ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΡΕΥΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΘΕΡΜΙΚΩΝ ΣΤΡΟΒΙΛΟΜΗΧΑΝΩΝ

Διπλωματική Εργασία :

ΜΗΧΑΝΕΣ ΔΙΑΝΥΣΜΑΤΩΝ ΥΠΟΣΤΗΡΙΞΗΣ (SUPPORT VECTOR MACHINES – SVM): ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΛΟΓΙΣΜΙΚΟΥ ΚΑΙ ΕΦΑΡΜΟΓΗ ΣΤΗ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

του

Δημητρίου Ι. Μητροδήμα

Επιβλέπων Καθηγητής:

Κυριάκος Χ. Γιαννάκογλου

Μάρτιος 2006

Εργαστήριο Θερμικών Στροβιλομηχανών
Τομέας Ρευστών,
Σχολή Μηχανολόγων Μηχανικών
Εθνικό Μετσόβιο Πολυτεχνείο

Διπλωματική Εργασία

"Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVM): Προγραμματισμός λογισμικού και εφαρμογή στη βελτιστοποίηση "

Του Δημητρίου Ι. Μητροδήμα

Μάρτιος 2006

Περίληψη

Παρότι η επεξεργαστική ισχύς έχει εμφανίσει ραγδαία αύξηση και το κόστος των υπολογιστών έχει μειωθεί πολύ, παραμένει στο επίκεντρο του ενδιαφέροντος η εξοικονόμηση υπολογιστικού χρόνου σε απαιτητικές εφαρμογές. Τυπικό παράδειγμα τέτοιων εφαρμογών είναι οι στοχαστικοί αλγόριθμοι βελτιστοποίησης, με κυριότερο εκπρόσωπο τους Εξελικτικούς Αλγορίθμους (ΕΑ). Τα τελευταία χρόνια, η χρήση διαφόρων μεθόδων επεξεργασίας δεδομένων για την επιτάχυνση της σύγκλισης των ΕΑ έχει αποτελέσει αντικείμενο μεγάλου ερευνητικού ενδιαφέροντος. Στην παρούσα εργασία, διερευνήθηκε μια μέθοδος επεξεργασίας δεδομένων με το όνομα Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – *SVM*). Η μέθοδος αυτή, που αναπτύχθηκε την τελευταία δεκαετία, έχει εμφανίσει αξιοσημείωτες επιδόσεις σε προβλήματα ταξινόμησης δεδομένων, ενώ την παρούσα στιγμή υπάρχει ανοιχτό ερευνητικό μέτωπο για την επιτυχή επέκτασή της και σε προβλήματα παλινδρόμησης. Βασίζεται στις αρχές της στατιστικής θεωρίας εκμάθησης, που ανέπτυξε ο Vapnik περί τα 1990, μέσω της οποίας παρέχεται μια στέρεα θεωρητική θεμελίωση για την απόδοσή της. Η μέθοδος προγραμματίστηκε σε γλώσσα Fortran 90 και εντάχθηκε στο λογισμικό Εξελικτικής

Βελτιστοποίησης του Εργαστηρίου Θερμικών Στροβιλομηχανών ΕΜΠ, σαν εργαλείο προεπιλογής υποψήφιων λύσεων. Η λειτουργία της δοκιμάστηκε τόσο σε μαθηματικά προβλήματα όσο και σε προβλήματα παραμετροποίησης στο αρχικό στάδιο της αεροδυναμικής βελτιστοποίησης, ενώ έγινε και μια πρώτη διερεύνηση των παραμέτρων που καθορίζουν τη βέλτιστη απόδοσή της.

Lab. of Thermal Turbomachines,
Fluids Section,
School of Mechanical Engineering,
National Technical University of Athens,

Diploma Thesis

"Support Vector Machines (SVM): Software programming and use in Optimization"

By Dimitris J.Mitrodimas

March 2006

Abstract

Although the power of processing systems has been greatly enhanced, while their cost is being continuously reduced, there is still a need to save computational time in demanding real-world applications. Stochastic optimization methods are typical examples of such applications, with their most well-known representative being Evolutionary Algorithms (EAs). Over the last few years there has been a great interest in applying data processing methods to optimization in order to reduce the cost of EAs. In the present diploma thesis, one such method, named Support Vector Machines (*SVM*), was investigated. Being developed over the last decade, *SVM* has shown remarkable performance in classification tasks, while there is still in progress research in application of the method to regression problems. *SVM* is based on the principles of statistical learning theory, developed by Vapnik around 1990, which gives rigid theoretical foundations to its performance. The method was programmed in Fortran 90 and incorporated in the Evolutionary Optimization software of the Thermal Turbomachines Laboratory of NTUA, as a candidate solution preselection tool applied within each generation. Furthermore, its performance was tested over mathematical and aerodynamic optimization problems, while

fine-tuning of the parameters defining its optimum performance was also examined.

Ευχαριστίες:

Επιβλέπων στην εκπόνηση της διπλωματικής μου εργασίας ήταν ο Αναπληρωτής Καθηγητής του Ε.Μ.Π. κ. Κυριάκος Χ. Γιαννάκογλου. Θα ήθελα να τον ευχαριστήσω θερμά για την ανάθεση του θέματος και για τη συνεχή καθοδήγηση και υποστήριξή του καθ' όλη τη διάρκεια εκπόνησης της εργασίας μου.

Θα ήθελα να εκφράσω επίσης τις θερμές μου ευχαριστίες και προς τους υποψήφιους διδάκτορες Ιωάννη Καμπόλη και Μάριο Καρακάση, για το χρόνο που αφιέρωσαν βοηθώντας με και προσφέροντάς μου πολύτιμες συμβουλές σε θέματα βελτιστοποίησης.

Δημήτρης Ι. Μητροδήμας

Περιεχόμενα:

Κεφάλαιο 1: Εξελικτική βελτιστοποίηση και χρήση της μεθόδου SVM

1.1 Βελτιστοποίηση μέσω Εξελικτικών Αλγορίθμων

Οι **Εξελικτικοί Αλγόριθμοι (Evolutionary Algorithms – EAs)** είναι μια από τις ευρύτερα χρησιμοποιούμενες στοχαστικές μεθόδους στη βελτιστοποίηση. Βασικό τους πλεονέκτημα που οδήγησε στην επικράτησή τους είναι το μη μαθηματικό τους υπόβαθρο και η ευκολία τους να προσαρμόζονται εύκολα σε κάθε νέο πρόβλημα, καθώς επίσης και η δυνατότητά τους να μην εγκλωβίζονται σε τοπικά ακρότατα, πράγμα που συμβαίνει με πολλές αιτιοκρατικές μεθόδους. Αρχικά αναπτύχθηκαν για την επίλυση προβλημάτων ενός στόχου, ωστόσο με κατάλληλες μετατροπές μπορούν να αντιμετωπίσουν και προβλήματα πολλαπλών στόχων.

Κύριο γνώρισμα των ΕΑ είναι ότι χειρίζονται **πληθυσμούς** υποψήφιων λύσεων, σε αντίθεση με αιτιοκρατικές αλλά και άλλες στοχαστικές μεθόδους (όπως π.χ. η μέθοδος της προσομοιούμενης ανόπτησης), που χειρίζονται μια μεμονωμένη λύση σε κάθε επανάληψη του αλγορίθμου. Οι πληθυσμοί αυτοί των λύσεων υπόκεινται σε διαδικασία εξέλιξης που προσομοιάζει με την εξέλιξη των φυσικών πληθυσμών στους ζωντανούς οργανισμούς, μέσω της οποίας τα γονίδια των επιτυχημένων ατόμων ή των ατόμων που έχουν προσαρμοσθεί καλύτερα στο περιβάλλον επιβιώνουν, και με την πάροδο των γενεών τα χαρακτηριστικά τους

μεταφέρονται σε μεγαλύτερο αριθμό απογόνων (Δαρβίνος, 1960). Μέσω αυτής της διαδικασίας εξέλιξης, οδηγούμαστε σε απογόνους που είναι ενδεχομένως καλύτεροι στα χαρακτηριστικά τους από ότι οι γονείς. Παρόμοια φιλοσοφία ακολουθείται και από τους ΕΑ. Οι υποψήφιες λύσεις του τρέχοντος πληθυσμού (τρέχουσα γενιά) αξιολογούνται με βάση κάποιο αντικειμενικό κριτήριο, υπόκεινται σε διαδικασία εξέλιξης, και οι ισχυρότερες λύσεις δίνουν περισσότερους απογόνους στις επόμενες γενιές. Με αυτό τον τρόπο, όσο προχωράει ο αλγόριθμος, θα έχουμε υποψήφιες λύσεις που στο σύνολό τους θα ικανοποιούν σε μεγαλύτερο βαθμό το κριτήριο βελτιστοποίησης, έως ότου, στη σύγκλιση της μεθόδου, οδηγηθούμε σε μια βέλτιστη λύση ή σε ένα σύνολο βέλτιστων λύσεων (αν πρόκειται για πρόβλημα πολλών στόχων).

1.2 *Περιγραφή ενός Εξελικτικού Αλγορίθμου*

Όπως προαναφέραμε οι ΕΑ χειρίζονται πληθυσμούς υποψήφιων λύσεων. Κατά την εξέλιξή τους, ένα σύνολο από μ άτομα (γονείς) εξελίσσεται σε ένα πληθυσμό λ απογόνων. Αυτοί οι απόγονοι είναι νέες λύσεις που προκύπτουν από τους μ γονείς, και που πιθανόν εμφανίζουν καλύτερα χαρακτηριστικά. Από τους λ απογόνους επιλέγονται με κριτήριο την καταλληλότητά τους, αυτοί που θα δώσουν τους μ γονείς της επόμενης γενιάς. Η διαδικασία αυτή συνεχίζεται έως ότου ικανοποιηθεί κάποιο κριτήριο σύγκλισης. Τέτοια κριτήρια στους ΕΑ υπάρχουν αρκετά, μεταξύ των οποίων είναι : (α) το να μην βελτιώνεται περαιτέρω η λύση για ένα αριθμό αξιολογήσεων ή γενιών, (β) το να έχει ομογενοποιηθεί επαρκώς ο πληθυσμός και (γ) το να έχει αναλωθεί ο υπολογιστικός χρόνος που επιτρέπει ο χρήστης.

Στα παρακάτω, αναφέρονται συνοπτικά τα βήματα που ακολουθούνται κατά την εφαρμογή ενός ΕΑ *[GIA03]*.

Βήμα 1: Επιλέγονται **τυχαία** τα μ μέλη του αρχικού πληθυσμού γονέων και τα λ μέλη του αρχικού πληθυσμού απογόνων. Η γενιά αυτή ονομάζεται και μηδενική γενιά.

Βήμα 2: Οι λ απόγονοι αξιολογούνται με βάση κάποια αντικειμενική συνάρτηση, και αποδίδεται έτσι σε κάθε έναν από αυτούς μια τιμή καταλληλότητας (fitness value).

Εν προκειμένω, για πρόβλημα βελτιστοποίησης μορφής σώματος (π.χ. αεροτομής, πτέρυγας ή πτερυγίου κτλ.) στην αεροδυναμική, απαιτείται για την αξιολόγηση κάθε ατόμου η αριθμητική επίλυση του πεδίου ροής, συνεπώς απαιτούνται λ κλήσεις του λογισμικού αριθμητικής επίλυσης των εξισώσεων Navier-Stokes. Αυτό το βήμα είναι που φέρει και το μεγαλύτερο υπολογιστικό κόστος.

Βήμα 3:

Με κάποιους μηχανισμούς, των οποίων η αναλυτική παρουσίαση ξεφεύγει από τους σκοπούς του παρόντος, επιλέγονται από το σύνολο των γονέων και των απογόνων, τα άτομα που θα στελεχώσουν το πληθυσμό γονέων της νέας γενιάς.

Βήμα 4:

Οι γονείς της νέας γενιάς υπόκεινται σε διαδικασία αναπαραγωγής, για να δώσουν τους απογόνους της νέας γενιάς. Επιλέγονται τυχαία δύο ή περισσότεροι γονείς, και εφαρμόζονται σε αυτούς διαδοχικά διάφοροι

τελεστές, μεταξύ των οποίων οι πιο γνωστοί είναι οι τελεστές διασταύρωσης και μετάλλαξης. Η εφαρμογή των τελεστών στους γονείς παράγει τους λ απογόνους της νέας γενιάς.

Βήμα 5:

Εφαρμόζεται το κριτήριο σύγκλισης και αν η μέθοδος θεωρείται ότι δεν έχει συγκλίνει αρχίζει μια νέα γενιά και επαναλαμβάνονται τα βήματα 2 έως 4. Αλλιώς, ο ΕΑ τερματίζει.

Οι ΕΑ έχουν βρει ευρεία εφαρμογή σε διάφορους τομείς βελτιστοποίησης. Στο Εργαστήριο Θερμικών Στροβιλομηχανών του ΕΜΠ υπάρχει μεγάλη εξοικείωση και εμπειρία στη χρήση τους στην αεροδυναμική βελτιστοποίηση. Για περισσότερες πληροφορίες, ο αναγνώστης παραπέμπεται ενδεικτικά στα *[GIA99],[GIO99],[GIA01],[KAR01],[EMM02],[KAMP04],[KAR04].* Έχει επιπλέον αναπτυχθεί από το εργαστήριο λογισμικό εξελικτικής βελτιστοποίησης, με την ονομασία E.A.SY (Evolutionary Algorithm SYstem).

1.3 Το μειονέκτημα των ΕΑ – τρόποι μείωσης υπολογιστικού χρόνου

Ωστόσο, βασικό μειονέκτημα των ΕΑ αποτελεί το γεγονός ότι σε κάθε επανάληψη της μεθόδου (δηλαδή σε κάθε νέα γενιά), απαιτείται η αξιολόγηση του συνόλου του τρέχοντος πληθυσμού (Βήμα 2 της μεθόδου). Ως αποτέλεσμα, αν ο υπολογισμός της τιμής της αντικειμενικής συνάρτησης (objective function evaluation – OFE) για ένα μεμονωμένο

άτομο είναι «ακριβή» (χρονοβόρος υπολογιστικά), ο συνολικός υπολογιστικός χρόνος που απαιτεί ο ΕΑ θα είναι απαγορευτικά μεγάλος. Το τελευταίο γίνεται ιδιαιτέρως εμφανές στην περίπτωση της αεροδυναμικής βελτιστοποίησης ή της αντίστροφης σχεδίασης μιας αεροτομής, όπου για την αξιολόγηση κάθε υποψήφιας λύσης απαιτείται κλήση λογισμικού αριθμητικής επίλυσης των εξισώσεων Navier – Stokes, δηλαδή κώδικα Υπολογιστικής Ρευστοδυναμικής (CFD) (στον οποίο συμπεριλαμβάνεται η γένεση του υπολογιστικού πλέγματος και κάθε σχετικό λογισμικό προ και μετ-επεξεργασίας), που ως γνωστόν κοστίζει πολύ σε χρόνο CPU.

Συνέπεια αυτού του γεγονότος είναι η στροφή των προσπαθειών στον τομέα της βελτιστοποίησης στην εύρεση τρόπων εξοικονόμησης υπολογιστικού χρόνου. Η γενική φιλοσοφία των μεθόδων εξοικονόμησης χρόνου είναι να περιοριστεί ο αριθμός των ακριβών αξιολογήσεων, δηλαδή για την περίπτωση της αεροδυναμικής ο αριθμός των μελών για τα οποία καλείται να τρέξει ο κώδικας Navier-Stokes σε κάθε γενιά, ή γενικότερα αυτό το οποίο στη συνέχεια θα αποκαλείται **«λογισμικό ακριβούς αξιολόγησης» (exact evaluation model)**. Είναι προφανές από τη διαδικασία της εξέλιξης σε έναν ΕΑ, ότι σε κάθε γενιά αξιολογούνται τόσο οι υποσχόμενες όσο και οι μη υποσχόμενες λύσεις, πράγμα που συνεπάγεται σπατάλη υπολογιστικού χρόνου σε λύσεις που τελικά δεν θα ανήκουν στις βέλτιστες.

Οι μέθοδοι που έχουν εφαρμοστεί για τον περιορισμό του κόστους των ΕΑ, μπορούν γενικά να καταταχθούν σε δύο κατηγορίες *[EMM02]*. Η πρώτη κατηγορία αφορά σε μεθόδους που χρησιμοποιούν **χαμηλότερης ακρίβειας μοντέλα αξιολόγησης (low-level evaluation models)** στη θέση του ακριβούς, με μικρότερη ακρίβεια στη μοντελοποίηση του φυσικού

φαινομένου προς βελτιστοποίηση, και συνεπώς με χαμηλότερο κόστος ανά αξιολόγηση σε σχέση με το ακριβές μοντέλο. Χαρακτηριστικό παράδειγμα είναι λ.χ. να υποκαθίσταται ο επιλύτης Navier-Stokes με μια γρήγορη μέθοδο επίλυσης των εξισώσεων του οριακού στρώματος σε ολοκληρωματική μορφή, συνοδευόμενη από μια διαδικασία αλληλεπίδρασης με την εξωτερική – ατριβή ροή όπου λύνονται, σε αραιό πλέγμα, οι εξισώσεις Euler ή αυτές της δυναμικής ροής. Στην κατηγορία αυτή, όλα τα μοντέλα αξιολόγησης επιλύουν με μεγαλύτερη ή μικρότερη ακρίβεια το ίδιο φυσικό πρόβλημα, με διαφορετικές παραδοχές το καθένα. Η δεύτερη κατηγορία αφορά σε μεθόδους που χρησιμοποιούν **υποκατάστατα μοντέλα προσέγγισης – εκτίμησης (surrogate or approximation models)** *[KAR04]* της επίδοσης κάθε υποψήφιας λύσης, λαμβάνοντας υπόψη τις επιδόσεις που είχαν εμφανίσει αντίστοιχες με αυτή λύσεις σε παρόμοια προβλήματα. Αυτά τα μοντέλα προσέγγισης ονομάζονται και **μεταπρότυπα**, και δρουν υποβοηθητικά στον ΕΑ, «φιλτράροντας» σε κάθε γενιά τα άτομα με (εκτιμώμενη) φτωχή επίδοση, και επιτρέποντας μόνο στα πλέον υποσχόμενα άτομα της γενιάς να υποστούν ακριβή αξιολόγηση.

Τα μεταπρότυπα απαιτούν προϋπάρχουσα γνώση πάνω στο πρόβλημα, δηλαδή την ύπαρξη μιας βάσης δεδομένων που να περιέχει προηγούμενα αξιολογηθείσες λύσεις και τις αντίστοιχες επιδόσεις τους. Βασιζόμενα σε αυτή τη γνώση, παράγουν μια εκτίμηση για την επίδοση μιας νέας υποψήφιας λύσης. Ένας συχνά χρησιμοποιούμενος τύπος μεταπροτύπου στην αεροδυναμική βελτιστοποίηση, είναι τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks – *ANNs*). Τα *ANN* συχνά αντικαθίστανται με άλλες μεθόδους από τα μαθηματικά ή τη στατιστική (λ.χ. μεθόδους παλινδρόμησης – regression models - , πολυωνυμικής παρεμβολής,

μοντέλα Gauss με κυριότερο εκπρόσωπο τη μέθοδο Kriging, κτλ.). Γενικά, η εφαρμογή ενός μεταπροτύπουυ στον ΕΑ βελτιστοποίησης έχει ως εξής :

1) Δημιουργείται ο αρχικός πληθυσμός της μηδενικής γενιάς, από τυχαία επιλεγμένα άτομα.

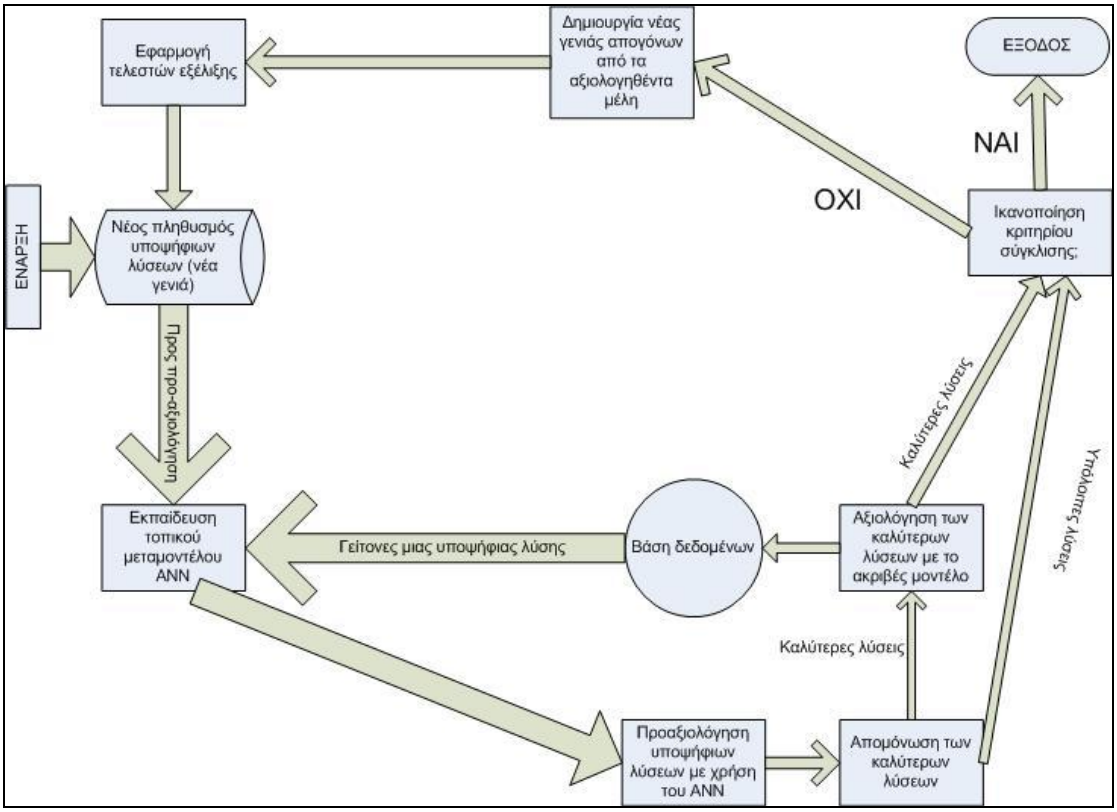
2) Ο πληθυσμός υπόκειται σε εξέλιξη για μερικές γενιές (συνήθως 2-3), κατά την οποία όλα τα άτομα κάθε γενιάς υφίστανται αξιολογήσεις με το λογισμικό ακριβούς αξιολόγησης. Τα άτομα αυτά και οι αντίστοιχες επιδόσεις τους αποθηκεύονται σε μία βάση δεδομένων (database – DB).

3) Στις επόμενες γενιές, και για κάθε μέλος της γενιάς, ανασύρονται από τη βάση δεδομένων οι «γείτονες» του μέλους αυτού, δηλαδή προηγούμενα αξιολογηθείσες λύσεις που συναντώνται στη «γειτονιά» του υπό εξέταση μέλους. Οι «γείτονες» αυτοί και οι αντίστοιχες (γνωστές) επιδόσεις τους, χρησιμοποιούνται για την εκπαίδευση ενός τοπικού μεταπροτύπουυ (local metamodel) στη γειτονιά του υπό εξέταση μέλους. Το εκπαιδευμένο μεταπρότυπο μπορεί τώρα να παρέχει μια εκτίμηση για την επίδοση του μέλους αυτού, βασισμένη στις επιδόσεις των «γειτόνων» του. Η ανωτέρω διαδικασία ονομάζεται **προσεγγιστική προ-αξιολόγηση (Inexact Pre-Evaluation – IPE)**, ακριβώς για να τονισθεί ότι η αξιολόγηση του μέλους δεν γίνεται με το ακριβές μοντέλο, αλλά εκτιμάται προσεγγιστικά με βάση τις επιδόσεις των γειτόνων του στη βάση δεδομένων. Σημειώνεται δε, ότι αυτή η διαδικασία κοστίζει ελάχιστα σε σχέση με την αξιολόγηση του μέλους από το λογισμικό ακριβούς αξιολόγησης.

4) Αφού εκπαιδευθεί ένα τοπικό μεταπρότυπο για κάθε μέλος της γενιάς, και συνεπώς γίνει μια προ-αξιολόγηση για όλα τα μέλη, επιλέγονται τα μέλη εκείνα με τις καλύτερες επιδόσεις, και αυτά μόνο στέλνονται για ακριβή αξιολόγηση, ενώ οι επιδόσεις τους καταχωρούνται στη βάση δεδομένων. Τα υπόλοιπα μέλη μένουν ως έχουν.

5) Στη συνέχεια εφαρμόζονται τα κλασικά βήματα ενός ΕΑ (επιλογή νέου πληθυσμού γονέων, εφαρμογή τελεστών εξέλιξης για δημιουργία των νέων απογόνων κτλ.) οπότε παράγεται η νέα γενιά απογόνων, και επαναλαμβάνεται η παραπάνω διαδικασία.

Η εφαρμογή των μεταμοντέλων στη βελτιστοποίηση μέσω ΕΑ, φαίνεται στο σχήμα 1:



Σχήμα 1: Εφαρμογή των μεταμοντέλων για προσεγγιστική προ-αξιολόγηση υποψήφιων λύσεων κατά την εφαρμογή ενός ΕΑ. Τονίζεται ο ρόλος της βάσης δεδομένων από την οποία αντλούνται πληροφορίες για τις ήδη αξιολογηθείσες λύσεις και τις επιδόσεις τους.

Περισσότερες λεπτομέρειες σχετικά με τα *ANNs* ή οποιοδήποτε από τα προαναφερθέντα μεταπρότυπα και την εφαρμογή τους ως εργαλείο προσεγγιστικής προ-αξιολόγησης υποψήφιων λύσεων στους ΕΑ, μπορεί κανείς να αναζητήσει σε αντίστοιχες εργασίες και δημοσιεύσεις του Εργαστηρίου Θερμικών Στροβιλομηχανών. Ενδεικτικά αναφέρουμε τις *[GIA01],[GIO99],[GGP00],[KAR01],[GIO01],[KGK05].*

1.4 Η προτεινόμενη μέθοδος – Προεπιλογή υποψήφιων λύσεων

Στην παρούσα εργασία διερευνήθηκε μια εναλλακτική αντιμετώπιση στο πρόβλημα της εξοικονόμησης του υπολογιστικού χρόνου που απαιτεί ο ΕΑ βελτιστοποίησης. Χρησιμοποιείται και πάλι ένα μεταπρότυπο με στόχο να «φιλτραριστούν» οι μη-υποσχόμενες λύσεις, στη βάση της ιδέας της προσεγγιστικής προ-αξιολόγησης που παρουσιάσθηκε προηγουμένως. Ωστόσο, ο τρόπος με τον οποίο «φιλτράρονται» αυτές είναι τώρα διαφορετικός. Αντί να προ-αξιολογούνται τα λ μέλη της νέας γενιάς και να βρίσκουμε έτσι μια εκτίμηση της επίδοσης κάθε μέλους, γίνεται τώρα μια απλή ταξινόμησή τους σε «καλά» και «κακά» μέλη. Η ταξινόμηση αυτή δεν προϋποθέτει την εύρεση μιας προσεγγιστικής τιμής για την επίδοση κάθε μέλους, αλλά απλά την κατάταξη κάθε μέλους σε μια εκ των δύο κατηγοριών («καλό» - «κακό»). Βέβαια, κριτήριο για την κατάταξη αυτή θα είναι, όπως και πριν, η προϋπάρχουσα γνώση που προσφέρει η βάση δεδομένων. Έτσι:

- Για κάθε μέλος από τα λ συνολικά της νέας γενιάς, εντοπίζονται μέσα στη βάση δεδομένων οι «γείτονές» του. Για το σκοπό αυτό χρησιμοποιούνται κριτήρια Ευκλείδειας απόστασης στο χώρο των μεταβλητών σχεδίασης.

- Οι «γείτονες» αυτοί, κατηγοριοποιούνται σε «καλούς» (δείκτης 1) και «κακούς» (δείκτης 0), ανάλογα με το αν η επίδοσή τους (η τιμή της αντικειμενικής συνάρτησης) βρίσκεται πάνω (μεγιστοποίηση) ή κάτω (ελαχιστοποίηση) από ένα προκαθορισμένο όριο.

- Η κατηγοριοποίηση αυτή των «γειτόνων» του υπό εξέταση μέλους, χρησιμεύει στην εκπαίδευση ενός τοπικού μεταπροτύπου, στη γειτονιά του υπό εξέταση μέλους. Το μεταπρότυπο «μαθαίνει» να ξεχωρίζει την «καλή»-υποσχόμενη από την «κακή»-μη υποσχόμενη

υποψήφια λύση. Διευκρινίζεται ότι, για τη φάση αυτή επαρκεί ένα μεταπρότυπο (όπως το *SVM* που θα μας απασχολήσει στη συνέχεια) που να μπορεί να επιλέγει αν το νέο μέλος είναι «καλό» ή «κακό», χωρίς να χρειάζεται το ίδιο το μεταπρότυπο να επιστρέφει και μια προσεγγιστική τιμή της συνάρτησης κόστους για το μέλος αυτό (όπως δηλαδή συνέβαινε με τα *ANNs).*

- Το υπό εξέταση μέλος της νέας γενιάς, αξιολογείται τώρα από το εκπαιδευμένο μεταπρότυπο ως «καλό» ή «κακό», ανάλογα με το αν τα χαρακτηριστικά του προσομοιάζουν τα χαρακτηριστικά των «καλών» ή «κακών» γειτόνων του στη βάση δεδομένων. Επιτυγχάνεται έτσι η απόδοση μιας ετικέτας («καλό»-«κακό») σε κάθε μέλος από τα λ συνολικά της υπό εξέταση γενιάς.

- Από τα λ μέλη της τρέχουσας γενιάς, επιλέγονται τα «καλά» μέλη, και τα καλύτερα από αυτά στέλνονται στο λογισμικό ακριβούς αξιολόγησης, ενώ τα υπόλοιπα («κακά») μέλη απορρίπτονται. Η επιλογή των καλύτερων μελών από τα μέλη που χαρακτηρίστηκαν ως «καλά» γίνεται αφού πρώτα τους αποδοθεί (με κάποια διαδικασία που θα περιγραφεί στο Κεφ.3) μια προσεγγιστική τιμή επίδοσης.

Η παραπάνω διαδικασία ονομάζεται **προεπιλογή υποψήφιων λύσεων**, ακριβώς για να τονιστεί το γεγονός ότι οι προς εξέταση λύσεις δεν προ-αξιολογούνται προκειμένου να ιεραρχηθούν ως προς την εκτιμώμενη επίδοσή τους (όπως γίνεται με τα *ANNs*), αλλά απλώς ταξινομούνται σε «καλές» και «κακές» και επιλέγονται μόνο οι «καλές» για ακριβή αξιολόγηση.

Για την προεπιλογή υποψήφιων λύσεων απαιτείται προφανώς κάποιο μεταπρότυπο που να κάνει ταξινόμηση σε δεδομένα. Ως τέτοιο, επιλέξαμε μια νέα σχετικά μέθοδο επεξεργασίας δεδομένων που ονομάζεται: **Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – *SVM*)**.

1.5 Σύντομη παρουσίαση της μεθόδου SVM και εφαρμογές

Όπως αναφέραμε και στην προηγούμενη παράγραφο, η μέθοδος των Μηχανών Διανυσμάτων Υποστήριξης (Support Vector Machines) είναι ένα εργαλείο για την επεξεργασία δεδομένων. Βασίζεται στη στατιστική θεωρία εκμάθησης, και αναπτύχθηκε από το Vapnik *[VAP95]* περί τα 1990.

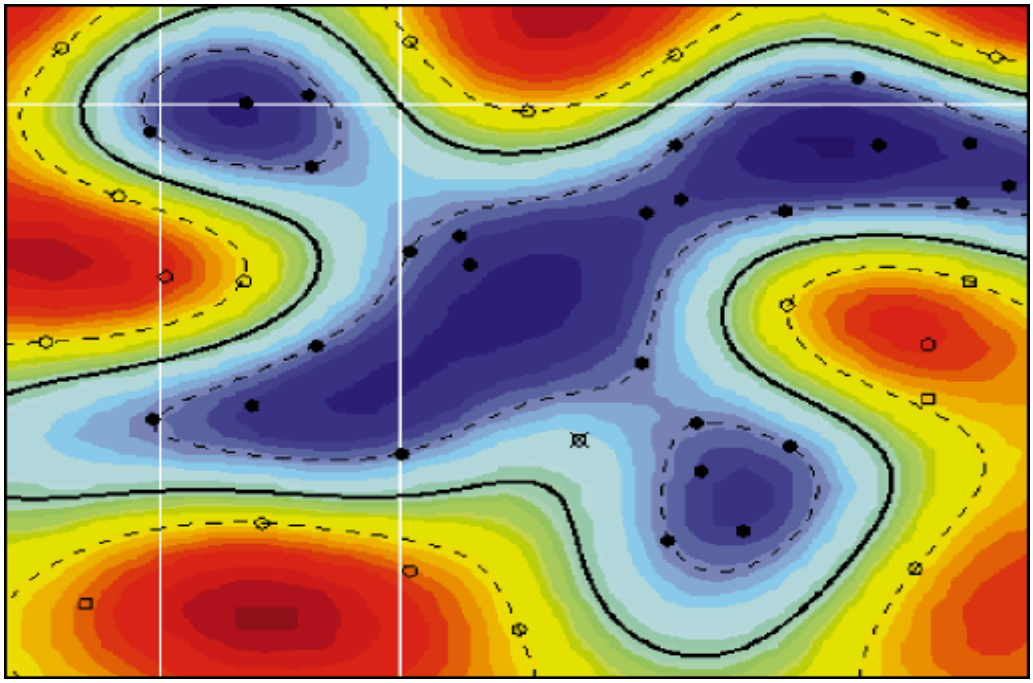
Η μέθοδος υπάγεται σε μια κατηγορία τεχνικών που είναι γνωστές με το όνομα **μηχανές εκμάθησης**. Στις μηχανές εκμάθησης ανήκουν και τα *ANN* που εξετάσαμε στην παράγραφο 1.3 . Όμοια με τα *ANN*, για να λειτουργήσει ο *SVM*, απαιτεί κάποια αρχική εκπαίδευση. Δηλαδή ένα σύνολο γνωστών εισόδων-εξόδων (inputs-outputs), το οποίο θα χρησιμοποιήσει για να «μάθει» την αντιστοίχιση μελλοντικών (αγνώστων) εισόδων στις σωστές εξόδους τους. Οι έξοδοι βέβαια στην περίπτωση της ταξινόμησης είναι στη μορφή «καλού»-«κακού» (1 ή 0). Ως αποτέλεσμα της εκπαίδευσης, ο αλγόριθμος θα μπορεί να παρέχει εκτιμήσεις για τις εξόδους μελλοντικών, αγνώστων εισόδων.

Η μέθοδος έχει χρησιμοποιηθεί με επιτυχία σε διάφορες περιπτώσεις ταξινόμησης δεδομένων (data classification), αναγνώρισης προτύπων (pattern recognition), και εκτίμησης συναρτήσεων παλινδρόμησης (regression

function estimation). Στα παρακάτω αναφέρουμε ενδεικτικά κάποια παραδείγματα εφαρμογής της μεθόδου *SVM*.

- Ταξινόμηση δεδομένων

Το ακόλουθο παράδειγμα *[HEA98]* επιδεικνύει τη χρήση της μεθόδου *SVM* στην περίπτωση της ταξινόμησης ενός συνόλου δεδομένων. Δίνεται ένα σύνολο σημείων στο 2Δ-χώρο, που ανήκουν σε δύο διαφορετικές κλάσεις: «κύκλοι» (σημεία χωρίς γέμισμα) και «δίσκοι» (σημεία με μαύρο γέμισμα), (βλ. σχήμα 2). Τα σημεία αυτά με τους αντίστοιχους χαρακτηρισμούς τους («κύκλος» ή «δίσκος») δίνονται στον *SVM* ως σύνολο εκπαίδευσης, από το οποίο καλείται να «μάθει» την αντιστοίχιση κάθε σημείου με την κλάση του, ώστε να μπορεί να ταξινομεί σωστά και αυτά, αλλά και νέα σημεία του 2Δ χώρου. Στόχος είναι να διαχωριστούν οι δύο κλάσεις, δηλαδή να απομονωθούν τα σημεία της μιας κλάσης από τα σημεία της άλλης. Όπως φαίνεται και από το σχήμα 2, με εφαρμογή της μεθόδου *SVM* χαράχθηκαν στο 2Δ επίπεδο τα διαχωριστικά όρια των δύο κλάσεων και έτσι απομονώθηκαν οι «κύκλοι» από τους «δίσκους». Αν τώρα δοθεί ένα νέο σημείο στο 2Δ χώρο το οποίο εμπίπτει στην περιοχή που ορίζουν τα όρια της κλάσης «κύκλος», αυτό θα χαρακτηριστεί από τον *SVM* ως «κύκλος». Αντίστοιχα, αν ένα νέο σημείο εμπίπτει στην περιοχή της κλάσης «δίσκος», θα χαρακτηριστεί ως «δίσκος».



Σχήμα 2: Το παράδειγμα του σχήματος δείχνει τα αποτελέσματα από την εφαρμογή του SVM στο πρόβλημα ταξινόμησης :"Διαχώρισε τους κύκλους από τους δίσκους". Όπως φαίνεται στο σχήμα, έχουν παραχθεί από τον *SVM* (μετά από την εκπαίδευση που του έγινε από τα εν λόγω σημεία), τα διαχωριστικά όρια της κάθε κλάσης. Έτσι, απομονώθηκαν οι «κύκλοι» από τους «δίσκους»

Με τον τρόπο αυτό, επιτυγχάνεται η ταξινόμηση νέων σημείων σε κλάσεις, με βάση την εκπαίδευση που έλαβε ο *SVM* από τα υπάρχοντα σημεία του σχήματος.

