνητά Δεδομένα

4.2.1 Μεθοδολογία Δημιουργίας Τεχνητών Δεδομένων

Στόχος της μεθόδου [27] που αναλύουμε παρακάτω είναι να δημιουργήσει ένα μονότονο σύνολο δεδομένων από n στοιχεία από ένα δοσμένο χώρο δειγμάτων X, και ένα σύνολο τιμών αποτελεσμάτων Y, τέτοιο ώστε η συχνότητα της παρατήρησης των αποτελεσμάτων να ακολουθεί ένα προκαθορισμένο πρότυπο. Αυτό το σύνολο δεδομένων θα είναι τυχαίο, υπό την έννοια ότι είναι ένα τυχαίο δείγμα από το χώρο όλων των δυνατών συνόλων δεδομένων που ικανοποιούν τις προκαθορισμένες συνθήκες. Ωστόσο, δεν εξασφαλίζεται ότι όλα τα δυνατά σύνολα δεδομένων έχουν τις ίδιες πιθανότητες να επιλεγούν. Για να περιγράψουμε την τεχνική, θα εισάγουμε πρώτα μια χαρτογράφηση ανάμεσα στα σύνολα δεδομένων και τους κύριους γράφους, που θα εξυπηρετήσει την περιγραφή.

Έστω $X = X_1 \times X_1 \times \ldots \times X_p$ ο χώρος συμβάντων, με p χαρακτηριστικά, Y ένα σύνολο από τιμές αποτελεσμάτων, και D ένα σύνολο δειγμάτων (x, y) από το χώρο $X \times Y$. Με το D θα συνδέσουμε έναν κατευθυνόμενο γράφο από ετικέτες (labeled directed graph) $G(D)$, ως ακολούθως: D_x, το σύνολο των τιμών του διανύσματος x από το σύνολο δεδομένων D, θα είναι το σύνολο των κορυφών του κατευθυνόμενου γράφου, και $[\mathrm{x},\mathrm{x}']$ ένα τόξο του γράφου εάν και μόνο εάν $\mathrm{x} \succ \mathrm{x}'$ (Τονίζουμε ότι ο γράφος $G(D)$ θα είναι πάντοτε ακυκλικός). Επιπρόσθετα, ας είναι κάθε κορυφή x του γράφου χαρακτηρισμένη με την κατηγορία y του δείγματος δεδομένων (x,y).

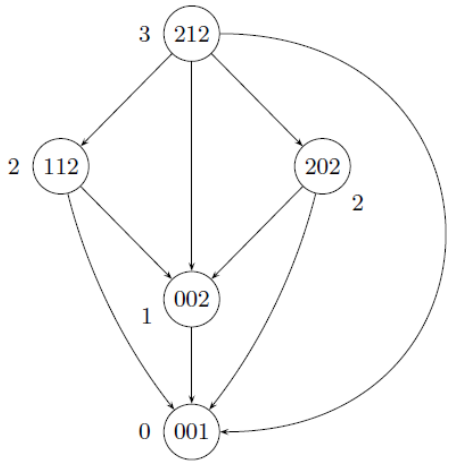
Για την κατανόηση των παραπάνω χρήσιμο είναι το παρακάτω παράδειγμα χρηματοπιστωτικής αξιολόγησης πελάτη, στο οποίο έχουμε αξιολόγηση σε τρία χαρακτηριστικά και βάσει αυτής προκύπτει ταξινόμηση κάθε αιτούντα σε 4 δυνατές ομάδες. Ο πίνακας 4.1 δείχνει το αρχικό σύνολο δεδομένων, με την τελική εκτίμηση, ενώ στον πίνακα 4.2 παρουσιάζεται με δυο διαφορετικούς τρόπους η δυνατή κωδικοποίηση των επί μέρους βαθμολογιών σε κάθε χαρακτηριστικό, όπως και η αντίστοιχη κατηγορία στην οποία κατατάσσεται. Τέλος, στο σχήμα 4.1 φαίνεται ο γράφος που προκύπτει για τα συγκεκριμένα δεδομένα.

Πελάτης	Εισόδημα	Εκπαίδευση	Ποινικό Μητρώο	Εκτίμηση Δανείου
No 1	Χαμηλό	Χαμηλή	Ικανοποιητικό	Απόρριψη
No 2	Χαμηλό	Χαμηλή	Άψογο	Χαμηλή
No 3	Μεσαίο	Μεσαία	Άψογο	Μεσαία
No 4	Υψηλό	Χαμηλή	Άψογο	Μεσαία
No 5	Υψηλό	Μεσαία	Άψογο	Υψηλή

Πίνακας 4.1 Δεδομένα αξιολόγησης για εκτίμηση δανείου

x_1	x_1	x_1	y	x	y
0	0	1	0	001	0
0	0	2	1	002	1
1	1	2	2	112	2
2	0	2	2	202	2
2	1	2	3	212	3

Πίνακας 4.2 Κωδικοποίηση δεδομένων παραδείγματος



Σχήμα 4.1 Γράφος G παραδείγματος

Θα ονομάσουμε κάθε δρόμο του γράφου $G(D)$ μη αυξανόμενο (non-increasing), εφόσον οι κατηγορίες στις οποίες ταξινομούνται εναλλακτικές σε διαφορετικές κορυφές του, διαμορφώνουν μια μη αύξουσα ακολουθία. Έτσι, στο προηγούμενο παράδειγμα, η διαδρομή [212, 112, 002, 001] είναι μη αύξουσα ακολουθία, καθώς οι ετικέτες δημιουργούν την μη αύξουσα ακολουθία [3,2,1,0].

Ακόμα και εάν αλλάζαμε την ετικέτα από 002 σε 2, το μονοπάτι θα έμενε μη αυξανόμενο. Ωστόσο, εάν αλλάζαμε την ετικέτα 001 σε 2, το μονοπάτι θα έχανε την ιδιότητα. Το ακόλουθο λήμμα συνδέει την μονοτονία ενός συνόλου δεδομένων με την μη αύξουσα φύση των διαδρομών στο γράφημα.

Λήμμα 1: Έστω D ένα σύνολο δεδομένων στο $X \times Y$, και $G(D)$ ο αντίστοιχος γράφος του. Τότε ισχύει ότι

$$D \text{ μονότονο} \Leftrightarrow \text{όλοι οι δρόμοι στο } G(D) \text{ είναι μη αύξοντες}$$

Συνεχίζουμε με μια περιγραφή του αλγορίθμου που παράγει μονότονα σύνολα δεδομένων. Έστω ότι τα απαιτούμενα σύνολα δεδομένων πρέπει να έχουν N δείγματα. Ξεκινάμε επιλέγοντας N διαφορετικά διανύσματα $\mathrm{x}_1, \mathrm{x}_1, \ldots, \mathrm{x}_N$, τυχαία από το χώρο X. Έπειτα επιλέγουμε N, όχι απαραίτητα διαφορετικές ετικέτες $y_1, y_2, \ldots, y_N$ από το Y. Τώρα, κάθε μια από τις ετικέτες $y_1, y_2, \ldots, y_N$ πρέπει να αντιστοιχηθεί με ακριβώς ένα από τα διανύσματα $\mathrm{x}_1, \mathrm{x}_1, \ldots, \mathrm{x}_N$ για να διαμορφωθεί το απαραίτητο σύνολο δεδομένων. Αυτό θα γίνει με τον ακόλουθο τρόπο:

- Δημιουργία του γράφου που συνδέεται με τα $\mathrm{x}_1, \mathrm{x}_1, \ldots, \mathrm{x}_N$
- Κατάταξη της ακολουθίας $y_1, y_2, \ldots, y_N$, έτσι ώστε $y_1 \le y_2 \le \ldots \le y_N$
- Τέλος, αναπαράσταση του γράφου όπως ένα διάγραμμα ροής: κάθε μια από τις ετικέτες $y_1, y_2, \ldots, y_N$ επιτρέπεται να «μετακινηθεί» μέσα στο δίκτυο, ακολουθώντας ένα τυχαίο δρόμο: κάθε διαδρομή ξεκινά σε ένα τυχαία επιλεγμένο κόμβο-πηγή (source node, ένα κόμβο χωρίς εισερχόμενα τόξα) και καταλήγει σε ένα τερματικό κόμβο (sink node, χωρίς εξερχόμενα από αυτόν τόξα). Να σημειώσουμε ότι το κάθε μονοπάτι είναι πεπερασμένο, καθώς ο γράφος είναι ακυκλικός. Η ετικέτα στόχου αποδίδεται στην τελευταία κορυφή της διαδρομής. Ακολούθως, η κορυφή αφαιρείται από το γράφο, και η επόμενη ετικέτα «μετακινείται» μέσα στο νέο γράφο. Αυτό συμβαίνει, μέχρι η κάθε κορυφή του αρχικού γράφου να αντιστοιχηθεί σε μια ετικέτα.

Στο προηγούμενο επομένως παράδειγμα, αν τα τυχαία διανύσματα είναι τα 001, 002, 112, 202 και 212 και οι ετικέτες στόχοι είναι 0,1,2,2,3, καταλήγουμε με τον γράφο G του σχήματος 4.1. Ωστόσο, αν στο ίδιο σύνολο δεδομένων θέλουμε να αντιστοιχίσουμε τις ετικέτες 0,1,1,2,3, καταλήγουμε σε ένα από τα δυο ακόλουθα σύνολα δεδομένων του πίνακα 4.3, το καθένα με πιθανότητα ½.

x	y	x	y
001	0	001	0
002	1	002	1
112	1	112	2
202	2	202	1
212	3	212	3

Πίνακας 4.3

Μια τυπική περιγραφή του συγκεκριμένου αλγορίθμου περιγράφεται στην εργασία του Potharst [27].

Έχοντας προχωρήσει στην παραπάνω υλοποίηση του αλγορίθμου, έχουμε τη δυνατότητα να δημιουργούμε δεδομένα στα οποία να δίδουμε ετικέτες με τέτοιο τρόπο, ώστε να είναι μονότονα. Αυτό μας επιτρέπει λοιπόν, σε ένα πίνακα 100*5 (100 αντικείμενα με 5 χαρακτηριστικά το καθένα), να μπορούμε να δώσουμε μια ετικέτα, ώστε τα δεδομένα που κατηγοριοποιούνται να παραμένουν μεταξύ τους μονότονα. Αν υποθέσουμε ότι η ετικέτα λαμβάνει τιμές στο διάστημα [0,1], τότε για να τα κατηγοριοποιήσουμε σε δυο τάξεις, μπορούμε να θεωρήσουμε ότι όσα είναι πάνω από 0.5 ανήκουν στη μια κατηγορία, και τα υπόλοιπα στην άλλη.