- Κατηγοριοποίηση κειμένου (text categorization)

Καθώς ο όγκος των ηλεκτρονικά διακινούμενων πληροφοριών ολοένα και αυξάνεται, υπάρχει επιτακτική ανάγκη να αναπτυχθούν εργαλεία που θα βοηθούν στην αποτελεσματικότερη διαχείρισή τους, την ταχύτερη εύρεση ενός εγγράφου και την ορθολογικότερη ταξινόμησή τους ανάλογα με το περιεχόμενό τους. Η ταξινόμηση εγγράφων σε κατηγορίες με βάση το περιεχόμενό τους (συγκεκριμένες – λέξεις κλειδιά που περιέχονται σε ένα έγγραφο), γίνεται μέχρι και τις μέρες μας από ανθρώπους. Μόλις τα τελευταία χρόνια έχουν αρχίσει να εφαρμόζονται μέθοδοι τεχνητής νοημοσύνης για την αυτοματοποίηση του εν λόγω έργου, και στον τομέα

αυτό, η συνεισφορά μηχανών εκμάθησης όπως η *SVM* μπορεί να αποδειχτεί ιδιαίτερα σημαντική.

Ο στόχος της αυτόματης κατηγοριοποίησης κειμένου είναι να αντιστοιχίζει νέα κείμενα σε μια ή περισσότερες από κάποιες προκαθορισμένες κατηγορίες, με βάση το περιεχόμενό τους και συγκεκριμένα κάποιες λέξεις-κλειδιά. Αυτές οι λέξεις αποτελούν τις εισόδους (inputs) της μηχανής εκμάθησης, και οι αποδιδόμενες κατηγορίες τις εξόδους (outputs). Οι *[JOA]* εκπαίδευσαν ένα *SVM*, δίνοντάς του σαν παραδείγματα κάποια δημοσιογραφικά κείμενα του πρακτορείου Reuters (υπό τη μορφή λέξεων-κλειδιά που περιέχονταν στα κείμενα), και τις αντίστοιχες κατηγορίες στις οποίες αυτά ανήκαν (π.χ. πολιτική, οικονομία, επικαιρότητα κτλ.). Για παράδειγμα, σε ένα κείμενο που αφορά στην οικονομία, ως λέξεις κλειδιά μπορούν να χρησιμοποιηθούν οι «τόκος», «πληθωρισμός», «κεφάλαια», ενώ σε ένα πολιτικό κείμενο οι λέξεις «βουλή», «υπουργός», «εξαγγελίες». Ο *SVM* «έμαθε» έτσι την αντιστοίχιση «λέξεις κλειδιά» - «κατηγορία». Στη συνέχεια δόθηκαν στον εκπαιδευμένο *SVM* κάποια νέα κείμενα, και του ζητήθηκε να τα αντιστοιχίσει σε μια από τις παραπάνω κατηγορίες. Τα αποτελέσματα ήταν ιδιαίτερα ενθαρρυντικά, καθώς σε σύνολο 118 κατηγοριών και 3.299 κειμένων, το 85.5% των κειμένων ταξινομήθηκαν στη σωστή τους κατηγορία. Σε σύγκριση δε με αντίστοιχες μεθόδους που είχαν χρησιμοποιηθεί στο παρελθόν για την κατηγοριοποίηση κειμένου, ο *SVM* εμφάνισε την καλύτερη επίδοση, και μάλιστα χρειάστηκε από 20 έως 50 φορές λιγότερο χρόνο CPU για να εκπαιδευθεί!

- Ανίχνευση προσώπων (face detection)

Το πρόβλημα της ανίχνευσης προσώπων ανήκει στην κατηγορία των προβλημάτων αναγνώρισης προτύπων (pattern recognition problems), και διατυπώνεται ως εξής: Δεδομένης μιας τυχαίας εικόνας (είτε στατικής είτε δυναμικής), να διαπιστωθεί αν στην εικόνα υπάρχουν ανθρώπινα πρόσωπα. Στον πυρήνα κάθε μεθόδου αναγνώρισης προσώπων βρίσκεται μια μηχανή εκμάθησης, η οποία εκπαιδεύεται στο να αναγνωρίζει αν σε μια εικόνα υπάρχει ανθρώπινο πρόσωπο ή όχι. Σε μια εφαρμογή του MIT Center for Biological and Computational Learning *[BUR],[ROW],[YAN]*, χρησιμοποιήθηκε ένας *SVM*, όπου δόθηκαν ως είσοδοι (inputs) κάποιες εικόνες, στις οποίες είχε αντιστοιχιστεί μια έξοδος (output): «face» ή «non-face» (αντίστοιχα 1 ή 0), ανάλογα με το αν αυτές οι εικόνες περιείχαν ή όχι ανθρώπινα πρόσωπα. Οι εικόνες αυτές με τις αντιστοιχίσεις τους εκπαίδευσαν τον *SVM* στο να αναγνωρίζει αν σε μια νέα εικόνα που θα του δοθεί, περιέχεται ή όχι ανθρώπινο πρόσωπο, και συνεπώς να της αποδίδει την αντίστοιχη ετικέτα «face» (1) ή «non-face» (0). Εν συνεχεία, στη φάση δοκιμής της μεθόδου, δόθηκαν στο σύστημα αναγνώρισης προσώπων κάποιες νέες εικόνες, οι οποίες χωρίστηκαν σε περιοχές. Κάθε περιοχή δόθηκε ως είσοδος στον εκπαιδευμένο *SVM*, όπου της αποδόθηκε από τον *SVM* μια έξοδος (ετικέτα): «face» ή «non-face». Στις περιοχές δε που χαρακτηρίστηκαν ως «face», σχεδιάστηκε γύρω τους και ένα τετράγωνο ώστε να απομονώσει την εικόνα του προσώπου από την υπόλοιπη εικόνα.

Σχήμα 3: Στις παραπάνω εικόνες, που δόθηκαν στον εκπαιδευμένο *SVM* στη φάση της δοκιμής του, έχουν ανιχνευθεί επιτυχώς τα ανθρώπινα πρόσωπα που περιέχονται σε αυτές, και έχουν περικλειστεί από ένα τετράγωνο.

Τα αποτελέσματα φαίνονται στις εικόνες του σχήματος 3, και καταδεικνύουν την αποτελεσματικότητα της μεθόδου *SVM* στο να αναγνωρίζει πρότυπα (patterns), εν προκειμένω ανθρώπινα πρόσωπα.

Τα παραπάνω παραδείγματα αποτελούν ορισμένες μόνο από τις πολλές εφαρμογές που έχει βρει η μέθοδος *SVM* σε προβλήματα επεξεργασίας δεδομένων. Στην παρούσα εργασία, όπως προαναφέραμε, η μέθοδος χρησιμοποιήθηκε ως μεταπρότυπο ταξινόμησης υποψήφιων λύσεων σε «καλές»-υποσχόμενες και «κακές»-μη υποσχόμενες, στα πλαίσια της αεροδυναμικής βελτιστοποίησης μέσω Εξελικτικών Αλγορίθμων. Στο επόμενο κεφάλαιο παρατίθενται οι λεπτομέρειες της μεθόδου, η θεωρητική της θεμελίωση και ο ακριβής τρόπος με τον οποίο λειτουργεί, με έμφαση στο πρόβλημα της ταξινόμησης δεδομένων, το οποίο και μας απασχόλησε.

Κεφάλαιο 2: Μηχανές Διανυσμάτων Υποστήριξης

Support Vector Machines

2.1 Εισαγωγή

Οι Μηχανές Διανυσμάτων Υποστήριξης, ή συντομογραφικά *SVM* (από την αγγλική τους ονομασία: Support Vector Machines), είναι μια τεχνική για την επεξεργασία δεδομένων η οποία έχει βρει τα τελευταία χρόνια εφαρμογή σε προβλήματα ταξινόμησης δεδομένων, παλινδρόμησης και εκτίμησης συναρτησιακών εξαρτήσεων. Η μέθοδος *SVM* υπάγεται σε μια

ευρύτερη κατηγορία τεχνικών, που είναι γνωστές ως **μηχανές εκμάθησης (learning machines)**.

Γενικά, ο στόχος μιας μηχανής εκμάθησης είναι ο εξής: Δίνεται ένα σύνολο l σημείων $\vec{x}_i \in \Re^N$, καθώς και οι αντίστοιχες τιμές $y_i \in \Re$ που παίρνει η άγνωστη και προς εκμάθηση συνάρτηση $y = y(\vec{x})$ σε κάθε ένα από αυτά τα σημεία. Ζητούμενο είναι η μηχανή να «μάθει» «στοιχεία» αυτής της συνάρτησης, δηλαδή να μπορέσει να εξάγει από τη δεδομένη πληροφορία τη συναρτησιακή εξάρτηση των $\vec{x}_i, y_i$ που κρύβεται από πίσω. Με άλλα λόγια, η μηχανή καλείται να μάθει την αντιστοίχιση:

$$\vec{x}_i \to y_i$$

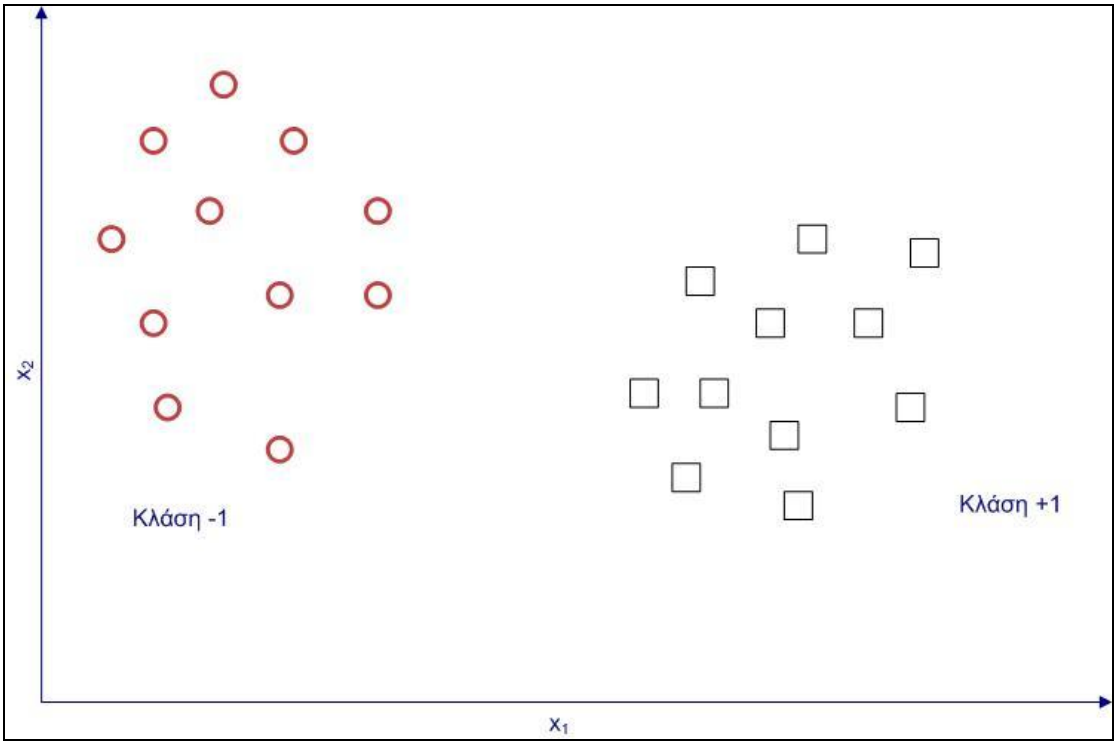
με μόνα δεδομένα το σύνολο $\{(\vec{x}_i, y_i), i = 1...l\}$. Το σύνολο σημείων $\vec{x}_i$ - τιμών της συνάρτησης y_i που δίνουμε στο *SVM* για να εκπαιδευτεί ονομάζεται **σύνολο εκπαίδευσης (training set)**. Τα σημεία $\vec{x}_i$ του συνόλου, ονομάζονται και **πρότυπα εκπαίδευσης (training patterns)**, ενώ οι τιμές y_i **στόχοι εκπαίδευσης (training targets)**. Τα πρότυπα εκπαίδευσης $\vec{x}_i$ αποτελούν τις εισόδους (inputs) της μηχανής εκμάθησης, ενώ οι αντίστοιχοι στόχοι εκπαίδευσης y_i τις επιθυμητές εξόδους (outputs). Η μηχανή καλείται να μάθει την ανωτέρω αντιστοιχία εισόδων-εξόδων, ώστε να χρησιμοποιήσει τη γνώση αυτή σε νέες άγνωστες εισόδους που θα της παρουσιαστούν. Κατ' αυτόν τον τρόπο, όταν εμφανιστεί μια νέα είσοδος $\vec{x}*$, θα μπορεί να παρέχει μια αντιστοίχιση για την τιμή της εξόδου της $y*$.

Η ανωτέρω διατύπωση του **προβλήματος εκμάθησης (learning task)** αφορά στη γενική περίπτωση που οι τιμές y_i της συνάρτησης ανήκουν στο σύνολο $\Re$ των πραγματικών αριθμών. Αυτό είναι γνωστό και ως «**γενικό πρόβλημα παλινδρόμησης**» **(general regression problem)**. Στην παρούσα εργασία, ασχοληθήκαμε με μια ειδική περίπτωση του προβλήματος εκμάθησης, αυτήν της ταξινόμησης δεδομένων, στην οποία η μηχανή εκμάθησης καλείται να μάθει την αντιστοίχιση :

$$\vec{x}_i \to y_i$$

όπου όμως $y_i \in \{-1, +1\}$.

Δηλαδή, στην περίπτωση αυτή, οι τιμές που μπορεί να πάρει η προς εκμάθηση συνάρτηση δεν ανήκουν στο σύνολο των πραγματικών αριθμών, αλλά είναι μόνο δύο: +1 ή -1. Σε κάθε πρότυπο $\vec{x}_i$ αντιστοιχίζεται μια ακέραια τιμή y_i, η οποία δείχνει σε ποια κλάση – κατηγορία ανήκει αυτό (+1 ή -1). Ζητούμενο από τη μηχανή εκμάθησης (εν προκειμένω από τον *SVM*) είναι τότε, με δεδομένα αυτά τα σημεία, να μάθει την αντιστοίχιση κάθε προτύπου με την κλάση του, με άλλα λόγια να μάθει να τα ταξινομεί σωστά. Το σχήμα 1 δείχνει μια απλή περίπτωση προβλήματος ταξινόμησης:

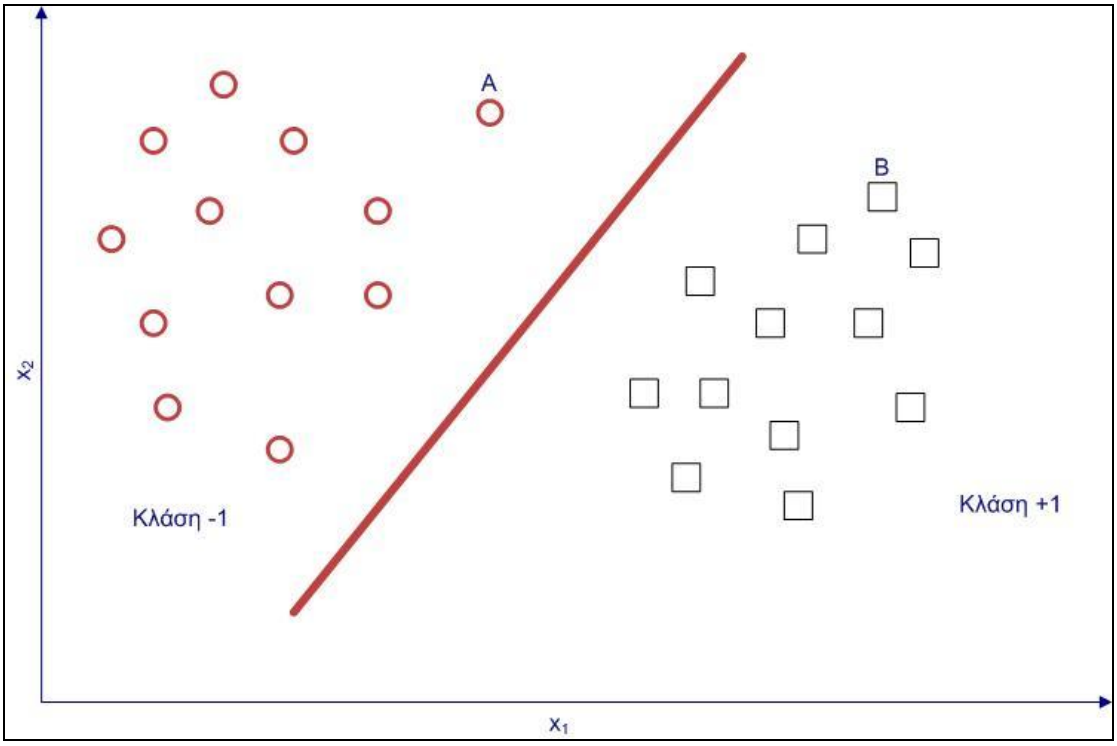


Σχήμα 1: Τυχαίο σύνολο σημείων εκπαίδευσης (training set) στο 2-Δ χώρο

Δίνονται κάποια σημεία στο 2-Δ χώρο, για τα οποία εμείς ξέρουμε εκ των προτέρων σε ποια κλάση ανήκει το καθένα. Συγκεκριμένα, κάποια από αυτά είναι μαρκαρισμένα ως «τετράγωνα», και τα υπόλοιπα ως «κύκλοι». Χωρίς βλάβη της γενικότητας, θεωρούμε ότι τα «τετράγωνα» ανήκουν στην κλάση +1 ενώ οι «κύκλοι» στην -1. Οι συντεταγμένες αυτών των σημείων αποτελούν τα πρότυπα εκπαίδευσης $\vec{x}_i$ του προβλήματος εκμάθησης. Οι τιμές εξόδου +1 ή -1, που δείχνουν σε ποια κατηγορία («τετράγωνο» ή «κύκλος») ανήκει το κάθε σημείο, αποτελούν τους στόχους εκπαίδευσης. Ειδικά για το πρόβλημα της ταξινόμησης, οι τιμές αυτές y_iονομάζονται και **ετικέτες εκπαίδευσης (training labels)**, ενώ τα διανύσματα $\vec{x}_i$ ονομάζονται και **διανύσματα παραμέτρων (attribute vectors).** Το σύνολο $\{(\vec{x}_i, y_i), i = 1...l\}$ αποτελεί το σύνολο εκπαίδευσης της μηχανής.

Ζητούμενο τώρα από τον *SVM*, είναι να μπορέσει να χρησιμοποιήσει το ανωτέρω σύνολο εκπαίδευσης, για να μάθει την αντιστοίχιση $\vec{x}_i \to y_i$. Αποτέλεσμα αυτής της διαδικασίας εκμάθησης θα είναι, όταν δοθεί ένα νέο σημείο του οποίου την κλάση δεν γνωρίζουμε, να μπορέσει ο *SVM* να προβλέψει σωστά σε ποια κλάση ανήκει αυτό, δηλαδή να το αντιστοιχίσει +1 ή -1.

Ο χώρος στον οποίο κείνται τα διανύσματα παραμέτρων $\vec{x}_i$ ονομάζεται και **χώρος εισόδων (input space)**. Στην περίπτωση αυτή ο χώρος εισόδων είναι το επίπεδο $\Re^2$. Είναι προφανές ότι για να γίνει ο διαχωρισμός των σημείων του συνόλου εκπαίδευσης, πρέπει να χαραχθεί στο χώρο εισόδων μια διαχωριστική γραμμή, από τη μια πλευρά της οποίας θα βρίσκονται τα σημεία της πρώτης κλάσης, ενώ από την άλλη πλευρά θα βρίσκονται τα σημεία της δεύτερης. Μια τέτοια γραμμή φαίνεται στο σχήμα 2. Ως αποτέλεσμα, ο χαρακτηρισμός ενός νέου σημείου ως «τετράγωνο» ή «κύκλος» από τον *SVM*, θα σχετίζεται με το ημιεπίπεδο εκατέρωθεν της ευθείας, στο οποίο κείται το σημείο αυτό.



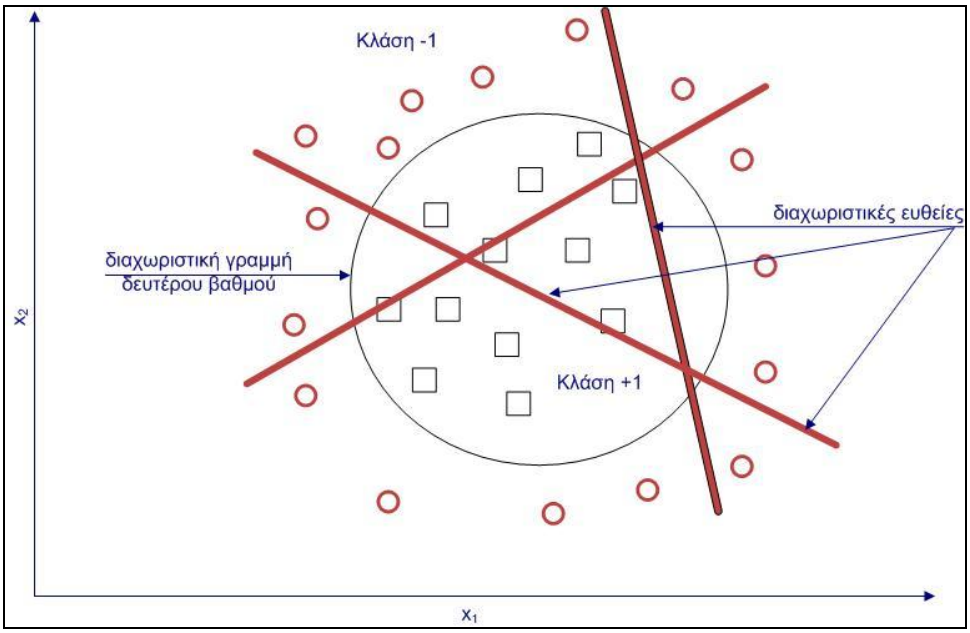
Σχήμα 2: Διαχωρισμός των σημείων του συνόλου εκπαίδευσης σε δύο κλάσεις με μια ευθεία. Η ταξινόμηση ενός νέου σημείου εξαρτάται από την πλευρά της διαχωριστικής ευθείας στην οποία αυτό κείται.

Για παράδειγμα, στο σχήμα 2 φαίνονται και δύο νέα σημεία, τα Α και Β. Για αυτά τα σημεία γνωρίζουμε τα διανύσματα παραμέτρων τους, αλλά όχι και τις ετικέτες τους, και ζητάμε από τον *SVM* να τους τις αποδώσει, με βάση την εκπαίδευση που του έγινε πριν. Αποτέλεσμα της εκπαίδευσης ήταν να χαραχθεί στο χώρο εισόδων η διαχωριστική γραμμή του σχήματος 2.

Έτσι, επειδή το Α βρίσκεται αριστερά της διαχωριστικής γραμμής (εκεί που βρίσκονται δηλαδή και οι «κύκλοι» του συνόλου εκπαίδευσης), του αποδίδεται ετικέτα -1 και θεωρείται «κύκλος». Αντίστοιχα, επειδή το Β βρίσκεται δεξιά της διαχωριστικής γραμμής (όπου βρίσκονται και τα «τετράγωνα» του συνόλου εκπαίδευσης), του αποδίδεται ετικέτα +1 και θεωρείται «τετράγωνο».

Στο σημείο αυτό υπεισέρχονται ορισμένα ερωτήματα για τη διαδικασία διαχωρισμού:

- Τι μορφή θα έχει η διαχωριστική γραμμή; Στο προηγούμενο παράδειγμα ήταν εμφανές ότι τα σημεία του συνόλου εκπαίδευσης μπορούσαν εύκολα να διαχωριστούν με μια ευθεία, ωστόσο αυτό δεν είναι κανόνας. Για παράδειγμα, το παρακάτω σύνολο σημείων (σχήμα 3) δεν μπορεί να διαχωριστεί γραμμικά:

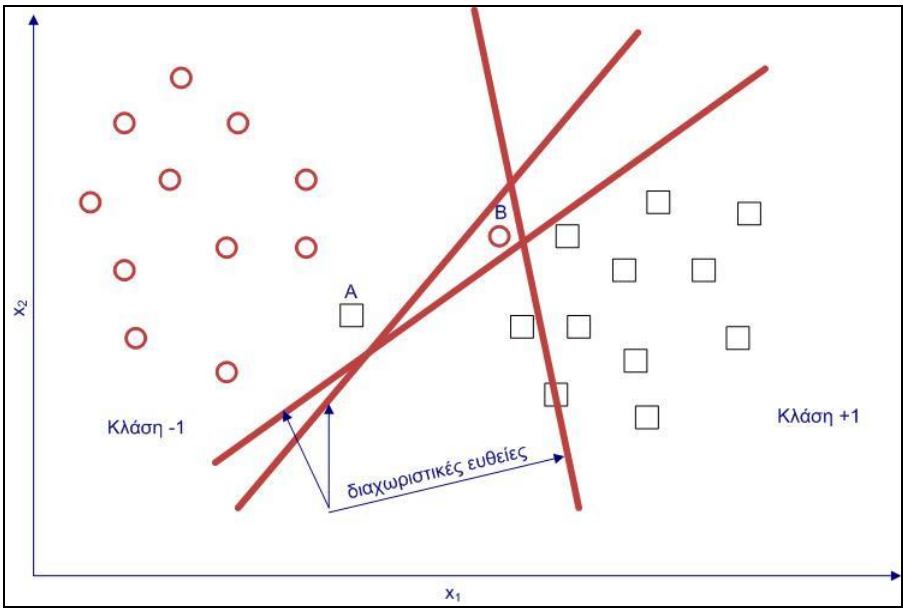


Σχήμα 3: Περίπτωση μη- γραμμικά διαχωρίσιμου συνόλου εκπαίδευσης

Στο συγκεκριμένο παράδειγμα, οποιαδήποτε πιθανή διαχωριστική ευθεία κι αν φανταστούμε, όπως π.χ. αυτές του σχήματος, δεν μπορεί να διαχωρίσει επιτυχώς τις δύο κλάσεις. Απαιτείται διαχωριστική γραμμή μεγαλυτέρου βαθμού (εν προκειμένω κύκλος ή περίπου κύκλος), όπως για παράδειγμα αυτή που φαίνεται στο σχήμα. Τίθεται λοιπόν ένα θέμα επιλογής της μορφής που θα έχει η διαχωριστική γραμμή. Όπως

θα δούμε, ο *SVM* προσφέρει αρκετές εναλλακτικές επιλογές της μορφής που θα έχει η συνάρτηση διαχωρισμού.

- Τα σημεία είναι σίγουρο ότι μπορούν να διαχωριστούν <u>όλα</u> επιτυχώς, για δεδομένη επιλογή συνάρτησης; Για παράδειγμα τα παρακάτω σημεία (σχήμα 4) δεν μπορούν να διαχωριστούν όλα επιτυχώς αν επιλέξουμε γραμμική συνάρτηση (ευθεία) για το διαχωρισμό τους.



Σχήμα 4: Το ανωτέρω σύνολο σημείων δεν είναι απόλυτα διαχωρίσιμο από μια ευθεία, κι αυτό λόγω της παρουσίας των σημείων Α και Β που υπεισέρχονται στα όρια της αντίθετης από τη δική τους κλάσης. Τέτοια σύνολα ονομάζονται μη-απόλυτα γραμμικά διαχωρίσιμα.

Κι αυτό γιατί τα σημεία Α και Β υπεισέρχονται στα όρια της αντίθετης κλάσης από αυτή στην οποία ανήκουν, και συνεπώς καμία γραμμική συνάρτηση (ευθεία) δεν θα μπορέσει να επιτύχει απόλυτο διαχωρισμό των δύο κλάσεων. Αν ωστόσο έλειπαν τα δύο αυτά σημεία από το σύνολο εκπαίδευσης, ο

διαχωρισμός των υπολοίπων από μια ευθεία θα ήταν απόλυτα επιτυχής, και δεν θα απαιτείτο διαχωριστική γραμμή μεγαλυτέρου βαθμού, όπως π.χ. αυτή του σχήματος 3. Το ερώτημα που προκύπτει είναι: είναι προτιμότερο να επιλεγεί ευθεία για το διαχωρισμό του εν λόγω συνόλου, αδιαφορώντας έτσι για τη μη σωστή ταξινόμηση δύο μόλις σημείων; Ή μήπως η παρουσία των δύο αυτών σημείων καθιστά επιτακτική την επιλογή διαχωριστικής γραμμής μεγαλυτέρου βαθμού;

- Το παραπάνω ερώτημα διατυπώνεται σε αυστηρότερη γλώσσα ως εξής: Ποια είναι η συνάρτηση που μπορεί να εξασφαλίσει την καλύτερη **γενίκευση (generalization)**; Με τον όρο γενίκευση, εννοούμε την ικανότητα της μηχανής να διαχωρίσει επιτυχώς <u>νέα</u> σημεία, των οποίων τις ετικέτες δεν γνωρίζουμε, με το μικρότερο δυνατό σφάλμα ταξινόμησης στα σημεία αυτά.

Θα ξεκινήσουμε απαντώντας στο τελευταίο ερώτημα, αφού είναι και το πιο καθοριστικό για την αποδοτική λειτουργία μιας μηχανής ταξινόμησης. Για την ελαχιστοποίηση του σφάλματος ταξινόμησης έχουν κατά περιόδους εφαρμοστεί διάφορες θεωρητικές μέθοδοι. Στα παρακάτω θα αναλυθεί μια τέτοια μέθοδος, γνωστή ως **Αρχή Ελαχιστοποίησης Κατασκευαστικού Ρίσκου**, η οποία προκύπτει από τη **στατιστική θεωρία εκμάθησης** που ανέπτυξε ο Vapnik *[VAP98]*, εφευρέτης της μεθόδου SVM. Η αρχή αυτή εφαρμόζεται από τη μέθοδο SVM, και επιτυγχάνει καλύτερα αποτελέσματα απ' ότι προγενέστερες μέθοδοι ελαχιστοποίησης του σφάλματος ταξινόμησης, όπως για παράδειγμα η Αρχή Ελαχιστοποίησης Εμπειρικού Ρίσκου, η οποία επίσης θα αναλυθεί.

2.2 Σφάλμα γενίκευσης μιας μηχανής ταξινόμησης

Έστω ότι δίνεται ένα σύνολο γνωστών εισόδων-εξόδων (σύνολο εκπαίδευσης):

$$\{(\vec{x}_i, y_i), i = 1...l\}$$

Τα δεδομένα του συνόλου αυτού υποθέτουμε ότι προέρχονται από μια (άγνωστη σε μας) κατανομή πιθανότητας $P(\vec{x}, y)$. Το πρόβλημα εκμάθησης τότε της μηχανής ταξινόμησης μπορεί να διατυπωθεί ως εξής: Να βρεθεί μια συνάρτηση $f : \Re^N \to \pm 1$, χρησιμοποιώντας τα δεδομένα εισόδου-εξόδου

$$(x_1, y_1), ..., (x_i, y_i), ..., i = 1, ...l$$

τέτοια ώστε η f θα ταξινομεί σωστά νέα δεδομένα $\vec{x}$, δηλαδή θα ισχύει $f(\vec{x}) = y$, για νέα σημεία που έχουν προέλθει από την ίδια κατανομή πιθανότητας $P(\vec{x}, y)$ με τα σημεία του συνόλου εκπαίδευσης. Η $f(\vec{x})$ θα ανήκει γενικά σε μια οικογένεια συναρτήσεων $f(\vec{x}, \vec{a})$, όπου $\vec{a}$ είναι το διάνυσμα ρυθμιστικών παραμέτρων της οικογένειας. Για παράδειγμα, αν εξετάζουμε την οικογένεια των πολυωνύμων

$$f(\vec{x}, a) = \sum_{i=0}^{a} c_i x^i$$

η ρυθμιστική παράμετρος a είναι ο βαθμός του πολυωνύμου.

Ορίζονται δύο ποσότητες που έχουν να κάνουν με την επίδοση της μηχανής ταξινόμησης που χρησιμοποιεί την οικογένεια συναρτήσεων $f(\vec{x}, \vec{a})$. Η πρώτη ονομάζεται **αναμενόμενο ή πραγματικό ρίσκο (expected/actual risk)** και ορίζεται ως:

$$R(\vec{a}) = \int \frac{1}{2} | y - f(\vec{x}, \vec{a}) | dP(\vec{x}, y) \qquad \textbf{(2-1)}$$

Το ολοκλήρωμα λαμβάνεται στο χώρο εισόδων, όπου η $P(\vec{x}, y)$ εκφράζει την πιθανότητα να βρεθεί ένα νέο σημείο στη θέση $\vec{x}$ που να έχει αντίστοιχη ετικέτα y. Η ποσότητα $R(\vec{a})$ είναι ένα μέτρο της απόκλισης που έχει η εκτίμηση $f(\vec{x}, \vec{a})$ της μηχανής ταξινόμησης, από την πραγματική έξοδο y ενός νέου σημείου. Είναι δηλαδή ο μέσος όρος του **σφάλματος ταξινόμησης** της μηχανής, υπολογισμένος στο σύνολο των νέων σημείων που κείνται στο χώρο εισόδων. Αναφέρεται δε και ως **σφάλμα δοκιμής (testing error).**

Η δεύτερη ποσότητα ονομάζεται **εμπειρικό ρίσκο (empirical risk)**, και ορίζεται ως:

$$R_{emp}(\vec{a}) = \frac{1}{2l}\sum_{i=1}^{l} | y_i - f(\vec{x}_i, \vec{a}) | \qquad \textbf{(2-2)}$$

Είναι ουσιαστικά ο μέσος όρος του σφάλματος ταξινόμησης, υπολογισμένος στο σύνολο των l σημείων του συνόλου εκπαίδευσης. Η ποσότητα $\frac{1}{2} | y_i - f(\vec{x}_i, \vec{a}) |$ ονομάζεται ***απώλεια* (loss)**. Εκφράζει τη διαφορά που έχει η εκτίμηση της μηχανής για την ετικέτα του i σημείου του συνόλου εκπαίδευσης ($f(\vec{x}_i, \vec{a})$), από την πραγματική του ετικέτα, y_i. Για την περίπτωση της ταξινόμησης (όπου $y_i, f(\vec{x}_i, \vec{a}) \in \{-1, +1\}$) είναι εμφανές ότι αυτή η ποσότητα παίρνει μόνο δύο τιμές, 0 ή 1.

Η σχέση ***(2-1)*** είναι γενική, και παρέχει ένα μέτρο της αναμενόμενης επίδοσης της μηχανής ταξινόμησης, δηλαδή της ικανότητάς της να αποδώσει τις σωστές ετικέτες y_i σε νέα σημεία $\vec{x}_i$ (πέραν αυτών του συνόλου εκπαίδευσης) που θα βρεθούν στο χώρο εισόδων. Αντίθετα, η

σχέση ***(2-2)*** αφορά στο συγκεκριμένο σύνολο εκπαίδευσης που έχει δοθεί στη μηχανή, και εφαρμόζεται μόνο στα l σημεία αυτού του συνόλου.

Στο πρόβλημα της ταξινόμησης μας ενδιαφέρει η ελαχιστοποίηση του πραγματικού ρίσκου της σχέσης ***(2-1)***, το οποίο πρακτικά σημαίνει ότι η συνάρτηση $f(\vec{x},\vec{a})$ που εφαρμόζεται από τη μηχανή εκμάθησης, κάνει όσο το δυνατόν λιγότερα σφάλματα στην ταξινόμηση νέων σημείων που θα βρεθούν με κάποια (άγνωστη) πιθανότητα στο χώρο εισόδων. Ωστόσο, παρατηρώντας τη σχέση αυτή, διαπιστώνουμε ότι στον υπολογισμό του $R(\vec{a})$ υπεισέρχεται η άγνωστη σε μας κατανομή πιθανότητας $P(\vec{x},y)$ από την οποία προκύπτουν τα νέα σημεία. Συνεπώς, η ποσότητα αυτή δεν μπορεί να υπολογιστεί.

Αντίθετα, το εμπειρικό ρίσκο που δίνεται από τη σχέση ***(2-2)***, μπορεί να υπολογιστεί διότι δεν υπεισέρχεται σε αυτό η $P(\vec{x},y)$, καθώς αφορά μόνο στα γνωστά σημεία του δοθέντος συνόλου εκπαίδευσης. Σε πολλές προγενέστερες μεθόδους ταξινόμησης, θεωρείτο ότι η ελαχιστοποίηση του εμπειρικού ρίσκου, θα μπορούσε να εξασφαλίσει και ελαχιστοποίηση του πραγματικού ρίσκου. Με άλλα λόγια, θεωρείτο ότι η μηχανή που θα κάνει τα λιγότερα σφάλματα στο σύνολο εκπαίδευσης, θα κάνει και τα λιγότερα σφάλματα σε νέα, άγνωστα σημεία. Για το λόγο αυτό, και επειδή ο υπολογισμός του $R(\vec{a})$ δεν είναι εφικτός, πολλές μηχανές εκμάθησης χρησιμοποιούν οικογένειες συναρτήσεων $f(\vec{x},\vec{a})$ τέτοιες, που να ελαχιστοποιούν το εμπειρικό ρίσκο $\mathrm{R}_{emp}(\vec{a})$.

Η ανωτέρω θεώρηση είναι γνωστή και ως **Αρχή Ελαχιστοποίησης Εμπειρικού Ρίσκου (Empirical Risk Minimization Principle – ERM)**, και εφαρμόζεται συνήθως και από μια ιδιαίτερα δημοφιλή μηχανή

εκμάθησης, τα Τεχνητά Νευρωνικά Δίκτυα. Η διατύπωση της είναι η ακόλουθη:

ERM Principle:

Προκειμένου να ελαχιστοποιήσουμε το πραγματικό ρίσκο μιας μηχανής ταξινόμησης, αρκεί να επιλέξουμε μια συνάρτηση που να ελαχιστοποιεί το εμπειρικό της ρίσκο, δηλαδή το σφάλμα που εμφανίζει αυτή κατά τη φάση της εκπαίδευσης.

Μαθηματικά εκφρασμένη, η αρχή ERM γράφεται:

$$\min : \mathrm{R}_{emp}(\vec{a}) \Rightarrow \min : R(\vec{a}) \qquad \textbf{(2-3)}$$

Στην πραγματικότητα, η συσχέτιση του εμπειρικού ρίσκου με το πραγματικό, παρέχεται από μια ανισότητα της μορφής:

$$R(\vec{a}) \le \mathrm{R}_{emp}(\vec{a}) + (\Pi) \qquad \textbf{(2-4)}$$

όπου *(Π)* μια ποσότητα που μειώνεται όσο αυξάνεται το μέγεθος *l* του συνόλου εκπαίδευσης. Για επαρκώς μεγάλα σύνολα εκπαίδευσης (μεγάλο *l*), η ποσότητα *(Π)* θεωρείται αμελητέα συγκρινόμενη με το εμπειρικό ρίσκο, και συνεπώς η ελαχιστοποίηση του εμπειρικού ρίσκου μπορεί πράγματι να εξασφαλίσει και ελαχιστοποίηση του πραγματικού ρίσκου. Δηλαδή, για σχετικά μεγάλα σύνολα εκπαίδευσης, η σχέση **(2-3)** της αρχής ERM είναι αληθής. Ωστόσο, για μικρά σύνολα εκπαίδευσης, ο όρος *(Π)* αποκτά βαρύνουσα σημασία. Εδώ εστιάζεται το μειονέκτημα της αρχής ERM, που δεν λαμβάνει καθόλου υπόψη τον όρο αυτό, αλλά επιδιώκει ελαχιστοποίηση μόνο του $\mathrm{R}_{emp}(\vec{a})$.

Ο όρος *(Π)* εξαρτάται από ένα μέγεθος που χαρακτηρίζει την πολυπλοκότητα μιας οικογένειας συναρτήσεων $f(\vec{x},\vec{a})$, και το οποίο ονομάζεται **χωρητικότητα** της οικογένειας.

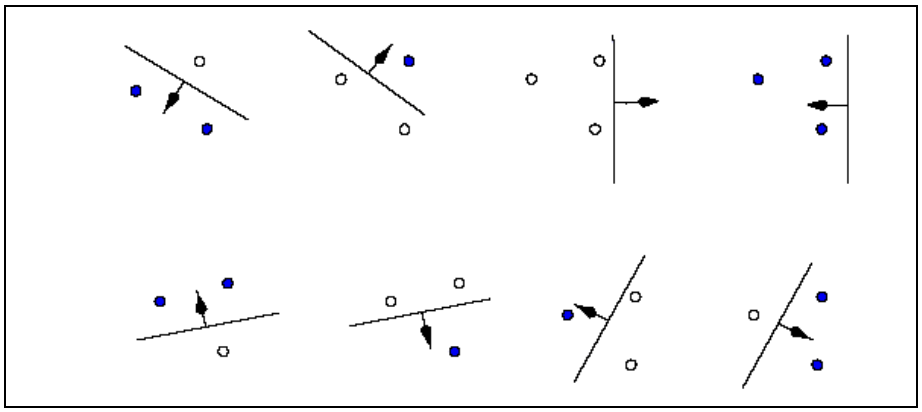
2.3 Χωρητικότητα μιας οικογένειας συναρτήσεων

Θεωρούμε ένα σύνολο l σημείων $\vec{x}_i, i=1,...l$. Σε κάθε σημείο αυτού του συνόλου επιδιώκεται να αποδοθεί μια ετικέτα (+1 ή -1), προκειμένου το σύνολο να ταξινομηθεί σε δύο κλάσεις. Οι πιθανοί συνδυασμοί ετικετών που μπορούν να αποδοθούν στα σημεία του εν λόγω συνόλου είναι προφανώς 2^l.

Αν μια οικογένεια συναρτήσεων $f(\vec{x},\vec{a})$ μπορεί να αποδώσει στο σύνολο αυτό όλους τους πιθανούς συνδυασμούς ετικετών, δηλαδή για κάθε ένα συνδυασμό από αυτούς υπάρχει έστω και μια συνάρτηση της οικογένειας, $f(\vec{x},\vec{a}^*)$, για κάποιο $\vec{a}^*$, η οποία να αποδίδει τις επιθυμητές ετικέτες των σημείων με τον ορθό τρόπο (δηλαδή να ισχύει $f(\vec{x}_i,\vec{a}^*)=y(\vec{x}_i),\forall i$), τότε λέμε ότι το σύνολο αυτό των l σημείων **σαρώνεται επιτυχώς** από την οικογένεια συναρτήσεων $f(\vec{x},\vec{a})$, ή αλλιώς ότι **η χωρητικότητα (capacity)** της οικογένειας είναι l.

Η χωρητικότητα δηλαδή μιας οικογένειας συναρτήσεων είναι ο μέγιστος αριθμός σημείων που μπορούν να διαχωριστούν με όλους τους πιθανούς τρόπους από την εν λόγω οικογένεια. Στη βιβλιογραφία, η χωρητικότητα αναφέρεται συχνά και ως **VC-διάσταση (VC-dimension)**, από τους Vapnik-Chervonenkis που εισήγαγαν τον όρο (*[VAP98],[VAP95)]*, συμβολίζεται δε συνήθως με το γράμμα h.

Για να αποσαφηνιστεί πλήρως η έννοια της χωρητικότητας (VC-διάστασης) , χρησιμοποιούμε το παράδειγμα του σχήματος 5. Θεωρούμε 3 σημεία στο 2-Δ χώρο (σχήμα 5), καθώς και την οικογένεια συναρτήσεων που αποτελείται από προσανατολισμένες ευθείες. Δηλαδή ευθείες που τους έχει αποδοθεί ένα κάθετο διάνυσμα προσανατολισμού, έτσι ώστε όσα σημεία «βλέπει» το διάνυσμα να ανήκουν στη μια κλάση, και όσα δεν «βλέπει» να ανήκουν στην άλλη.



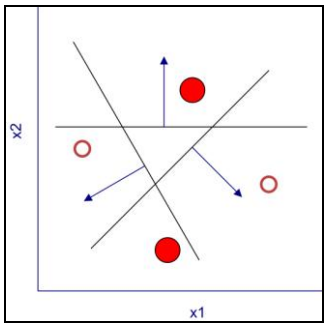
Σχήμα 5: 3 σημεία στο 2Δ χώρο, που διαχωρίζονται με όλους τους πιθανούς τρόπους από την οικογένεια των προσανατολισμένων ευθειών. Λέμε τότε ότι η εν λόγω οικογένεια σαρώνει επιτυχώς τα 3 αυτά σημεία.

Η οικογένεια αυτή χρησιμοποιείται για το σάρωμα των τριών αυτών σημείων. Δηλαδή, κάθε μέλος της διαχωρίζει τα τρία αυτά σημεία, με διαφορετικό όμως τρόπο το κάθε μέλος, αποδίδοντας έτσι διαφορετικό κάθε φορά συνδυασμό ετικετών στα σημεία αυτά.

Η έννοια του διαχωρισμού είναι όπως είπαμε συνυφασμένη με τη σχετική θέση ενός σημείου ως προς την ευθεία: Όσα σημεία βρίσκονται από τη μια πλευρά της ευθείας θεωρούνται «+1» , ενώ αυτά που βρίσκονται από την άλλη πλευρά θεωρούνται «-1». Αυτό φυσικά γενικεύεται και στην περίπτωση που έχουμε οποιαδήποτε διαχωριστική γραμμή και όχι μόνο ευθεία, έτσι ώστε να μπορούμε να διατυπώσουμε την πρόταση:

Όσα σημεία βρίσκονται από τη μια πλευρά της διαχωριστικής γραμμής, τους αποδίδεται ετικέτα +1 (-1), ενώ σε αυτά που βρίσκονται στην άλλη πλευρά αποδίδεται ετικέτα -1(+1).

Όπως μπορεί να φανεί και από το σχήμα 5, οποιαδήποτε κι αν είναι η επιθυμητή απόδοση ετικετών στα σημεία του εν λόγω συνόλου, υπάρχει πάντοτε ένα μέλος της οικογένειας ευθειών που μπορεί να την αποδώσει ορθά. Με διαφορετική διατύπωση, η οικογένεια αυτή των προσανατολισμένων ευθειών μπορεί να αποδώσει ετικέτες στο εν λόγω σύνολο με $2^3 = 8$ διαφορετικούς τρόπους. Αν τώρα προσθέσουμε και ένα τέταρτο σημείο, και θελήσουμε να αποδώσουμε στα 4 συνολικά σημεία το συνδυασμό ετικετών του σχήματος 6:



Σχήμα 6: 4 σημεία που δεν μπορούν να διαχωριστούν επιτυχώς από την οικογένεια των προσανατολισμένων ευθειών.

τότε εύκολα διαπιστώνουμε ότι δεν είναι δυνατόν τα σημεία αυτά να διαχωριστούν από κανένα μέλος της οικογένειας προσανατολισμένων ευθειών. Με διαφορετική διατύπωση, οι προσανατολισμένες ευθείες δεν μπορούν να σαρώσουν επιτυχώς 4 σημεία. Λέμε τότε ότι η VC-διάσταση (χωρητικότητα) της οικογένειας προσανατολισμένων ευθειών είναι 3, αφού 3 είναι ο μέγιστος αριθμός σημείων που μπορεί να σαρωθεί επιτυχώς από αυτήν.

2.4 Ελαχιστοποίηση Κατασκευαστικού Ρίσκου (Structural Risk Minimization – SRM)

Αφού εξετάσαμε την έννοια της χωρητικότητας μιας οικογένειας συναρτήσεων, ερχόμαστε τώρα σε μια πιο αυστηρή και πληρέστερη διατύπωση του άνω ορίου του πραγματικού ρίσκου ταξινόμησης. Στην παράγραφο 1.2 αναφέραμε ότι το άνω όριο του πραγματικού ρίσκου ταξινόμησης εξαρτάται από μια ποσότητα *(Π)* (***(2-4)***), που είναι συνάρτηση της χωρητικότητας της οικογένειας που εφαρμόζει η μηχανή εκμάθησης. Ο Vapnik *[VAP98]* παρείχε μια έκφραση για αυτή την ποσότητα, που την ονόμασε **VC- εμπιστοσύνη (VC-confidence).** Η έκφραση αυτή φαίνεται στη σχέση ***(2-5)***.

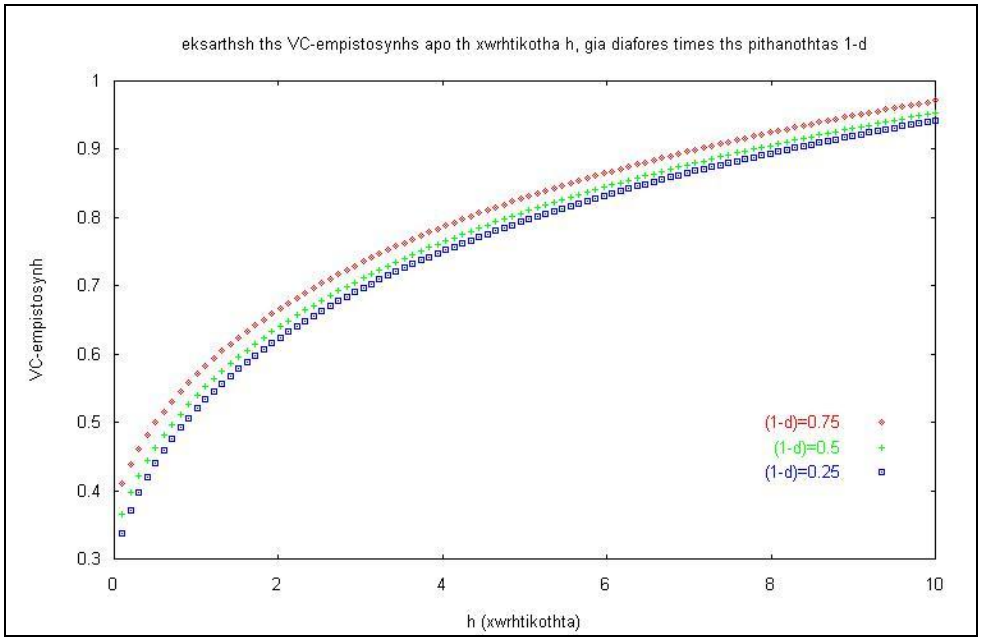
$$R(a) \le R_{emp}(a) + \sqrt{\frac{h \cdot (\log(2l/h)+1) - \log(\delta/4)}{l}} \qquad \textbf{(2-5)}$$

όπου ο δεύτερος όρος του δεξιού σκέλους είναι η VC- εμπιστοσύνη, l είναι το μέγεθος του συνόλου εκπαίδευσης και h η χωρητικότητα της οικογένειας συναρτήσεων που εφαρμόζει η μηχανή εκμάθησης. Ο αριθμός δ που υπεισέρχεται στη σχέση κυμαίνεται στο εύρος [0,1], και η ανωτέρω σχέση ισχύει με πιθανότητα τουλάχιστον 1-δ *[VAP98]*.

Η σχέση ***(2-5)*** παρέχει ένα <u>άνω όριο</u> για το πραγματικό ρίσκο ταξινόμησης, το οποίο είναι γνωστό ως **όριο-VC (VC-bound)**. Προσεκτική παρατήρησή της μας οδηγεί στα εξής συμπεράσματα: Πρώτον, είναι αξιοσημείωτο ότι το όριο-VC δεν εξαρτάται από την κατανομή $P(\vec{x}, y)$. Υποθέτει μόνο ότι τα σημεία του συνόλου εκπαίδευσης και οι αντίστοιχες ετικέτες τους προέρχονται από <u>κάποια</u> κατανομή, χωρίς να απαιτείται να

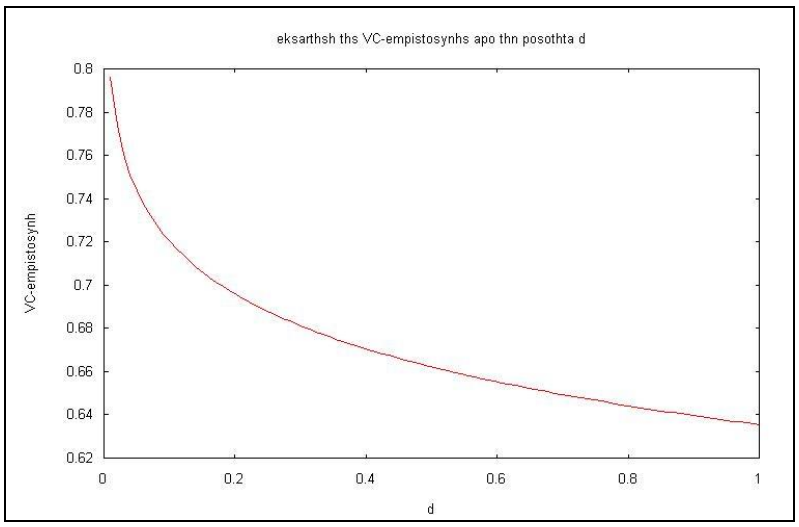
τη γνωρίζουμε για να το υπολογίσουμε. Δεύτερον, όπως προείπαμε, δεν είναι συνήθως δυνατόν να υπολογιστεί το αριστερό σκέλος της ανισότητας (δηλαδή το $R(\vec{a})$). Είναι ωστόσο δυνατόν να υπολογιστεί το όριο-VC, αν γνωρίζουμε το h (VC-διάσταση) και το εμπειρικό ρίσκο, το οποίο μπορεί να υπολογιστεί αφού το σύνολο εκπαίδευσης είναι γνωστό. Εφόσον στο όριο αυτό εμπεριέχεται η ποσότητα VC-εμπιστοσύνη, διερευνούμε την εξάρτηση αυτής της ποσότητας από τις παραμέτρους l, δ, h.

Στο σχήμα 7 φαίνεται η VC-εμπιστοσύνη συναρτήσει της χωρητικότητας h της οικογένειας συναρτήσεων που εφαρμόζει η μηχανή εκμάθησης, για διάφορες τιμές της πιθανότητας 1-δ. Παρατηρούμε ότι η VC-εμπιστοσύνη είναι μονότονη αύξουσα συνάρτηση της χωρητικότητας. Επίσης, όσο αυξάνεται η πιθανότητα (1-δ) με την οποία ισχύει το όριο-VC της σχέσης ***(2-5)***, τόσο μεγαλώνουν οι τιμές που παίρνει η VC-εμπιστοσύνη.



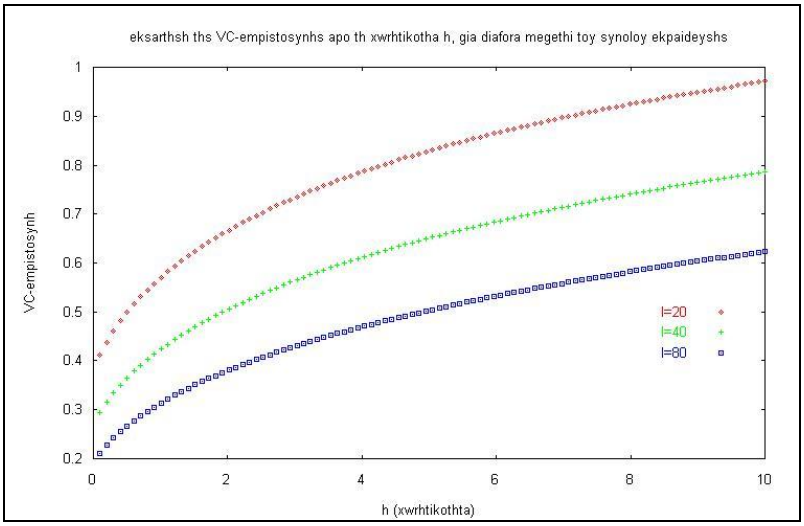
Σχήμα 7: Εξάρτηση της VC-εμπιστοσύνης από τη χωρητικότητα, για διάφορες τιμές της πιθανότητας (1-δ)

Δηλαδή, όσο μικρότερη είναι η ποσότητα δ (για την ίδια χωρητικότητα h), τόσο πιο «χαλαρό» είναι το άνω όριο του Vapnik. Με διαφορετική διατύπωση, το όριο-VC του Vapnik είναι πιο «αυστηρό» όσο επιλέγεται μικρότερη τιμή (1-δ) της πιθανότητας ισχύος του. Το σχήμα 8 δείχνει την εξάρτηση της VC-εμπιστοσύνης από τον αριθμό δ.



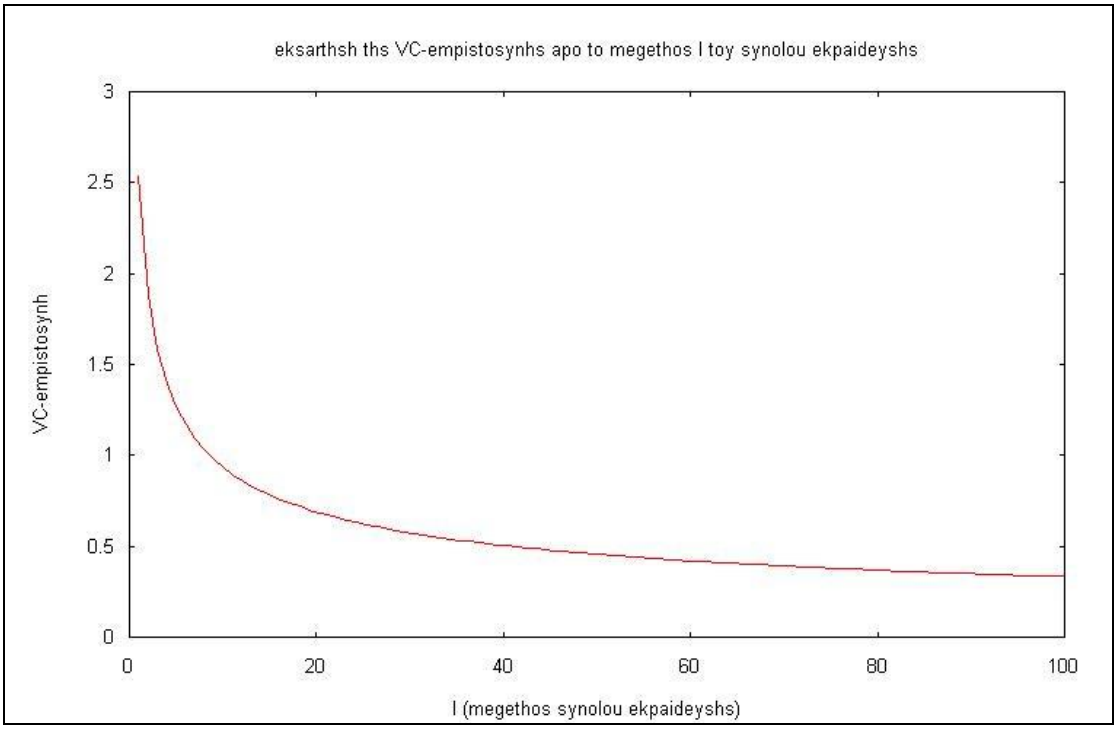
Σχήμα 8: Εξάρτηση της VC-εμπιστοσύνης από τον αριθμό δ

Διερευνούμε επίσης την επίδραση του μεγέθους *l* του συνόλου εκπαίδευσης στη VC-εμπιστοσύνη. Όπως παρατηρούμε από το σχήμα 9 , όσο μικρότερο είναι το σύνολο εκπαίδευσης, τόσο πιο απότομα αυξάνει η VC-εμπιστοσύνη με το h.



Σχήμα 9: Εξάρτηση της VC-εμπιστοσύνης από τη χωρητικότητα, για διάφορα μεγέθη του συνόλου εκπαίδευσης

Αυτό πρακτικά σημαίνει ότι οικογένειες με μεγάλη χωρητικότητα, είναι ακατάλληλες για μικρά σύνολα εκπαίδευσης, καθώς αντιστοιχούν σε μεγάλη τιμή της VC-εμπιστοσύνης, και συνεπώς σε μεγάλο όριο-VC. Αντίθετα, ένα πιο μεγάλο σύνολο εκπαίδευσης μπορεί να δεχτεί με επιτυχία μια οικογένεια με μεγάλη χωρητικότητα. Η γραφική εξάρτηση της VC-εμπιστοσύνης από το μέγεθος του συνόλου εκπαίδευσης φαίνεται στο σχήμα 10.



Σχήμα 10: Εξάρτηση της VC-εμπιστοσύνης από το μέγεθος του συνόλου εκπαίδευσης

Από την ανωτέρω παρατήρηση προκύπτει ένας γενικός κανόνας: Η χωρητικότητα της οικογένειας συναρτήσεων που εφαρμόζει η μηχανή εκμάθησης θα πρέπει να είναι κατάλληλη για το σύνολο εκπαίδευσης που καλείται αυτή να διαχωρίσει. Έτσι διατυπώνεται η αρχή που είναι γνωστή με το όνομα **Ελαχιστοποίηση Κατασκευαστικού Ρίσκου (Structural Risk Minimization -SRM)**, και η οποία έχει ως εξής:

<u>SRM principle:</u>

Προκειμένου να ελαχιστοποιήσουμε το άνω όριο του πραγματικού ρίσκου μιας μηχανής ταξινόμησης, πρέπει να ελαχιστοποιήσουμε το εμπειρικό σφάλμα της μηχανής $R_{emp}(\vec{a})$, αλλά αυτό να γίνει από μια οικογένεια συναρτήσεων που να έχει κατάλληλη χωρητικότητα για το εν λόγω σύνολο εκπαίδευσης.

Για να έχουν πρακτική χρησιμότητα η αρχή SRM και το όριο-VC που παρέχεται από τη σχέση ***(2-5)***, θα πρέπει να χρησιμοποιείται μια οικογένεια συναρτήσεων της οποίας η χωρητικότητα να μπορεί να υπολογιστεί. Οι Vapnik – Chervonenkis χρησιμοποίησαν μια τέτοια οικογένεια συναρτήσεων, τους λεγόμενους **γραμμικούς διαχωριστές με διάκενο**, των οποίων η χωρητικότητα όπως θα δούμε μπορεί να υπολογιστεί. Κατασκεύασαν έτσι έναν αλγόριθμο ταξινόμησης με το όνομα **Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)**, που εφαρμόζει την ανωτέρω οικογένεια συναρτήσεων και ικανοποιεί την αρχή SRM, πετυχαίνοντας καλύτερη γενίκευση σε προβλήματα ταξινόμησης από ότι προγενέστεροι αλγόριθμοι εκμάθησης (όπως π.χ. ΤΝΔ, δέντρα αποφάσεων, κτλ.)

2.5 *Ανάλυση της μεθόδου SVM*

Στα παρακάτω θα επιχειρήσουμε σε βάθος ανάλυση του ακριβούς τρόπου με τον οποίο λειτουργεί η μέθοδος *SVM* σε προβλήματα ταξινόμησης.

2.5.1 Η απλή περίπτωση: Γραμμική συνάρτηση διαχωρισμού, σύνολο εκπαίδευσης απόλυτα διαχωρίσιμο

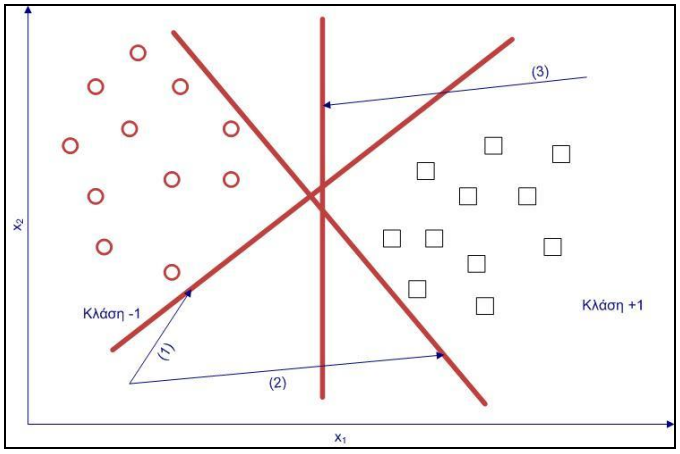
Θα ξεκινήσουμε από την απλούστερη περίπτωση διαχωρισμού: Η συνάρτηση που εφαρμόζει ο *SVM* είναι γραμμική, και το σύνολο εκπαίδευσης είναι πλήρως διαχωρίσιμο από αυτήν. Αργότερα θα δούμε πώς γίνεται η γενίκευση στην περίπτωση μη-γραμμικής συνάρτησης διαχωρισμού, αλλά και σε αυτήν που το σύνολο εκπαίδευσης δεν είναι απόλυτα διαχωρίσιμο.

Ας θεωρήσουμε πάλι την περίπτωση όπου έχουμε ένα σύνολο σημείων στο 2-Δ χώρο (Η γενίκευση για σημεία στο Ν-διάστατο χώρο θα γίνει αμέσως μετά). Όπως μπορεί να φανεί και στο σχήμα 11, υπάρχουν πολλές εναλλακτικές ευθείες που μπορούν να διαχωρίσουν επιτυχώς τα σημεία αυτά (π.χ. οι ευθείες *(1),(2),(3)* του σχήματος).

Αυτές οι διαχωριστικές ευθείες έχουν τη γενική μορφή :

$$< \vec{w}, \vec{x} > -b = 0 \qquad \textbf{(2-6)}$$

όπου $\vec{x}$ είναι το διάνυσμα παραμέτρων (συντεταγμένες) ενός σημείου της ευθείας (εδώ $\vec{x} \in \Re^2$), $\vec{w}$ είναι το κάθετο στην ευθεία διάνυσμα (διάνυσμα προσανατολισμού, εδώ πάλι $\vec{w} \in \Re^2$), $b \in \Re$ είναι μια σταθερά, ενώ το σύμβολο $< \_, \_ >$ αντιπροσωπεύει εσωτερικό γινόμενο των δύο διανυσμάτων, $\vec{w}$ και $\vec{x}$.



Σχήμα 11: Διάφορες εναλλακτικές ευθείες διαχωρισμού ενός συνόλου εκπαίδευσης

Για την 2-Δ περίπτωση του σχήματος, η σχέση ***(2-6)*** γράφεται:

$$w_1 x_1 + w_2 x_2 - b = 0$$

που αποτελεί εξίσωση ευθείας στο επίπεδο $x_1 x_2$. Με βάση τη σχέση ***(2-6)***, τα σημεία που ανήκουν στη μια κλάση είναι αυτά για τα οποία :

$$< \vec{w}, \vec{x} > -b \geq 0 \qquad \textbf{(2-7}^{\alpha}\textbf{)}$$

ενώ αυτά που ανήκουν στην αντίθετη κλάση είναι αυτά για τα οποία :

$$< \vec{w}, \vec{x} > -b \leq 0 \qquad \textbf{(2-7}^{\beta}\textbf{)}$$

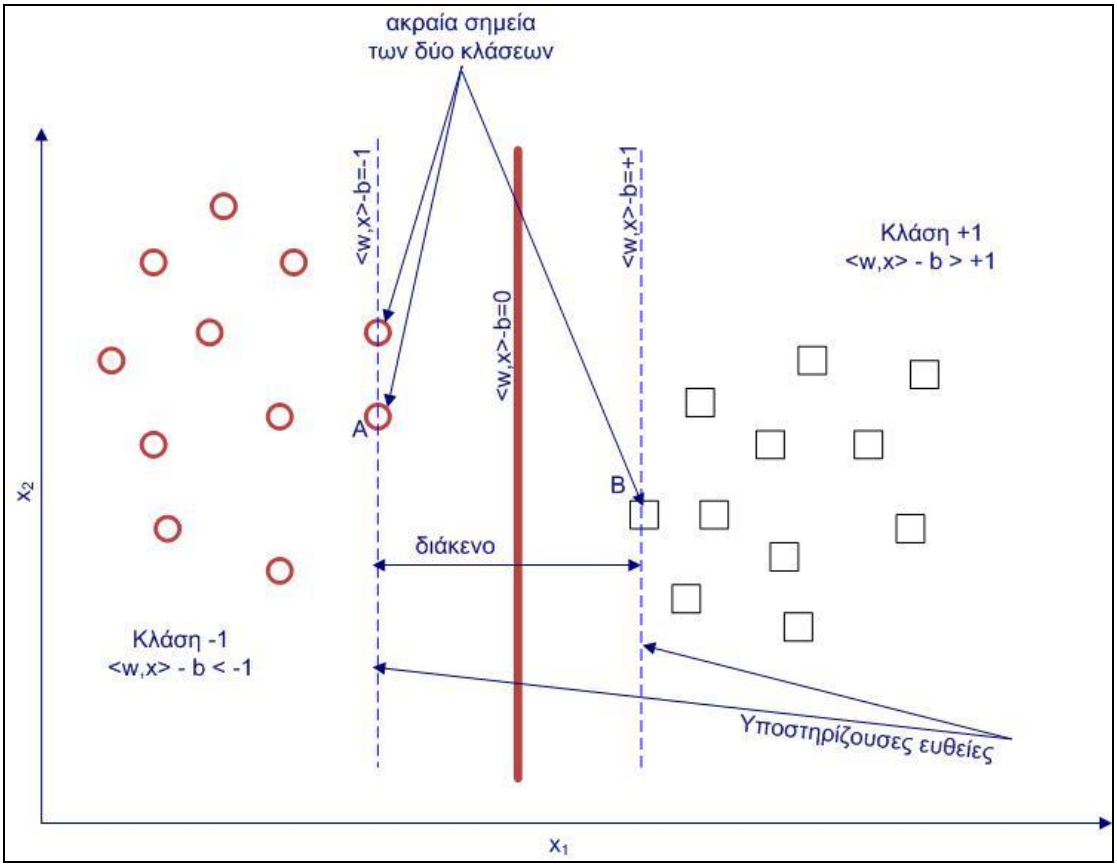
Το ζητούμενο που προκύπτει τώρα είναι πώς θα επιλέξουμε κατάλληλα τα $\vec{w}, b$ ούτως ώστε να βρούμε εκείνη τη διαχωριστική ευθεία με την καλύτερη γενίκευση (generalization). Με διαφορετική διατύπωση, ζητούμενο είναι να βρούμε εκείνη την ευθεία που θα ικανοποιεί την αρχή SRM του Vapnik, δηλαδή εκείνη που θα εξασφαλίζει την ελαχιστοποίηση του ορίου-VC.

Διαισθητικά και μόνο, υποπτευόμαστε ότι ευθείες *(1)* και *(2)* του σχήματος 11 δε θα μπορέσουν να επιτύχουν καλή γενίκευση, διότι βρίσκονται πολύ κοντά στα ακραία σημεία κάθε κλάσης. Αυτό συνεπάγεται ότι ένα <u>νέο σημείο</u> που θα απέχει έστω και λίγο από τα ακραία αυτά σημεία, θα ταξινομηθεί πιθανόν ως ανήκον στην αντίθετη κλάση, αφού θα περνάει ίσως τη διαχωριστική γραμμή. Αντίθετα, η ευθεία *(3)* αναμένεται να πετύχει καλύτερη γενίκευση, καθώς είναι σχετικά ισαπέχουσα από τις δύο κλάσεις σε σχέση με τις *(1),(2)*. Στην επόμενη παράγραφο θα παρουσιαστεί μια συστηματική μεθοδολογία για την εύρεση της ευθείας εκείνης που επιτυγχάνει την καλύτερη γενίκευση από όλες τις άλλες.

<u>*2.5.1.1 Η μέθοδος των υποστηριζουσών ευθειών*</u>

Στο σύνολο εκπαίδευσης του σχήματος 11, θεωρούμε μια διαχωριστική ευθεία, καθώς και δύο άλλες, **παράλληλες** προς τη διαχωριστική, οι οποίες φαίνονται στο σχήμα 12 με διακεκομμένη γραμμή. Οι

διακεκομμένες ευθείες έχουν την ιδιότητα να αφήνουν τα σημεία κάθε κλάσης από τη μια μεριά τους, και ταυτόχρονα να διέρχονται από τα **ακραία σημεία** κάθε κλάσης (ακραία προς την πλευρά της διαχωριστικής), δηλαδή από τα σημεία αυτά που βρίσκονται κοντινότερα στη διαχωριστική ευθεία.



Σχήμα 12: Επεξήγηση της μεθόδου των υποστηριζουσών ευθειών. Στο σχήμα φαίνονται οι υποστηρίζουσες ευθείες των δύο κλάσεων, η διαχωριστική τους γραμμή και τα διανύσματα υποστήριξης

Λέμε ότι αυτές οι ευθείες αποτελούν τις **υποστηρίζουσες ευθείες (supporting lines)** των δύο κλάσεων, ενώ τα σημεία που κείνται πάνω τους τα ονομάζουμε **υποστηρικτικά διανύσματα ή διανύσματα υποστήριξης (support vectors).**

Οι εξισώσεις αυτών των δύο ευθειών είναι :

$$< \vec{w}, \vec{x} > -b = +1 \qquad \textbf{(2-8}^{\alpha}\textbf{)}$$

για την υποστηρίζουσα της μιας κλάσης, και

$$< \vec{w}, \vec{x} > -b = -1 \qquad \textbf{(2-8}^{\beta}\textbf{)}$$

για την υποστηρίζουσα της άλλης κλάσης.

Οι δύο υποστηρίζουσες ευθείες δημιουργούν έτσι μια **διαχωριστική ζώνη** ανάμεσα στις δύο κλάσεις σημείων. Τα σημεία που βρίσκονται από τη μια πλευρά της διαχωριστικής ζώνης ανήκουν στην πρώτη κλάση, ενώ αυτά που βρίσκονται από την άλλη πλευρά ανήκουν στη δεύτερη κλάση. Τυχόντα σημεία που εμπίπτουν στο εσωτερικό της διαχωριστικής ζώνης, ονομάζονται σημεία-παραπλάνησης. Προς το παρόν θεωρούμε ότι δεν υπάρχουν τέτοια σημεία στο σύνολο εκπαίδευσης. Θεωρούμε δηλαδή σύνολο εκπαίδευσης απόλυτα διαχωρίσιμο από την εν λόγω διαχωριστική ζώνη. Στην επόμενη παράγραφο θα δοθεί η γενίκευση της μεθόδου στην περίπτωση ύπαρξης σημείων παραπλάνησης.

Με βάση τα παραπάνω, μπορεί να οριστεί το εύρος της διαχωριστικής ζώνης, ή αλλιώς το **διάκενο (margin)** των δύο κλάσεων, ως η κάθετη απόσταση μεταξύ των υποστηριζουσών ευθειών. Η απόσταση αυτή όπως εύκολα αποδεικνύεται είναι $\gamma = \frac{2}{\| \vec{w} \|_2}$, όπου $\| \vec{w} \|_2$ είναι η ευκλείδεια νόρμα του διανύσματος προσανατολισμού $\vec{w}$ των ευθειών.