Τα δεδομένα που δημιουργήθηκαν για τους σκοπούς της εργασίας ακολούθησαν την πολυμεταβλητή κανονική κατανομή με μέση τιμή 0 και πίνακα συνδιακύμανσης

$$sigma = \begin{bmatrix} 1 & 0.1 & \cdots & 0.1 \\ 0.1 & 1 & 0.1 & 0.1 \\ \vdots & 0.1 & \ddots & \vdots \\ 0.1 & \cdots & 0.1 & 1 \end{bmatrix}$$

(στον οποίο τα στοιχεία της κυρίας διαγωνίου είναι ίσα με 1 και όλα τα υπόλοιπα ίσα με 0.1)

Για τη σύγκριση των αλγορίθμων δημιουργήθηκαν 6 μονότονα σύνολα, για 5 και 10 μεταβλητές και για σύνολα εκπαίδευσης των 100, 200 και 300 παρατηρήσεων. Έτσι μας δίδεται η δυνατότητα να μπορέσουμε να εκτιμήσουμε πόσο «ευαίσθητοι» είναι οι υπό μελέτη αλγόριθμοι στην αύξηση των μεταβλητών αλλά και το πόσο «γρήγορα» εκπαιδεύονται, υπό την έννοια του πόσο μεγάλο απαιτείται να είναι το δείγμα εκπαίδευσης για να λειτουργήσει ικανοποιητικά ένας αλγόριθμος. Το δείγμα ελέγχου, που δημιουργήθηκε με την ίδια μέθοδο, είχε σε όλες τις περιπτώσεις 200 παρατηρήσεις. Τα δεδομένα ανήκουν με πιθανότητα 0.5 σε κάθε μια από τις δύο κατηγορίες ταξινόμησης, τόσο για το σύνολο εκπαίδευσης όσο και για το σύνολο ελέγχου. Αξίζει να σημειωθεί ότι η ακρίβεια των υπό μελέτη αλγορίθμων που αναλύεται παρακάτω, προκύπτει ως η μέση τιμή της ακρίβειας ταξινόμησης στην κάθε μια κατηγορία, που ωστόσο είναι κατά μεγάλη προσέγγιση ίση στην περίπτωση αυτή με την συνολική ακρίβεια ταξινόμησης, εφόσον τα σύνολα είναι εξίσου μοιρασμένα στις δυο κατηγορίες.

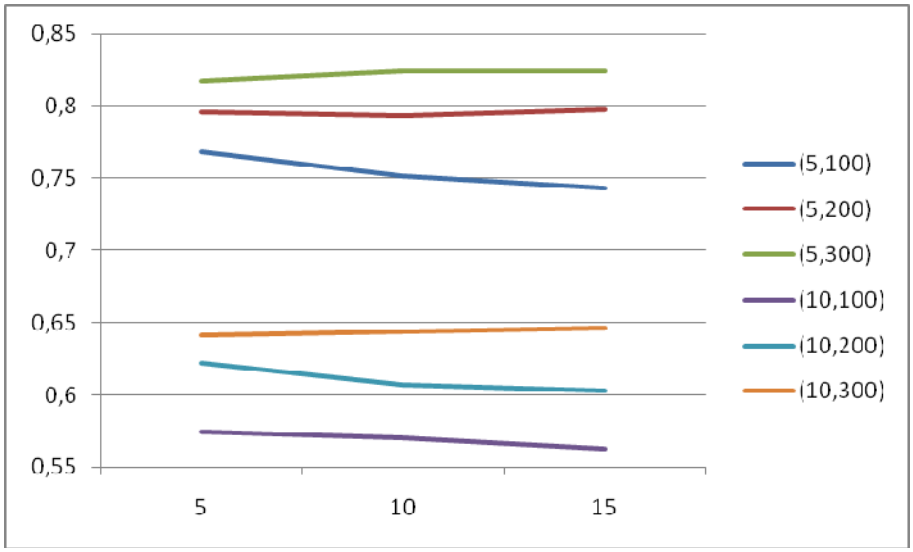
4.2.2 Αποτελέσματα

4.2.2.1 Τεχνητά νευρωνικά δίκτυα

Ο πιο «απλός» από τους υπό μελέτη αλγορίθμους, ταυτόχρονα όμως «state-of-the-art» μεθοδολογία με πλήθος εφαρμογών, χρησιμοποιήθηκε ως αρχική αναφορά για τα τεχνητά δεδομένα, και η ακρίβειά του στο δείγμα ελέγχου παρουσιάζεται στον ακόλουθο πίνακα και αντίστοιχο γράφημα.

Αριθμός μεταβλητών	Αριθμός νευρώνων	Μέγεθος δείγματος εκπαίδευσης		
		100	200	300
5	5	0,7680	0,7962	0,8175
	10	0,7518	0,7929	0,8242
	15	0,7431	0,7973	0,8244
10	5	0,5746	0,6225	0,6414
	10	0,5705	0,6075	0,6439
	15	0,5627	0,6026	0,6467

Πίνακας 4.4 Ακρίβεια νευρωνικών δικτύων σε τεχνητά δεδομένα



Σχήμα 4.2 Ακρίβεια νευρωνικών δικτύων σε τεχνητά δεδομένα

Μπορούμε να εξάγουμε τα παρακάτω συμπεράσματα:

- Συμπεριφορά του ΤΝΔ ως προς το πλήθος των νευρώνων.

Δεν υπάρχει εμφανής διαφορά στην απόδοση κατά τη μεταβολή του πλήθους των νευρώνων. Ωστόσο, αυτό που μπορεί κανείς να παρατηρήσει στους πίνακες (σχετικά μικρή τάση, που ωστόσο επιβεβαιώνεται διαισθητικά), είναι ότι για σχετικά μικρό δείγμα εκπαίδευσης είναι συγκριτικά αποδοτικότερος αλγόριθμος με λίγους (δηλ. 5) νευρώνες, ενώ για μεγαλύτερο (δηλ. 300) δείγμα εκπαίδευσης είναι ελάχιστα αποδοτικότερος ο αλγόριθμος με τους περισσότερους -

15 – νευρώνες. Η εξήγηση είναι προφανής: όταν το δείγμα εκπαίδευσης είναι μικρό και χρησιμοποιούνται πολλοί νευρώνες, οι συνθήκες ευνοούν την υψηλή προσαρμογή στα δεδομένα εκπαίδευσης. Το δείγμα δεν επαρκεί για την κατάλληλη εκπαίδευση τόσο μεγάλου πλήθους νευρώνων. Αντίθετα, όταν το δείγμα είναι μεγαλύτερο, και τότε οι 15 νευρώνες αποδίδουν καλύτερα για ένα σχετικά πολύπλοκο μοντέλο (των 5 ή 10 μεταβλητών), όπως αυτό των τεχνητών δεδομένων που δημιουργήσαμε.

- Συμπεριφορά ως προς το πλήθος των ανεξάρτητων μεταβλητών.

Κατά την αύξηση του πλήθους των μεταβλητών (από 5 σε 10) παρατηρείται μείωση της ακρίβειας (από 0,16 μέχρι και 0,18 ανά περίπτωση). Το συγκεκριμένο γεγονός επιβεβαιώνεται και από τα αποτελέσματα κατά την εκτέλεση των υπολοίπων αλγορίθμων, και εξηγείται από το ότι η αύξηση των διαστάσεων του προβλήματος (που προκύπτει από την αύξηση των μεταβλητών) κάνει το προς επίλυση πρόβλημα περισσότερο πολύπλοκο, καθώς η σχέση κυριαρχίας γίνεται περισσότερο σύνθετη με δεδομένο ότι τα δεδομένα διαμορφώθηκαν θεωρώντας ότι οι μεταβλητές παρουσιάζουν χαμηλό βαθμό συσχέτισης.

- Αλληλεπίδραση πλήθους μεταβλητών και μεγέθους δείγματος εκπαίδευσης.

Εντοπίζεται βελτίωση στην απόδοση του αλγορίθμου τόσο στην περίπτωση του συνόλου δεδομένων με 5 όσο και του συνόλου με 10 μεταβλητές, κατά την αύξηση του πλήθους του συνόλου εκπαίδευσης. Αξίζει επίσης να σημειωθεί ότι κατά την αύξηση του δείγματος εκπαίδευσης από 100 σε 200 και τέλος σε 300 εγγραφές, δεν παρουσιάστηκε σύγκλιση της απόδοσης των διαφορετικών συνόλων, γεγονός που σημαίνει ότι ενδέχεται η περεταίρω αύξηση π.χ. σε 400 εγγραφές να οδηγήσει σε ακόμα καλύτερη εκπαίδευση και επομένως

απόδοση του νευρωνικού δικτύου. Ωστόσο η συγκεκριμένη περίπτωση δε μελετήθηκε στην παρούσα εργασία, καθώς ξεφεύγει από τους σκοπούς αυτής.

4.2.2.2 Μονότονα νευρωνικά δίκτυα

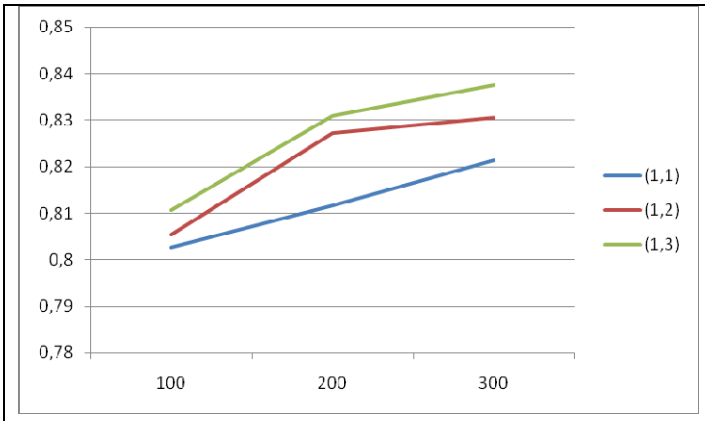
Τα αποτελέσματα στην περίπτωση του αλγορίθμου νευρωνικού δικτύου μονοτονίας φαίνονται στον παρακάτω πίνακα, όπου (x,y) είναι (πλήθος παραμέτρων, μέγεθος δείγματος εκπαίδευσης).