Το ανωτέρω σύνολο παραλλήλων ευθειών ανήκει σε μια οικογένεια συναρτήσεων, τους **γραμμικούς διαχωριστές με διάκενο (linear margin separators)**. Ο Vapnik απέδειξε *[VAP98]* ότι στην οικογένεια αυτή, όσο μεγαλύτερο είναι το διάκενο που δημιουργούν οι υποστηρίζουσες ευθείες του διαχωριστή ανάμεσα στις δύο κλάσεις, τόσο μικρότερη είναι η χωρητικότητα h της οικογένειας. Συνεπώς, η VC-εμπιστοσύνη της εν

λόγω οικογένειας μειώνεται και αυτή όσο αυξάνεται το διάκενο (αφού αποτελεί αύξουσα συνάρτηση της χωρητικότητας, βλ. σχήμα 7). Επιπρόσθετα, στην προκειμένη περίπτωση όπου έχουμε απόλυτα διαχωρίσιμο σύνολο εκπαίδευσης, είναι προφανές ότι μπορεί να βρεθεί διαχωριστής που να εξασφαλίζει μηδενικό εμπειρικό ρίσκο, δηλαδή επιτυχή ταξινόμηση όλων των σημείων του συνόλου εκπαίδευσης στη σωστή τους κλάση. Ως εκ τούτου, με μηδενικό $R_{emp}(\vec{a})$, η ελαχιστοποίηση της VC-εμπιστοσύνης εξασφαλίζει και ελαχιστοποίηση ολόκληρου του ορίου-VC (σχέση ***(2-5)***). Συνεπώς, για να ικανοποιηθεί η αρχή SRM του Vapnik, πρέπει να χρησιμοποιηθεί εκείνη η διαχωριστική ευθεία που να εξασφαλίζει το μέγιστο διάκενο, και άρα την ελάχιστη χωρητικότητα και VC-εμπιστοσύνη.

Η ευθεία που ικανοποιεί την αρχή SRM και που υπόσχεται την καλύτερη γενίκευση είναι εκείνη που μεγιστοποιεί το διάκενο γ των δύο κλάσεων.

Ισοδύναμα, αυτή η ευθεία θα έχει την ελάχιστη νόρμα $\|\vec{w}\|_2$, αφού το διάκενο είναι $\gamma = \frac{2}{\|\vec{w}\|_2}$. Συνεπώς, για να βρούμε τα $\vec{w}, b$ της βέλτιστης αυτής ευθείας, διαμορφώνουμε το ακόλουθο πρόβλημα ελαχιστοποίησης :

$$\min : \frac{1}{2}\|\vec{w}\|_2^{\,2} = \frac{1}{2} < \vec{w}, \vec{w} > = \frac{1}{2}\sum_{i=1}^{N} w_i \cdot w_i \qquad \textbf{(2-9)}$$

Επιπλέον πρέπει να ικανοποιούνται οι προφανείς περιορισμοί: κάθε σημείο να ταξινομείται σωστά στην κλάση του. Δηλαδή να ισχύει:

$$< \vec{w}, \vec{x}^{(i)} > -b \geq +1 \ , \ \alpha\nu \ \ y_i = +1 \quad , \text{ενώ}$$

$$< \vec{w}, \vec{x}^{(i)} > -b \leq -1 \ , \ \alpha\nu \ \ y_i = -1$$

Οι δύο παραπάνω μπορούν να συμπτυχθούν σε ένα και μόνο περιορισμό :

$$y_i \cdot (< \vec{w}, \vec{x}^{(i)} > -b) \geq 1 \ , \ i = 1, \ldots l \qquad \textbf{(2-10)}$$

Το πρόβλημα ***(2-9)*** με τους περιορισμούς ***(2-10)*** είναι ένα πρόβλημα **τετραγωνικού προγραμματισμού (quadratic programming – QP)** με αγνώστους τα $\vec{w}, b$: η αντικειμενική συνάρτηση είναι τετραγωνική συνάρτηση του $\vec{w}$, και οι περιορισμοί γραμμικοί. Μια μέθοδος επίλυσής του είναι με χρήση πολλαπλασιαστών Lagrange *[BUR98]*, την οποία και θα εφαρμόσουμε στη συνέχεια. Ο λόγος που επιλέγουμε να περάσουμε από το πρωτεύον πρόβλημα ***(2-9)*** στο δυαδικό του κατά Lagrange, είναι ότι οι περιορισμοί γίνονται πολύ απλούστεροι, όπως θα φανεί παρακάτω.

Έτσι, εισάγονται l το πλήθος μη αρνητικοί πολλαπλασιαστές λ_i (όσοι και οι περιορισμοί ***(2-10)***). Η θεωρία της δυαδικότητας επιτάσσει ότι στην περίπτωση που οι αρχικοί περιορισμοί ***(2-10)*** είναι ανισοτικοί της μορφής $c_i(\vec{w}) \geq 0$, τότε οι λ_i πολλαπλασιάζονται με τους περιορισμούς, και αφαιρούνται από την αντικειμενική συνάρτηση, προκειμένου να προκύψει η Lagrangian του προβλήματος. Συνεπακόλουθα, η Lagrangian θα έχει την παρακάτω μορφή:

$$L(\vec{w},b,\lambda)=\frac{1}{2}\|\vec{w}\|_2^{\,2}-\sum_{i=1}^{l}\lambda_i\cdot[y_i\cdot(<\vec{w},\vec{x}^{(i)}>-b)-1]=$$
$$=\frac{1}{2}\sum_{i=1}^{N}w_i\cdot w_i-\sum_{i=1}^{l}\lambda_i\cdot[y_i\cdot(\sum_{j=1}^{N}w_j x_j^{(i)}-b)-1] \qquad \textbf{(2-11)}$$

Επιπλέον, η βέλτιστη λύση θα πρέπει να ικανοποιεί τις συνθήκες Karush-Kuhn-Tucker (KKT conditions), οι οποίες είναι:

$$\frac{\partial L}{\partial w_j}=w_j-\sum_{i}^{l}\lambda_i y_i x_j^{(i)}=0\Rightarrow$$
$$w_j=\sum_{i}^{l}\lambda_i y_i x_j^{(i)},\ j=1,\ldots,N \qquad \textbf{(2-12)}$$

$$\frac{\partial L}{\partial b}=-\sum_{i=1}^{l}\lambda_i y_i=0\Rightarrow\sum_{i=1}^{l}\lambda_i y_i=0 \qquad \textbf{(2-13)}$$

$$\lambda_i\cdot(y_i\cdot(<w,x_i>+b)-1)=0,\forall i \qquad \textbf{(2-14)}$$

Η Λαγκρανζιανή λόγω των ***(2-12),(2-13)*** γίνεται:

$$L(\vec{w},b,\lambda)=\frac{1}{2}\sum_{i=1}^{N}w_i\cdot w_i-\sum_{i=1}^{l}\lambda_i\cdot[y_i\cdot(\sum_{j=1}^{N}w_j x_j^{(i)}-b)-1]=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{l}w_i\lambda_j y_j x_i^{(j)}-$$
$$-\sum_{i=1}^{l}\sum_{j=1}^{N}\lambda_i y_i w_j x_j^{(i)}+b\cancel{\sum_{i=1}^{l}\lambda_i y_i}+\sum_{i=1}^{l}\lambda_i=\sum_{i=1}^{l}\lambda_i+\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{l}w_i\lambda_j y_j x_i^{(j)}-\sum_{i=1}^{N}\sum_{j=1}^{l}w_i\lambda_j y_j x_i^{(j)}\Rightarrow$$

$$L(\vec{w},b,\vec{\lambda})=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{l}w_i\lambda_j y_j x_i^{(j)}=\sum_{i=1}^{l}\lambda i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{l}\sum_{k=1}^{l}\lambda_k y_k x_i^{(k)}\lambda_j y_j x_i^{(j)}=$$
$$=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{j=1}^{l}\sum_{k=1}^{l}\lambda_k\lambda_j y_k y_j\cdot<\vec{x}_k,\vec{x}_j>$$

Με κατάλληλη αλλαγή των δεικτών άθροισης καταλήγουμε στην εξής ισοδύναμη μορφή της:

$$L(\vec{w},b,\vec{\lambda}) = \sum_{i=1}^{l}\lambda_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y_i y_j \cdot <\vec{x}_i,\vec{x}_j> \qquad \textbf{(2-15)}$$

Σύμφωνα με τη δυαδική θεωρία του Wolfe *[FLE87]*, η ελαχιστοποίηση της Langrangian της σχέσης ***(2-9)*** με τους περιορισμούς ***(2-10),*** είναι ισοδύναμη με τη μεγιστοποίηση της Lagrangian της σχέσης ***(2-15)*** με τους περιορισμούς :

$$\sum_{i=1}^{l}\lambda_i y_i = 0 \qquad \textbf{(2-16)}$$

όπου οι πολλαπλασιαστές $\lambda_i \geq 0$

Με βάση τα παραπάνω, το πρόβλημα βελτιστοποίησης διαμορφώνεται ως εξής:

$$\max : L(\vec{w},b,\vec{\lambda}) = \sum_{i=1}^{l}\lambda_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y_i y_j \cdot <\vec{x}_i,\vec{x}_j> \qquad \textbf{(2-15}^{\beta}\textbf{)}$$

με περιορισμούς: $\sum_{i=1}^{l}\lambda_i y_i = 0$

$\lambda_i \geq 0, i = 1,\ldots,l$

<u>*Παρατηρήσεις για το πρόβλημα βελτιστοποίησης :*</u>

Οι συνθήκες ΚΚΤ προσφέρουν πολύ ενδιαφέρουσες πληροφορίες για τη μεθοδολογία του *SVM*. Συγκεκριμένα, η σχέση ***(2-12)*** μπορεί να γραφεί σε διανυσματική μορφή ως εξής:

$$\vec{w} = \sum_{i=1}^{l}\lambda_i y_i \vec{x}_i \qquad \textbf{(2-17)}$$

Από την ανωτέρω σχέση εξάγεται το συμπέρασμα ότι το διάνυσμα $\vec{w}$ της βέλτιστης διαχωριστικής ευθείας (δηλαδή η κλίση - μορφή της ευθείας αυτής) δίνεται από μια σειρά, στην οποία περιέχονται μόνο εκείνα τα σημεία $\vec{x}_i$ που αντιστοιχούν σε μη-μηδενικούς πολλαπλασιαστές λ_i. Όμως, όπως μπορεί εύκολα να φανεί από τη σχέση ***(2-14)***, όταν για το σημείο $\vec{x}_i$ ισχύει $\lambda_i > 0$, τότε αναγκαστικά ισχύει και $y_i \cdot (< w, x_i > +b) - 1 = 0$, δηλαδή το σημείο αυτό ικανοποιεί:

$$< \vec{w}, \vec{x}_i > +b = 1, \text{ αν } \vec{x}_i \in class+1 \quad \text{ή}$$
$$< \vec{w}, \vec{x}_i > +b = -1, \text{ αν } \vec{x}_i \in class-1 \qquad \textbf{(2-18)}$$

Οι σχέσεις ***(2-18)*** ικανοποιούνται μόνο για τα ακραία σημεία κάθε κλάσης, δηλαδή τα διανύσματα υποστήριξης (support vectors). Με άλλα λόγια, το διάνυσμα προσανατολισμού $\vec{w}$ της βέλτιστης ευθείας διαχωρισμού αναλύεται σε μια σειρά που περιέχει μόνο τα διανύσματα υποστήριξης. Αυτά είναι και τα μόνα σημεία που επηρεάζουν την μορφή που θα έχει η ευθεία διαχωρισμού.

Το διάνυσμα προσανατολισμού της διαχωριστικής ευθείας αναλύεται σε μια σειρά με όρους μόνο τα διανύσματα υποστήριξης κάθε κλάσης.

Με γενικότερη διατύπωση, μπορούμε να πούμε ότι: **Η μορφή της διαχωριστικής ευθείας καθορίζεται αποκλειστικά και μόνο από τα διανύσματα υποστήριξης κάθε κλάσης.** Όλα τα υπόλοιπα σημεία με $\lambda_i = 0$, δεν υπεισέρχονται στο άθροισμα της σχέσης ***(2-17)*** , και συνεπώς δεν επηρεάζουν το $\vec{w}$ (τη μορφή της ευθείας). Είτε υπάρχουν στο σύνολο

εκπαίδευσης, είτε όχι, το αποτέλεσμα (η βέλτιστη διαχωριστική γραμμή) θα είναι το ίδιο.

Αποτέλεσμα του διαχωρισμού – συνάρτηση απόφασης

Η επίλυση του προβλήματος τετραγωνικής βελτιστοποίησης ***(2-9)*** παρέχει τα λ_i που αντιστοιχούν στο διάνυσμα προσανατολισμού $\vec{w}_o$ της βέλτιστης ευθείας διαχωρισμού. Η παράμετρος b_o της ευθείας υπολογίζεται από τη συνθήκη ΚΚΤ ***(2-14)***. Επιλέγουμε ένα λ_i και λύνουμε την ***(2-14)*** ως προς b_o[1]. Το κριτήριο τώρα για την ταξινόμηση ενός νέου σημείου σε μια από τις δύο κλάσεις (+1 ή -1), ή αλλιώς η **συνάρτηση απόφασης (decision function)** που εφαρμόζεται από τον SVM, έχει τη μορφή *[VAP98]*:

$$f(\vec{x}, \vec{w}_o) = sign(< \vec{w}_o, \vec{x} > -b_o) \qquad \textbf{(2-19)}$$

η οποία όπως είπαμε και στην αρχή του κεφαλαίου μπορεί να πάρει μόνο δύο τιμές : +1 ή -1. Αυτή είναι η συνάρτηση που παρέχει την εκτίμηση του *SVM* για την ετικέτα ενός νέου σημείου. Αντικαθιστώντας το $\vec{w}_o$ από τη σχέση ***(2-17)***, η συνάρτηση απόφασης του *SVM* γράφεται και:

$$f(\vec{x}, \vec{\lambda}) = sign(\sum_{i \in SVs} \lambda_i y_i < x_i, \vec{x} > -b_o)$$

όπου η άθροιση γίνεται στο σύνολο των διανυσμάτων υποστήριξης. Έτσι, αν για ένα νέο σημείο $\vec{x}^*$ ισχύει

$$f(\vec{x}^*, \vec{\lambda}) = sign(\sum_{i \in SVs} \lambda_i y_i < x_i, \vec{x}^* > -b_o) = +1$$

[1] Στην πράξη και για λόγους αριθμητικής ευστάθειας, οι ***(2-14)*** επιλύονται ως προς b_o για όλα τα λ_i, και η τελική τιμή του b_o προκύπτει ως ο μέσος όρος των τιμών που υπολογίστηκαν.

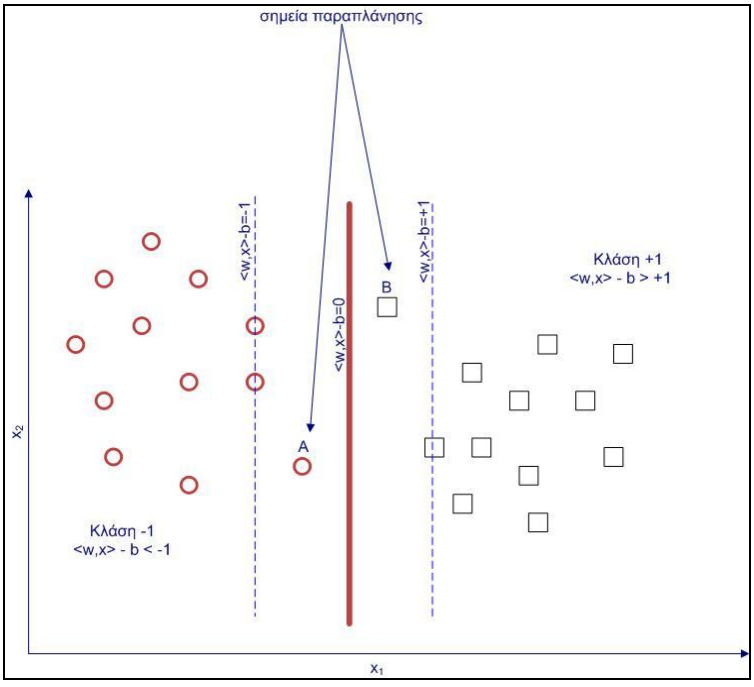
τότε το $\vec{x}^*$ θα ταξινομηθεί στην κλάση +1 , ενώ αν ισχύει

$$f(\vec{x}^*, \vec{\lambda}) = sign(\sum_{i \in SVs} \lambda_i y_i < x_i, \vec{x}^* > -b_o) = -1$$

τότε το $\vec{x}^*$ θα ταξινομηθεί στην κλάση -1.

2.5.1.2. Η περίπτωση του μη-απόλυτα διαχωρίσιμου συνόλου εκπαίδευσης

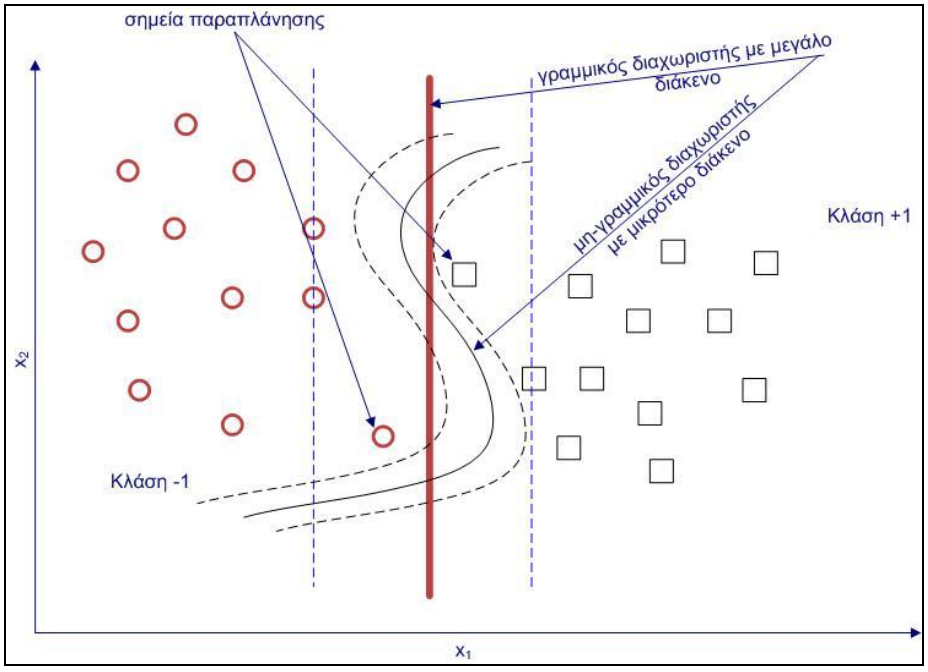
Μέχρι στιγμής εξετάσαμε την περίπτωση όπου το σύνολο εκπαίδευσης ήταν απόλυτα διαχωρίσιμο από μια γραμμική συνάρτηση. Πράγματι, τα σημεία του συνόλου εκπαίδευσης του σχήματος 12 βρίσκονταν όλα έξω από τη διαχωριστική ζώνη, δεν υπήρχαν δηλαδή σημεία παραπλάνησης που να εμπίπτουν στη ζώνη. Ας θεωρήσουμε τώρα την περίπτωση όπου στο σύνολο εκπαίδευσης του σχήματος 12, έχουν προστεθεί δύο ακόμα σημεία, όπως φαίνεται στο σχήμα 13 (σημεία A και B).



Σχήμα 13: Περίπτωση μη-απόλυτα διαχωρίσιμου συνόλου εκπαίδευσης, όπου δύο σημεία του (σημεία A και B) εμπίπτουν στο εσωτερικό της διαχωριστικής ζώνης.

Τα σημεία A και B βρίσκονται εντός των ορίων της διαχωριστικής ζώνης, και συνεπώς πρόκειται για σημεία παραπλάνησης. Στην περίπτωση αυτή,

η ανωτέρω διαχωριστική ζώνη αποτυγχάνει στο να διαχωρίσει επιτυχώς τις δύο κλάσεις. Για να επεκτείνουμε τη χρήση της μεθόδου *SVM* και στην περίπτωση συνόλων εκπαίδευσης όπως το ανωτέρω (τα οποία εφεξής θα ονομάζουμε **μη-γραμμικά διαχωρίσιμα** *[VAP95]*), υπάρχουν δύο εναλλακτικές προσεγγίσεις. Η πρώτη προσέγγιση είναι να χρησιμοποιηθεί διαχωριστική ζώνη μεγαλυτέρου βαθμού, π.χ. πολυωνυμικής μορφής 2ου ή 3ου βαθμού, που θα έχει μεγαλύτερη γεωμετρική ευελιξία και συνεπώς θα μπορεί να «τοποθετηθεί» στο χώρο εισόδων με τέτοιο τρόπο ώστε κανένα σημείο να μην εμπίπτει στο εσωτερικό της. Ωστόσο, η αντιμετώπιση αυτή δεν είναι πάντα κατάλληλη, καθώς με το να αυξήσει κανείς την πολυπλοκότητα της οικογένειας συναρτήσεων διαχωρισμού, δηλαδή τη χωρητικότητα αυτής, μειώνει παράλληλα το διάκενο του διαχωριστή και πιθανόν να μην επιτύχει καλή γενίκευση σε μελλοντικά σημεία. Ένα παράδειγμα που αποδεικνύει την ανωτέρω σκέψη είναι αυτό του σχήματος 14.



Σχήμα 14: Ο μη-γραμμικός διαχωριστής έχει μικρότερο διάκενο από τον αντίστοιχο γραμμικό, και συνεπώς δεν θα επιτύχει καλή γενίκευση στο εν λόγω σύνολο εκπαίδευσης.

Είναι προφανές ότι ο πολύπλοκος μη-γραμμικός διαχωριστής που επιλέχτηκε για το συγκεκριμένο σύνολο εκπαίδευσης, θα έχει χειρότερη γενίκευση από έναν πιο απλό γραμμικό διαχωριστή, ακόμα κι αν ο τελευταίος δεν εξασφαλίζει απόλυτο διαχωρισμό των σημείων του συνόλου εκπαίδευσης. Κι αυτό γιατί το διάκενο (το εύρος της μη-γραμμικής ζώνης διαχωρισμού) είναι σημαντικά μειωμένο σε σχέση με αυτό του γραμμικού διαχωριστή.

Μια εναλλακτική αντιμετώπιση *[COR]* είναι να χρησιμοποιηθεί και πάλι γραμμικός διαχωριστής, αλλά να επιβληθεί με κάποιο τρόπο ποινή στα σημεία παραπλάνησης, εισάγοντας στην αντικειμενική συνάρτηση ***(2-9)*** έναν επιπλέον όρο. Ο όρος αυτός είναι της μορφής: $C\sum_{i=1}^{l}\xi_i$, όπου οι μεταβλητές $\xi_i \in \Re, 0 \le \xi_i \le 1$ που ονομάζονται **μεταβλητές χάρης (slack variables)**, είναι ένα μέτρο του σφάλματος ταξινόμησης που γίνεται στο κάθε σημείο παραπλάνησης, ενώ $C \in \Re$ είναι μια σταθερά που καθορίζεται από το χρήστη. Επιδιώκεται έτσι η ελαχιστοποίηση της ακόλουθης ποσότητας:

$$\frac{1}{2}\| w \|_2^{\,2} + C\sum_{i=1}^{l}\xi_i$$

Η ελαχιστοποίηση του πρώτου όρου της ανωτέρω αντικειμενικής συνάρτησης έχει να κάνει με τη μεγιστοποίηση του διακένου του διαχωριστή, όπως συνέβαινε και στην περίπτωση του απόλυτα διαχωρίσιμου συνόλου του σχήματος 12. Ο δεύτερος όρος έχει να κάνει με την ελαχιστοποίηση των σφαλμάτων ταξινόμησης στο σύνολο εκπαίδευσης. Με άλλα λόγια, ο πρώτος όρος ελαχιστοποιεί τη VC-εμπιστοσύνη, ενώ ο δεύτερος ελαχιστοποιεί το εμπειρικό ρίσκο. Οι δύο αυτοί όροι είναι αντικρουόμενοι. Πράγματι, για να μειωθεί ο πρώτος όρος και συνεπώς να εξασφαλισθεί μικρή VC-εμπιστοσύνη, θα πρέπει ο

διαχωριστής να έχει μικρή χωρητικότητα. Αυτό όμως συνεπάγεται περισσότερα σφάλματα στο σύνολο εκπαίδευσης (μεγαλύτερο εμπειρικό ρίσκο), άρα μεγαλύτερη τιμή του δεύτερου όρου. Η σχετική βαρύτητα των δύο όρων στο πρόβλημα ελαχιστοποίησης καθορίζεται από τη ρυθμιστική παράμετρο C. Όταν η τιμή του C επιλεγεί μεγάλη, δίνεται βαρύτητα στην ελαχιστοποίηση των σφαλμάτων ταξινόμησης του συνόλου εκπαίδευσης. Από την άλλη, όταν η τιμή του C επιλεγεί μικρή, δίνεται βαρύτητα στη μεγιστοποίηση του διακένου του γραμμικού διαχωριστή. Η τιμή του C δηλαδή δρα ως επιπρόσθετη παράμετρος ελέγχου της χωρητικότητας του διαχωριστή.

Για να μπορεί να λειτουργήσει η μέθοδος *SVM* στην περίπτωση ύπαρξης σημείων παραπλάνησης, θα πρέπει οι περιορισμοί ***(2-10)*** να γίνουν πιο ελαστικοί, ώστε να δέχονται και τέτοια σημεία στο σύνολο εκπαίδευσης. Για να γίνει αυτό, εισάγονται στους ανισοτικούς περιορισμούς οι μεταβλητές ξ_i, και οι τελευταίοι γίνονται *[COR],[VAP98]*:

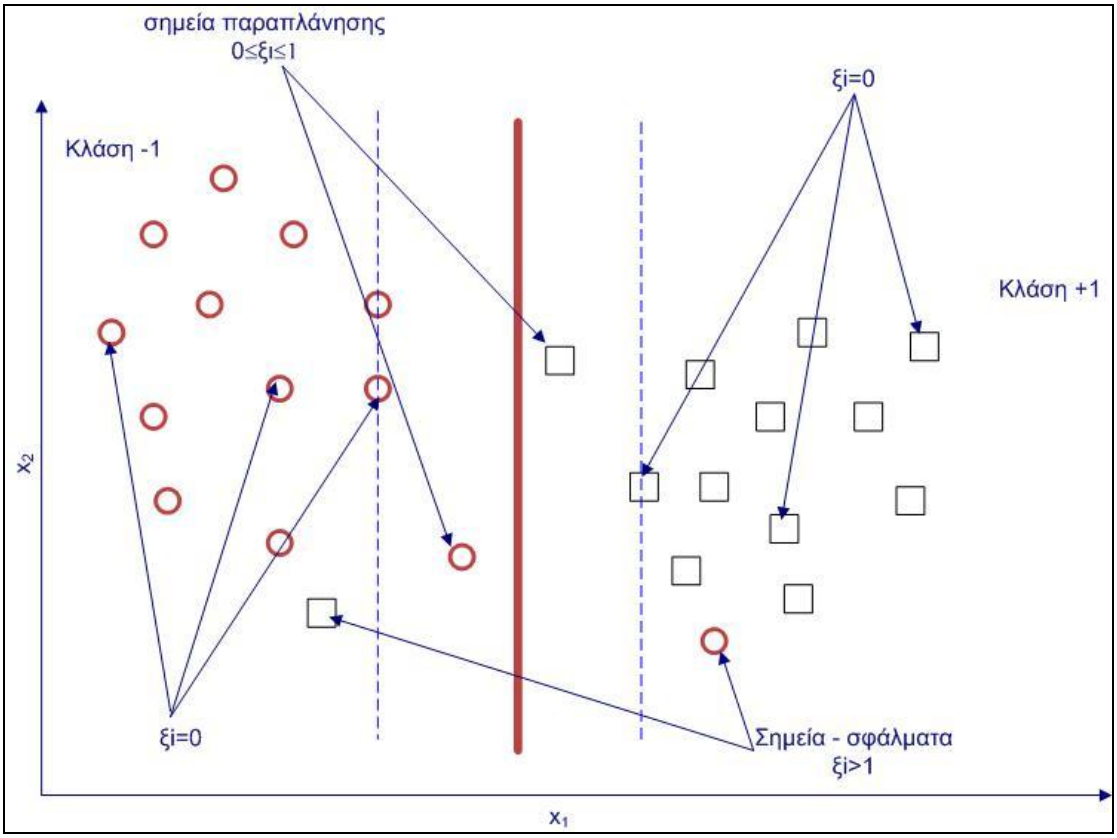
$$< \vec{w}, \vec{x}_i > -b \geq 1 - \xi_i, \quad \alpha\nu \ y_i = +1 \qquad \textbf{(2-20}^{\alpha}\textbf{)}$$

$$< \vec{w}, \vec{x}_i > -b \leq \xi_i - 1, \quad \alpha\nu \ y_i = -1 \qquad \textbf{(2-20}^{\beta}\textbf{)}$$

Με αναφορά στο σχήμα 15, διακρίνουμε τρεις κατηγορίες σημείων στο σύνολο εκπαίδευσης:

1) Σημεία με $\xi_i = 0$, που είναι τα πλήρως διαχωρίσιμα σημεία του συνόλου εκπαίδευσης, δηλαδή βρίσκονται έξω από τα όρια της διαχωριστικής ζώνης και συνεπώς δεν αποτελούν σημεία παραπλάνησης. Στην προηγούμενη παράγραφο που εξετάζαμε πλήρως διαχωρίσιμα σύνολα, αυτά τα σύνολα απαρτίζονταν μόνο

από τέτοια σημεία, αφού ικανοποιούσαν όλα τους περιορισμούς ***(2-10)***.



Σχήμα 15: Οι τρεις κατηγορίες σημείων ενός συνόλου εκπαίδευσης, με τις αντίστοιχες μεταβλητές ξ_i

2) Σημεία με $0 \le \xi_i \le 1$, που βρίσκονται εντός της ζώνης διαχωρισμού, αλλά από την ίδια πλευρά της διαχωριστικής γραμμής με τα υπόλοιπα σημεία της κλάσης τους. Πρόκειται δηλαδή για σημεία παραπλάνησης. Εφόσον τα σημεία αυτά εξακολουθούν να ικανοποιούν τους (πιο ελαστικούς πλέον) περιορισμούς ***(2-20)***, θα ταξινομηθούν και αυτά σωστά από τον *SVM*.

3) Σημεία με $\xi_i > 1$, που όχι μόνο βρίσκονται μέσα στο εύρος της ζώνης διαχωρισμού, αλλά και στην απέναντι πλευρά της διαχωριστικής γραμμής από αυτή που βρίσκονται τα υπόλοιπα σημεία της κλάσης τους. Πρόκειται για **σημεία-σφάλματα**, τα οποία θα ταξινομηθούν

λάθος από τον *SVM*. Κι αυτό γιατί, αν στους περιορισμούς ***(2-20)*** ισχύσει $\xi_i > 1$, τότε ικανοποιούνται οι ανισοτικοί περιορισμοί της αντίθετης κλάσης από αυτήν που ανήκει το σημείο, δηλαδή το σημείο θα θεωρηθεί ως ανήκον στην αντίθετη κλάση.

Με βάση τα παραπάνω, το πρόβλημα ελαχιστοποίησης στην περίπτωση μη-απόλυτα διαχωρίσιμου συνόλου εκπαίδευσης γίνεται:

$$\min : \frac{1}{2} \| \vec{w} \|_2^2 + C \sum_{i=1}^{l} \xi_i \qquad \textbf{(2-21)}$$

με : $y_i \cdot (< \vec{w}, \vec{x}_i > - b) \geq 1 - \xi_i \quad , \quad \xi_i \geq 0$

Για την επίλυσή του χρησιμοποιείται και πάλι η μέθοδος Lagrange. Σχηματίζουμε έτσι την Lagrangian του ανωτέρω προβλήματος, και καταλήγουμε, όπως και στην περίπτωση του απόλυτα διαχωρίσιμου συνόλου εκπαίδευσης, στην εξής δυαδική μορφή:

$$\max : L(\vec{w}, \vec{\lambda}) = \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j \cdot < \vec{x}_i, \vec{x}_j > \qquad \textbf{(2-22)}$$

με τους περιορισμούς : $0 \leq \lambda_i \leq C$, και : $\sum_{i=1}^{l} \lambda_i y_i = 0$ ***(2-23)***

2.5.1.3 Ενοποίηση των δύο περιπτώσεων (διαχωρίσιμο και μη-διαχωρίσιμο σύνολο εκπαίδευσης)

Συγκρίνοντας το πρόβλημα ***(2-22)***, με το πρόβλημα βελτιστοποίησης στην περίπτωση απόλυτα διαχωρίσιμου συνόλου εκπαίδευσης ***(2-15$^{\beta}$)***, παρατηρούμε ότι πρόκειται για δύο κατά βάση ίδια προβλήματα, με μόνη διαφορά το άνω όριο C των πολλαπλασιαστών Lagrange, που αντί για $+\infty$ είναι τώρα πεπερασμένο. Έτσι, αυτά τα δύο προβλήματα μπορούν να

ενοποιηθούν σε ένα και μόνο πρόβλημα τετραγωνικού προγραμματισμού, που θα αφορά σε όλες τις περιπτώσεις, το εξής:

$$\max : L(\vec{w}, \vec{\lambda}) = \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j \cdot < \vec{x}_i, \vec{x}_j > \qquad \textbf{(2-24)}$$

με περιορισμούς: $0 \le \lambda_i \le C$

$$\sum_{i=1}^{l} \lambda_i y_i = 0 \qquad \textbf{(2-25)}$$

όπου C=+∞ όταν έχουμε απόλυτα διαχωρίσιμο σετ εκπαίδευσης, και C πεπερασμένο όταν το σετ δεν είναι απόλυτα διαχωρίσιμο.

Παρατηρήσεις:

Η ανωτέρω μεθοδολογία για το διαχωρισμό ενός μη-απόλυτα διαχωρίσιμου συνόλου εκπαίδευσης, προϋποθέτει κατάλληλη επιλογή της παραμέτρου C. Η επιλογή της παραμέτρου θα πρέπει να γίνεται με κριτήριο την προϋπάρχουσα γνώση που έχει ο χρήστης για το πρόβλημα ταξινόμησης που καλείται να λύσει. Για παράδειγμα, αν τα σημεία του συνόλου εκπαίδευσης προέρχονται από μετρήσεις που εμπεριέχουν θόρυβο, τότε είναι πιθανό ότι η ύπαρξη μερικών σημείων παραπλάνησης οφείλεται στο θόρυβο. Στην περίπτωση αυτή, δεν είναι απαραίτητο ότι πρέπει να επιλεγεί μεγάλη τιμή C ώστε ο διαχωριστής να διαχωρίσει ακόμα και τα σημεία θορύβου, καθώς τα σημεία αυτά δεν αντιπροσωπεύουν το φυσικό πρόβλημα. Απεναντίας, στην περίπτωση αυτή μια μικρή τιμή C στο διαχωριστή μπορεί να εξασφαλίσει πιθανότατα καλύτερη γενίκευση, καθώς αφήνει ανεπηρέαστο το διάκενο του διαχωριστή. Από την άλλη, αν τα σημεία «θορύβου» είναι πολλά, ίσως αυτό να σημαίνει ότι το εν λόγω σύνολο εκπαίδευσης χρειάζεται

πράγματι διαχωριστική γραμμή με μικρότερο εύρος ή και μη-γραμμική. Πάντως, μέχρι στιγμής δεν έχει αναπτυχθεί κάποιος συστηματικός τρόπος με τον οποίο να καθορίζεται εκ των προτέρων η βέλτιστη τιμή της παραμέτρου C για ένα συγκεκριμένο πρόβλημα. Στην πράξη χρησιμοποιούνται συγκριτικές δοκιμές της μεθόδου *SVM* με διάφορες τιμές του C, και επιλέγεται μετά η μηχανή που πέτυχε την καλύτερη γενίκευση σε ένα συγκεκριμένο σύνολο σημείων δοκιμής. Περισσότερες λεπτομέρειες για το θέμα αυτό θα δοθούν στην παράγραφο 2.6 .

2.5.2 Επέκταση για το N-διάστατο χώρο

Η επέκταση της μεθόδου από το 2-Δ στο N-διάστατο χώρο γίνεται άμεσα. Η μόνη διαφορά με πριν είναι ότι τα διανύσματα $\vec{x}, \vec{w}$ κείνται στο N-διάστατο χώρο και όχι στο επίπεδο $\Re^2$, αλλά αυτό δεν επηρεάζει σε τίποτα τη διαδικασία διαχωρισμού. Αυτό οφείλεται στην πολύ βολική μορφή που έχουν τόσο η γραμμή διαχωρισμού όσο και η συνάρτηση απόφασης: Τα διανύσματα $\vec{w}, \vec{x}$ εμπλέκονται μόνο σε πράξεις εσωτερικού γινομένου, δηλαδή απλούς πολλαπλασιασμούς και προσθέσεις των συντεταγμένων τους. Οι πράξεις αυτές δεν επηρεάζονται από τη διάσταση που έχουν τα διανύσματα. Το μόνο που αλλάζει όσο προσθέτουμε διαστάσεις στα διανύσματα, είναι το πλήθος των πράξεων και όχι η πολυπλοκότητά τους ή το είδος τους. Συνεπώς, μπορούμε να επεκτείνουμε την προγενέστερη ανάλυση από το 2-Δ στον πολυδιάστατο χώρο, χωρίς να χρειαστεί να αλλάξουμε καμία μαθηματική λεπτομέρεια της μεθόδου, παρά μόνο την ορολογία.

Πλέον, δεν μιλάμε για διαχωριστική ευθεία, αλλά για **διαχωριστικό υπερεπίπεδο (separating hyperplane)** *[VAP95],[VAP98]* το οποίο έχει την ίδια ακριβώς μορφή:

$$< \vec{w}, \vec{x} > -b = 0$$

μόνο που τώρα $\vec{w}, \vec{x} \in \Re^N$. Ακολουθώντας παρόμοια ορολογία, σε κάθε κλάση αντιστοιχίζεται ένα **υποστηρίζον υπερεπίπεδο (supporting hyperplane)** – κατ' αντιστοιχία με τις υποστηρίζουσες ευθείες. Η φιλοσοφία της μεθόδου παραμένει κατά τα άλλα η ίδια: Λύνουμε το πρόβλημα τετραγωνικής βελτιστοποίησης και υπολογίζουμε τα $\vec{w}_o, b_o$ του βέλτιστου διαχωριστικού υπερεπιπέδου **(optimal separating hyperplane)**: $< \vec{w}_o, \vec{x} > -b_o = 0$. Η συνάρτηση απόφασης του SVM θα είναι όμοια με πριν

$$f(\vec{x}, \vec{w}_o) = sign(< \vec{w}_o, \vec{x} > -b) = sign(\sum_{i \in SVs} \lambda_i y_i < \vec{x}_i, \vec{x} > -b_o)$$

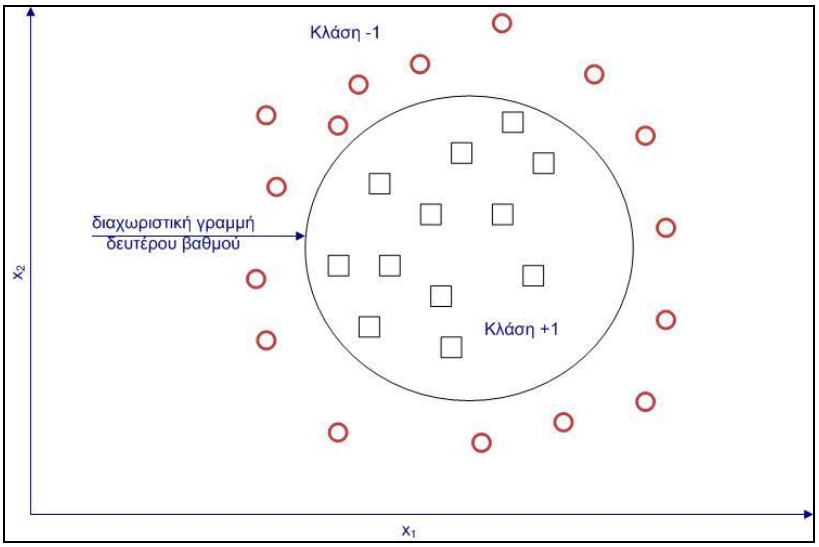
με $\vec{w}, \vec{x} \in \Re^N$.

2.5.3 Επέκταση για μη-γραμμικές συναρτήσεις διαχωρισμού

2.5.3.1 Χρησιμοποιώντας μη-γραμμικές απεικονίσεις

Η ανάλυση που προηγήθηκε αφορούσε σε περιπτώσεις που η συνάρτηση διαχωρισμού είναι γραμμική ως προς το $\vec{x}$. Αυτό όπως είπαμε και στην αρχή του κεφαλαίου δε δίνει ικανοποιητικά αποτελέσματα παρά μόνο για σύνολα εκπαίδευσης τα οποία όντως μπορούν να διαχωριστούν γραμμικά. Υπάρχει συνεπώς η ανάγκη για επέκταση της μεθόδου, ώστε αυτή να εφαρμόζει και μη-γραμμικές συναρτήσεις διαχωρισμού.

Έστω ότι δίνεται κάποιο σύνολο εκπαίδευσης $(\vec{x}_i \in \Re^2, y_i \in \{-1,+1\}), i = 1,...l$, όπως αυτό του σχήματος 16. Τα σημεία αυτού του συνόλου δεν είναι δυνατόν να διαχωριστούν με μια ευθεία. Ακόμα κι αν εφαρμόσουμε την μεθοδολογία που περιγράφηκε στην προηγούμενη παράγραφο για μη-απόλυτα διαχωρίσιμα σύνολα εκπαίδευσης, είναι προφανές ότι καμία



Σχήμα 16: Το ανωτέρω σύνολο εκπαίδευσης δεν μπορεί να διαχωριστεί από ευθεία, αλλά απαιτεί διαχωριστική γραμμή δευτέρου βαθμού για το διαχωρισμό του.

γραμμική διαχωριστική ζώνη δεν μπορεί να εξασφαλίσει καλή γενίκευση στο συγκεκριμένο παράδειγμα. Το εν λόγω σύνολο εκπαίδευσης απαιτεί τετραγωνική συνάρτηση. Δηλαδή συνάρτηση της οποίας η μορφή δεν εξαρτάται από το $< \vec{x}_i, \vec{x}_j >$, αλλά από το $< \vec{x}_i, \vec{x}_j >^2$. Στην περίπτωση αυτή δεν μπορεί να εφαρμοστεί η θεωρία του γραμμικού διαχωρισμού, εφόσον η επιδιωκόμενη συνάρτηση διαχωρισμού δεν είναι γραμμική. Το ζητούμενο είναι να επιτύχουμε να διατηρήσουμε την μαθηματική απλότητα του γραμμικού διαχωρισμού, και ταυτόχρονα να παράγουμε την απαιτούμενη μη-γραμμική συνάρτηση.

Για να το επιτύχουμε αυτό, εκμεταλλευόμαστε και πάλι την ιδιότητα του *SVM* να εξαρτάται αποκλειστικά από πράξεις εσωτερικού γινομένου *[BOZ]*. Χρησιμοποιούμε μια μη-γραμμική απεικόνιση Φ *[AIZ64]*, η οποία «περνάει» τα διανύσματα παραμέτρων $\vec{x}_i$ από το χώρο εισόδων L, σε έναν άλλο (πιθανώς απειρο-διάστατο) χώρο H. Ο χώρος H ονομάζεται και **χώρος χαρακτηριστικών (feature space).**

Σχηματικά:

$$\Phi : L \to H$$

Ο στόχος αυτής της απεικόνισης είναι, η μη-γραμμική πράξη μεταξύ των $\vec{x}_i, \vec{x}_j$ στο χώρο L, να μετατραπεί σε πράξη εσωτερικού γινομένου μεταξύ των $\Phi(\vec{x}_i), \Phi(\vec{x}_j)$ στο χώρο H. Ο λόγος είναι ότι η εφαρμογή του γραμμικού *SVM* που αναπτύξαμε παραπάνω, απαιτεί πράξεις εσωτερικού γινομένου μεταξύ των διανυσμάτων παραμέτρων.

Η ανωτέρω απεικόνιση πρακτικά επιτυγχάνεται εισάγοντας στα ήδη υπάρχοντα διανύσματα παραμέτρων νέες συντεταγμένες, που είναι μη-γραμμικοί συνδυασμοί των υπαρχουσών *[BUR98],[BEN]*. Έτσι, αν $\vec{x}_i = (\vec{x}_{i1}, \vec{x}_{i2})$ είναι τα αρχικά διανύσματα παραμέτρων του χώρου εισόδων, τότε για την περίπτωση π.χ. που θέλουμε να εφαρμόσουμε την ανωτέρω συνάρτηση διαχωρισμού (πολυωνυμική $2^{ου}$ βαθμού) τα επαυξημένα διανύσματα παραμέτρων θα είναι

$$\Phi(\vec{x}_i) = (x_{i1}{}^2, \sqrt{2}x_{i1} \cdot x_{i2}, x_{i2}{}^2) \ ,$$

και συνεπώς ο χώρος στον οποίο κείνται αυτά θα έχει διάσταση 3, δηλαδή $H = \Re^3$.

Πράγματι, αποδεικνύεται τότε με απλές αλγεβρικές πράξεις ότι

$$< \vec{x}_i, \vec{x}_j >^2 = (x_{i1}x_{j1} + x_{i2}x_{j2})^2 = (x_{i1}x_{j1})^2 + (x_{i2}x_{j2})^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} =$$
$$=< ({x_{i1}}^2, {x_{i2}}^2, \sqrt{2}x_{i1}x_{i2}), ({x_{j1}}^2, {x_{j2}}^2, \sqrt{2}x_{j1}x_{j2}) >=< \Phi(\vec{x}_i), \Phi(\vec{x}_j) >$$

Παρατηρούμε δηλαδή ότι, η μη γραμμική πράξη μεταξύ των $\vec{x}_i, \vec{x}_j$ στο χώρο L, έγινε πράξη εσωτερικού γινομένου των $\Phi(\vec{x}_i), \Phi(\vec{x}_j)$ στο νέο χώρο H. Εφόσον η συνάρτηση διαχωρισμού και η συνάρτηση απόφασης του *SVM* εξαρτώνται μόνο από πράξεις εσωτερικού γινομένου μεταξύ των σημείων του συνόλου εκπαίδευσης, μπορούμε να εφαρμόσουμε το γραμμικό *SVM* στο νέο χώρο H, όπου θα έχουμε πράξεις εσωτερικού γινομένου μεταξύ των απεικονίσεων $\Phi(\vec{x}_i)$ των σημείων του συνόλου εκπαίδευσης. Με άλλα λόγια, για να επιτύχουμε μη- γραμμικό διαχωρισμό δεδομένων, εφαρμόζουμε και πάλι τον γραμμικό *SVM*, αφού όμως πρώτα περάσουμε τα σημεία του συνόλου εκπαίδευσης σε διαφορετικό χώρο (τον H)[2], μέσω μιας κατάλληλης απεικόνισης Φ. Έτσι παράγεται η εξίσωση του βέλτιστου γραμμικού διαχωριστή των απεικονίσεων $\Phi(\vec{x}_i)$ στο χώρο H, η οποία είναι:

$$< \vec{w}_o, \Phi(\vec{x}) > - b_o = 0 \Leftrightarrow \sum_{i \in SVs} \lambda_i y_i \cdot < \Phi(\vec{x}_i), \Phi(\vec{x}) > - b_o = 0 \qquad \textbf{(2-26)}$$

Στο χώρο L ωστόσο, όπου κείνται τα αρχικά διανύσματα παραμέτρων $\vec{x}_i$, η ανωτέρω διαχωριστική γραμμή θα είναι τετραγωνική, αφού όπως δείξαμε:

$$< \Phi(\vec{x}_i), \Phi(\vec{x}_j) >=< \vec{x}_i, \vec{x}_j >^2 \quad .$$

---

[2] *Παρατήρηση: Είναι προφανές ότι εφόσον περάσαμε τα διανύσματα παραμέτρων $\vec{x}_i$ στο νέο χώρο H, το διάνυσμα $\vec{w}$ της διαχωριστικής γραμμής θα κείται επίσης στο χώρο H.*

2.5.3.2 Η συνάρτηση κελύφους

Το μειονέκτημα που προκύπτει κατά την ανωτέρω διαδικασία είναι ότι πρέπει να γνωρίζουμε ποια ακριβώς είναι αυτή η απεικόνιση Φ, που θα καταφέρει να μετατρέψει την εκάστοτε μη γραμμική πράξη του χώρου L, σε πράξη εσωτερικού γινομένου στο χώρο H. Με άλλα λόγια, θα πρέπει να γνωρίζουμε κάθε φορά ποιες νέες παραμέτρους να εισαγάγουμε στα $\vec{x}_i$ για να πάρουμε τα κατάλληλα $\Phi(\vec{x}_i)$. Τη δυσκολία αυτή την παρακάμπτουμε χρησιμοποιώντας αυτό που εφεξής θα ονομάζουμε **συνάρτηση κελύφους (kernel function)**, K. Η συνάρτηση αυτή υλοποιεί αυτήν ακριβώς την απεικόνιση. Δηλαδή ισχύει για τη συνάρτηση K :

$$K(\vec{x}_i, \vec{x}_j) = <\Phi(\vec{x}_i), \Phi(\vec{x}_j)> \qquad \textbf{(2-27)}$$

Αντικαθιστούμε τώρα το εσωτερικό γινόμενο $<\Phi(\vec{x}_i), \Phi(\vec{x}_j)>$, με τη συνάρτηση κελύφους K. Έτσι, η εξίσωση της διαχωριστικής γραμμής γίνεται:

$$\sum_{i \in SVs} \lambda_i y_i K(\vec{x}_i, \vec{x}) - b = 0 \qquad \textbf{(2-28)}$$

και εξαρτάται άμεσα πλέον μόνο από τη συνάρτηση κελύφους, και όχι από την απεικόνιση Φ, την οποία δε χρειάζεται καν να γνωρίζουμε. Για την προκειμένη περίπτωση π.χ., που θέλουμε συνάρτηση διαχωρισμού πολυωνυμικής μορφής 2ου βαθμού, η εξίσωση της συνάρτησης κελύφους θα είναι: $K(x_i, x_j) = < x_i, x_j >^2$.

Με αυτό το «έξυπνο» τέχνασμα, πετυχαίνουμε η εφαρμογή του *SVM* να είναι ακριβώς η ίδια με τη γραμμική περίπτωση, ενώ η μη- γραμμικότητα δεν μας απασχολεί καθόλου, αφού υλοποιείται έμμεσα με χρήση της συνάρτησης κελύφους. Η γνώση της απεικόνισης Φ είναι περιττή. Συνοψίζοντας, η γενίκευση του *SVM* σε μη-γραμμικές συναρτήσεις διαχωρισμού, είναι η εξής:

Εξίσωση διαχωριστικής γραμμής: $\sum_{i \in SVs} \lambda_i y_i K(\vec{x}_i, \vec{x}) - b_o = 0$ ***(2-29)***

Συνάρτηση απόφασης: $f(x) = sign(\sum_{i \in SVs} \lambda_i y_i K(\vec{x}_i, \vec{x}) - b_o)$ ***(2-30)***

<u>*Παρατηρήσεις για τη συνάρτηση κελύφους:*</u>

Στα παραπάνω, αφήσαμε να εννοηθεί ότι η συνάρτηση κελύφους $K(\vec{x}_i, \vec{x}_j)$ ισοδυναμεί με μια πράξη εσωτερικού γινομένου σε κάποιον (πιθανόν απειρο-διάστατο) χώρο $\boldsymbol{H}$. Για να ισχύει ωστόσο κάτι τέτοιο, θα πρέπει η συνάρτηση κελύφους να ικανοποιεί κάποιες προϋποθέσεις, ώστε να είναι κατάλληλη για χρήση ως «απεικόνιση εσωτερικού γινομένου» στον *SVM*. Οι προϋποθέσεις αυτές παρέχονται από το παρακάτω θεώρημα *[COU],[VAP98]*.

<u>*2.5.3.3 Θεώρημα του Mercer*</u>

Μια συνάρτηση K μπορεί να γραφεί ως εσωτερικό γινόμενο κάποιας απεικόνισης Φ, δηλαδή ως :

$$K(\vec{x}, \vec{y}) = \sum_i \Phi(\vec{x})_i \cdot \Phi(\vec{y})_i$$

αν και μόνο αν ικανοποιούνται οι εξής προϋποθέσεις: Για κάθε συνάρτηση g(x), τέτοια ώστε

$$\int g(x)^2 dx \in \Re$$

να ισχύει

$$\int K(x,y)g(x)g(y)dxdy \geq 0$$

2.5.3.4. Διερεύνηση του θεωρήματος του Mercer – ιδιότητες των συναρτήσεων κελύφους

Κατά την επίλυση του προβλήματος τετραγωνικής βελτιστοποίησης, υπολογίζεται η τιμή της συνάρτησης κελύφους για κάθε πιθανό συνδυασμό $\vec{x}_i, \vec{x}_j$ των σημείων του συνόλου εκπαίδευσης. Οι τιμές $K(\vec{x}_i, \vec{x}_j)$ ή για συντομία $K(i, j)$ της συνάρτησης κελύφους, σχηματίζουν έναν πίνακα που ονομάζεται **μήτρα κελύφους (kernel matrix)**.

Από το θεώρημα του Mercer συνάγεται ότι για να είναι μια συνάρτηση ***K*** έγκυρη συνάρτηση κελύφους στον *SVM*, θα πρέπει η αντίστοιχη μήτρα κελύφους της να είναι συμμετρική και θετικά ορισμένη. Ισχύει ωστόσο και το αντίστροφο: Κάθε συμμετρική και θετικά ορισμένη μήτρα, είναι έγκυρη μήτρα κελύφους για τον *SVM*, δηλαδή μπορεί να χρησιμοποιηθεί ως μήτρα εσωτερικών γινομένων σε κάποιο χώρο ***H.***

Ακόμα, αποδεικνύονται εύκολα τα εξής: Αν οι $K_1(\vec{x}, \vec{z})$, $K_2(\vec{x}, \vec{z})$ είναι έγκυρες συναρτήσεις κελύφους, και $c \in \Re$ μια σταθερά, τότε και οι ακόλουθες συναρτήσεις είναι κι αυτές έγκυρες συναρτήσεις κελύφους :

- $K(\vec{x}, \vec{z}) = c \cdot K_1(\vec{x}, \vec{z})$
- $K(\vec{x}, \vec{z}) = c + K_1(\vec{x}, \vec{z})$

- $K(\vec{x},\vec{z}) = K_1(\vec{x},\vec{z}) + K_2(\vec{x},\vec{z})$
- $K(\vec{x},\vec{z}) = K_1(\vec{x},\vec{z}) \cdot K_2(\vec{x},\vec{z})$

Στην εφαρμογή του SVM, χρησιμοποιούνται διάφορες συναρτήσεις κελύφους, εκ των οποίων οι σημαντικότερες είναι οι εξής:

- Κέλυφος εσωτερικού γινόμενου (dot product kernel):
  $K(\vec{x}_i,\vec{x}_j) = <\vec{x}_i,\vec{x}_j>$. Πρόκειται για το κέλυφος που χρησιμοποιείται στην απλή περίπτωση του γραμμικού διαχωρισμού.
- Πολυωνυμικό κέλυφος (polynomial kernel):
  $K(\vec{x}_i,\vec{x}_j) = (<\vec{x}_i,\vec{x}_j>+1)^p$, όπου p ο βαθμός του πολυωνύμου.
- Κέλυφος ακτινικής βάσης (RBF ή Gaussian kernel):
  $K(\vec{x}_i,\vec{x}_j) = e^{-\|\vec{x}_i-\vec{x}_j\|^2/(2\sigma^2)}$, όπου σ μια παράμετρος που ονομάζεται τυπική απόκλιση.
- Σιγμοειδές κέλυφος (sigmoid kernel):
  $K(\vec{x}_i,\vec{x}_j) = \tanh(k\cdot <\vec{x}_i,\vec{x}_j> -\delta)$

2.6 Βέλτιστη επιλογή των παραμέτρων του SVM

2.6.1 Η μέθοδος της διασταυρωτικής επιλογής(cross validation) και οι παραλλαγές της

Για την εφαρμογή του *SVM*, θα πρέπει αρχικά να επιλέξουμε τη συνάρτηση κελύφους που θα χρησιμοποιηθεί για το διαχωρισμό του συνόλου εκπαίδευσης, αλλά και μετέπειτα για την ταξινόμηση νέων σημείων. Στην πλειοψηφία των περιπτώσεων, δεν είμαστε σε θέση να

γνωρίζουμε εκ των προτέρων ποια από όλες τις συναρτήσεις κελύφους υπόσχεται την καλύτερη γενίκευση. Ένας γενικός κανόνας είναι ότι η μορφή της συνάρτησης κελύφους θα πρέπει να είναι τοπολογικά συμβατή με την τοπολογία των σημείων του συνόλου εκπαίδευσης που αυτή καλείται να διαχωρίσει. Ωστόσο, αυτός ο κανόνας είναι περισσότερο διαισθητικός και δεν παρέχει ακριβείς πληροφορίες για την καταλληλότητα ή μη μιας συνάρτησης. Επιπλέον, σε όλες σχεδόν τις οικογένειες συναρτήσεων κελύφους (εκτός αυτής του εσωτερικού γινομένου), υπάρχουν μια ή περισσότερες σταθερές τις οποίες και πρέπει να καθορίσουμε. Για παράδειγμα, αν χρησιμοποιηθεί πολυωνυμικό κέλυφος, της μορφής

$$K(x_i, x_j) = (< x_i, x_j > +1)^p$$

πρέπει να καθοριστεί η τιμή του p, ενώ αν χρησιμοποιηθεί κέλυφος ακτινικής βάσης (RBF):

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

πρέπει να καθοριστεί η τιμή του σ. Τέλος, μια ακόμα βασική επιλογή που πρέπει να γίνει, είναι αυτή της παραμέτρου C του προβλήματος βελτιστοποίησης, αφού αυτή θα καθορίσει την ελαστικότητα- ανοχή της διαχωριστικής ζώνης σε σημεία παραπλάνησης.

Δημιουργείται συνεπώς ένα **πρόβλημα βέλτιστης επιλογής παραμέτρων** για την εφαρμογή του *SVM*. Είναι γεγονός ότι η προϋπάρχουσα γνώση πάνω στο συγκεκριμένο κάθε φορά πρόβλημα ταξινόμησης προσφέρει χρήσιμη καθοδήγηση για την επιλογή αυτών των

παραμέτρων, και συνεπώς μπορεί να συντελέσει στη βέλτιστη επιλογή τους. Στις περιπτώσεις όμως όπου η γνώση αυτή δεν υπάρχει, είναι σύνηθες να ακολουθείται η παρακάτω τακτική για την επίλυσή του προβλήματος βέλτιστης επιλογής παραμέτρων *[MOO]*:

Αρχικά επιλέγεται ένας τύπος κελύφους (η επιλογή είναι μάλλον διαισθητική, από την εμπειρία, και λιγότερο ορθολογική). Για το κέλυφος αυτό, σαρώνεται το πεδίο ορισμού των ελεύθερων σταθερών του με κάποιο συγκεκριμένο βήμα, όπως επίσης σαρώνεται και το πεδίο ορισμού της σταθεράς C με κάποιο βήμα. Δημιουργούνται έτσι ένας ή περισσότεροι επαναληπτικοί βρόχοι, και κάθε επανάληψη αντιστοιχεί σε ένα συγκεκριμένο συνδυασμό παραμέτρων. Ο αλγόριθμος του *SVM* τρέχει για κάθε έναν από αυτούς τους συνδυασμούς, οπότε και παράγεται η αντίστοιχη λύση του προβλήματος βελτιστοποίησης, δηλαδή η μορφή της διαχωριστικής γραμμής. Για κάθε μια από τις λύσεις αυτές, δίνεται ένα νέο σύνολο σημείων (σύνολο δοκιμής) προς ταξινόμηση στην αντίστοιχη διαχωριστική γραμμή και καταγράφεται ο αριθμός των σφαλμάτων ταξινόμησης στα σημεία αυτά. Με αυτό τον τρόπο, για κάθε πιθανό συνδυασμό παραμέτρων, είναι διαθέσιμη και η «επίδοση» της αντίστοιχης μηχανής ταξινόμησης. Είναι προφανές τώρα ότι κριτήριο για την επιλογή των βέλτιστων τιμών των παραμέτρων αποτελεί το ελάχιστο από όλα τα σφάλματα ταξινόμησης που εμφανίστηκαν στην πορεία, δηλαδή η βέλτιστη των επιδόσεων.

Η ανωτέρω διαδικασία αναζήτησης του βέλτιστου συνόλου παραμέτρων ονομάζεται **διασταυρωτική επιλογή(cross validation)**, και εφαρμόζεται συχνά σε προβλήματα βέλτιστης επιλογής παραμέτρων. Παραπάνω εξετάσαμε την απλούστερη δυνατή μορφή της, όπου απλά σαρώνεται το πεδίο τιμών όλων των σταθερών μέχρις ότου βρεθεί ο βέλτιστος

συνδυασμός τους. Η τεχνική αυτή είναι γνωστή και ως **πλεγματική αναζήτηση (grid search)**, και οδηγεί με ασφάλεια στη βέλτιστη λύση, αλλά είναι αρκετά χρονοβόρα.

Μια εναλλακτική μορφή της διασταυρωτικής επιλογής, είναι και η ακόλουθη (που ονομάζεται **n-fold διασταυρωτική επιλογή**) *[MOO]*: Το αρχικό σύνολο εκπαίδευσης χωρίζεται σε n, ίσα κατά το δυνατόν, υποσύνολα. Το ένα εξ αυτών απομονώνεται από το αρχικό σύνολο εκπαίδευσης, και γίνεται πλεγματική αναζήτηση για το βέλτιστο καθορισμό των παραμέτρων, χρησιμοποιώντας ως σύνολο εκπαίδευσης τα εναπομείναντα n-1 υποσύνολα του αρχικού συνόλου. Σε κάθε βήμα της πλεγματικής αναζήτησης, αντί να χρησιμοποιείται ένα νέο σύνολο δοκιμής για μέτρηση της επίδοσης της μηχανής, χρησιμοποιείται εκείνο το υποσύνολο που αφαιρέθηκε από το αρχικό σύνολο εκπαίδευσης. Το ίδιο επαναλαμβάνεται και για τα υπόλοιπα υποσύνολα. Δηλαδή, κάθε φορά αφαιρείται ένα υποσύνολο από το αρχικό σύνολο εκπαίδευσης, γίνεται εκπαίδευση του *SVM* με τα υπόλοιπα n-1 υποσύνολα, και μέτρηση της επίδοσής του με το αφαιρεθέν. Το πλεονέκτημα της n-fold διασταυρωτικής επιλογής έναντι της απλής, είναι ότι δε χρειάζεται να έχουμε και επιπρόσθετο σύνολο δοκιμής, ενώ σε κάθε βήμα του αλγορίθμου η εκπαίδευση είναι λιγότερο χρονοβόρα, αφού χρησιμοποιούνται n-1 υποσύνολα , και όχι ολόκληρο το σύνολο εκπαίδευσης.

Ακραία μορφή της n-fold διασταυρωτικής επιλογής, αποτελεί η περίπτωση στην οποία αφαιρείται κάθε φορά από το σύνολο εκπαίδευσης <u>ένα και μόνο</u> σημείο, και γίνεται εκπαίδευση του *SVM* με τα εναπομείναντα, και δοκιμή του με το σημείο που αφαιρέθηκε. Η τεχνική αυτή ονομάζεται και **Leave One Out Cross Validation (LOOCV)** *[MOO]*, και κατ' ουσίαν είναι

μια n-fold διασταυρωτική επιλογή, όπου όμως n=N , με N τον αριθμό των σημείων του συνόλου εκπαίδευσης.

Η n-fold διασταυρωτική επιλογή πλεονεκτεί στο ότι κοστίζει λιγότερο σε υπολογιστικό χρόνο. Μειονεκτεί ωστόσο στο ότι δεν γίνεται εκπαίδευση της μηχανής από ολόκληρο το σύνολο εκπαίδευσης, δηλαδή «πετά» σε κάθε βήμα χρήσιμα δεδομένα. Ως αποτέλεσμα, αν το σύνολο εκπαίδευσης είναι μικρό, δεν μπορούμε να είμαστε σίγουροι για τη σωστή ή μη εκπαίδευση της μηχανής. Από την άλλη, η Leave One Out κοστίζει μεν αρκετά υπολογιστικά, αλλά δεν «πετά» δεδομένα (παρά μόνο ένα κάθε φορά), και συνεπώς παράγει πιο ασφαλή συμπεράσματα για τις βέλτιστες τιμές των παραμέτρων του *SVM*.

Το κοινό χαρακτηριστικό πάντως, τόσο της απλής διασταυρωτικής επιλογής, όσο και των δύο παραλλαγών της (n-fold,LOOCV), είναι ότι χρησιμοποιούν πλεγματική αναζήτηση (grid search) για την εύρεση των βέλτιστων τιμών των παραμέτρων. Η τεχνική αυτή, παρότι γενικά είναι ασφαλής και απλή, μπορεί να καταστεί αρκετά χρονοβόρα, καθώς απαιτείται εφαρμογή του αλγορίθμου *SVM* τόσες φορές όσοι και οι πιθανοί συνδυασμοί παραμέτρων. Ας θεωρήσουμε για παράδειγμα την περίπτωση της βέλτιστης επιλογής παραμέτρων σε ένα πολυωνυμικό κέλυφος:

$$K(x_i, x_j) = (< x_i, x_j > +1)^p$$

όπου ως ελεύθερες παραμέτρους έχουμε τον εκθέτη p (βαθμός του πολυωνύμου) και το άνω όριο των πολλαπλασιαστών Lagrange, C. Έστω ότι περιορίζουμε το πεδίο ορισμού της p στο εύρος [1,5] των ακεραίων

αριθμών, και το σαρώνουμε με βήμα 1. Δηλαδή θεωρούμε πολυώνυμα το πολύ 5ου βαθμού. Περιορίζουμε επίσης το πεδίο ορισμού της C από 10^{-3} έως 10^{20}, και το σαρώνουμε με βήμα έστω 10^2

$$C=10^{-3},10^{-1},10,10^{2},10^{4},10^{6},10^{8},...,10^{20} \quad (13 \text{ τιμές συνολικά})$$

Αυτό σημαίνει ότι δημιουργούνται 5x13=65 διαφορετικοί πιθανοί συνδυασμοί παραμέτρων. Δηλαδή ο αλγόριθμος πρέπει να τρέξει 65 φορές, κάτι που μπορεί να είναι απαγορευτικό υπολογιστικά, ειδικά για πολύ μεγάλα σύνολα εκπαίδευσης. Ειδικά δε στην περίπτωση της n-fold διασταυρωτικής επιλογής, η ανωτέρω διαδικασία θα γίνει $65 \cdot n$ φορές!

Για το σκοπό αυτό, και για να περιοριστεί ο υπολογιστικός χρόνος κατά τη διαδικασία της βέλτιστης επιλογής παραμέτρων, έχουν αναπτυχθεί διάφορες μέθοδοι που βασίζονται σε ευριστικές τεχνικές και στη θεωρία των πιθανοτήτων (*[SCHI],[BOUG],[CAW],[CHEMA])* με στόχο να γίνει πιο ντετερμινιστική και όχι τόσο διερευνητική η διαδικασία επιλογής των παραμέτρων του *SVM*. Ωστόσο, η εφαρμογή τους σε πραγματικά προβλήματα είναι ακόμα περιορισμένη, και επειδή ξεφεύγουν από τους σκοπούς του παρόντος, δεν θα αναφερθούμε περισσότερο σε αυτές.

2.6.2 Βέλτιστη επιλογή παραμέτρων – προτεινόμενη μέθοδος

Μια ενδιαφέρουσα πρόταση έκαναν οι *[HSU]*. Προτείνουν να χρησιμοποιείται αρχικά το κέλυφος ακτινικής βάσης (RBF):

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

Υπάρχουν πολλοί λόγοι που υπαγορεύουν αυτή την επιλογή. Πρώτον, επειδή επιτυγχάνει μη-γραμμική απεικόνιση των δεδομένων, μπορεί να χειριστεί τις περιπτώσεις όπου το σύνολο εκπαίδευσης δε διαχωρίζεται από ευθεία, κάτι που δεν συμβαίνει προφανώς με το γραμμικό κέλυφος. Έπειτα, το γραμμικό κέλυφος αποτελεί υποπερίπτωση του κελύφους RBF, καθώς όπως αποδεικνύουν οι *[KEE]*, είναι δυνατόν ένας *SVM* με παράμετρο C* που χρησιμοποιεί γραμμικό κέλυφος, να παράξει τα ίδια αποτελέσματα με τον *SVM* που χρησιμοποιεί κέλυφος RBF με παραμέτρους (C,σ), για κάποιες δεδομένες τιμές των C,σ, αλλά και με το σιγμοειδές κέλυφος. Τέλος, ένας ακόμα σημαντικός λόγος που προτιμάται το RBF κέλυφος, είναι ότι εμφανίζει λιγότερα αριθμητικά προβλήματα σε σχέση με τα υπόλοιπα μη-γραμμικά κελύφη. Οι τιμές της μήτρας κελύφους του RBF ικανοποιούν πάντα

$$0 \le K_{ij} \le 1$$

σε αντίθεση με το πολυωνυμικό κέλυφος, που μπορούν να φτάσουν και στο +∞, ενώ το σιγμοειδές κέλυφος μπορεί να καταστεί μη έγκυρη συνάρτηση κελύφους, για συγκεκριμένη επιλογή παραμέτρων.

Αφού καθοριστεί ως κέλυφος το RBF, εν συνεχεία οι *[HSU]* προτείνουν να εκτελείται μια πλεγματική αναζήτηση σε όλο το εύρος τιμών των παραμέτρων C και σ, με ένα αρχικά αραιό πλέγμα (μεγάλο βήμα τιμών των C και σ). Μόλις εντοπιστεί μια βέλτιστη περιοχή τιμών, εκεί πυκνώνει το πλέγμα (μειώνεται το βήμα), και επαναλαμβάνεται η πλεγματική αναζήτηση στην εν λόγω περιοχή. Η διαδικασία συνεχίζεται μέχρις ότου η μέθοδος συγκλίνει στο βέλτιστο σύνολο παραμέτρων (C_o, σ_o).

2.7 Κωδικοποίηση των δεδομένων στον SVM – κλιμάκωση των δεδομένων

Για να λειτουργήσει ο μαθηματικός αλγόριθμος του *SVM*, πρέπει τα σημεία του συνόλου εκπαίδευσης, αλλά και οποιοδήποτε άλλο νέο σημείο του δοθεί προς ταξινόμηση, να βρίσκονται υπό την μορφή ενός διανύσματος αριθμητικών παραμέτρων. Για παράδειγμα, το διάνυσμα παραμέτρων $\vec{x} = (0.1, 151.28, 0.004423)$ είναι έγκυρο, ενώ ένα διάνυσμα «περιγραφικών» παραμέτρων όπως το $\vec{x} = ("καλό", "ωραίο", "τετράγωνο")$ δε θα μπορέσει ο *SVM* προφανώς να το μεταχειριστεί.

Αυτό σημαίνει ότι στην περίπτωση που τα σημεία προς ταξινόμηση απαρτίζονται από ιδιότητες που δεν είναι αριθμητικά εκφρασμένες, πρέπει πρώτα να βρούμε μια κατάλληλη αριθμητική κωδικοποίηση για αυτές *[HSU]*, προτού εφαρμόσουμε τον *SVM*. Για παράδειγμα, έστω ότι μας δίνεται ένα σύνολο εκπαίδευσης που απαρτίζεται από σημεία με διαφορετική απόχρωση το καθένα, και μας ζητάνε να διαχωριστούν σε σημεία με θερμό χρώμα (π.χ. κόκκινο) και σε άλλα με ψυχρό (π.χ. μπλε). Τότε είναι προφανές ότι στην εκπαίδευση του *SVM* θα πρέπει να ληφθεί υπόψη η ιδιότητα «χρώμα» του κάθε σημείου, και να υπάρχει ως αριθμητική έκφραση στο διάνυσμα παραμέτρων του σημείου. Μια καλή κωδικοποίηση που μπορούμε να χρησιμοποιήσουμε στην περίπτωση αυτή, είναι η RGB. Δηλαδή σε κάθε χρώμα αντιστοιχίζονται τρεις παράμετροι, που κάθε μια δείχνει την αναλογία των τριών βασικών χρωμάτων $(κόκκινο, πράσινο, μπλε)$ στο εν λόγω χρώμα. Π.χ. , το κόκκινο είναι το $(1,0,0)$, το μπλε είναι το $(0,0,1)$ ενώ το μαύρο είναι το $(0,0,0)$, το άσπρο το $(1,1,1)$, και το γκρι είναι το $(0.5,0.5,0.5)$. Κατ' αυτό τον τρόπο, η ιδιότητα «χρώμα» κάθε σημείου εκφράστηκε από ένα διάνυσμα αριθμητικών

παραμέτρων, το οποίο ενσωματώνεται στο συνολικό διάνυσμα παραμέτρων του εν λόγω σημείου.

Γενικά, θεωρείται καλή πρακτική να χρησιμοποιούνται για την κωδικοποίηση τόσες αριθμητικές τιμές, όσες είναι και οι κατηγορίες που ορίζουν την εν λόγω ιδιότητα. Εν προκειμένω, είχαμε τρεις κατηγορίες (κόκκινο, πράσινο και μπλε) που ορίζουν ένα χρώμα, γι' αυτό και κάθε χρώμα αντιστοιχίστηκε με ένα διάνυσμα τριών παραμέτρων. Μια εναλλακτική επιλογή θα ήταν να αντιστοιχίζαμε ένα και μόνο νούμερο σε κάθε χρώμα, και να ενσωματώναμε αυτό το νούμερο ως ιδιότητα στο διάνυσμα παραμέτρων κάθε σημείου. Προτιμάται ωστόσο η πρώτη μέθοδος, γιατί έχει αποδειχτεί στην πράξη ότι είναι περισσότερο ευσταθής και δίνει καλύτερα αποτελέσματα.

Εκτός από την κωδικοποίηση των ιδιοτήτων των σημείων του συνόλου εκπαίδευσης, ένας άλλος παράγοντας που πρέπει να ληφθεί υπόψη, είναι η **κλιμάκωσή τους (scaling)** *[HSU]*. Εάν έχουμε σημεία, των οποίων οι ιδιότητες εμφανίζουν μεγάλες αριθμητικές διαφορές μεταξύ τους (π.χ. μια είναι $3 \cdot 10^4$ και μια άλλη 0.0005), τότε υπάρχει κίνδυνος οι ιδιότητες που κινούνται σε εύρος μεγαλύτερων τιμών, να υπερκαλύψουν αυτές που κινούνται σε μικρότερες τιμές. Αυτό συμβαίνει διότι δημιουργούνται αριθμητικές δυσκολίες κατά τον υπολογισμό των εσωτερικών γινομένων μεταξύ των σημείων στον υπολογιστή (σφάλματα στρογγυλοποίησης). Γι' αυτό, και όταν το σύνολο εκπαίδευσης εμπεριέχει ιδιότητες με μεγάλες αριθμητικές διαφορές, γίνεται στην αρχή μια γραμμική κλιμάκωση κάθε ιδιότητας στο εύρος [-1,+1], ή πιο σύνηθες στο εύρος [0,1], αλλά και οποιαδήποτε άλλη κλιμάκωση κρίνουμε εμείς κατάλληλη.