		Πλήθος ομάδων		
(5,100)		1	2	3
Επίπεδα	1	0,8026	0,797	0,7942
	2	0,8054	0,8006	0,7968
	3	**0,8106**	0,8018	0,8009
(5,200)		1	2	3
Επίπεδα	1	0,8117	0,8099	0,8092
	2	0,8273	0,819	0,8181
	3	**0,831**	0,8206	0,8239
(5,300)		1	2	3
Επίπεδα	1	0,8213	0,8188	0,8184
	2	0,8306	0,8227	0,8261
	3	**0,8376**	0,8329	0,8345

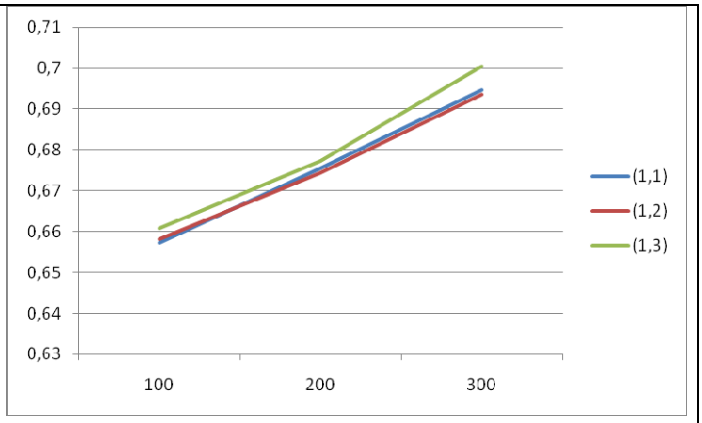
		Πλήθος ομάδων		
(10,100)		1	2	3
Επίπεδα	1	0,6573	0,6576	0,6569
	2	0,6582	0,6553	0,6536
	3	0,6607	0,6614	**0,6633**
(10,200)		1	2	3
Επίπεδα	1	0,6754	**0,6802**	0,6744
	2	0,6745	0,6761	0,6751
	3	0,6774	0,6798	0,6724
(10,300)		1	2	3
Επίπεδα	1	0,6947	0,698	0,6919
	2	0,6935	0,6956	0,6986
	3	**0,7004**	0,701	0,6966

Πίνακας 4.5 Ακρίβεια μονότονου νευρωνικού δικτυου σε τεχνητά δεδομένα

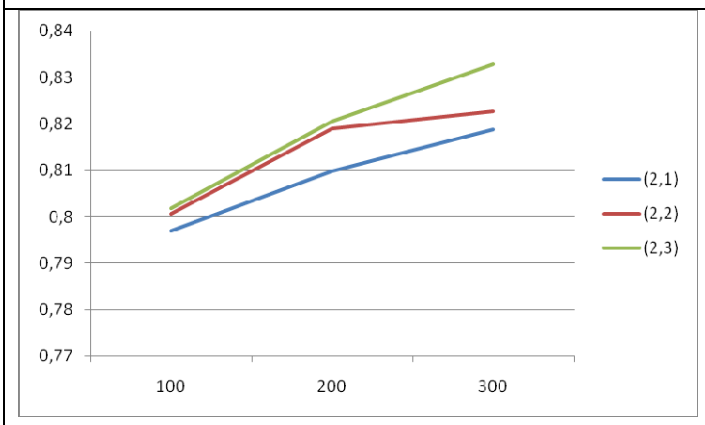
Καθώς είναι προφανές ότι ανάμεσα στην περίπτωση των δεδομένων που έχουν 5 και 10 μεταβλητές, η απόδοση κυμαίνεται σε διαφορετικά επίπεδα, για να προκύψουν περισσότερο ευδιάκριτα γραφήματα, απεικονίζουμε χωριστά τα αποτελέσματα για τις δυο περιπτώσεις, στον οριζόντιο άξονα των οποίων απεικονίζεται το μέγεθος του δείγματος εκπαίδευσης του αλγορίθμου.



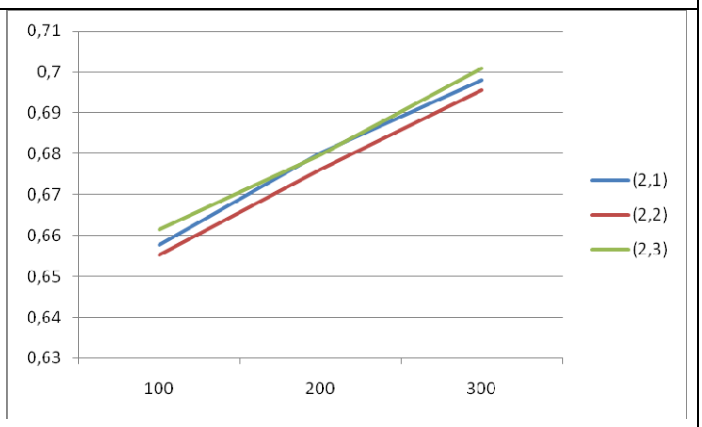
Ακρίβεια για 5 μεταβλητές, 1 ομάδα, μεταβάλλοντας αριθμό των επιπέδων



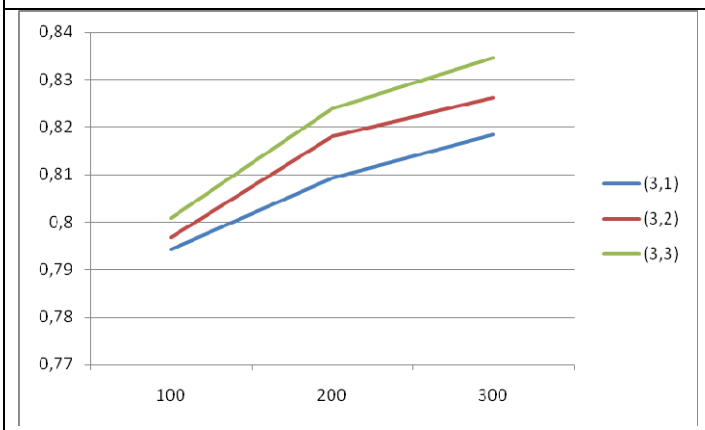
Ακρίβεια για 10 μεταβλητές, 1 ομάδα, μεταβάλλοντας αριθμό των επιπέδων



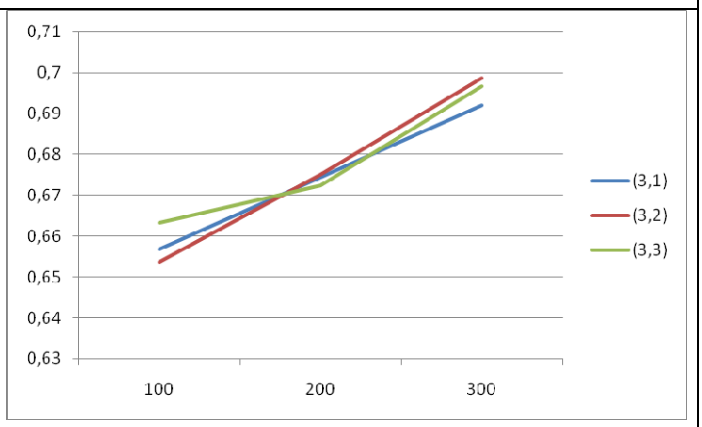
Ακρίβεια για 5 μεταβλητές, 2 ομάδες, μεταβάλλοντας αριθμό των επιπέδων



Ακρίβεια για 10 μεταβλητές, 2 ομάδες, μεταβάλλοντας αριθμό των επιπέδων



Ακρίβεια για 5 μεταβλητές, 3 ομάδες, μεταβάλλοντας αριθμό των επιπέδων



Ακρίβεια για 10 μεταβλητές, 3 ομάδες, μεταβάλλοντας αριθμό των επιπέδων

Από τους πίνακες γίνεται αντιληπτό ότι με την αύξηση της τιμής των επιπέδων, υπάρχει μια μικρή αύξηση της ακρίβειας, γεγονός που επιβεβαιωνεται και από τα παρακάτω γραφήματα, ειδικά στην περίπτωση των δεδομένων που περιγράφονται από 5 μεταβλητές. Δεν ισχύει ωστόσο το ίδιο συμπέρασμα και κατά την αύξηση της τιμής των ομάδων, όπου τα συμπεράσματα είναι ασαφή. Σε όλες τις περιπτώσεις ωστόσο, μπορούμε να καταλήξουμε στο συμπέρασμα ότι γενικά ο αλγόριθμος αποδίδει στα ίδια επίπεδα, και ενδεχομένως η δομή του να είναι αυτό που του εξασφαλίζει την ποιοτική απόδοση, και όχι τόσο η πολυπλοκότητα που εισάγουμε αυξάνοντας τις τιμές των ομάδων και των επιπέδων.

Προφανής είναι για μια ακόμα φορά η σταδιακή βελτίωση της ακρίβειας του αλγορίθμου όταν γίνεται χρήση μεγαλύτερου δείγματος εκπαίδευσης, αλλά και η μείωσή της κατά την αύξηση του πλήθους των ανεξάρτητων μεταβλητών του προβλήματος

4.2.2.3 Μηχανές διανυσμάτων υποστήριξης

Ο έλεγχος με τα πραγματικά δεδομένα για τον αλγόριθμο SVM ξεκίνησε αναζητώντας τις τιμές εκείνες των παραμέτρων C και σ, για τις οποίες έχουμε τη βέλτιστη δυνατή απόδοση. Για το σκοπό αυτό, για το σύνολο των τεχνητών δεδομένων που δημιουργήθηκαν, χρησιμοποιήθηκε ο αλγόριθμος SVM, με αποτελέσματα τα οποία παρουσιάζονται στους αμέσως επόμενους πίνακες.

Όπως είναι προφανές από τους Πίνακες, δεν υπάρχει ουσιαστική μεταβολή – ευαισθησία της ακρίβειας στην τιμή της παραμέτρου C. Ωστόσο, υπάρχει εξάρτηση από την τιμή της παραμέτρου σ όπου παρατηρείται ότι όσο η τιμή μειώνεται, τόσο βελτιώνεται η προβλεπτική ικανότητα του μοντέλου.

Η συγκεκριμένη παρατήρηση μας ώθησε στην περεταίρω μείωση της τιμής της παραμέτρου σ, για μελέτη της συμπεριφοράς του αλγορίθμου, κρατώντας όμως αυτή τη φορά την τιμή της παραμέτρου C σταθερή (C=2), ώστε να είναι δυνατή η

οπτική παρουσίαση των όποιων αποτελεσμάτων. Η τιμή του σ μειώθηκε μέχρι τα επίπεδα εκείνα στα οποία δεν εντοπίζαμε επιπλέον βελτίωση από επιπλέον μεταβολή, και επομένως προέκυψε το σχήμα 4.3.