Βέβαια, η ίδια κλιμάκωση θα πρέπει να χρησιμοποιηθεί και στα νέα (προς ταξινόμηση) σημεία του συνόλου δοκιμής. Ας υποθέσουμε για παράδειγμα, ότι κάποια ιδιότητα των σημείων του συνόλου εκπαίδευσης κείται στο διάστημα [-10,10] και με κατάλληλη κλιμάκωση την φέρουμε στο διάστημα [-1,1]. Τότε, όταν στα νέα σημεία του συνόλου δοκιμής, η αντίστοιχη ιδιότητά τους κείται στο διάστημα π.χ. [-11,+8], θα πρέπει να γίνει ανάλογη κλιμάκωσή της, ώστε να έρθει στο [-1.1,+0.8].

2.8 *Ανακεφαλαίωση – Βήματα εφαρμογής του SVM*

Στα παρακάτω συνοψίζουμε τα βασικά βήματα που ακολουθούνται για την εφαρμογή του *SVM*. Έστω ότι δίνεται ένα σύνολο εκπαίδευσης $(\vec{x}_i \in \Re^N, y_i \in \{-1,+1\}), i = 1,...l$, και ζητείται να διαχωριστεί. Τότε:

- Αν υπάρχουν μη – αριθμητικές ιδιότητες στα σημεία του συνόλου εκπαίδευσης, βρίσκεται μια κατάλληλη αριθμητική κωδικοποίηση για αυτές. Έτσι, όλα τα σημεία του συνόλου εκπαίδευσης εκφράζονται ως διανύσματα αριθμητικών παραμέτρων.

- Γίνεται κλιμάκωση στις παραμέτρους των διανυσμάτων, δηλαδή στις ιδιότητες των σημείων, εάν και εφόσον κάτι τέτοιο πράγματι χρειάζεται.

- Επιλέγεται ένα κέλυφος το οποίο θεωρούμε ότι προσεγγίζει καλύτερα τη μορφή της διαχωριστικής γραμμής. Εάν δεν

προϋπάρχει καμία γνώση για τα σημεία του συνόλου εκπαίδευσης, και γενικά για το εν λόγω πρόβλημα ταξινόμησης, προτείνεται να επιλέγεται το κέλυφος RBF.

- Χρησιμοποιείται διασταυρωτική επιλογή (cross-validation) για την εύρεση των βέλτιστων τιμές των παραμέτρων C και σ(τυπική απόκλιση του κελύφους RBF). Εναλλακτικά, μπορούν απλά να δοκιμαστούν κάποιες τιμές του σ, κρατώντας το C πολύ μεγάλο (τείνον στο άπειρο, π.χ. $C = 10^{30}$), και αν δεν προκύψουν ικανοποιητικά αποτελέσματα στη μετέπειτα φάση της δοκιμής (σύνολο δοκιμής), να μειωθεί σταδιακά το C με κάποιο βήμα της επιλογής μας.
- Όποια διαδικασία κι αν ακολουθηθεί, επιλύεται τελικά το πρόβλημα τετραγωνικής βελτιστοποίησης ***(2-24)***, και ευρίσκονται οι πολλαπλασιαστές Lagrange που αντιστοιχούν στα $\vec{w}_o, b_o$ του βέλτιστου διαχωριστικού υπερεπιπέδου:

$$< \vec{w}_o, \vec{x} > - b_o = 0$$

- Εν συνεχεία, περνάμε στη φάση δοκιμής της μηχανής ταξινόμησης *SVM*. Δηλαδή της δίνεται ένα σύνολο νέων σημείων (σύνολο δοκιμής), στα οποία η μηχανή αποδίδει μια ετικέτα (+1 ή -1). Προτού δοθούν τα σημεία, γίνεται και πάλι κατάλληλη κωδικοποίηση και κλιμάκωση των ιδιοτήτων τους, όπου χρειάζεται, χρησιμοποιώντας την ίδια κλιμάκωση με τα σημεία του συνόλου εκπαίδευσης. Η συνάρτηση απόφασης του *SVM* που καθορίζει την κλάση στην οποία θα ταξινομηθεί ένα νέο σημείο $\vec{x}^*$ παρέχεται τότε από τη σχέση ***(2-30)*** :

$$f(\vec{x}^*) = sign(\sum_{i \in SVs} \lambda_i y_i K(\vec{x}_i, \vec{x}^*) - b_o)$$

ούτως ώστε, αν

$$f(\vec{x}^*) = 1$$

το σημείο $\vec{x}^*$ θα ταξινομηθεί στην κλάση +1, ενώ αν

$$f(\vec{x}^*) = -1$$

το σημείο $\vec{x}^*$ θα ταξινομηθεί στην κλάση -1.

2.9 Επέκταση της μεθόδου SVM σε προβλήματα παλινδρόμησης

Στην αρχή του κεφαλαίου, αναφέραμε ότι η μέθοδος *SVM* μπορεί να χρησιμοποιηθεί και για το γενικό **πρόβλημα παλινδρόμησης (regression problem)**, όπου δίνεται ένα σύνολο σημείων και ζητείται να βρεθεί η συνάρτηση παλινδρόμησης που τα προσεγγίζει βέλτιστα.

Εξετάζουμε καταρχήν την περίπτωση της γραμμικής παλινδρόμησης (*[VAP98],[GUN98],[SBS98]*). Η διατύπωση αυτού του προβλήματος έχει τότε ως εξής: Δίνεται ένα σύνολο σημείων $\{(\vec{x}_i, y_i)\}, i = 1,...,l$, και στόχος είναι να βρεθεί μια γραμμική συνάρτηση $f(\vec{x})$ που να προσεγγίζει κατά βέλτιστο τρόπο την συναρτησιακή εξάρτηση $y = y(\vec{x})$ που κρύβεται πίσω από τα $\vec{x}_i, y_i$. Η συνάρτηση αυτή θα έχει τη γενική μορφή :

$$f(\vec{x}, \vec{w}) = <\vec{w}, \vec{x}> + b \qquad \textbf{(2-31)}$$

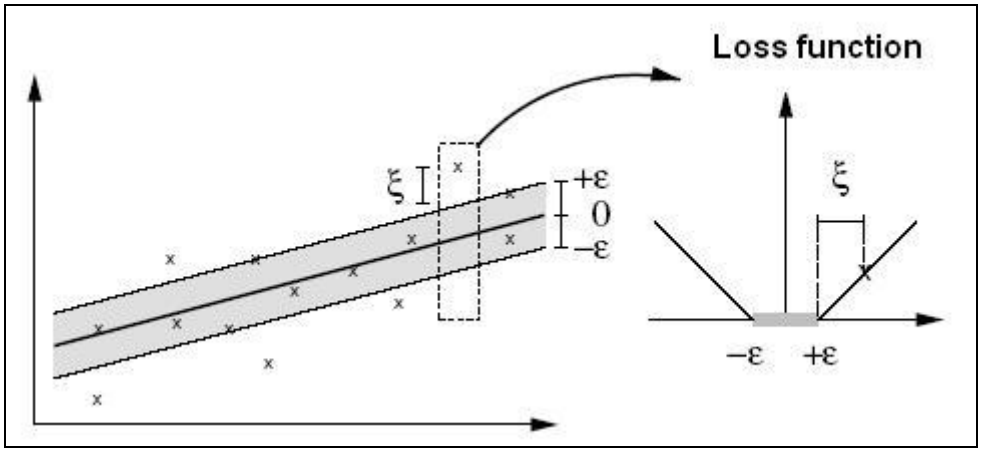
όπου η $f(\vec{x})$ θα παίρνει αυτή τη φορά τιμές στο σύνολο των πραγματικών αριθμών, και όχι μόνο +1 ή -1, όπως στην ταξινόμηση. Και πάλι ζητούμενο είναι να προσδιοριστούν τα βέλτιστα $\vec{w}, b$, δηλαδή να προσδιοριστεί η ευθεία που προσεγγίζει κατά βέλτιστο τρόπο τα δεδομένα. Όμοια με την ταξινόμηση, η ευθεία αυτή που θα βρεθεί από τον *SVM* είναι εκείνη που αναμένεται να έχει την καλύτερη γενίκευση σε μελλοντικά σημεία. Στο πρόβλημα της παλινδρόμησης η ικανότητα γενίκευσης μετριέται από την απόκλιση που θα έχει η τιμή $f(\vec{x}^*)$ που θα δώσει η παραχθείσα γραμμή παλινδρόμησης, από την πραγματική τιμή $y(\vec{x}^*)$ του σημείου $\vec{x}^*$.

Είναι προφανές ότι η έννοια του διακένου, το οποίο επιδιωκόταν να μεγιστοποιηθεί στο πρόβλημα της ταξινόμησης, δεν έχει νόημα εδώ. Η έννοια που λειτουργεί ως ανάλογο του διακένου στο πρόβλημα της παλινδρόμησης, είναι η λεγόμενη **συνάρτηση απώλειας (loss function)**. Στην εφαρμογή οιουδήποτε αλγορίθμου παλινδρόμησης, υπάρχει πάντα μια τέτοια συνάρτηση, η οποία εκφράζει την απόκλιση που έχει η εκτιμηθείσα τιμή από την πραγματική. Στον *SVM* χρησιμοποιείται συχνά η λεγόμενη **ε-αδιάφορη συνάρτηση απώλειας (ε-insensitive loss function)** του Vapnik *[VAP98]*, η οποία ορίζεται ως εξής :

$$|y(\vec{x}) - f(\vec{x})|_{\varepsilon} := \max\{0, |y(\vec{x}) - f(\vec{x}) - \varepsilon|\} \qquad \textbf{(2-32)}$$

Η γραφική της παράσταση φαίνεται στο σχήμα 17 (δεξί γράφημα). Με χρήση αυτής της συνάρτησης απώλειας, ο στόχος ενός προβλήματος παλινδρόμησης είναι να βρεθεί μια συνάρτηση $f(\vec{x})$ τέτοια που να έχει

απόκλιση το πολύ ε από τις πραγματικές τιμές $y(\vec{x}_i), i = 1,...l$ των σημείων του συνόλου εκπαίδευσης.



Σχήμα 17: Η ε-αδιάφορη συνάρτηση απώλειας του Vapnik, που χρησιμοποιείται στον SVM σε προβλήματα παλινδρόμησης

Στο σχήμα 17, φαίνονται στα αριστερά κάποια σημεία, στα οποία προσαρμόζεται μια γραμμική συνάρτηση παλινδρόμησης (ευθεία με έντονο μαύρο χρώμα). Για οποιαδήποτε σημείο που εμφανίζει απόκλιση από τη γραμμή παλινδρόμησης μεγαλύτερη του ε, η συνάρτηση απώλειας θα πάρει μη-μηδενική τιμή. Για όλα τα υπόλοιπα σημεία που περικλείονται μέσα στο ορθογώνιο με εύρος ε εκατέρωθεν της γραμμής παλινδρόμησης (ορίζεται από τις δύο παράλληλες διακεκομμένες ευθείες) , η συνάρτηση απώλειας είναι μηδενική. Η περίπτωση αυτή προσομοιάζει με εκείνη της ταξινόμησης για μη-απόλυτα διαχωρίσιμο σύνολο εκπαίδευσης: Δεν μας ενδιαφέρει να γίνουν κάποια σφάλματα στη προσέγγιση του συνόλου εκπαίδευσης, αρκεί αυτά να μην ξεπερνάνε ένα προκαθορισμένο όριο. Εν προκειμένω για την παλινδρόμηση, να μην υπερβαίνει η απόκλιση της πραγματικής από την εκτιμηθείσα τιμή το ε.

Η συνάρτηση απώλειας υπεισέρχεται τώρα ως όρος ποινής στην αντικειμενική συνάρτηση προς βελτιστοποίηση, η οποία στο πρόβλημα της παλινδρόμησης διαμορφώνεται ως εξής:

$$\min : \frac{1}{2} \| \vec{w} \|_2^{\,2} + C \cdot \sum_{i=1}^{l} | \, y(\vec{x}_i) - f(\vec{x}_i) \, |_{\varepsilon} \qquad \textbf{(2-33)}$$

Ισοδύναμη διατύπωση της ***(2-33)*** με χρήση των αποκλίσεων ξ_i, ξ_i^*, είναι η ακόλουθη:

$$\min : \frac{1}{2} \| \vec{w} \|_2^{\,2} + C \cdot \sum_{i=1}^{l} (\xi_i + \xi_i^*) \qquad \textbf{(2-34)}$$

$$\text{με περιορισμούς :} \quad \begin{aligned} & (< \vec{w}, \vec{x}_i > + b) - \vec{y}_i \le \varepsilon + \xi_i \\ & y_i - (< \vec{w}, \vec{x}_i > + b) \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0 \end{aligned} \qquad \textbf{(2-35)}$$

Παρατηρούμε ότι σύμφωνα με τους περιορισμούς ***(2-35)***, οποιαδήποτε απόκλιση της $y(\vec{x})$ από την $f(\vec{x})$ μικρότερη από ε, δεν απαιτεί μη-μηδενικές τιμές των ξ_i, ξ_i^* και συνεπώς δεν υπεισέρχεται στην αντικειμενική συνάρτηση.

Τα ανωτέρω ισχύουν για την περίπτωση της γραμμικής παλινδρόμησης. Η γενίκευση στη μη-γραμμική παλινδρόμηση γίνεται σε απόλυτη αντιστοιχία με το πρόβλημα της ταξινόμησης, χρησιμοποιούνται δηλαδή και εδώ συναρτήσεις κελύφους.

Εισάγοντας και εδώ μη αρνητικούς πολλαπλασιαστές Lagrange, καταλήγουμε στην ακόλουθη δυαδική μορφή του προβλήματος βελτιστοποίησης :

$$\max : L(\vec{\lambda}, \vec{\lambda}^*) = -\varepsilon \sum_{i=1}^{l} (\lambda_i + \lambda_i^*) + \sum_{i=1}^{l} (\lambda_i^* - \lambda_i) y_i - \\ -\frac{1}{2} \sum_{i,j=1}^{l} (\lambda_i^* - \lambda_i)(\lambda_j^* - \lambda_j) \cdot K(\vec{x}_i, \vec{x}_j) \qquad \textbf{(2-36)}$$

με περιορισμούς :

$$0 \le \lambda_i, \lambda_i^* \le C \\ \sum_{i=1}^{l} (\lambda_i - \lambda_i^*) = 0 \qquad \textbf{(2-37)}$$

Η βέλτιστη συνάρτηση παλινδρόμησης του *SVM* έχει τότε τη μορφή *[VAP98]*:

$$f(\vec{x}) = \sum_{i=1}^{l} (\lambda_{io}^* - \lambda_{io}) K(\vec{x}_i, \vec{x}) + b_o \qquad \textbf{(2-38)}$$

όπου οι πολλαπλασιαστές $\lambda_{io}, \lambda_{io}^*$ ευρίσκονται από την επίλυση του ανωτέρω προβλήματος βελτιστοποίησης. Το b_o υπολογίζεται από τις ***(2-35)*** θέτοντας αντίστοιχα $\xi_i = 0$ και $\xi_i^* = 0$.

Κεφάλαιο 3: Περιπτώσεις εφαρμογών

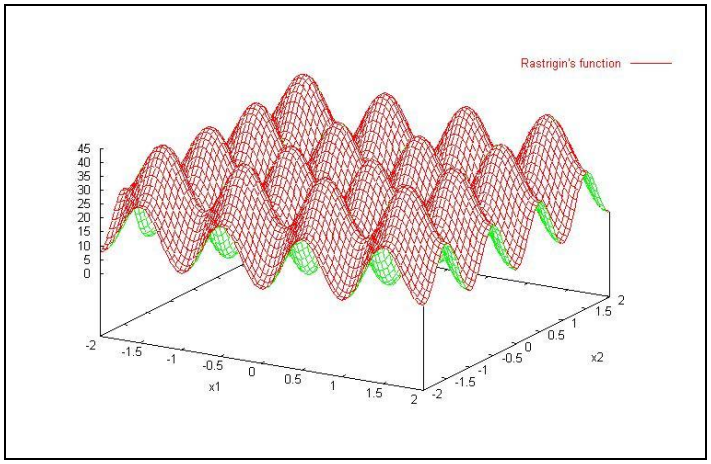
3.1 Εφαρμογή του SVM στη συνάρτηση του Rastrigin

Για μια πρώτη δοκιμή των δυνατοτήτων του *SVM* να επιτυγχάνει καλή ταξινόμηση, χρησιμοποιούμε την ακόλουθη συνάρτηση:

$$Ras(\vec{x}) = 10 \cdot N + \sum_{i=1}^{N} [x_i^{\,2} - 10 \cdot \cos(2 \cdot \pi \cdot x_i)]$$

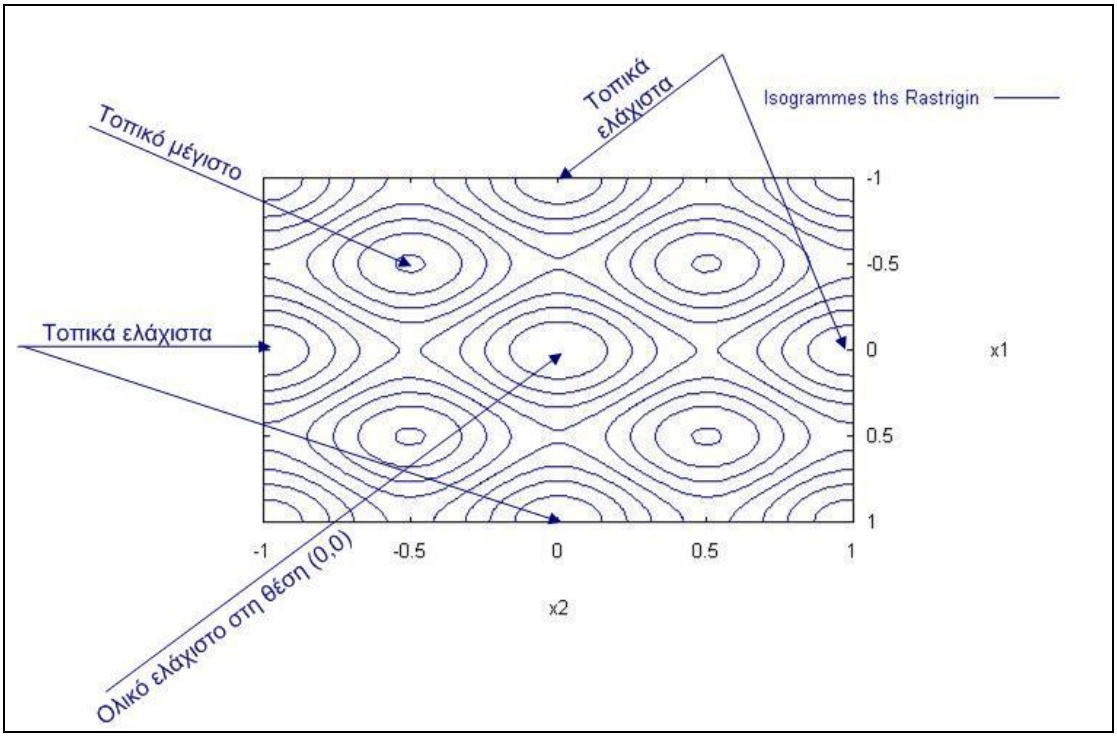
που ονομάζεται συνάρτηση του Rastigin. N είναι το πλήθος των μεταβλητών x_i, που καθορίζει τη διάσταση του χώρου στην οποία κείται η συνάρτηση. Για 2-Δ πρόβλημα ($N = 2$), η συνάρτηση γράφεται απλά (βλ. γραφική της παράσταση στο σχήμα 1):

$$Ras(x_1, x_2) = 20 + x_1^{\,2} + x_2^{\,2} - 10 \cdot (\cos(2\pi \cdot x_1) + \cos(2\pi \cdot x_2))$$



Σχήμα 1: Η συνάρτηση του Rastrigin σε δύο διαστάσεις. Είναι εμφανής η παρουσία πολλών τοπικών ακροτάτων. Η συνάρτηση εμφανίζει το ολικό της ελάχιστο στη θέση (0,0).

Όπως φαίνεται και από το σχήμα 1, η συνάρτηση εμφανίζει πλήθος τοπικών ακροτάτων. Προβάλλοντάς την στο επίπεδο x_1x_2 παίρνουμε τις ισογραμμές (contour lines) που φαίνονται στο σχήμα 2.

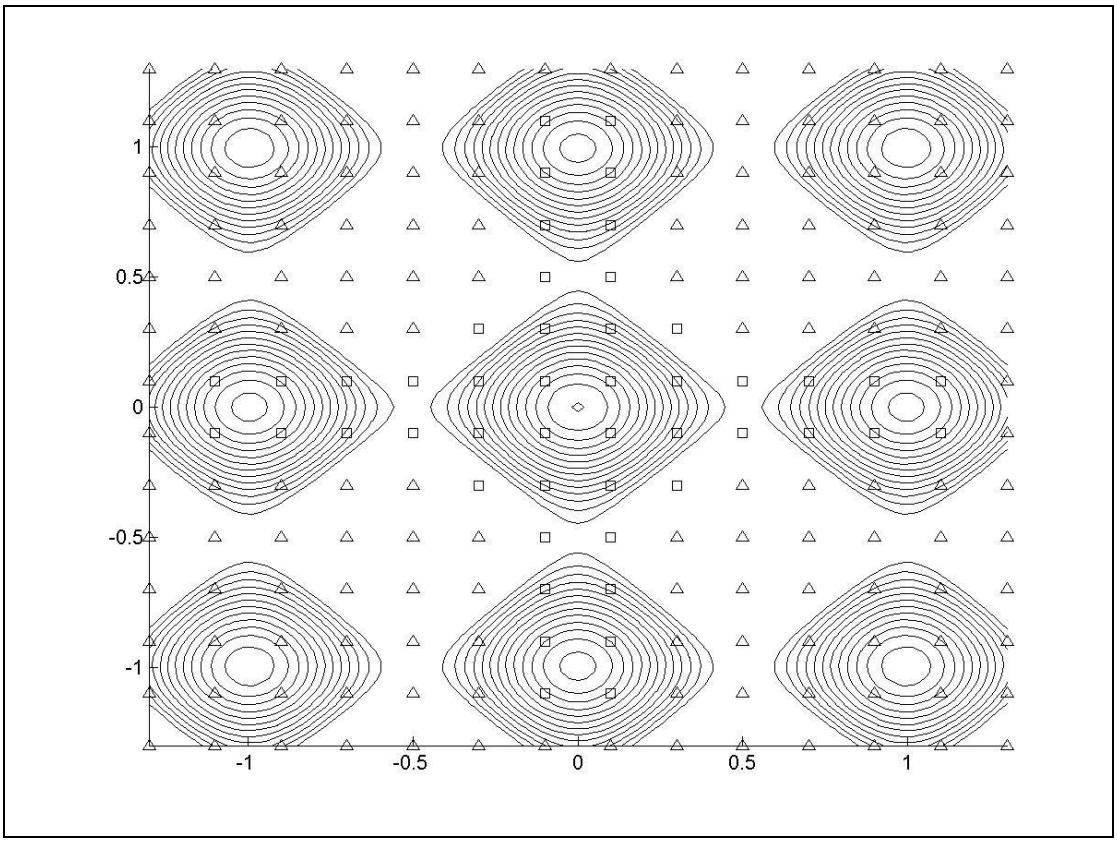


Σχήμα 2: Ισογραμμές της συνάρτησης του Rastrigin. Στο σχήμα δείχνονται οι θέσεις κάποιων τοπικών ακροτάτων, καθώς και η θέση του ενός ολικού ελαχίστου της συνάρτησης.

Το ολικό ελάχιστο εντοπίζεται στη θέση (0,0), ενώ σε συμμετρικές θέσεις γύρω από αυτό βρίσκονται τα τοπικά ελάχιστα και μέγιστα της συνάρτησης. Μάλιστα, όσο πιο μακριά βρίσκεται οποιοδήποτε άλλο τοπικό ελάχιστο από το σημείο (0,0), τόσο μεγαλύτερη τιμή έχει εκεί η συνάρτηση.

Εστιάζουμε σε μια περιοχή γύρω από το ολικό ελάχιστο και θεωρούμε σε αυτή την περιοχή ένα πλέγμα σημείων. Όπως φαίνεται και στο σχήμα 3, κάποια από αυτά τα σημεία (μαρκαρισμένα ως «τετράγωνα») βρίσκονται στην περιοχή ελαχίστων τιμών της συνάρτησης, ενώ τα υπόλοιπα (μαρκαρισμένα ως «τρίγωνα») βρίσκονται σε περιοχές μεγαλύτερων

τιμών. Χρησιμοποιούμε τώρα αυτό το σύνολο σημείων ως σύνολο εκπαίδευσης, για να κατασκευάσουμε το ακόλουθο πρόβλημα εκμάθησης για τον *SVM*: Διαχώρισε τα «καλά» από τα «κακά» σημεία.

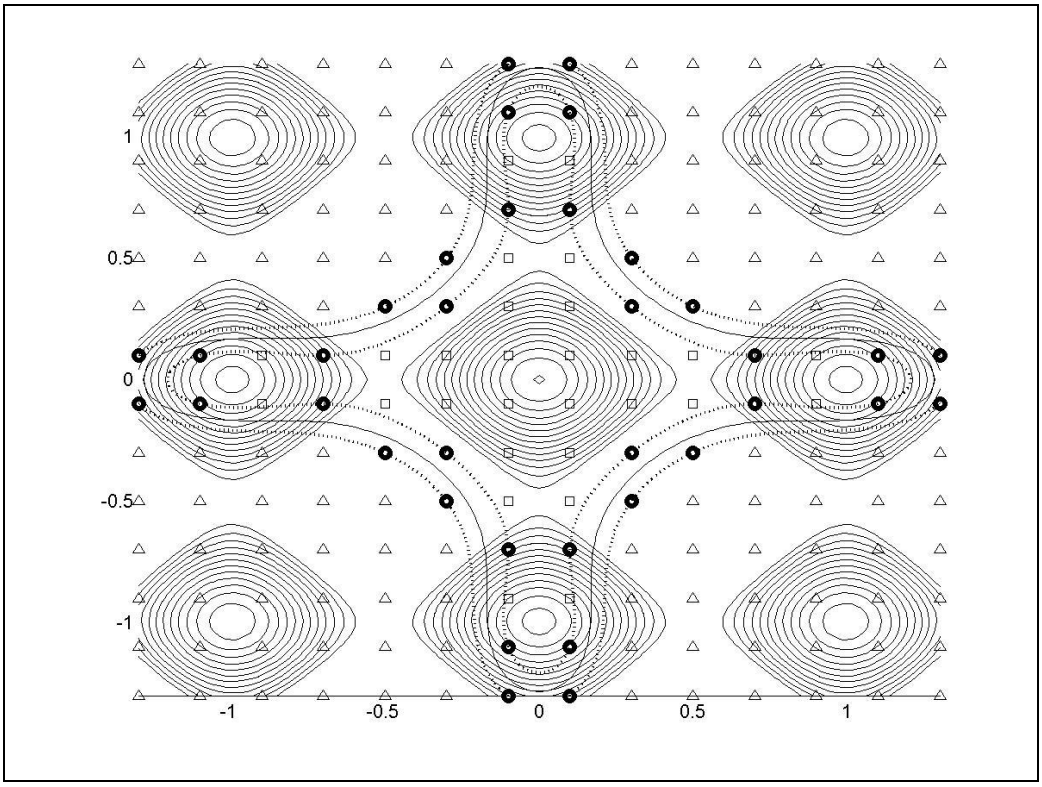


Σχήμα 3: Πλέγμα σημείων στη γειτονιά του σημείου ολικού ελαχίστου της συνάρτησης του Rastrigin. Τα σημεία αυτά με τις αντίστοιχες ετικέτες τους («τετράγωνα»-«τρίγωνα» στο σχήμα) αποτελούν το σύνολο εκπαίδευσης του *SVM*.

Για το σκοπό αυτό, γίνεται σε κάθε σημείο μια αξιολόγηση, δηλαδή υπολογίζεται η τιμή που έχει εκεί η συνάρτηση του Rastrigin, και με βάση την αξιολόγηση αυτή ιεραρχούνται σε σειρά από το χειρότερο προς το καλύτερο, όπου καλύτερο θεωρείται εκείνο όπου η συνάρτηση εμφανίζει μικρότερη τιμή. Επιλέγονται έτσι τα K σημεία με τις μικρότερες τιμές της συνάρτησης, και σε αυτά αποδίδεται ετικέτα +1 («καλά»). Στα υπόλοιπα σημεία αποδίδεται ετικέτα -1 («κακά»). Δημιουργούνται έτσι δύο κλάσεις: Αυτή των σημείων όπου η συνάρτηση εμφανίζει τις ελάχιστες τιμές της (κλάση +1) και αυτή των σημείων όπου η συνάρτηση παίρνει μεγαλύτερες

τιμές (κλάση -1). Ο *SVM* καλείται να χρησιμοποιήσει την πληροφορία αυτού του συνόλου εκπαίδευσης, προκειμένου να παράξει τη διαχωριστική γραμμή των δύο κλάσεων, που θα αποτελεί το κριτήριο για την ταξινόμηση νέων σημείων.

Για το διαχωρισμό του συνόλου επιλέχθηκε κέλυφος ακτινικής βάσης (RBF) με τυπική απόκλιση σ=0.9, ενώ το άνω όριο των πολλαπλασιαστών Lagrange επιλέχθηκε $C=\infty$. Το αποτέλεσμα της εκπαίδευσης του *SVM* φαίνεται στο σχήμα 4[α], όπου έχει χαραχθεί στο χώρο εισόδων η μη-γραμμική διαχωριστική ζώνη των δύο κλάσεων, αποτελούμενη από τη γραμμή διαχωρισμού (συνεχής γραμμή) και τις υποστηρίζουσες γραμμές των δύο κλάσεων (διακεκομμένες).



Σχήμα 4[α]: Εφαρμογή της μεθόδου *SVM* για την ταξινόμηση σημείων της συνάρτησης του Rastrigin. Στο σχήμα φαίνεται η διαχωριστική γραμμή, καθώς και οι υποστηρίζουσες γραμμές των δύο κλάσεων, ενώ με έντονο κύκλο έχουν μαρκαριστεί τα αντίστοιχα διανύσματα υποστήριξης.

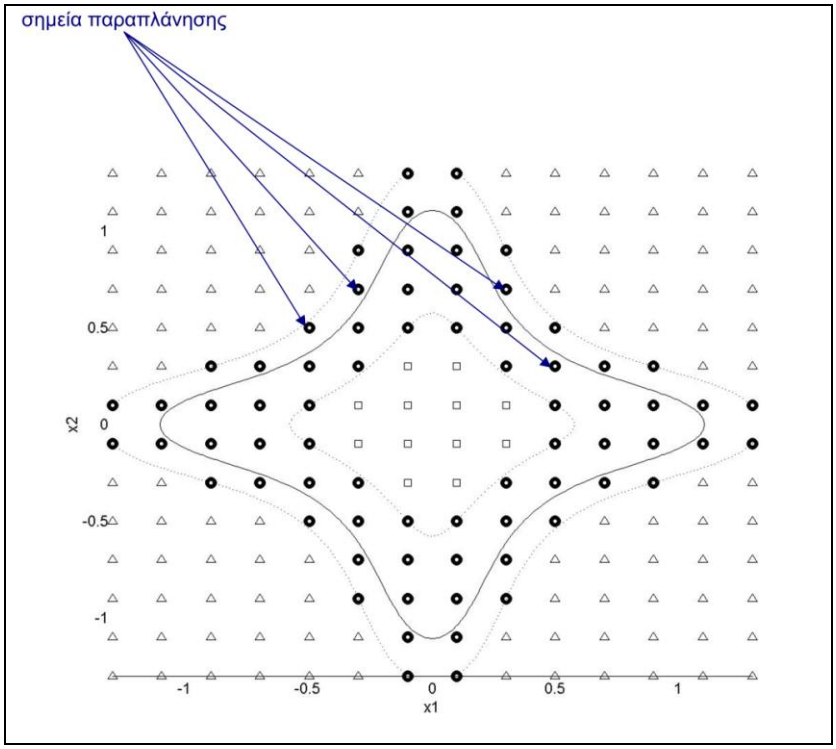
Παρατηρούμε ότι οι υποστηρίζουσες γραμμές διέρχονται από κάποια από τα σημεία του συνόλου εκπαίδευσης (κυκλωμένα σημεία με έντονη σήμανση στο σχ.4[α]). Αυτά αποτελούν τα ακραία σημεία κάθε κλάσης, δηλαδή βρίσκονται κοντινότερα στη διαχωριστική γραμμή από όλα τα υπόλοιπα σημεία της κλάσης τους. Είναι δηλαδή τα διανύσματα υποστήριξης των δύο κλάσεων, και είναι τα μόνα σημεία που καθορίζουν τη μορφή της διαχωριστικής ζώνης. Η παρουσία όλων των υπολοίπων σημείων δεν επηρεάζει σε τίποτα την παραγόμενη από τον *SVM* γραμμή διαχωρισμού.

Ο αλγόριθμος κατάφερε να απομονώσει την «καλή» περιοχή (περιοχή ελαχίστων) της συνάρτησης, που αποτελείται από τις 4 «κοιλότητες» συμμετρικά τοποθετημένες γύρω από την «κοιλότητα» ολικού ελαχίστου της αρχής (0,0). Ως αποτέλεσμα, ένα νέο σημείο που θα δοθεί, θα χαρακτηριστεί ως «καλό» αν εμπίπτει στην περιοχή ελαχίστων, ή ως «κακό» αν εμπίπτει στη γύρω περιοχή. Με άλλα λόγια, η διαχωριστική γραμμή που παρήγαγε ο *SVM*, ως αποτέλεσμα της εκπαίδευσης που του έγινε, είναι σε θέση να ταξινομήσει επιτυχώς ένα νέο σημείο σε μια από τις δύο προαναφερθείσες κλάσεις.

Διερεύνηση της τιμής του άνω ορίου των πολλαπλασιαστών Lagrange, C

Όπως παρατηρούμε από το σχήμα 4[α], η επιλογή μεγάλης τιμής για το άνω όριο C των πολλαπλασιαστών Lagrange, επέβαλε στον *SVM* να βρει διαχωριστική ζώνη που να απομονώνει πλήρως τα σημεία των δύο κλάσεων. Δεν επετράπη δηλαδή η ύπαρξη σημείων παραπλάνησης ή σημείων-σφάλματα στο σύνολο εκπαίδευσης. Αυτό είχε ως αποτέλεσμα το εύρος της διαχωριστικής ζώνης να είναι σχετικά μειωμένο, ώστε να μπορέσει να «περάσει» από όλα τα σημεία.

Δοκιμάζουμε τώρα να μειώσουμε το άνω όριο C, θεωρώντας την άλλη ακραία περίπτωση, δηλαδή αποδίδουμε στο C μια πολύ μικρή τιμή, π.χ. C=100. Επαναλαμβάνουμε την εκπαίδευση του SVM, οπότε παράγεται η διαχωριστική ζώνη του σχήματος 4β. Όπως παρατηρούμε η νέα διαχωριστική ζώνη έχει ευρύτερα όρια (ισοδύναμα ο διαχωριστής έχει μεγαλύτερο διάκενο), κι αυτό συνεπάγεται ότι κάποια σημεία του συνόλου εκπαίδευσης βρίσκονται στο εσωτερικό της. Ο διαχωριστής αυτός δηλαδή είναι πιο «ελαστικός» στην ύπαρξη σημείων παραπλάνησης σε σχέση με τον διαχωριστή $C=\infty$. Επιπρόσθετα παρατηρούμε ότι η διαχωριστική γραμμή είναι τώρα λιγότερο «αυστηρή» στην ταξινόμηση ενός σημείου στην κλάση των «καλών». Δηλαδή τα όρια της κλάσης +1 διευρύνθηκαν, με αποτέλεσμα κάποια νέα σημεία τα οποία ο διαχωριστής $C=\infty$ θα χαρακτήριζε ως «κακά» να χαρακτηρίζονται τώρα ως «καλά».



Σχήμα 4β: Διαχωρισμός των σημείων του ανωτέρω συνόλου εκπαίδευσης από ένα πιο ελαστικό διαχωριστή με C=100. Το διάκενο είναι μεγαλύτερο από την περίπτωση $C=\infty$, και η περιοχή της κλάσης +1 διευρύνθηκε. Στο σχήμα δείχνονται και κάποια σημεία παραπλάνησης.

Από την παραπάνω διερεύνηση προκύπτει το ερώτημα: Ποια από τις δύο είναι η πιο κατάλληλη τιμή του άνω ορίου C; Η απάντηση σε αυτό το ερώτημα εξαρτάται από το πόσο αυστηροί θα είμαστε στο χαρακτηρισμό ενός σημείου ως «καλό». Αν θέλουμε τα σημεία που θα χαρακτηριστούν ως «καλά» να εμπίπτουν αυστηρά στην περιοχή των κοιλοτήτων ελαχίστου της Rastrigin, τότε ο διαχωριστής με $C=\infty$ θα επιτύχει καλύτερη γενίκευση στο αντίστοιχο πρόβλημα ταξινόμησης, αφού εφαρμόζει μια πιο «αυστηρή» ζώνη διαχωρισμού με στενά όρια στις δύο κλάσεις. Από την άλλη, αν δεν μας ενδιαφέρει η αυστηρή εγγύτητα των «καλών» σημείων στις περιοχές ελαχίστου, τότε ένα σημείο που βρίσκεται κοντά και όχι απαραίτητα μέσα σε μια κοιλότητα ελαχίστου θα θέλουμε να θεωρείται κι αυτό ως «καλό». Στην περίπτωση αυτή, θα πρέπει να προτιμηθεί ο διαχωριστής $C=100$ καθώς η διαχωριστική του ζώνη είναι πιο ελαστική και συνεπώς επιτρέπει σε σημεία που βρίσκονται στην ευρύτερη «γειτονιά» των κοιλοτήτων ελαχίστου να χαρακτηριστούν κι αυτά ως «καλά», χωρίς απαραίτητα αυτά τα σημεία να εμπίπτουν στο εσωτερικό μιας κοιλότητας. Με άλλα λόγια, η επιλογή μικρού ή μεγάλου C έχει να κάνει με την ακρίβεια με την οποία έχει οριστεί το πρόβλημα ταξινόμησης, ή αλλιώς με την ακρίβεια με την οποία προϋπάρχει πληροφόρηση για το ποια είναι τα «καλά» και ποια τα «κακά» σημεία. Η πληροφόρηση αυτή έχει να κάνει με το μέγεθος του συνόλου εκπαίδευσης που χρησιμοποιεί ο *SVM*.

Πράγματι, όσο περισσότερα σημεία έχει το σύνολο εκπαίδευσης, τόση περισσότερη πληροφορία παρέχεται στον αλγόριθμο για το πρόβλημα ταξινόμησης που καλείται να επιλύσει. Αυτό έρχεται σε συμφωνία με τα όσα είπαμε στο κεφ.2 περί συσχέτισης του μεγέθους του συνόλου εκπαίδευσης με τη χωρητικότητα της οικογένειας συναρτήσεων που χρησιμοποιείται. Όταν έχουμε μικρό σύνολο εκπαίδευσης και

χρησιμοποιηθεί οικογένεια με μεγάλη χωρητικότητα (μικρό διάκενο → μεγάλο C) τότε η γενίκευση θα είναι χειρότερη απ' ότι όταν χρησιμοποιηθεί οικογένεια με μεγάλο διάκενο→ μικρό C. Η κατάσταση αντιστρέφεται όταν το σύνολο εκπαίδευσης περιέχει επαρκή πληροφορία για την κατάταξη των «καλών» σημείων. Στην περίπτωση αυτή, ο διαχωριστής με μεγάλο C επιτυγχάνει πιο ακριβή ταξινόμηση των νέων σημείων, δηλαδή καλύτερη γενίκευση.

Το τελευταίο έρχεται σε συμφωνία με τα όσα είχαμε πει στο κεφ.2 περί ύπαρξης σημείων θορύβου: Όταν δεν διαθέτουμε επαρκή πληροφορία για το φυσικό πρόβλημα που θέλουμε να ταξινομήσει ο *SVM*, τότε έχουμε πιο ελαστικά κριτήρια επιλογής της κλάσης ενός σημείου, κι ως εκ τούτου επιτρέπουμε στο διαχωριστή να θεωρεί κάποια σημεία (που είχαμε χαρακτηρίσει στο κεφ.2 *σημεία θορύβου*) ως σημεία παραπλάνησης. Αντίθετα, όταν είμαστε απολύτως σίγουροι για τα όρια διαχωρισμού των δύο κλάσεων (δηλαδή έχουμε επαρκή πληροφορία για το ποια σημεία είναι «καλά»), τότε επιβάλλουμε στον *SVM* να χρησιμοποιήσει ένα διαχωριστή με μεγάλο άνω όριο C, επιδιώκοντας αυστηρά το διαχωρισμό όλων των σημείων από την παραγόμενη ζώνη.

Στην επόμενη παράγραφο, όπου θα γίνει εφαρμογή της μεθόδου *SVM* σαν εργαλείο προεπιλογής υποψήφιων λύσεων στους Εξελικτικούς Αλγορίθμους, θα φανεί άμεσα η επίδραση της παραμέτρου C στην αποτελεσματικότητα του *SVM*, και συνεπώς στην ταχύτητα σύγκλισης του ΕΑ που τον χρησιμοποιεί.

3.2 Συνδυασμός του SVM με Εξελικτικούς Αλγορίθμους για προβλήματα βελτιστοποίησης

Όπως αναφέρθηκε και στο Κεφ.1, η μέθοδος *SVM* χρησιμοποιήθηκε στην παρούσα εργασία ως εργαλείο προεπιλογής υποψήφιων λύσεων στα πλαίσια της βελτιστοποίησης μέσω Εξελικτικών Αλγορίθμων. Η μέθοδος ενσωματώθηκε σαν επιλογή στο λογισμικό βελτιστοποίησης του Εργαστηρίου Θερμικών Στροβιλομηχανών ΕΜΠ, το οποίο βασίζεται σε ΕΑ. Η ιδέα που εδώ δοκιμάζεται είναι το SVM να χρησιμοποιείται ως εργαλείο προσεγγιστικής αξιολόγησης (μεταπρότυπο) για εξοικονόμηση υπολογιστικού χρόνου όταν το λογισμικό ακριβούς αξιολόγησης είναι ιδιαίτερα ακριβό σε χρόνο CPU. Στα επόμενα εξετάζονται με μεγαλύτερη λεπτομέρεια τα βήματα εφαρμογής της μεθόδου *SVM* στον ΕΑ βελτιστοποίησης που αναφέρθηκαν και στο Κεφ.1:

Βήμα 1: Ξεκινάει ο ΕΑ, οπότε και δημιουργείται ο πληθυσμός της αρχικής (μηδενικής) γενιάς, από τυχαία επιλεγμένα άτομα.

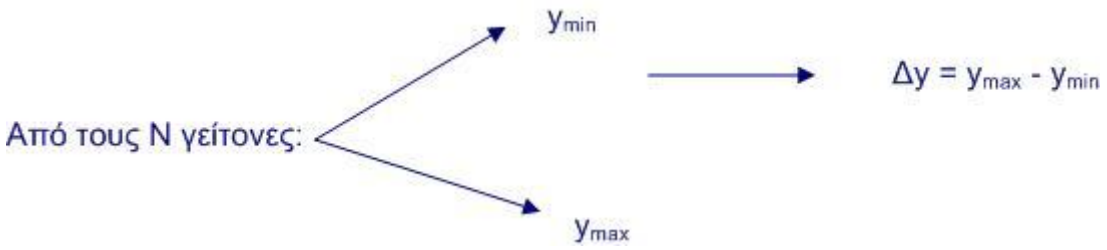
Βήμα 2: Ο αρχικός πληθυσμός υπόκειται σε εξέλιξη για ένα αριθμό ***M*** γενεών, κατά την οποία τα μέλη κάθε γενιάς αξιολογούνται από το ακριβές λογισμικό αξιολόγησης και οι επιδόσεις τους καταγράφονται σε μια βάση δεδομένων. Συνήθως ***M***=1÷4, αν και αυτή η επιλογή εξαρτάται από το πρόβλημα αλλά και το μέγεθος του πληθυσμού που διαχειρίζεται ο ΕΑ.

Βήμα 3: Κατά τις επόμενες γενιές, και για κάθε μέλος της τρέχουσας γενιάς, ανασύρονται από τη βάση δεδομένων οι ***N*** γείτονες του μέλους αυτού, όπου τους αποδίδεται μια ετικέτα «καλός»-«κακός», ή αλλιώς «+1»,

«-1». Η απόδοση ετικετών στους γείτονες γίνεται αφού πρώτα αυτοί βαθμονομηθούν ως προς τις επιδόσεις τους (τιμή της αντικειμενικής συνάρτησης), οπότε και επιλέγονται οι ***K*** «καλύτεροι» (αυτοί που έχουν καλύτερη τιμή αντικειμενικής συνάρτησης), και αυτοί χαρακτηρίζονται ως «καλοί». Οι υπόλοιποι χαρακτηρίζονται ως «κακοί». (Σημ: θεωρούμε πρόβλημα ελαχιστοποίησης, οπότε οι καλύτεροι γείτονες θα είναι αυτοί που εμφανίζουν μικρότερη τιμή αντικειμενικής συνάρτησης.)

<u>*Βήμα 4*</u>: Οι γείτονες με τις αντίστοιχες ετικέτες τους, εκπαιδεύουν ένα τοπικό *SVM*, ο οποίος μαθαίνει την αντιστοιχία εισόδων-εξόδων, όπου είσοδος είναι οι τιμές των μεταβλητών σχεδίασης ή ελεύθερων παραμέτρων του υπό εξέταση μέλους μιας νέας γενιάς και η έξοδος είναι: +1 αν το μέλος θεωρηθεί «καλό» και -1 αν το μέλος θεωρηθεί «κακό».

<u>*Βήμα 5*</u>: Ο εκπαιδευμένος *SVM* χρησιμοποιείται για να μαντέψει την έξοδο του υπό εξέταση μέλους της τρέχουσας γενιάς. Αποκτάται έτσι μια τιμή : +1 ή -1 για το μέλος, ανάλογα με το αν αυτό κατατάσσεται στα «καλά» ή στα «κακά». Η τιμή αυτή (±1) μετασχηματίζεται σε μια προσεγγιστική τιμή κόστους με έναν απλό τρόπο: ευρίσκονται από τη βάση δεδομένων οι τιμές της αντικειμενικής συνάρτησης που εμφάνισαν οι γείτονές του υπό εξέταση μέλους, και καταγράφονται η μέγιστη και η ελάχιστη από αυτές. Σχηματικά:



Αν η εκτίμηση του *SVM* για το υπό εξέταση μέλος i είναι +1, τότε στο μέλος αποδίδεται και εκτίμηση της αντικειμενικής συνάρτησης:

$$y_i = y_{\min} - 0.03 \cdot \Delta y$$

ενώ αν η εκτίμηση του *SVM* είναι -1, τότε στο μέλος αποδίδεται εκτίμηση της αντικειμενικής συνάρτησης:

$$y_i = y_{\max} + 0.03 \cdot \Delta y \ (^3)$$

Κατ' αυτό τον τρόπο, σε κάθε μέλος της νέας γενιάς, έχει αποδοθεί εκτός από ένας χαρακτηρισμός («καλό»-«κακό»), και μια **προσεγγιστική** τιμή κόστους (fitness value).

Βήμα 6: Η ανωτέρω διαδικασία επαναλαμβάνεται για όλα τα μέλη της τρέχουσας γενιάς. Αποκτάται έτσι μια προσεγγιστική τιμή κόστους για κάθε ένα από αυτά.

Βήμα 7: Με βάση τα παραπάνω, από τα μέλη που χαρακτηρίστηκαν ως «καλά» επιλέγονται τα μέλη εκείνα με τις Λ καλύτερες προσεγγιστικές τιμές κόστους (Λ μικρότερες για πρόβλημα ελαχιστοποίησης, Λ μεγαλύτερες για πρόβλημα μεγιστοποίησης). Αυτά τα μέλη στέλνονται για αξιολόγηση στο λογισμικό ακριβούς αξιολόγησης, ενώ τα υπόλοιπα μένουν ως έχουν.

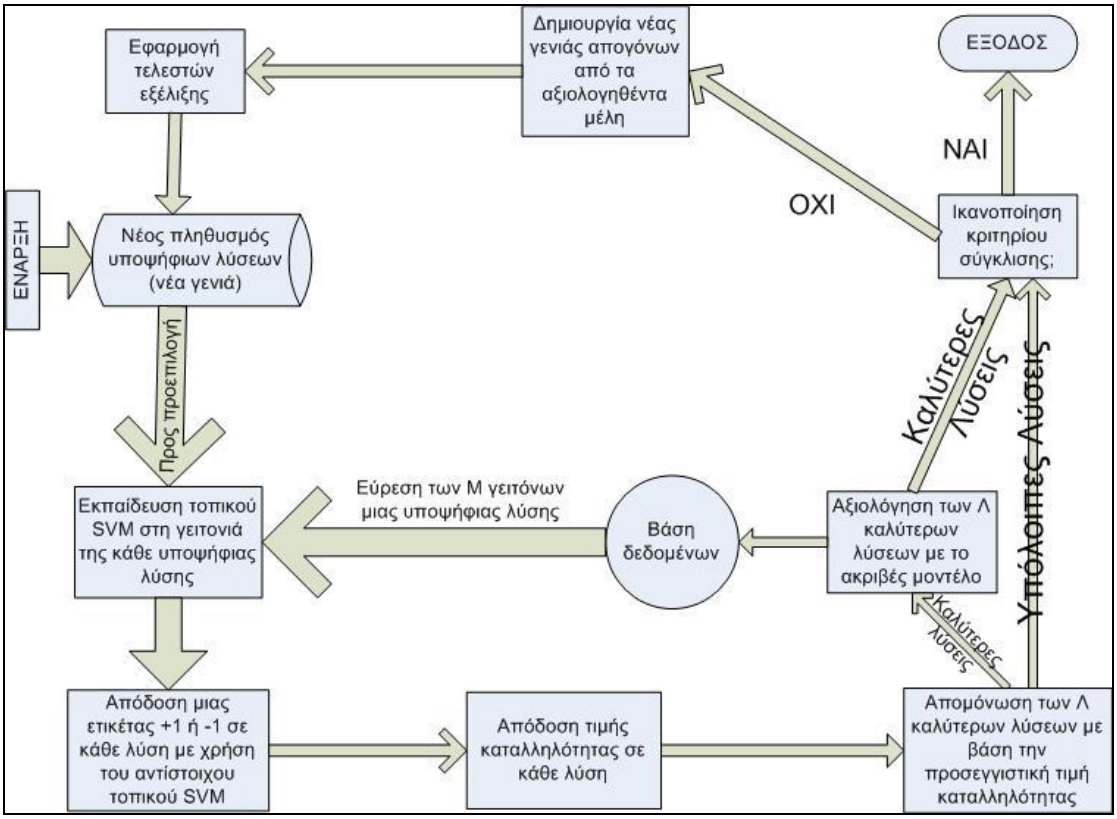
Βήμα 8: Στη συνέχεια ακολουθούνται τα κλασικά βήματα ενός ΕΑ (δημιουργία νέας γενιάς γονέων, εφαρμογή τελεστών εξέλιξης για τη

---

[3] Η ποσότητα $0.03 \cdot \Delta y$ χρησιμοποιείται για να πριμοδοτήσει τα εκτιμώμενα καλά μέλη της γενιάς. Έτσι, στο μαρκαρισμένο ως «καλό» μέλος αποδίδεται μια τιμή λίγο καλύτερη από αυτή του καλύτερου γείτονά του. Αντίστοιχα, στο «κακό» μέλος αποδίδεται μια τιμή λίγο χειρότερη από αυτή του χειρότερου γείτονά του. Με τον τρόπο αυτό καθοδηγούμε τον ΕΑ να κινηθεί πιο επιθετικά στην αναζήτηση καλύτερων λύσεων στην τρέχουσα γενιά.

δημιουργία απογόνων κτλ.) και η ανωτέρω διαδικασία επαναλαμβάνεται για την επόμενη γενιά κ.ο.κ.

Σχηματικά, η εμπλοκή της μεθόδου *SVM* ως μεταπροτύπου προεπιλογής υποψήφιων λύσεων σε ένα ΕΑ, φαίνεται στο σχήμα που ακολουθεί:



Σχήμα 5: Ενσωμάτωση του SVM στον ΕΑ βελτιστοποίησης, σαν εργαλείο προεπιλογής υποψήφιων λύσεων.

Είναι προφανές ότι οι παράμετροι ***Κ,Λ,Μ,Ν*** που αναφέρθηκαν παραπάνω επηρεάζουν την αποτελεσματικότητα του αλγορίθμου. Μια σωστή επιλογή παραμέτρων μπορεί να οδηγήσει τον ΕΑ σε πολύ αποτελεσματικότερη αξιοποίηση του *SVM* σαν εργαλείο προεπιλογής λύσεων και συνεπώς σε πολύ ταχύτερη σύγκλιση του ΕΑ στη βέλτιστη λύση. Ωστόσο, οι βέλτιστες τιμές των παραμέτρων αυτών δεν είναι δυνατόν να καθοριστούν εκ των προτέρων, καθώς διαφέρουν γενικά από

πρόβλημα σε πρόβλημα, και γι' αυτό απαιτούνται δοκιμές με διάφορους συνδυασμούς τους έως ότου βρεθούν αυτές που εξασφαλίζουν ικανοποιητική επίδοση του μεταπροτύπου *SVM*.

Παρακάτω ακολουθούν δύο περιπτώσεις προβλημάτων ελαχιστοποίησης που επιλύθηκαν με τη χρήση μιας απλοποιημένης έκδοσης του λογισμικού εξελικτικής βελτιστοποίησης του Εργαστηρίου Θερμικών Στροβιλομηχανών ΕΜΠ([4]). Η πρώτη αφορά σε ένα μαθηματικό πρόβλημα, και συγκεκριμένα στην εύρεση του ελαχίστου της συνάρτησης του Rastrigin. Η δεύτερη έχει να κάνει με τη βέλτιστη παραμετροποίηση μιας αεροτομής κατά Bezier.

Σε κάθε μια από τις περιπτώσεις αυτές έγινε αρχικά ένα τρέξιμο του Εξελικτικού Αλγορίθμου χωρίς τη συνεισφορά του *SVM* και καταγράφηκε η πορεία σύγκλισης αυτού στη βέλτιστη λύση. Στη συνέχεια ακολούθησε ένα τρέξιμο του ΕΑ με τη συνεισφορά του *SVM* και συγκρίθηκε η νέα πορεία σύγκλισης με την παλιά. Στα αποτελέσματα που θα παρουσιαστούν παρακάτω είναι εμφανής η διαφορά στους ρυθμούς σύγκλισης, που μεταφράζεται σε σημαντική μείωση υπολογιστικού χρόνου.

3.2.1 Ελαχιστοποίηση της συνάρτησης του Rastrigin

Η συνάρτηση του Rastrigin (που περιγράφηκε στην παράγραφο 3.1) έχει χρησιμοποιηθεί πολλές φορές στο παρελθόν για τη δοκιμή της αποτελεσματικότητας μεθόδων βελτιστοποίησης. Κι αυτό λόγω της

[4] Σημειώνεται, για λόγους πληρότητας, ότι ο πληθυσμός κάθε γενιάς που χρησιμοποιήθηκε στον ΕΑ αποτελείτο από μ=30 γονείς και λ=70 απογόνους.

ιδιάζουσας μορφής της, που εμφανίζει πλήθος τοπικών ακροτάτων αλλά ένα μόνο ολικό ελάχιστο, και συνεπώς αποτελεί πρόκληση για κάθε μέθοδο βελτιστοποίησης η αποφυγή του εγκλωβισμού σε ένα από τα τοπικά ελάχιστα και ο εντοπισμός με επιτυχία του ολικού ελαχίστου.

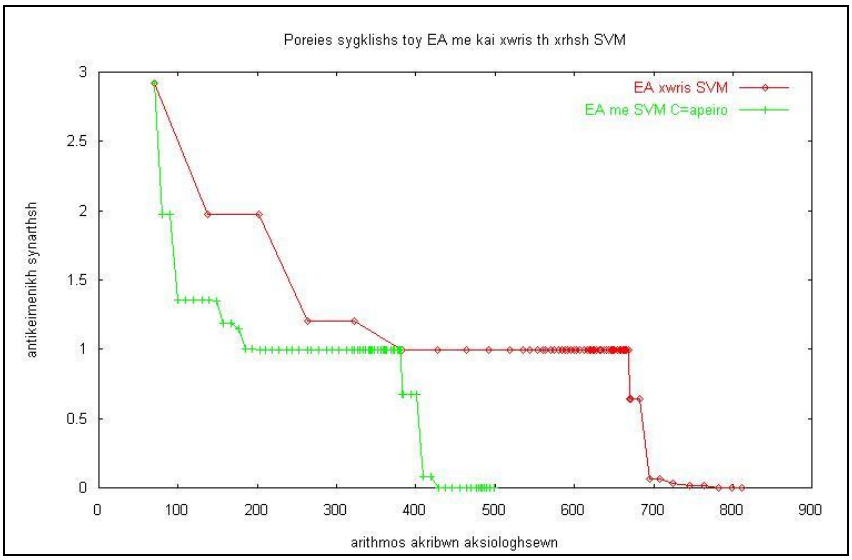
Η συνάρτηση δόθηκε στον ΕΑ ως αντικειμενική συνάρτηση προς ελαχιστοποίηση. Έγιναν δύο τρεξίματα του ΕΑ, ένα χωρίς χρήση της μεθόδου *SVM* και ένα με χρήση αυτής ως εργαλείου προεπιλογής υποψήφιων λύσεων. Σε κάθε εφαρμογή του *SVM* στον ΕΑ, έγινε διερεύνηση των παραμέτρων ***M,N,K,Λ*** και φάνηκε πώς αυτές επηρεάζουν το ρυθμό σύγκλισης του ΕΑ που χρησιμοποιεί τον *SVM*. Στον πίνακα που ακολουθεί υπενθυμίζεται η σημασία των συμβόλων που χρησιμοποιούμε παρακάτω.

M	Αριθμός γενεών που υφίστανται αποκλειστικά ακριβή αξιολόγηση πριν την εφαρμογή του *SVM*
N	Αριθμός γειτόνων ενός μέλους – σύνολο εκπαίδευσης- που χρησιμοποιούνται για την εκπαίδευση του τοπικού *SVM* στη γειτονιά του μέλους
K	Αριθμός μελών του συνόλου εκπαίδευσης που μαρκάρονται ως «+1»
Λ	Αριθμός μελών κάθε γενιάς που στέλνονται για ακριβή αξιολόγηση
C	Άνω όριο των πολλαπλασιαστών Lagrange στο πρόβλημα μεγιστοποίησης του *SVM*, το οποίο καθορίζει την ελαστικότητα της διαχωριστικής ζώνης στην ύπαρξη σημείων παραπλάνησης

Περίπτωση (Α): ***M=1,N=20,K=5,Λ=10***

Επιλέγονται κάποιες τιμές για τις παραμέτρους ***K,Λ,M,N*** και για αυτές τις παραμέτρους δοκιμάζεται η αποτελεσματικότητα του *SVM* για διάφορες τιμές του άνω ορίου των πολλαπλασιαστών Lagrange, C.

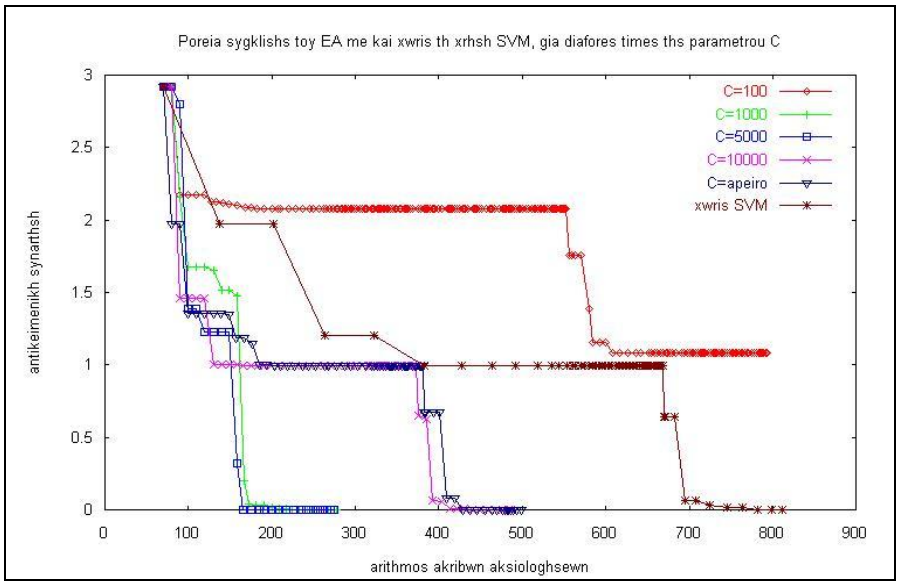
Οι τιμές της παραμέτρου C που δοκιμάστηκαν είναι οι εξής: $C=100, C=1000, C=5000, C=10000, C=\infty$. Στο σχήμα 6 φαίνεται η πορεία σύγκλισης του «απλού» ΕΑ, που δεν χρησιμοποιεί τον *SVM* σαν μεταπρότυπο, έναντι του ΕΑ που χρησιμοποιεί τον *SVM* με τιμή $C=\infty$. Είναι εμφανής η μείωση του αριθμού των ακριβών αξιολογήσεων για να οδηγηθεί ο ΕΑ σε σύγκλιση, όταν αυτός χρησιμοποιεί σαν μεταπρότυπο τον *SVM*.



Σχήμα 6: Σύγκριση της πορείας σύγκλισης του ΕΑ που δεν χρησιμοποιεί *SVM*, έναντι αυτού που χρησιμοποιεί *SVM* (για $C=\infty$).

Στο σχήμα 7 φαίνονται οι πορείες σύγκλισης του ΕΑ που χρησιμοποιεί τον *SVM*, για διάφορες τιμές της παραμέτρου C. Όπως παρατηρούμε, το βέλτιστο ρυθμό σύγκλισης επιτυγχάνει ο *SVM* με C=5000. Περαιτέρω μείωση του C προκαλεί χειρότερη σύγκλιση, ενώ το ίδιο συμβαίνει και με

αύξηση του C πάνω από την τιμή 5000. Σημειώνεται ότι τα παραπάνω ισχύουν για τις συγκεκριμένες τιμές των ***M,N,K,Λ*** που επιλέξαμε.



Σχήμα 7: Πορείες σύγκλισης ΕΑ που χρησιμοποιούν SVM, για διάφορες τιμές του άνω ορίου C των πολλαπλασιαστών Lagrange.

Στον πίνακα 2 φαίνονται, ενδεικτικά για μια γενιά, τα μέλη της γενιάς που υπέστησαν ακριβή αξιολόγηση από τον ΕΑ (οι αύξοντες αριθμοί τους) καθώς και οι αντίστοιχες τιμές κόστους τους (fitness values). Παρατηρούμε ότι στο σύνολό τους, τα μέλη που επέλεξε ο διαχωριστής C=5000 για ακριβή αξιολόγηση εμφανίζουν καλύτερες τιμές από ότι τα μέλη που επέλεξαν οι υπόλοιποι διαχωριστές. Αυτό συνεπάγεται ότι στάλθηκαν στον ΕΑ για ακριβή αξιολόγηση καλύτερες λύσεις, και συνεπώς ο ΕΑ συνέκλινε πιο γρήγορα στη βέλτιστη λύση. Έτσι δικαιολογείται η υπεροχή στο ρυθμό σύγκλισης του διαχωριστή C=5000 σε σχέση με τους άλλους διαχωριστές.

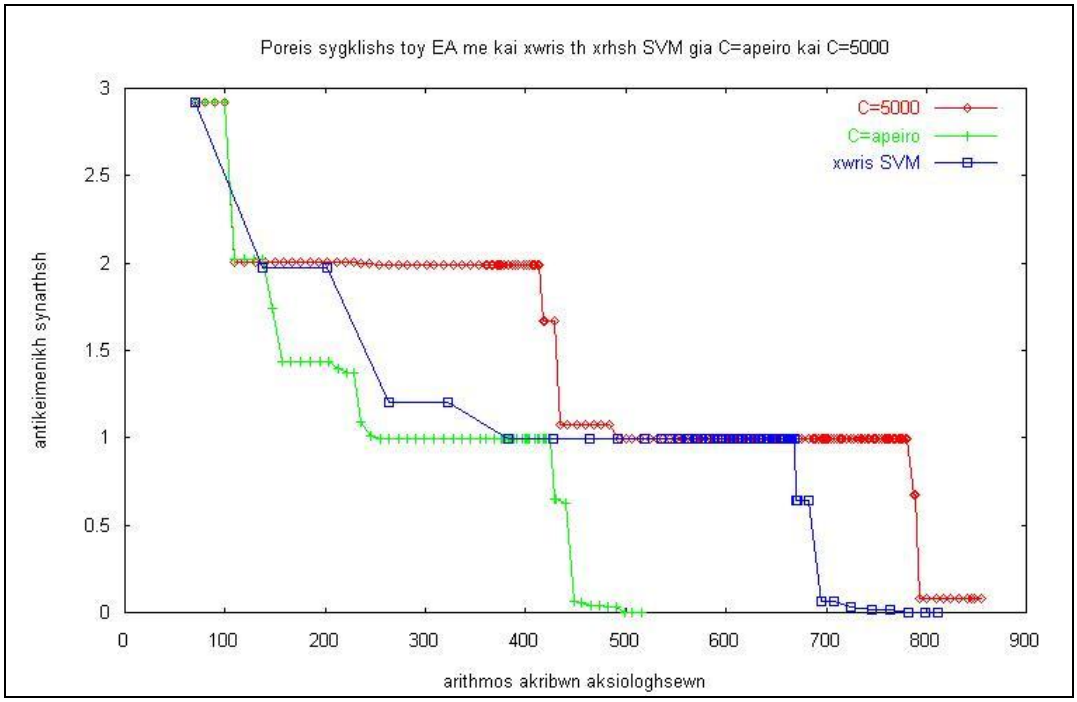
Διαχωριστής:	**C=100**	**C=1000**	**C=5000**	**C=10000**	**C=∞**
α/α μέλους – τιμή κόστους	59 16.2301385393	36 23.7163458918	***3*** ***10.0093524989***	51 39.2665236254	49 11.9536957872
	5 11.6658054757	19 23.3939576155	***34*** ***15.3024031998***	50 4.0549097007	4 32.3129225020
	60	50	***1***	22	14

	35.1138139784	3.1249582109	***9.3155628669***	3.0053856053	8.3911636244
	35 5.8173808073	47 3.4715393728	***37*** ***3.4025396401***	30 28.7876073899	55 20.8284676384
	45 19.0022142584	32 13.6242616046	***11*** ***6.5151473570***	56 2.8425371273	62 21.3962634513
	1 19.7047006964	45 9.5165895495	***58*** ***3.2722816400***	58 3.7713454733	20 28.8748780680
	43 10.9222198340	60 32.4540231438	***2*** ***8.6898802138***	48 18.5342604767	2 1.9287576572
	36 32.3052624033	65 23.9922262891	***67*** ***13.5557938496***	25 2.4788156824	31 23.2798176198
	3 10.5882488367	16 26.1882061945	***69*** ***10.6601375533***	47 2.5732886661	13 22.1555156464
	56 10.9045783247	6 28.8074582336	***38*** ***5.5195602803***	28 3.7802476790	69 34.3350785695

Πίνακας 2: Σύγκριση των τιμών κόστους των μελών μιας γενιάς που στέλνονται για ακριβή αξιολόγηση από τους διάφορους διαχωριστές.

<u>*Περίπτωση (Β):*</u> ***Μ=1,Ν=20,Κ=10,Λ=10***

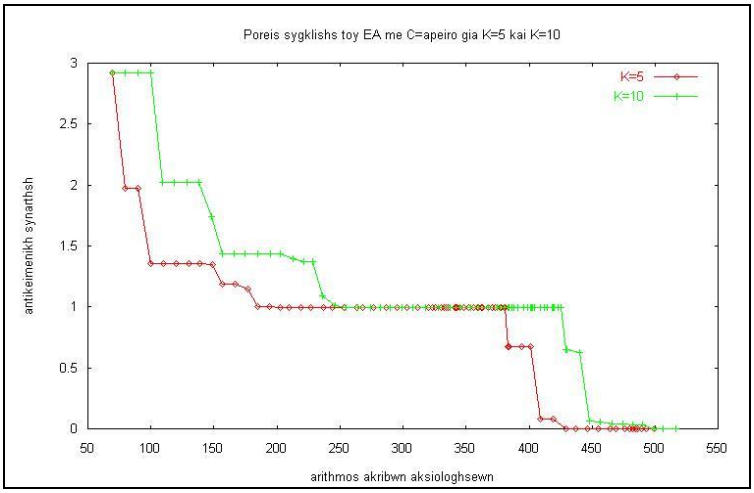
Δοκιμάζουμε να αυξήσουμε την παράμετρο ***Κ*** σε ***Κ=10***, δηλαδή τον αριθμό των μελών του τοπικού συνόλου εκπαίδευσης που μαρκάρονται ως «+1». Οι υπόλοιπες παράμετροι κρατούνται σταθερές. Οι αντίστοιχες πορείες σύγκλισης για $C=\infty$ και C=5000 φαίνονται στο σχήμα 8. Παρατηρούμε ότι στην περίπτωση αυτή, ο διαχωριστής με $C=\infty$ επιτυγχάνει καλύτερο ρυθμό σύγκλισης απ' ότι αυτός με C=5000. Δηλαδή η κατάσταση αντιστρέφεται υπέρ του πιο «αυστηρού» διαχωριστή, όταν αυξάνουμε το ***Κ.***



Σχήμα 8: Πορείες σύγκλισης των ΕΑ με διαχωριστές $C=\infty$ και C=5000, για *K*=10

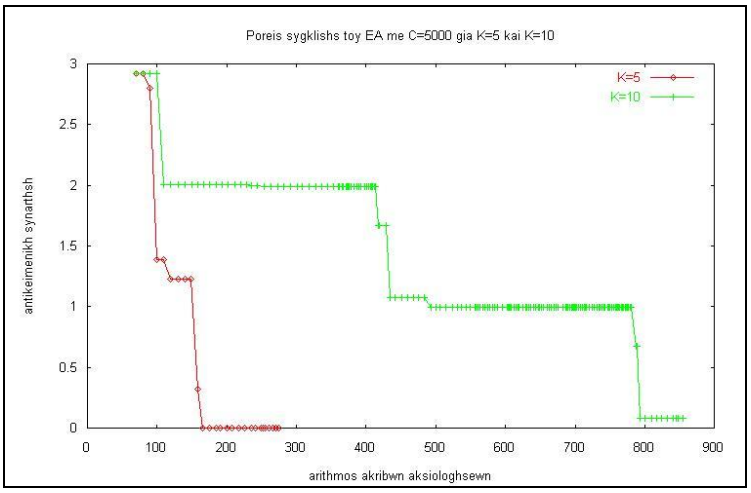
Το τελευταίο οφείλεται στο ότι αυξάνοντας το ***K***, αυξάνουμε την πληροφορία που έχει ο τοπικός *SVM* για το ποια είναι τα «καλά» μέλη της γενιάς. Σε συμφωνία λοιπόν με τα όσα ειπώθηκαν στην παράγραφο 3.1, όταν η πληροφορία αυτή είναι μεγαλύτερη, ο διαχωριστής με το μεγαλύτερο όριο C επιτυγχάνει καλύτερη γενίκευση. Αυτό εξηγεί τον ταχύτερο ρυθμό σύγκλισης του αντίστοιχου ΕΑ, έναντι αυτού που χρησιμοποιεί *SVM* με C=5000.

Ωστόσο και για τους δύο διαχωριστές η πορεία σύγκλισης χειροτέρεψε με αύξηση του ***K.*** Αυτό φαίνεται στα σχήματα 9,10 όπου συγκρίνονται οι πορείες σύγκλισης των διαχωριστών με ***K=5*** και ***K=10*** αντίστοιχα.



Σχήμα 9: Πορεία σύγκλισης του ΕΑ που χρησιμοποιεί SVM με $C=\infty$, για K=5 και K=10.

Το τελευταίο εξηγείται από το γεγονός ότι αυξάνοντας τον αριθμό των γειτόνων που μαρκάρονται ως «καλοί», διευρύνεται το πεδίο αναζήτησης της βέλτιστης λύσης του ΕΑ σε περιοχές που δεν ανήκουν αυστηρά στην περιοχή ελαχίστων τιμών. Ως αποτέλεσμα, ο ΕΑ χρειάζεται περισσότερες επαναλήψεις για να συγκλίνει στη βέλτιστη λύση.

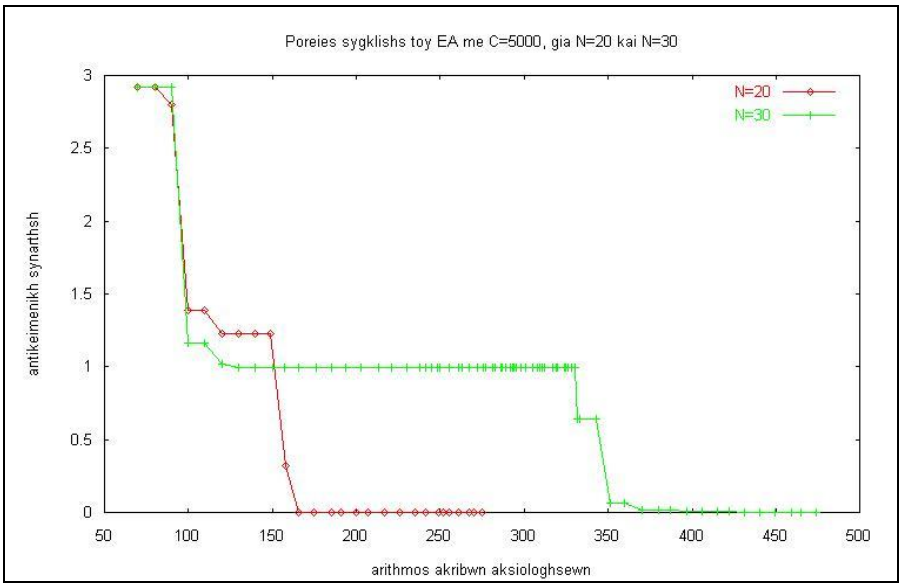


Σχήμα 10: Πορεία σύγκλισης του ΕΑ που χρησιμοποιεί SVM με C=5000, για K=5 και K=10.