Από το σχήμα αυτό μπορούμε να εξάγουμε τα παρακάτω συμπεράσματα:

- Συμπεριφορά του αλγορίθμου ως προς την παράμετρο της συνάρτησης πυρήνα.

Η μείωση της τιμής του σ οδηγεί σε αύξηση της ακρίβειας. Για μεγάλη τιμή του σ (περίπου 32) έχουμε ουσιαστικά αδρανοποίηση του αλγορίθμου, καθώς η απόδοσή του φθάνει σε τιμές περίπου στο 0,50, η οποία αντιστοιχεί ουσιαστικά σε μια τυχαία ταξινόμηση. Όσο μειώνεται όμως το σ, γίνεται πιο αποδοτικός ο αλγόριθμος SVM, ώστε να σταθεροποιείται στη μέγιστη απόδοσή του για τιμές του σ κοντά στο 0,02.

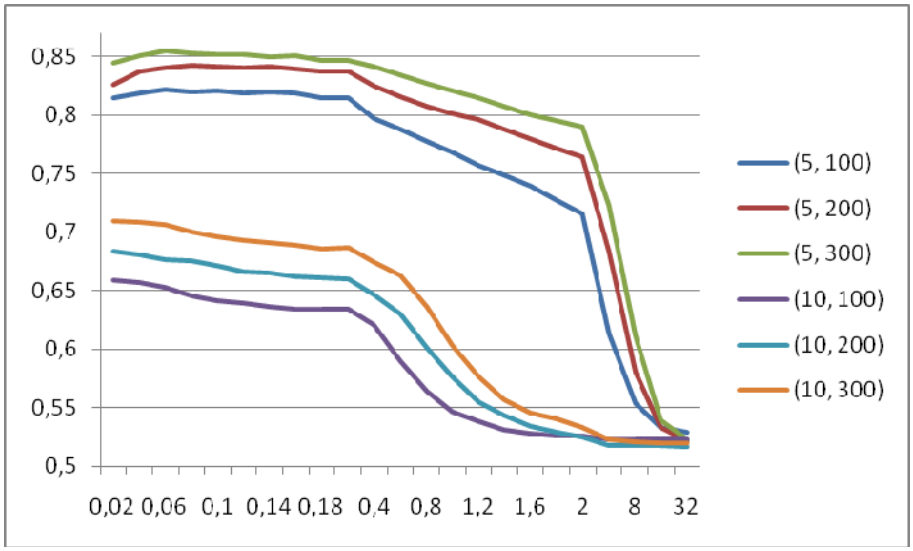
- Συμπεριφορά ως προς το πλήθος των ανεξάρτητων μεταβλητών.

Η αύξηση των μεταβλητών (από 5 σε 10) οδηγεί σε επακόλουθη μείωση της ακρίβειας (από επίπεδα του 0,85 σε αυτά του 0,71), κάτι το οποίο όπως αναφέραμε παρατηρείται γενικότερα. Επίσης η ευαισθησία του αλγορίθμου κατά την αύξηση των μεταβλητών αυξάνεται, καθώς παρατηρούμε από τις καμπύλες του σχήματος 4.3 ότι το εύρος των τιμών του σ για τις οποίες είναι αποδεκτή η ακρίβεια του αλγορίθμου μειώνεται κατά την αύξηση των παραμέτρων από 5 σε 10.

- Αλληλεπίδραση πλήθους μεταβλητών και μεγέθους δείγματος εκπαίδευσης.

Είναι ευδιάκριτη η βελτίωση στην ακρίβεια του αλγορίθμου τόσο στην περίπτωση του συνόλου δεδομένων με 5 όσο και του συνόλου με 10 μεταβλητές, καθώς αυξάνεται το πλήθος του συνόλου εκπαίδευσης.

Αξίζει επίσης να σημειωθεί ότι κατά την αύξηση του δείγματος εκπαίδευσης από 100 σε 200 και τέλος σε 300 εγγραφές, όπως συνέβη και κατά τη μελέτη των νευρωνικών δικτύων, δεν παρουσιάστηκε σύγκλιση της απόδοσης των διαφορετικών συνόλων, γεγονός που σημαίνει ότι ενδέχεται η περεταίρω αύξηση π.χ. σε 400 εγγραφές να οδηγήσει σε ακόμα καλύτερη εκπαίδευση και επομένως απόδοση του SVM.



Σχήμα 4.3. Ακρίβεια SVMs για διαφορετικά σύνολα τεχνητών δεδομένων, ως προς τιμή της παραμέτρου σ, όπου

(x,y) = (αριθμός μεταβλητών, μέγεθος συνόλου εκπαίδευσης)

4.2.2.4 Μηχανές διανυσμάτων υποστήριξης με ενσωμάτωση παραδειγμάτων μονοτονίας

Όπως και στην περίπτωση των απλών SVM αναζητούνται αρχικά οι τιμές των παραμέτρων C και σ, για τις οποίες έχουμε τη βέλτιστη δυνατή απόδοση. Τα αποτελέσματα του αλγορίθμου SVM που χρησιμοποιεί παραδείγματα μονοτονίας (hints) παρουσιάζονται στο σχήμα 4.4. Επισημαίνεται ότι για κάθε μια παρατήρηση

(είτε σε δείγμα ελέγχου, είτε σε δείγμα εκπαίδευσης) δημιουργήθηκαν για το κάθε χαρακτηριστικό 5πλάσιο πλήθος μονότονων παραδειγμάτων (δηλ. για εκπαίδευση με 100 παρατηρήσεις έγινε χρήση 500 μονότονων ζευγών).

Όπως είναι παρατηρείται στα αποτελέσματα, η τιμή της παραμέτρου C επηρεάζει σε πολύ μικρό βαθμό την απόδοση του αλγορίθμου (η μεταβολή της οδηγεί σε διακύμανση της απόδοσης σε επίπεδα της τάξης του 1-5%). Ειδικότερα, στην περίπτωση των δεδομένων με 5 μεταβλητές η αύξηση του C οδηγεί σε αύξηση (έως 5%) της απόδοσης, ενώ στην περίπτωση των δεδομένων με 10 μεταβλητές η αντίστοιχη μεταβολή του C οδηγεί σε μείωση (μέχρι 1%) της απόδοσης. Ωστόσο, υπάρχει εξάρτηση από την τιμή της παραμέτρου σ (όπως και στην περίπτωση των απλών SVM) όπου παρατηρείται ότι όσο η τιμή μειώνεται, τόσο βελτιώνεται η ακρίβεια.

Για το λόγο αυτό, όπως και στα απλά SVM για μελέτη της συμπεριφοράς του αλγορίθμου, κρατώντας την τιμή της παραμέτρου C σταθερή (C=2), ελέγξαμε τι συμβαίνει κατά τη μεταβολή του σ. Η τιμή του σ μειώθηκε μέχρι τα επίπεδα εκείνα στα οποία δεν εντοπίζαμε επιπλέον βελτίωση από επιπλέον μεταβολή, οπότε προέκυψε το γράφημα του σχήματος 4.4.

Από το γράφημα μπορούμε να εξάγουμε τα παρακάτω συμπεράσματα:

- Συμπεριφορά του αλγορίθμου ως προς την παράμετρο της συνάρτησης πυρήνα.

  Διαπιστώνεται αντίστοιχη συμπεριφορά όπως και στα απλά SVM, δηλαδή η μείωση της τιμής του σ οδηγεί σε αύξηση της ακρίβειας. Ωστόσο, παρατηρείται ότι το εύρος των τιμών του σ για τις οποίες παρουσιάζει ικανοποιητική απόδοση ο αλγόριθμος, είναι μικρότερο, δηλαδή δεν φτάνει μεγάλες τιμές μέχρι 32, αλλά μέχρι 1! Έτσι για σ περίπου ίσο με 1, έχουμε ουσιαστικά αδρανοποίηση του αλγορίθμου, (ακρίβεια 50%). Όσο μειώνεται όμως το σ, γίνεται πιο αποδοτικός ο

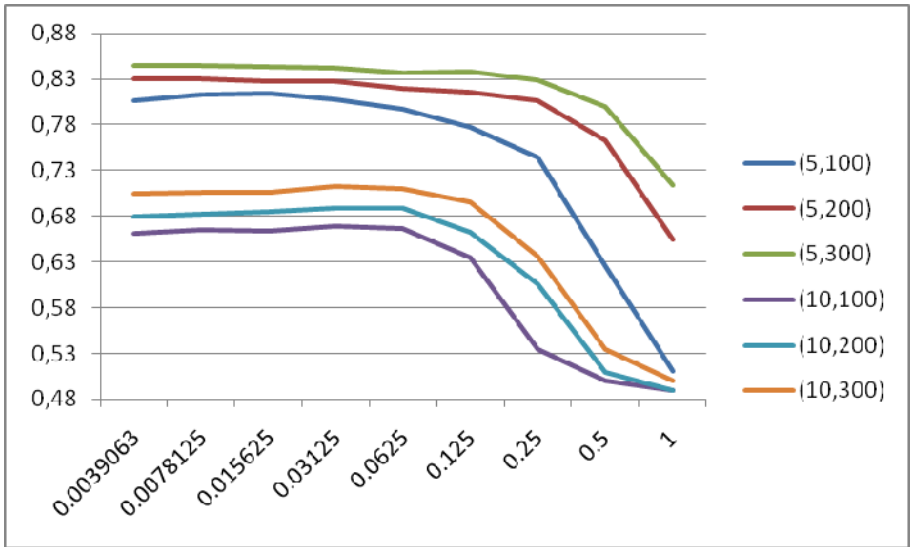
αλγόριθμος SVM, ώστε να σταθεροποιείται στη μέγιστη απόδοσή του για τιμές του σ κοντά στο 0,03.

- Συμπεριφορά ως προς το πλήθος των ανεξάρτητων μεταβλητών.

  Επιβεβαιώνεται το συμπέρασμα του ότι η αύξηση της τιμής των μεταβλητών (από 5 σε 10) οδηγεί σε μείωση της βέλτιστης απόδοσης του αλγορίθμου (από επίπεδα του 0,85 σε αυτά του 0,71), καθώς και η αύξηση της ευαισθησίας στις μεταβολές της παραμέτρου σ κατά την αύξηση του πλήθους των ανεξάρτητων μεταβλητών.

  - Αλληλεπίδραση πλήθους μεταβλητών και μεγέθους δείγματος εκπαίδευσης.

  Παρατηρούμε όπως και στις απλές μηχανές διανυσμάτων υποστήριξης, μια βελτίωση στην απόδοση του αλγορίθμου κατά την αύξηση του μεγέθους του συνόλου εκπαίδευσης.



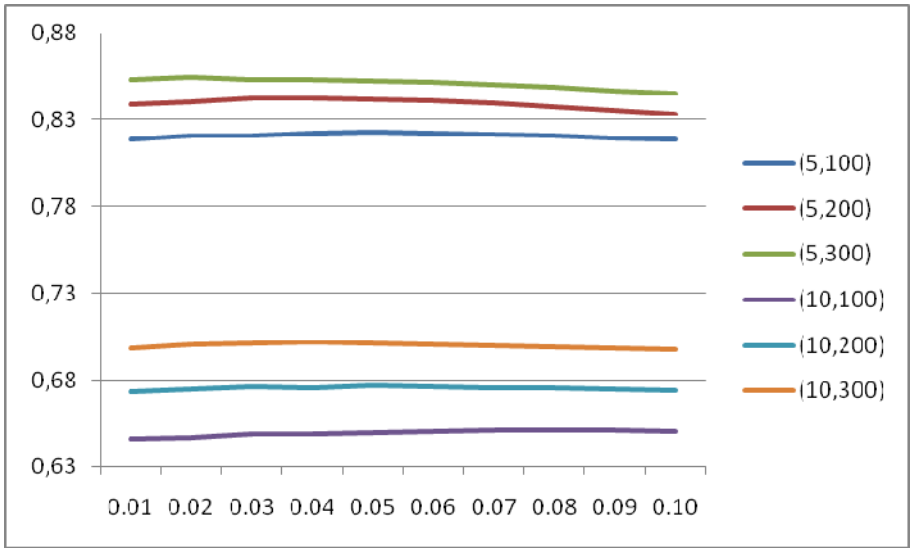
Σχήμα 4.4. Ακρίβεια SVMs με χρήση παραδειγμάτων μονοτονίας για διαφορετικά σύνολα τεχνητών δεδομένων, ως προς τιμή της παραμέτρου σ (με C=2), όπου

(x,y) = (αριθμός μεταβλητών, μέγεθος συνόλου εκπαίδευσης)

4.2.2.5 Μέθοδος UTADIS

Ο έλεγχος της μεθόδου UTADIS στα σύνολα τεχνητών δεδομένων παρουσιάζεται στο διάγραμμα του σχήματος 4.5, και έγινε με ρύθμιση – μεταβολή της παραμέτρου δ. Τα συμπεράσματα που προκύπτουν από τη μελέτη των σχημάτων είναι τα παρακάτω:

- Η ακρίβεια της μεθόδου εμφανίζεται να μην εξαρτάται από την τιμή της παραμέτρου δ, καθώς μένει ουσιαστικά σταθερή.
- Όσο αυξάνεται το μέγεθος του δείγματος εκπαίδευσης τόσο αυξάνεται η ακρίβεια στο δείγμα ελέγχου. Η συγκεκριμένη παρατήρηση είναι αναμενόμενη και συνδέεται άμεσα με το γεγονός ότι τα μικρά σχετικά δείγματα εκπαίδευσης είναι ανεπαρκή για την ικανοποιητική εκπαίδευση του αλγορίθμου.
- Και στο συγκεκριμένο αλγόριθμο εντοπίζεται μείωση της απόδοσης κατά την αύξηση των μεταβλητών του προβλήματος.



Σχήμα 4.5. Ακρίβεια πολυκριτήριας μεθόδου UTADIS για διαφορετικά σύνολα τεχνητών δεδομένων, για διαφορετικές τιμές παραμέτρου δ, όπου

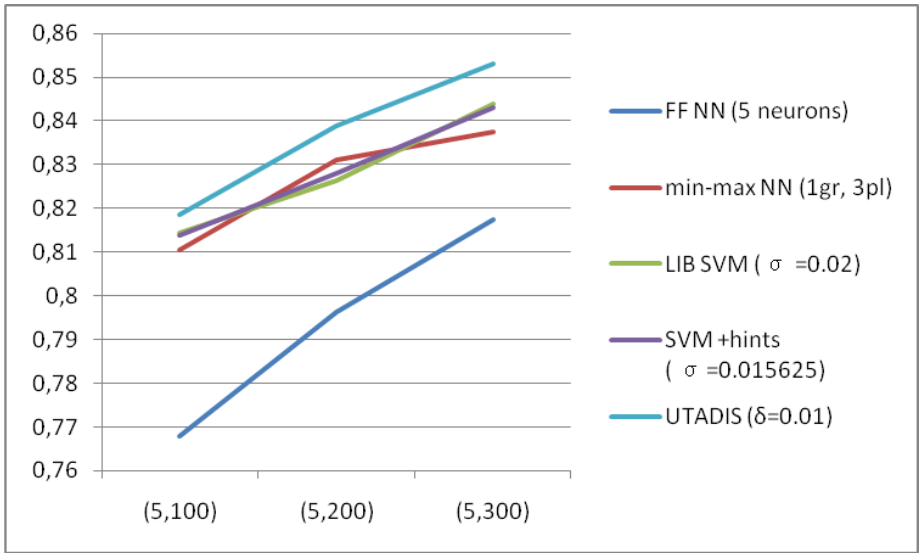
(x,y) = (αριθμός μεταβλητών, μέγεθος συνόλου εκπαίδευσης)

4.2.2.6 Συγκριτική αξιολόγηση μεθόδων

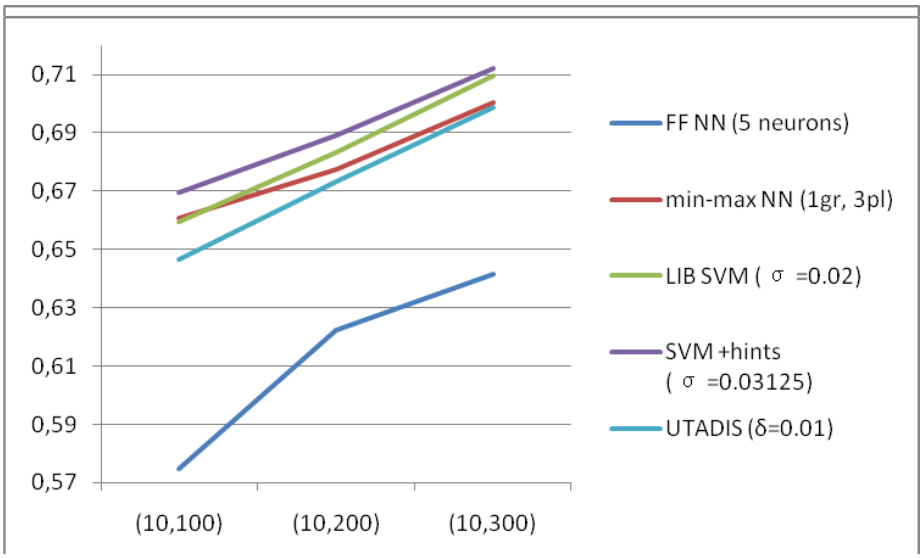
Έχοντας παρατηρήσει τη συμπεριφορά των αλγορίθμων ταξινόμησης για τα έξι διαφορετικά σύνολα δεδομένων εκπαίδευσης που δημιουργήσαμε, παραθέτουμε ακολούθως τον πίνακα 4.6 και τα αντίστοιχα γραφήματα στα σχήματα 4.6 και 4.7, που συνοψίζουν τα αποτελέσματα. Απο κάθε αλγόριθμο επιλέχθηκε η διαμόρφωση παραμέτρων που παρατηρήθηκε ότι είχε την μεγαλύτερη ακρίβεια.

Αλγόριθμος vs dataset	(5,100)	(5,200)	(5,300)
FF NN (5 neurons)	0,7680	0,7962	0,8175
min-max NN (1gr, 3pl)	0,8106	0,8310	0,8376
LIB SVM (σ=0.02)	0,8144	0,8263	0,8439
SVM +hints (σ=0.015625)	0,8138	0,8280	0,8430
UTADIS (δ=0.01)	0,8186	0,8389	0,8531
	(10,100)	(10,200)	(10,300)
FF NN (5 neurons)	0,5746	0,6225	0,6414
min-max NN (1gr, 3pl)	0,6607	0,6774	0,7004
LIB SVM (σ=0.02)	0,6593	0,6834	0,7096
SVM +hints (σ=0.03125)	0,6694	0,6892	0,7123
UTADIS (δ=0.01)	0,6465	0,6734	0,6987

Πίνακας 4.6 Σύγκριση αλγορίθμων σε τεχνητά δεδομένα



Σχήμα 4.6



Σχήμα 4.7

Από τα σχήματα μπορούν να εξαχθούν τα παρακάτω συμπεράσματα:

- Υπάρχει προφανής βελτίωση της απόδοσης του αλγορίθμου νευρωνικών δικτύων εξαιτίας της χρήσης της ιδιότητας της μονοτονίας (στην περίπτωση των 5 μεταβλητών από 2-5% σε ακρίβεια ταξινόμησης ενώ για 10 μεταβλητές ακόμα περισσότερο, από 6-8%).

- Παρουσιάζεται μικρή βελτίωση της ακρίβειας ταξινόμησης (της τάξεως του 1%) των μηχανών διανυσμάτων υποστήριξης κατά την χρήση παραδειγμάτων μονοτονίας για την εκπαίδευση τους, η οποία είναι περισσότερο εμφανής στην ταξινόμηση των συνόλων με τις 10 μεταβλητές.

- Η χρήση της πολυκριτήριας μεθόδου UTADIS (η οποία έχει μονότονη φύση) στις περισσότερες περιπτώσεις έχει μεγαλύτερη ακρίβεια ταξινόμησης από τις συμβατικές, μη μονότονες μεθόδους, εκτός των μηχανών υποστήριξης διανυσμάτων που αποδίδουν ελάχιστα καλύτερα (1%) στην περίπτωση των δεδομένων με 10 μεταβλητές. Μάλιστα στην περίπτωση των 5 μεταβλητών, η μέθοδος υπερισχύει όλων των υπολοίπων κατά τουλάχιστον 1% για ικανό δείγμα εκπαίδευσης (πλήθους πάνω από 100).

- Το σύνολο των μεθόδων έχει βελτίωση της ακρίβειας ταξινόμησης κατά την αύξηση του μεγέθους του συνόλου εκπαίδευσης, η οποία ωστόσο είναι περισσότερο εμφανής στα απλά νευρωνικά δίκτυα, τα οποία παρουσιάζουν και την χειρότερη απόδοση.

4.3 Πραγματικά Δεδομένα

4.3.1 Περιγραφή Πραγματικών Δεδομένων

Για την σύγκριση των αλγορίθμων με πραγματικά δεδομένα έγινε χρήση ενός συνόλου εγγραφών στοιχειών εμπορικών επιχειρήσεων της Ελλάδος, απαιτήθηκαν για αξιολόγηση έγκρισης δανείου, κατά τη χρονική περίοδο 1998-2003 (διαθέσιμα από την ICAP). Το σύνολο των χαρακτηριστικών αφορούσε την αξιολόγηση στα στοιχεία που παρουσιάζονται στον ακόλουθο πίνακα. Ο κάθε κριτήριο αξιολόγησης έχει θετική ή αρνητική σχέση μονοτονίας με την εκτίμηση για την αξιολόγηση της επιχείρησης, η οποία δηλώνεται με (+) ή (-) αντίστοιχα στον πίνακα 4.7.

Σχέση μονοτονίας	Χαρακτηριστικό
+	Ακίνητα
+	Αριθμός συνεργαζόμενων τραπεζών
+	Προσωπικό
+	Αποδοτικότητα συνολικών κεφαλαίων (προ φόρων)
+	Δείκτης αποσβέσεων παγίων
-	Συνολικές υποχρεώσεις/Παθητικό
-	Χρηματοοικονομικά έξοδα/Πωλήσεις
-	Μέσος όρος είσπραξης απαιτήσεων
-	Μέσος όρος εξόφλησης υποχρεώσεων
-	Κυκλοφοριακή ταχύτητα αποθεμάτων
+	Πωλήσεις/Ενεργητικό
+	Ειδική ρευστότητα
+	Λογάριθμος πωλήσεων
-	Αξία δυσμενών την τελευταία τριετία/Πλέον πρόσφατες πωλήσεις

Πίνακας 4.7 Χαρακτηριστικά αξιολόγησης επιχειρήσεων

Από το σύνολο των δεδομένων δημιουργήθηκαν:

- Το σύνολο εκπαίδευσης (1148 εγγραφές) με στοιχεία της περιόδου 1998-2000. Αποτελείται από όλες τις εγγραφές (574) επιχειρήσεων που αποδείχθηκαν ασυνεπείς, και από ισάριθμο πλήθος συνεπών επιχειρήσεων.
- Το σύνολο ελέγχου (23783 εγγραφές) με στοιχεία της περιόδου 2001-2003. Αποτελείται από 436 ασυνεπείς επιχειρήσεις και 23347 συνεπείς επιχειρήσεις.

Επισημαίνεται ότι για το σύνολο εκπαίδευσης δε χρησιμοποιήθηκαν όλες οι συνεπείς επιχειρήσεις της περιόδου, για σκοπούς ταχύτερης εκπαίδευσης των αλγορίθμων, καθώς ένα μεγαλύτερο δείγμα εκπαίδευσης δε κρίθηκε ότι θα εξυπηρετούσε σε βελτίωση της ακρίβειας των αλγορίθμων. Επίσης, η ακρίβεια των αλγορίθμων που υπολογίστηκε προέκυψε ως ο μέσος όρος της ακρίβειας ταξινόμησης σε κάθε μια από τις δυο ομάδες (συνεπείς / ασυνεπείς επιχειρήσεις), καθώς το πλήθος των συνεπών επιχειρήσεων στο δείγμα εκπαίδευσης είναι δυσανάλογα μεγάλο σε σχέση με το δείγμα των συνεπών.

4.3.2 Αποτελέσματα

4.3.2.1 Τεχνητά νευρωνικά δίκτυα και μονότονα νευρωνικά δίκτυα

Η ακρίβεια ταξινόμησης των τεχνητών νευρωνικών δικτύων στα πραγματικά δεδομένα παρουσιάζεται στον ακόλουθο πίνακα:

Αριθμός νευρώνων	5	10	15
Ακρίβεια	0,672	0,6954	0,6229

Παρατηρείται ότι για τα πραγματικά δεδομένα η βέλτιστη ακρίβεια ταξινόμησης εντοπίζεται για το νευρωνικό δίκτυο με 10 νευρώνες, και είναι κοντά στο 70%. Αντίθετα για ακόμα μεγαλύτερη αύξηση του πλήθους των νευρώνων

έχουμε μείωση κατά 7% περίπου της ακρίβειας (λόγω υπερβολικού ταιριάσματος στα δεδομένα εκπαίδευσης). Όταν το πλήθος των νευρώνων είναι μικρό (5) η απόδοση δεν είναι βέλτιστη, εξαιτίας της πολύπλοκης φύσης του μοντέλου που ο αλγόριθμος προσεγγίζει. Επομένως η συμπεριφορά του αλγορίθμου εξηγείται διαισθητικά σε ικανοποιητικό βαθμό.

Αντίστοιχα τα αποτελέσματα βάσει του αλγορίθμου μονότονων νευρωνικών δικτύων είναι αυτά του πίνακα 4.8.

		Αριθμός Επιπέδων ανά ομάδα		
		1	2	3
Αριθμός Ομάδων	1	0,6901	0,6918	0,6983
	2	0,6877	0,6985	0,7031
	3	0,6917	0,6992	0,7011

Πίνακας 4.8 Ακρίβεια μονότονων νευρωνικών δικτύων σε πραγματικά δεδομένα.

Από τα αποτελέσματα που παρουσιάζονται στον προηγούμενο πίνακα είναι εμφανής μια μικρή βελτίωση της ακρίβειας του αλγορίθμου κατά την αύξηση του αριθμού των επιπέδων ανά ομάδα. Δε μπορεί ωστόσο να προκύψει αντίστοιχο συμπέρασμα κατά τη μεταβολή του αριθμού των ομάδων που χρησιμοποιεί ο αλγόριθμος.

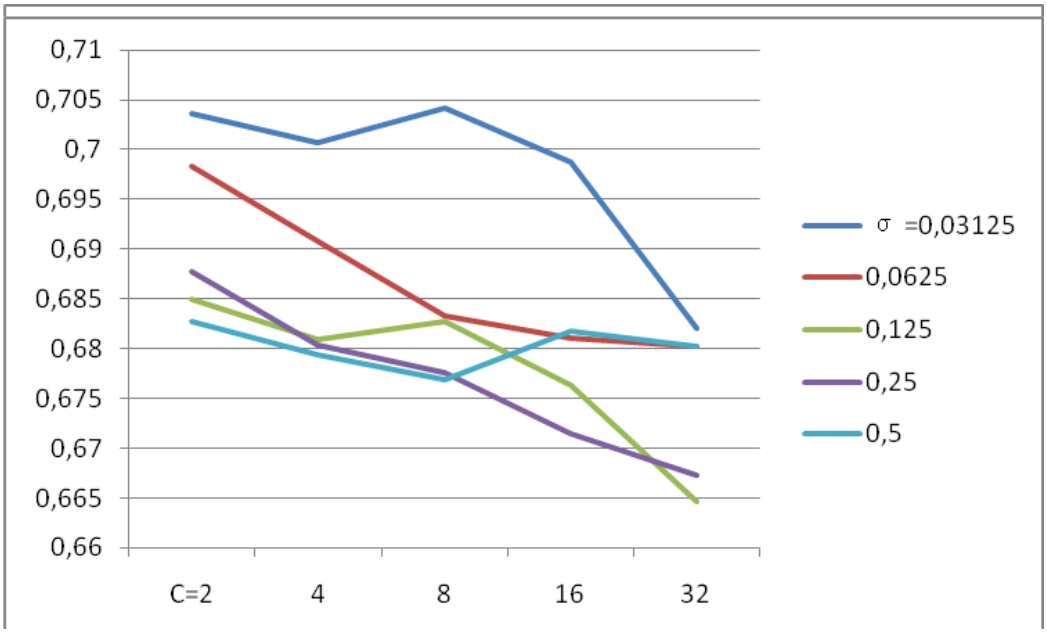
Συγκρίνοντας την συμπεριφορά των δυο αλγορίθμων προκύπτουν τα ακόλουθα:

- Τα μονότονα νευρωνικά δίκτυα έχουν μεγαλύτερη ακρίβεια ταξινόμησης, τουλάχιστον κατά 1%.
- Τα μονότονα νευρωνικά δίκτυα έχουν μικρότερη διακύμανση στην απόδοσή τους κατά τη μεταβολή των παραμέτρων τους, καθώς αυτή κυμαίνεται μεταξύ 69-70 %, σε αντίθεση με τα νευρωνικά δίκτυα που είναι περισσότερο ευαίσθητα κατά την αύξηση του πλήθους των νευρώνων.

4.3.2.2 Μηχανές διανυσμάτων υποστήριξης και μηχανές διανυσμάτων υποστήριξης με χρήση παραδειγμάτων μονοτονίας

Αφού έγινε έλεγχος της ακρίβειας του αλγορίθμου μηχανών υποστήριξης διανυσμάτων στα πραγματικά δεδομένα, καταλήξαμε στα ακόλουθα συμπεράσματα:

- Παρατηρείται μια μικρή εξάρτηση από την τιμή της παραμέτρου C. Η ακρίβεια του αλγορίθμου όταν γίνεται χρήση του δείγματος ελέγχου, έχει την τάση να μειώνεται για σχετικά μεγάλες τιμές του C.



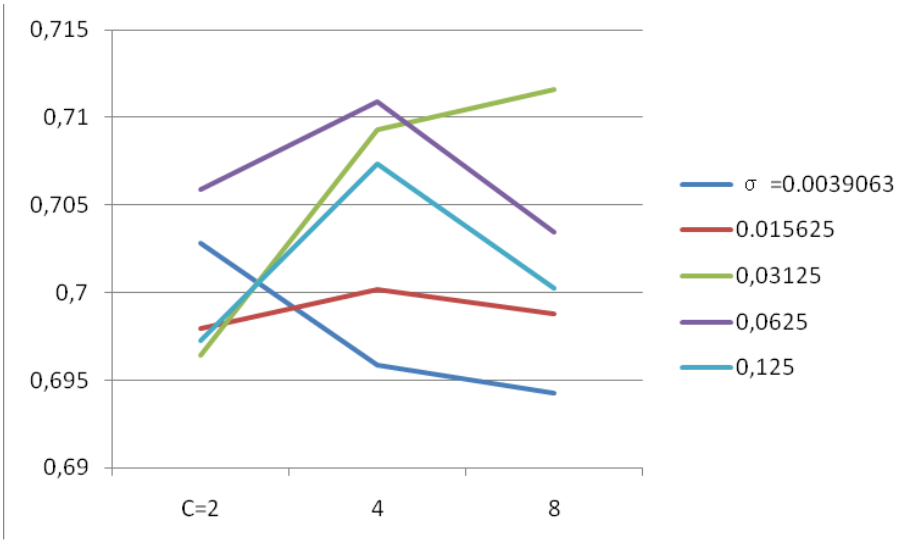
Σχήμα 4.8. Ακρίβεια SVM στο σύνολο ελέγχου.

- Εντοπίζεται σημαντική μεταβολή της ακρίβειας κατά την αλλαγή της παραμέτρου σ. Ειδικότερα, για αρκετά χαμηλές τιμές της σ (της τάξης του 0.03) έχουμε την καλύτερη ακρίβεια. Αντίθετα, για τις μεγαλύτερες τιμές (μέχρι και 32) παρατηρούμε ότι δεν έχει ικανότητα γενίκευσης ο αλγόριθμος, καθώς η ακρίβεια περιορίζεται κάτω από 60% για τα δεδομένα ελέγχου. Προφανώς, είναι επιθυμητή η πρώτη περίπτωση, οπότε εκτιμάται και ότι ο αλγόριθμος αποδίδει καλύτερα. Ταυτόχρονα όσο για χαμηλές τιμές της

παραμέτρου σ ο αλγόριθμος έχει αποδεκτή ακρίβεια, εντοπίζεται μια μικρή τάση μείωσής της κατά την αύξηση της τιμής της παραμέτρου C.

- Αξίζει να τονίσουμε ότι σχεδόν αντίστοιχα συμπεράσματα είχαν προκύψει και κατά τη χρήση της μεθόδου στα τεχνητά δεδομένα, δηλαδή ότι παρουσιαζόταν παρόμοια συμπεριφορά του αλγορίθμου κατά την μεταβολή των υπό μελέτη παραμέτρων.

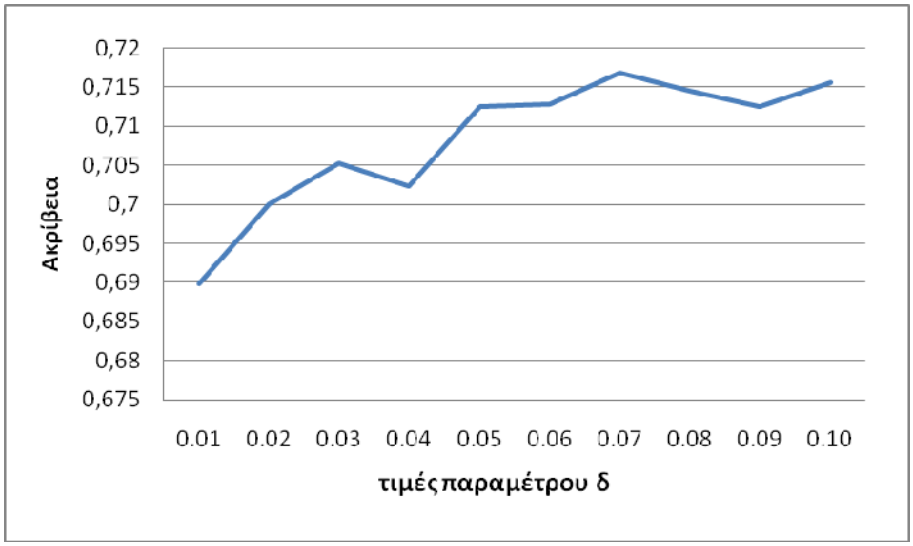
Ακολούθως, τα αποτελέσματα για τον αλγόριθμο SVM με χρήση παραδειγμάτων μονοτονίας που παρουσιάζονται στο σχήμα 4.9 επιβεβαιώνουν αντίστοιχη συμπεριφορά του αλγορίθμου με τα απλά SVM, ωστόσο εντοπίζεται μικρότερη ευαισθησία του αλγορίθμου κατά τη μεταβολή της παραμέτρου σ. Επιπλέον η ακρίβεια ταξινόμησης είναι καλύτερη κατά τουλάχιστον 0,5% από την αντίστοιχη που έχουμε όταν δε γίνεται εκμετάλλευση της μονότονης φύσης των πραγματικών δεδομένων.



Σχήμα 4.9. Ακρίβεια SVM με χρήση παραδειγμάτων μονοτονίας στο σύνολο ελέγχου.

4.3.2.3 Μέθοδος UTADIS

Στο διάγραμμα 4.10 απεικονίζεται η ακρίβεια του αλγορίθμου κατά τη χρήση του συνόλου ελέγχου. Παρατηρούμε ότι η ακρίβεια σταθεροποιείται για τιμές της παραμέτρου δ μεταξύ 0.05 και 0.1, ανάμεσα σε 70 και 72%.



Σχήμα 4.10. Ακρίβεια UTADIS στο σύνολο ελέγχου.

4.3.2.4 Συγκριτική αξιολόγηση μεθόδων

Με βάση τα προηγούμενα αποτελέσματα προκύπτει ο πίνακας 4.9 που παρουσιάζει για κάθε αλγόριθμο την βέλτιστη ακρίβεια ταξινόμησης που επετεύχθη καθώς και τις αντίστοιχες παραμέτρους.

Αλγόριθμος	Ακρίβεια
FF NN (10 neurons)	0,6954
min-max NN (2gr, 3pl)	0,7031
LIB SVM (σ=0,03125, C=8)	0,7041
SVM +hints (σ=0,0625, C=4)	0,7109
UTADIS (δ=0.07)	0,7169

Πίνακας 4.9 Σύγκριση αλγορίθμων στα δεδομένα αξιολόγησης του πιστωτικού κινδύνου

Είναι εμφανής η υπεροχή ως προς την ακρίβεια ταξινόμησης των μονότονων μεθόδων (νευρωνικά δικτύων και μηχανών διανυσμάτων υποστήριξης) ως προς τις απλούστερες μη μονότονες υλοποιήσεις τους. Τέλος, στα πραγματικά δεδομένα, κυριαρχεί όλων η πολυκριτήρια μέθοδος UTADIS, η οποία είναι από τη φύση της μονότονη.

4.4 Σύνοψη

Έχοντας αναλύσει τα αποτελέσματα των προσομοιώσεων για το σύνολο των αλγορίθμων και των συνόλων τεχνητών και πραγματικών δεδομένων, μπορούμε να συνοψίσουμε τα συμπεράσματα της ανάλυσης στα ακόλουθα σημεία:

- Η εκμετάλλευση της ιδιότητας της μονοτονίας οδηγεί σε βελτίωση της απόδοσης αλγορίθμων όπως τα νευρωνικά δίκτυα και οι μηχανές διανυσμάτων υποστήριξης.
- Τα συμπεράσματα είναι περισσότερο σαφή στην περίπτωση των τεχνητών δεδομένων, για τα οποία είναι ξεκάθαρη η μονότονή τους φύση λόγω του αλγορίθμου κατασκευής τους, ενώ επιβεβαιώνονται και στα πραγματικά δεδομένα.
- Η πολυκριτήρια μέθοδος UTADIS υπερισχύει των μη μονότονων μεθόδων των νευρωνικών δικτύων και μηχανών υποστήριξης αποφάσεων.
- Στο σύνολο των αλγορίθμων η αύξηση του μεγέθους του συνόλου εκπαίδευσης βελτίωνε την ακρίβεια ταξινόμησής τους.
- Η αύξηση του πλήθους των ανεξάρτητων μεταβλητών του προβλήματος οδηγεί σε μείωση της βέλτιστης ακρίβειας ταξινόμησης του συνόλου των αλγορίθμων.

ΚΕΦΑΛΑΙΟ 5ο ΕΠΙΛΟΓΟΣ & ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Η παρούσα εργασία επικεντρώθηκε στη μελέτη ενός προβλήματος ιδιαίτερα ενδιαφέροντος, με πλήθος εφαρμογών σε τομείς της επιστήμης, του προβλήματος της ταξινόμησης. Η συνεισφορά της κρίνεται σημαντική καθώς με κατά το δυνατόν πιο τεκμηριωμένο τρόπο, κάνοντας χρήση τόσο τεχνητών όσο και πραγματικών δεδομένων, σύγκρινε τους διαφόρους αλγόριθμους. Έτσι, έγινε αντιπαραβολή σύγχρονων μεθόδων, τόσο μονότονων όσο και μη, εξάχθηκαν ιδιαίτερα πολύτιμα συμπεράσματα, τα οποία καταδεικνύουν τη χρησιμότητα εκμετάλλευσης της εκ των προτέρων γνώσης της μονοτονίας, όταν αυτή υπάρχει στο υπό μελέτη πρόβλημα.

Βεβαίως πάντα υπάρχουν δυνατότητες περεταίρω έρευνας, οι οποίες ωστόσο ξεφεύγουν από τα πλαίσια της παρούσας εργασίας. Για λόγους μελλοντικής χρήσης παραθέτουμε ορισμένες επιπλέον διαστάσεις στις οποίες θα μπορούσε να επεκταθεί η προηγούμενη έρευνα.

→ Θα μπορούσε να διαμορφωθούν μικρότερα σύνολα πραγματικών δεδομένων από επιλεγμένες, κατά το δυνατόν ανεξάρτητες μεταξύ τους – βάσει θεωρίας - μεταβλητές και από διαφορετικές κατηγορίες χρημοτοοικονομικών δεικτών, ώστε να παρατηρήσουμε τη συμπεριφορά των αλγορίθμων ως προς το πλήθος των μεταβλητών και σε σύνολα πραγματικών δεδομένων.

→ Αντίστοιχη μελέτη μπορεί να γίνει και για το μέγεθος του δείγματος εκπαίδευσης, χρησιμοποιώντας μικρότερα ή μεγαλύτερα σύνολα για την εκπαίδευση των υπό μελέτη αλγορίθμων. Ωστόσο, επειδή η τυχαιότητα και ομοιόμορφη κατανομή δειγμάτων – εγγραφών στο σύνολο των πραγματικών δεδομένων δεν είναι εξασφαλισμένη (κάτι που δεν συμβαίνει στα τεχνητά δεδομένα για τα οποία γνωρίζουμε από ποια κατανομή προέρχονται) αυτό δεν θεωρήθηκε αναγκαίο και ενδεχομένως να μην οδηγούσε σε νέα συμπεράσματα πέρα από όσα ήδη επιβεβαιώσαμε.

→ Μια άλλη παράμετρος που μπορεί να εξεταστεί λεπτομερέστερα είναι η ευαισθησία των αλγορίθμων σχετικά με το πόσο μεταβάλλεται η ακρίβειά τους κατά την αλλαγή των παραμέτρων τους (οι οποίες είναι βέβαια διαφορετικές για τον κάθε αλγόριθμο). Όπως είναι προφανές είναι επιθυμητό αφενός να υπάρχει μικρή διακύμανση της ακρίβειας και αφετέρου να διατηρείται για όσο μεγαλύτερο εύρος τιμών των παραμέτρων. Επιπλέον μπορεί να επικεντρωθεί μια τέτοια μελέτη στις περιοχές στις οποίες κρίνεται αποδεκτή η ακρίβεια του κάθε αλγορίθμου, το οποίο είναι αντικείμενο που καθορίζεται σε μεγάλο βαθμό από την εκάστοτε εφαρμογή. Παρατηρήθηκε σε γενικές γραμμές να έχουν λιγότερη ευαισθησία στη μεταβολή των παραμέτρων τους οι αλγόριθμοι UTADIS και μονότονων νευρωνικών δικτύων ενώ ο αλγόριθμος νευρωνικών δικτύων είχε την μεγαλύτερη διακύμανση κατά τη μεταβολή του πλήθους των νευρώνων.

→ Μια τελευταία παράμετρος που δύναται εξεταστεί είναι αυτή του χρόνου εκπαίδευσης και εκτέλεσης του αλγορίθμου, η οποία αποτελεί βαρόμετρο σχετικά με το εύρος των εφαρμογών καθενός. Για παράδειγμα ένας αλγόριθμος με μεγάλο χρόνο εκτέλεσης δε μπορεί να χρησιμοποιηθεί για on-line εφαρμογές. Επίσης ένας αλγόριθμος με συγκριτικά μικρότερο χρόνο εκπαίδευσης ενδέχεται να επιτρέψει την συχνότερη ανανέωσή του με δεδομένα εκπαίδευσης για προσαρμογή σε πιθανές αλλαγές του μοντέλου που προσεγγίζει.

→ Τέλος, είναι σημαντικό να τονιστεί το γεγονός ότι υπάρχουν σημαντικά ακόμα περιθώρια στη βελτίωση των αλγορίθμων καθώς:

- Δεν έγινε βέλτιστη αρχικοποίηση των παραμέτρων στα νευρωνικά δίκτυα
- Δεν μελετήθηκαν εναλλακτικές μέθοδοι προ-επεξεργασίας των δεδομένων.

Βιβλιογραφία

1. Domingos et al (1997) "On the optimality of the simple Bayesian classifier under zero-one loss". Machine Learning, 29:103–137.

2. Rish I. (2001). "An empirical study of the naive Bayes classifier". IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence.

3. Keinosuke Fukunaga (1990). “Introduction to Statistical Pattern Recognition”. Academic Press, Second Edition

4. Rogovin M. (1980). “Three Mile Island: A report to the Commissioners and to the Public”, Volume I. Nuclear Regulatory Commission, Special Inquiry Group.

5. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). “Learning internal representations by error propagation”. In Rumelhart, D. E. and McClelland, J. L., editors, Parallel Distributed Processing, volume 1, pages 318-362. MIT Press.

6. Bishop C.M., Svensen M., and Williams C.K.I. (1997) “Gtm: A principle alternative to the self-organizing map”. In Advances in neural information processing systems, 5:354-360.

7. Vapnik, V. N., and Chervonenkis A. Ya., (1974). “Teoriya raspoznavaniya obrazov: Statisticheskie problemy obucheniya”. (Russian) [Theory of pattern recognition: Statistical problems of learning]. Moscow: Nauka.

8. Boser, B. E., Guyon I. M., and Vapnik V. N., 1992. “A training algorithm for optimal margin classifiers”. In: COLT ’92: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. New York, NY, USA: ACM Press, pp. 144–152.

9. Jacquet-Lagreze, E. & Siskos, J., (1982). "Assessing a set of additive utility functions for multicriteria decision-making, the UTA method," European Journal of Operational Research, Elsevier, vol. 10(2), pages 151-164, June.

10. Allwein E., Schapire R., and Singer Y. (2000). “Reducing multiclass to binary: A unifying approach for margin classifiers”, Journal of Machine Learning Research, vol. 1, pp113-141

11. Breiman L. (1996). “Bagging predictors. Machine Learning”, 26(2):123-140.

12. Jain A.K., Duin R., Mao J. (2000). “Statistical Pattern Recognition: A Review”, IEEE Transactions on Patern Analysis and Machine Intelligence, vol.22, No1, pp4-37

13. Michie D., Spiegelhalter D.J., Taylor C.C. (1994). “Machine Learning, Neural and Statistical Classification”, Prentice Hall, February 17

14. C. Burges (1998). “A Tutorial on Support Vector Machines for Pattern Recognition”. *Data mining and knowledge discovery*, Vol. 2, Nr. 2Springer, p. 121--167.

15. Doumpos M. and Zopounidis C., (2009). "Monotonic Support Vector Machines For Credit Risk Rating," New Mathematics and Natural Computation (NMNC), World Scientific Publishing Co. Pte. Ltd., vol. 5(03), pages 557-570.

16. Devaud, J.M., Groussaud, G. and Jacquet-Lagrèze, E. (1980). “UTADIS: Une méthode de construction de fonctions d'utilité additives rendant compte de jugements globaux”. European Working Group on Multicriteria Decision Aid. Bochum.

17. Vapnik V. N., (1998). “Statistical Learning Theory”, Wiley, New York

18. Zopounidis, C. and Doumpos, M. (1998). “Developing a multicriteria decision support system for financial classification problems: The FINCLAS system”. Optimization Methods and Software , 8, 277-304.

19. Zopounidis, C, (1999) . “Business failure prediction using the UTADIS multicriteria analysis method”. Journal of the Operational research Society, 50(11):1138.

20. Zopounidis, C. (1999). “A Multicriteria Decision Aid Methodology for Sorting Decision Problems: The Case of Financial Distress”.Computational Economics, 14(3):197-218.

21. Doumpos M. and Zopounidis C., (2002).“Multicriteria Decision Aid Classification Methods”. Kluwer Academic Publishers, Dordrecht.

22. Hellerstein, L. (1989). “On Characterizing and Learning some Classes of Read-Once Functions”. Doctoral Thesis. UMI Order Number: AAI9028868., University of California, Berkeley.

23. Karpf, J. (1991). “Inductive modelling in law: example based expert systems in administrative law”. In Proceedings of the 3rd international Conference on Artificial intelligence and Law (Oxford, England). ICAIL '91. ACM, New York, NY, 297-306.

24. Ben-David A., Sterling L., Pao Y.-H. (1989), “Learning and classification of monotonic ordinal concepts”, Computational Intelligence, v.5 n.1, p.45-49, February

25. Ben-David A., Sterling L., Tran T. (2009), “Adding monotonicity to learning algorithms may impair their accuracy”, Expert Systems with Applications: An International Journal, v.36 n.3, p.6627-6634, April

26. Moshkovich H. M. , Mechitov A. I., Olson D. L. (2002).“Ordinal judgments in multiattribute decision analysis”. Original Research Article European Journal of Operational Research, Volume 137, Issue 3, 16 March, Pages 625-641

27. Potharst, R. and Van Wezel, M., (2005). "Generating artificial data with monotonicity constraints," Econometric Institute Report EI 2005-06, Erasmus University Rotterdam, Econometric Institute.

28. Sill, J. and Abu-Mostafa Y. S. (1997), "Monotonicity Hints", Advances in Neural Information Processing Systems, vol 9, 634-640.

29. Sill, J. and Abu-Mostafa, Y. (1997). “Monotonicity: theory and implementation”. In intelligent Methods in Signal Processing and Communications Birkhauser Boston, Cambridge, MA, pages 129-146.

30. Abu-Mostafa, Y. (1990). “Learning from Hints in Neural Networks”. Journal of Complexity 6, pages 192-198.

31. Gunn, S. R., (1998). “Support Vector Machines for Classification and Regression”, Technical Report, UNIVERSITY OF SOUTHAMPTON, 10 May

32. Sill, J. (1998). “Monotonic networks”. In Proceedings of the 1997 Conference on Advances in Neural information Processing Systems 10 (Denver, Colorado, United States). M. I. Jordan, M. J. Kearns, and S. A. Solla, Eds. MIT Press, Cambridge, MA, 661-667.

33. Velikova, M. V., (2006). “Monotone Models for Prediction in Data Mining”, Nov 2006, Ph. D. Thesis, Tilburg University

34. Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput.10, 7 (Oct. 1998), 1895-1923.

35. Potharst, R. and Feelders, A. J. (2002). Classification trees for problems with monotonicity constraints. SIGKDD Explor. Newsl.4, 1, Jun.

36. Dembczyński K., Kotłowski W. and Słowińsk R. (2008)i, “Ensemble of Decision Rules for Ordinal Classification with Monotonicity Constraints”, ROUGH SETS AND KNOWLEDGE TECHNOLOGY, Lecture Notes in Computer Science, 2008, Volume 5009/2008, 260-267

37. Kotłowski, W. and Słowiński, R. (2009). “Rule learning with monotonicity constraints”. In Proceedings of the 26th Annual international Conference on Machine Learning (Montreal, Quebec, Canada, June 14 - 18). ICML '09, vol. 382. ACM, New York, NY, 537-544

38. Dembczyński, K., Kotłowski, W. and Słowiński, R. (2009). “Learning Rule Ensembles for Ordinal Classification with Monotonicity Constraints”, Fundamenta Informaticae 94(2): 163-178

39. Chih-Chung Chang and Chih-Jen Lin, “LIBSVM : a library for support vector machines, 2001”. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

40. Greco, S., and Matarazzo, B., and Slowinski, R. (1999), "Rough approximation of a preference relation by dominance relations", European Journal of Operational Research, 117, 63-83.

41. Greco, S., and Matarazzo, B., and Slowinski, R. (2001), "Rough sets theory for multicriteria decision analysis", European Journal of Operational Research, 129, 1-47.

42. Greco, S., and Matarazzo, B., and Slowinski, R. (2004), "Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules", European Journal of Operational Research, 158, 271-292.

43. Fortemps, P., and Greco, S., and Slowinski, R. (2008), "Multicriteria decision support using rules that represent rough-graded preference relations", European Journal of Operational Research, 188, 206-223.