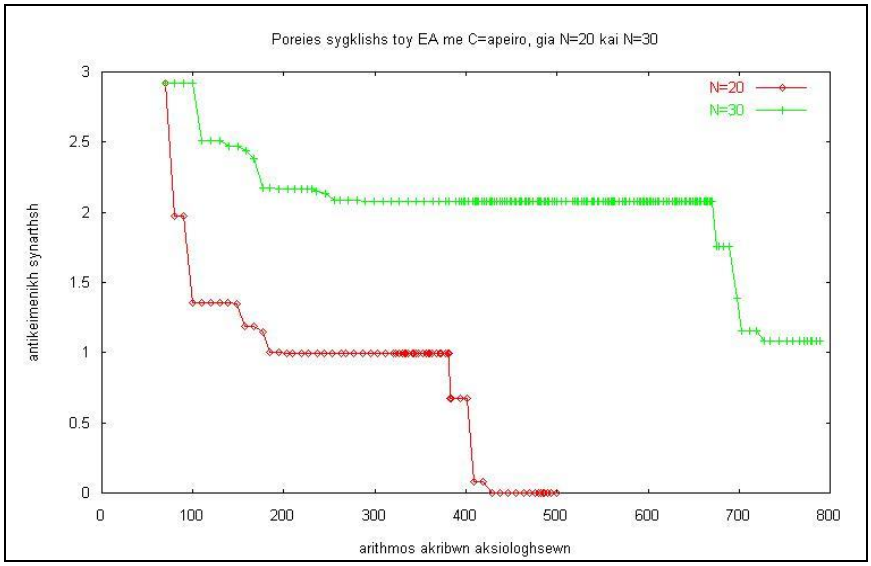
Περίπτωση (Γ): ***M=1,N=30,K=5,Λ=10***

Κρατούμε τις υπόλοιπες παραμέτρους σταθερές και αυξάνουμε το μέγεθος του συνόλου εκπαίδευσης των τοπικών *SVM* από ***N=20*** σε ***N=30.***

Οι αντίστοιχες πορείες σύγκλισης είναι χειρότερες σε σχέση με πριν, όπως φαίνεται από τα σχήματα 11,12 (ενδεικτικά για τους διαχωριστές C=5000, $C=\infty$).



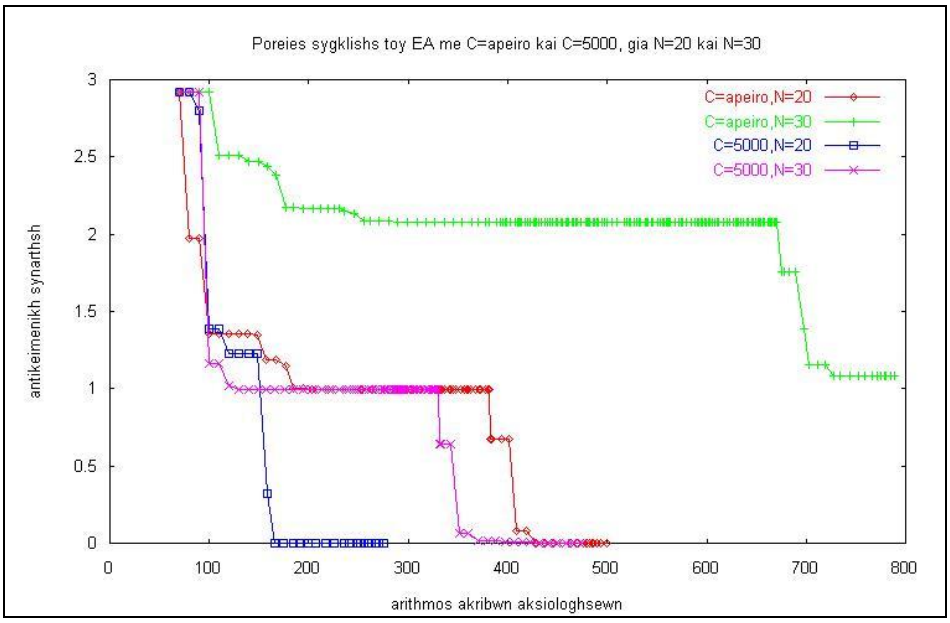
Σχήμα 11: Σύγκριση της πορείας σύγκλισης του ΕΑ που χρησιμοποιεί *SVM* με C=5000, για μεγέθη του συνόλου εκπαίδευσης *N*=20 και *N*=30.



Σχήμα 12: Σύγκριση της πορείας σύγκλισης του ΕΑ που χρησιμοποιεί *SVM* με $C=\infty$, για μεγέθη του συνόλου εκπαίδευσης *N*=20 και *N*=30.

Μάλιστα, όπως φαίνεται και από το σχήμα 13, η «ψαλίδα» ανάμεσα στις πορείες σύγκλισης των διαχωριστών C=5000, $C=\infty$ ανοίγει ακόμα

περισσότερο. Η αύξηση της παραμέτρου ***N*** έχει την ίδια επίδραση με την μείωση της παραμέτρου ***K***: Η πληροφορία που έχει ο διαχωριστής για τους «καλούς» γείτονες του συνόλου εκπαίδευσης μειώνεται.



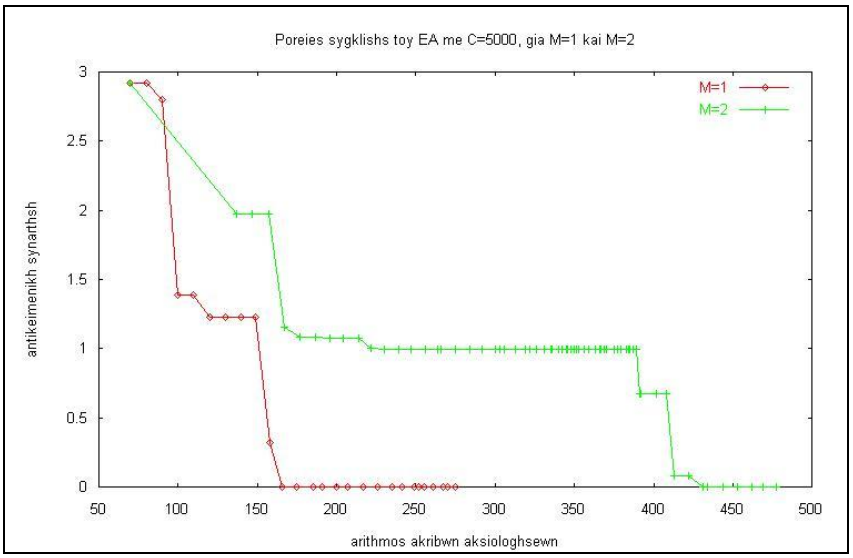
Σχήμα 13: Σύγκριση της πορείας σύγκλισης των ΕΑ με $C=\infty$ και C=5000, για τις δύο τιμές του *N*. Η «ψαλίδα» ανοίγει υπέρ του πιο ελαστικού διαχωριστή (C=5000) όσο αυξάνεται το *N*.

Πράγματι, για την περίπτωση (Β) ο αριθμός των «καλών» γειτόνων στο σύνολο εκπαίδευσης προς το μέγεθος του συνόλου εκπαίδευσης είναι ***K/N***=0.5 (50%). Ο λόγος αυτός στις περιπτώσεις (Α) και (Γ) είναι αντίστοιχα ***K/N***=0.25 (25%) και ***K/N***=0.166 (16.6%). Αυτό εξηγεί την χειρότερη απόδοση του διαχωριστή $C=\infty$ έναντι του C=5000 στις περιπτώσεις (Α) και (Γ).

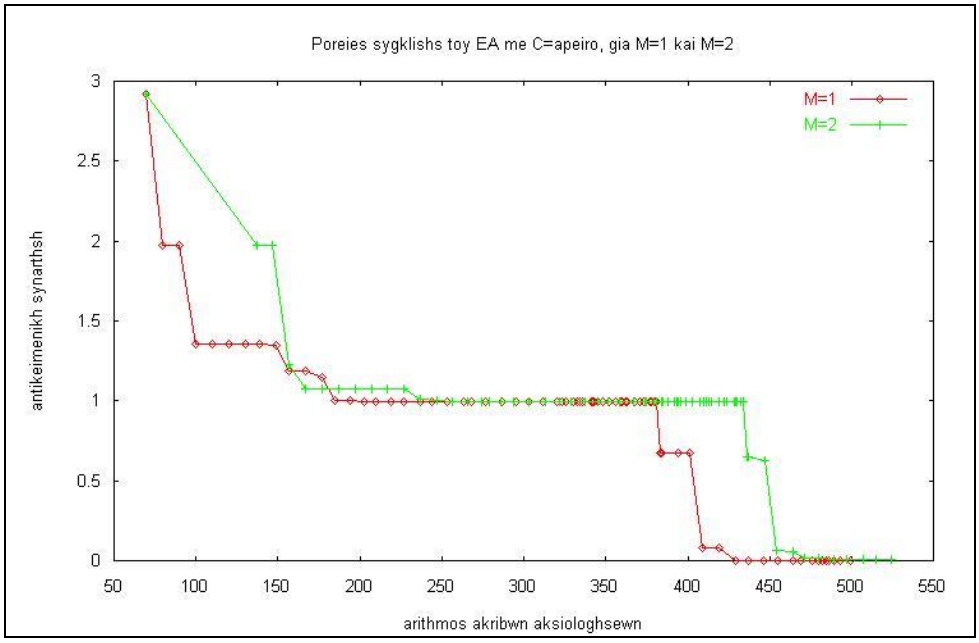
<u>*Περίπτωση (Δ):*</u> ***M=2,N=20,K=5,Λ=10*** και ***M=3,N=20,K=5,Λ=10***

Δοκιμάζουμε τώρα να αυξήσουμε τον αρχικό αριθμό γενεών που υφίστανται ακριβή αξιολόγηση πριν την εφαρμογή του *SVM* από ***M=1*** σε ***M=2*** και μετέπειτα σε ***M=3***, κρατώντας τις υπόλοιπες παραμέτρους ***K,Λ,N*** ίδιες με την περίπτωση *(Α)*. Οι πορείες σύγκλισης τότε είναι χειρότερες σε σχέση με πριν, όπως φαίνεται από τα σχήματα 14,15. Από τα σχήματα συνάγεται ότι η εφαρμογή του *SVM* από τα πρώτα στάδια του ΕΑ (μικρό

M), εκεί δηλαδή που γενικά ο ΕΑ έχει συνήθως τον πιο έντονο ρυθμό πτώσης, επιτυγχάνει ταχύτερους ρυθμούς σύγκλισης. Αυτό υποδηλώνει μια πιο «επιθετική» συμπεριφορά του ΕΑ, που φιλτράρει τις μη-υποσχόμενες λύσεις από την αρχή κιόλας της βελτιστοποίησης, και συνεπώς περιορίζει το εύρος αναζήτησης λύσεων στις πλέον βέλτιστες περιοχές.

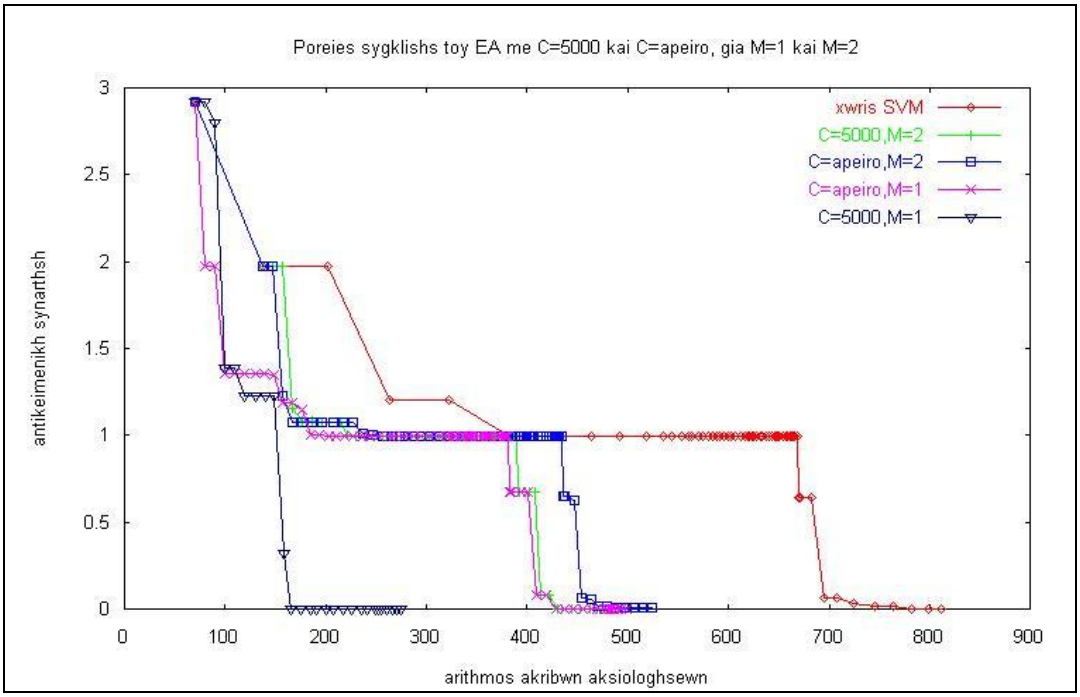


Σχήμα 14: Πορείες σύγκλισης του ΕΑ με C=5000, για *M=1* και *M=2* αντίστοιχα.

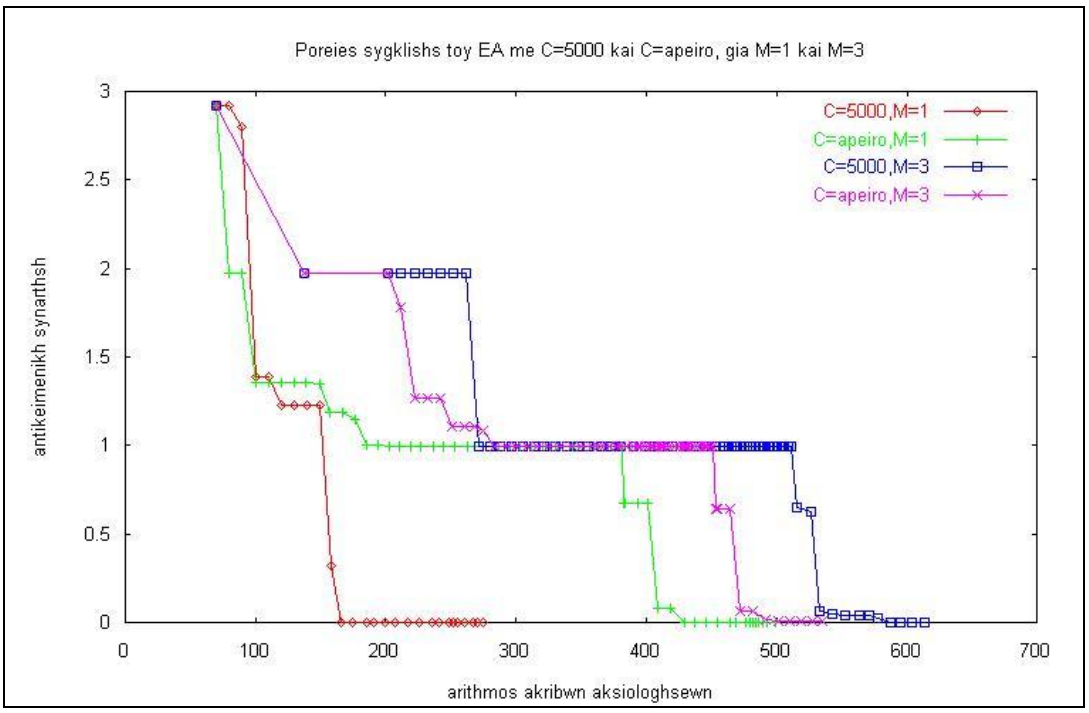


Σχήμα 15: Πορείες σύγκλισης του ΕΑ με $C=\infty$, για *M*=1 και *M*=2 αντίστοιχα.

Ως εκ τούτου, οι απαιτούμενες επαναλήψεις για να βρεθεί η βέλτιστη λύση από τον ΕΑ μειώνονται. Επίσης, παρατηρώντας τα σχήματα 16 και 17, όπου φαίνονται οι πορείες σύγκλισης του ΕΑ με $C=\infty$ και C=5000 για τις τρεις τιμές της παραμέτρου ***M***, διαπιστώνουμε ότι αυξάνοντας το ***M*** ο «αυστηρός» διαχωριστής $C=\infty$ φαίνεται να κερδίζει έδαφος έναντι του πιο «ελαστικού» C=5000. Αυτό οφείλεται στο ότι η αρχική πληροφορία που παρέχεται από τη βάση δεδομένων για την εκπαίδευση των τοπικών *SVM* είναι πιο «ακριβής» (περισσότερες αρχικές ακριβείς αξιολογήσεις, πιο ορθολογική απόδοση ετικετών στο σύνολο εκπαίδευσης, ή αλλιώς λιγότερος «θόρυβος»). Συνεπώς, ο «αυστηρός» διαχωριστής $C=\infty$ την εκμεταλλεύεται πιο σωστά, και παράγει διαχωριστική γραμμή με καλύτερη γενίκευση απ' ότι ο πιο ελαστικός C=5000.



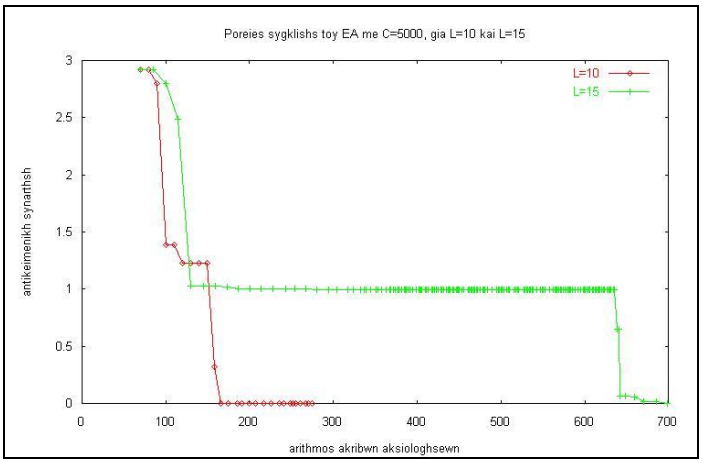
Σχήμα 16: Πορείες σύγκλισης των ΕΑ που χρησιμοποιούν τους δύο διαχωριστές, για *M*=1 και *M*=2. Η ψαλίδα κλείνει υπέρ του πιο αυστηρού διαχωριστή ($C = \infty$) όσο αυξάνεται το *M*.



Σχήμα 17: Πορείες σύγκλισης των ΕΑ που χρησιμοποιούν τους δύο διαχωριστές, για *M*=1 και *M*=3. Ο «αυστηρός» διαχωριστής έχει επιτύχει για *M*=3 καλύτερο ρυθμό σύγκλισης από τον πιο «ελαστικό» C=5000.

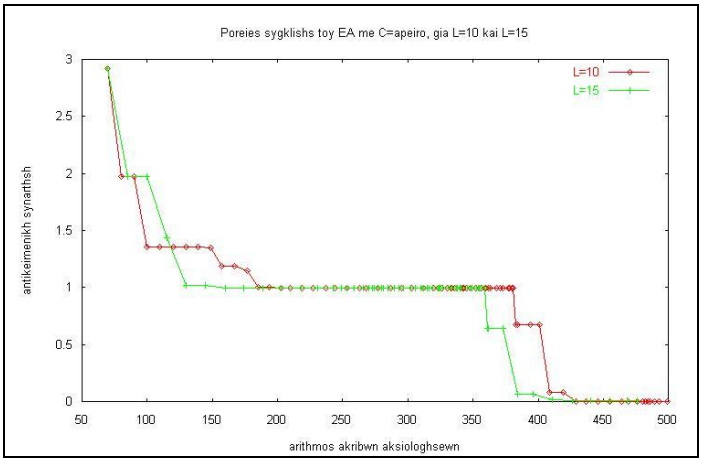
Περίπτωση (Ε): ***Μ=1,Κ=5,Ν=20,Λ=15***

Εξετάζουμε τέλος την πορεία σύγκλισης των ΕΑ που χρησιμοποιούν τους δύο διαχωριστές (C=5000 και $C=\infty$), κρατώντας σταθερές τις παραμέτρους ***Μ,Κ,Ν*** και αυξάνοντας τον αριθμό ***Λ*** των μελών κάθε γενιάς που στέλνονται για ακριβή αξιολόγηση από τους τοπικούς *SVM*, από ***Λ=10*** σε ***Λ=15***. Στα σχήματα 18,19 φαίνονται οι νέες πορείες σύγκλισης των δύο διαχωριστών, συγκρινόμενες με τις παλιές.



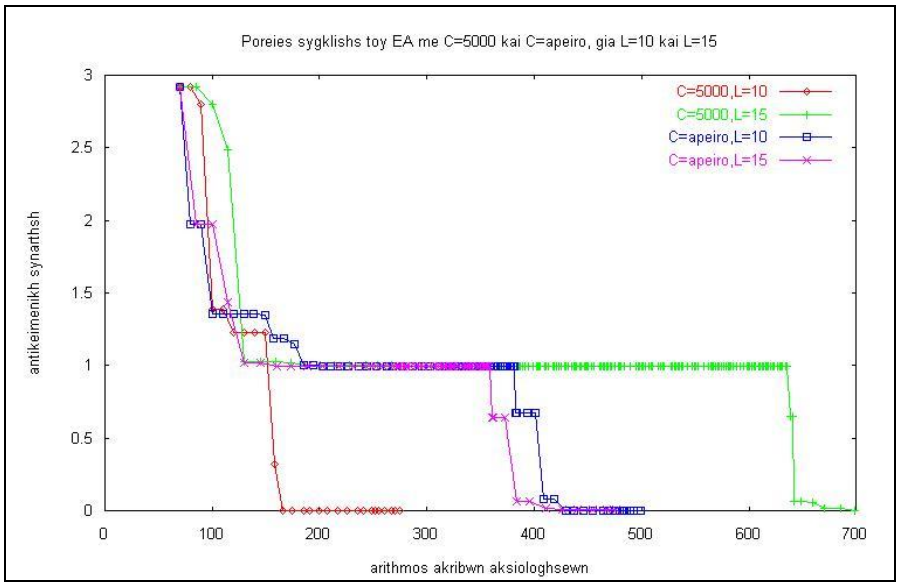
Σχήμα 18: Πορεία σύγκλισης του ΕΑ με C=5000, για αριθμό ακριβών αξιολογήσεων ανά γενιά *Λ*=10 και *Λ*=15 αντίστοιχα. Ο διαχωριστής αυτός δεν ευνοείται από την αύξηση της παραμέτρου *Λ* από 10 σε 15.

Όπως φαίνεται στο σχήμα 19, ο «αυστηρός» διαχωριστής $C=\infty$ ευνοείται από την αύξηση του αριθμού των ακριβών αξιολογήσεων ανά γενιά.



Σχήμα 19: Πορεία σύγκλισης του ΕΑ με $C=\infty$, για αριθμό ακριβών αξιολογήσεων ανά γενιά *Λ*=10 και *Λ*=15 αντίστοιχα. Ο διαχωριστής αυτός ευνοείται από την αύξηση της παραμέτρου *Λ*.

Αυτό οφείλεται στο ότι σε κάθε γενιά η βάση δεδομένων εμπλουτίζεται με περισσότερη πληροφορία απ' ότι πριν, την οποία ο διαχωριστής $C=\infty$ εκμεταλλεύεται κατάλληλα και παράγει διαχωριστική γραμμή με καλύτερη γενίκευση. Αντίθετα, ο πιο ελαστικός διαχωριστής C=5000 εμφανίζει χειρότερη απόδοση (σχήμα 19), για τους ίδιους λόγους που εξετάστηκαν και στην περίπτωση *(Δ)*. Στο σχήμα 20 παρατίθενται οι πορείες σύγκλισης των ΕΑ που χρησιμοποιούν τους δύο διαχωριστές, για ***Λ=10*** (αρχική τιμή) και ***Λ=15*** (νέα τιμή).



Σχήμα 20: Σύγκριση της πορείας σύγκλισης των ΕΑ που χρησιμοποιούν *SVM* με C=5000 και $C=\infty$, για τις δύο τιμές της παραμέτρου *Λ*. Είναι εμφανές ότι η αύξηση του αριθμού *Λ* των ακριβών αξιολογήσεων ανά γενιά επιδρά θετικά στη σύγκλιση του ΕΑ $C=\infty$ και αρνητικά στη σύγκλιση του ΕΑ C=5000.

Αξιολόγηση των αποτελεσμάτων – Συμπεράσματα και παρατηρήσεις:

- Από τα παραπάνω φάνηκε ότι η συνεισφορά του *SVM* στον ΕΑ βελτιστοποίησης επιτάχυνε το ρυθμό σύγκλισης σε όλες σχεδόν τις περιπτώσεις.
- Οι ρυθμιστικές παράμετροι ***K,Λ,M,N***, όπως επίσης και το άνω όριο C των πολλαπλασιαστών Lagrange, επηρεάζουν σημαντικά την απόδοση του *SVM*
- Για διαφορετικές τιμές του ορίου C που θα επιλέξουμε για το διαχωριστή, ο βέλτιστος συνδυασμός των παραμέτρων ***K,Λ,M,N*** διαφέρει. Μάλιστα, η αύξηση των τιμών αυτών ευνοεί κάποιους διαχωριστές και επιδρά αρνητικά στην απόδοση κάποιων άλλων.

Αναλυτικότερα για τις παραμέτρους ***K,Λ,M,N*** συμπεραίνουμε τα εξής:

- Όσο μεγαλύτερος είναι ο αριθμός ***M*** των γενεών που υφίστανται ακριβή αξιολόγηση πριν εφαρμοστεί ο *SVM*, τόσο πιο αργός είναι ο ρυθμός σύγκλισης. Αυτό ισχύει γενικά για όλους τους διαχωριστές, ανεξαρτήτως της τιμής C αυτών. Ωστόσο, ο πιο «αυστηρός» διαχωριστής δείχνει να ευνοείται έναντι του πιο «ελαστικού» διότι εμπλουτίζεται η αρχική βάση δεδομένων με περισσότερη πληροφορία.
- Η τιμή της παραμέτρου ***N*** θα πρέπει να ρυθμίζεται σε συνδυασμό με την παράμετρο ***K***, λαμβάνοντας υπόψη το λόγο τους. Μεγάλος λόγος ***K/N*** ευνοεί τον πιο «αυστηρό» διαχωριστή, ενώ το αντίθετο συμβαίνει για μικρό λόγο.
- Η παράμετρος ***Λ*** αυξανόμενη ευνοεί τον «αυστηρό» διαχωριστή και δρα αρνητικά στον «ελαστικό».

Συμπερασματικά, επιλέγοντας την τιμή ***M=1*** για να οδηγήσουμε τον ΕΑ πιο επιθετικά στην αναζήτηση των βέλτιστων λύσεων από τα πρώτα

κιόλας βήματα, προτιμάμε τον διαχωριστή C=5000, και οι τιμές των υπολοίπων παραμέτρων που επιλέγουμε με βάση την παραπάνω ανάλυση είναι:

N=20 για το μέγεθος του συνόλου εκπαίδευσης των τοπικών *SVM*

K=5 για το πλήθος των γειτόνων ενός μέλους που μαρκάρονται ως «καλοί»

Λ=10 για τον αριθμό των μελών κάθε γενιάς που στέλνονται για ακριβή αξιολόγηση.

Τονίζουμε τέλος ότι οι παραπάνω επιλογές παραμέτρων αφορούν στο συγκεκριμένο πρόβλημα βελτιστοποίησης, και δεν έχουν γενική εφαρμογή και σε άλλα προβλήματα. Για παράδειγμα, η επιλογή μικρής τιμής για την παράμετρο ***M*** στο συγκεκριμένο πρόβλημα δρα θετικά στη σύγκλιση του ΕΑ. Στην επόμενη παράγραφο ωστόσο, που αντιμετωπίζεται ένα πιο σύνθετο πρόβλημα ελαχιστοποίησης, η μείωση κάτω από ένα όριο του αριθμού ***M*** των γενεών που υφίστανται αρχικά ακριβή αξιολόγηση, δρα αρνητικά στην πορεία σύγκλισης. Το εν λόγω πρόβλημα είναι πιο απαιτητικό, και ο αντίστοιχος *SVM* χρειάζεται πιο επαρκή πληροφόρηση (μεγαλύτερη αρχική βάση δεδομένων) για να λειτουργήσει αποδοτικά. Επίσης, σε κάθε περίπτωση έγινε ένα μόνο τρέξιμο του ΕΑ. Κανονικά, θα έπρεπε να γίνουν περισσότερα (λόγω στοχαστικότητας του αλγορίθμου) και τα συμπεράσματα να εξάγονται από τη μέση συμπεριφορά του.

Παρατηρήσεις σχετικά με την εξοικονόμηση υπολογιστικού χρόνου:

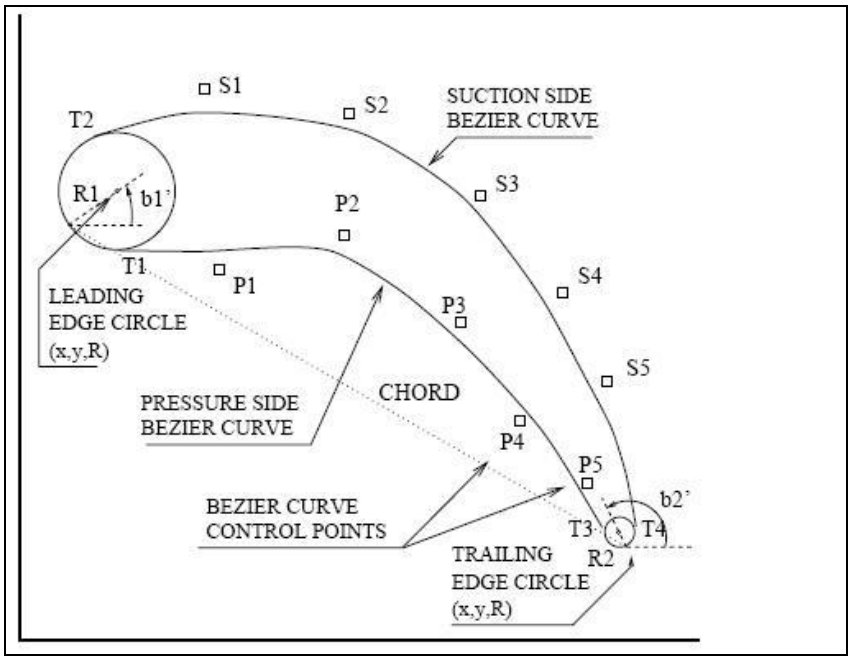
Η εφαρμογή του *SVM* υποβοηθητικά στον ΕΑ οδήγησε σε σημαντική μείωση του αριθμού των ακριβών αξιολογήσεων, και συνεπώς θα μπορούσε να πει κανείς και σε μείωση του υπολογιστικού χρόνου. Βέβαια,

αυτό είναι μέχρι ενός σημείου αληθές. Κι αυτό γιατί η εκπαίδευση του *SVM* απαιτεί κι αυτή κάποιο χρόνο CPU. Μάλιστα, στην προκειμένη περίπτωση, ο χρόνος που απαιτείται για την εκπαίδευση του *SVM* είναι συγκρίσιμος με το χρόνο που απαιτείται για την ακριβή αξιολόγηση μιας νέας γενιάς από τον «απλό» ΕΑ. Αυτό συμβαίνει διότι το συγκεκριμένο πρόβλημα ελαχιστοποίησης είναι σχετικά απλό, εφόσον εμπλέκονται μόνο δύο παράμετροι σε αυτό (οι συντεταγμένες x_1, x_2 των υπό εξέταση σημείων), και άρα η ακριβής αξιολόγηση κάθε υποψήφιας λύσης (εύρεση της τιμής της Rastrigin) είναι γρήγορη. Ωστόσο, σε ένα πιο σύνθετο πρόβλημα που θα απαιτεί πολύ περισσότερο χρόνο για την αξιολόγηση μιας λύσης, όπως συνήθως είναι τα προβλήματα αεροδυναμικής βελτιστοποίησης, ο χρόνος εκπαίδευσης του *SVM* μπορεί να θεωρηθεί αμελητέος συγκρινόμενος με το χρόνο της ακριβούς αξιολόγησης κάθε γενιάς. Συνεπώς, στις περιπτώσεις αυτές, η μείωση στον αριθμό των αξιολογήσεων μέσω του *SVM* μεταφράζεται άμεσα και σε μείωση του υπολογιστικού χρόνου που απαιτεί ο εξελικτικός αλγόριθμος, χωρίς να λαμβάνεται υπόψη ο λίγος χρόνος εκπαίδευσης του *SVM*.

3.2.2 Προσέγγιση της γεωμετρίας μιας δοθείσης αεροτομής

Δεδομένου ότι το κόστος σχεδιασμού μιας νέας αεροτομής είναι σχετικά υψηλό, είναι πολλές φορές προτιμητέο από μια βιομηχανία να εξελίσσονται οι υπάρχουσες αεροτομές, χωρίς να γίνεται επανασχεδιασμός τους από την αρχή. Η εξέλιξη – βελτιστοποίηση μιας αεροτομής προϋποθέτει γενικά ένα συστηματικό τρόπο περιγραφής της γεωμετρίας της. Μια μέθοδος για την περιγραφή της γεωμετρίας μιας

αεροτομής είναι με τη χρήση καμπυλών Bezier. Συγκεκριμένα, η αεροτομή θεωρείται ότι αποτελείται από τρία διακριτά τμήματα: την ακμή προσβολής, την ακμή εκφυγής και το κυρίως σώμα. Η ακμή προσβολής μοντελοποιείται ως ένας κύκλος με καθορισμένη ακτίνα και κέντρο, όπως επίσης και η ακμή εκφυγής. Το υπόλοιπο σώμα της αεροτομής ορίζεται από ένα πλήθος σημείων ελέγχου Bezier (Bezier control points), τόσο στην πλευρά πίεσης όσο και στην πλευρά υποπίεσης (pressure side/suction side). Στο σχήμα 21 φαίνεται η παραμετροποίηση της γεωμετρίας μιας αεροτομής με τη χρήση καμπυλών Bezier.



Σχήμα 21: Παραμετροποίηση μιας αεροτομής με καμπύλες Bezier. Διακρίνονται τα σημεία ελέγχου των πλευρών πίεσης και υποπίεσης (σημεία P_i και S_i στο σχήμα)

Με την παραπάνω παραμετροποίηση, αν γνωρίζει κανείς τα κέντρα και τις ακτίνες των κύκλων προσβολής και εκφυγής, καθώς και τα σημεία ελέγχου των καμπυλών Bezier των πλευρών πίεσης και υποπίεσης, μπορεί να παράξει το σχήμα της αεροτομής.

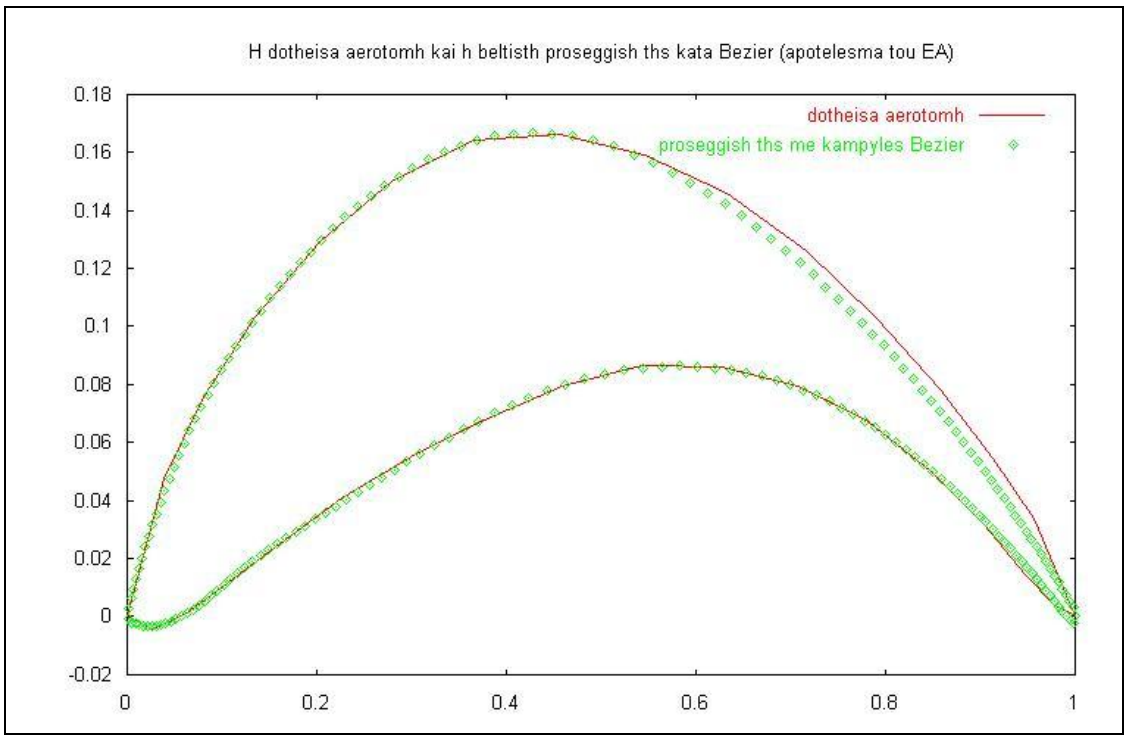
Ένα πρόβλημα που εμφανίζεται συχνά κατά την εξέλιξη υπαρχουσών αεροτομών, είναι ότι ο τρόπος περιγραφής της γεωμετρίας της υπάρχουσας αεροτομής δεν είναι διαθέσιμος. Συνήθως είναι διαθέσιμη η ίδια η γεωμετρία της αεροτομής με τη μορφή ενός συνόλου σημείων στο 2-Δ ή 3-Δ χώρο, χωρίς ωστόσο να είναι διαθέσιμη η παραμετροποίηση (αν όντως υπήρξε) από την οποία προέκυψαν αυτά τα σημεία. Ως εκ τούτου, υπάρχει η ανάγκη όταν δίνεται η γεωμετρία μιας αεροτομής, να βρεθεί μια βέλτιστη προσέγγισή της με τη χρήση καμπυλών Bezier, δηλαδή να βρεθούν οι θέσεις εκείνες των σημείων ελέγχου, ώστε η προκύπτουσα παραμετροποίηση κατά Bezier να είναι όσο το δυνατόν πλησιέστερη στην πραγματική γεωμετρία της αεροτομής. Ως αποτέλεσμα, θα μπορεί στη συνέχεια να εφαρμοστεί σε αυτή την παραμετροποιημένη κατά Bezier αεροτομή ένας ΕΑ βελτιστοποίησης της θέσης των σημείων ελέγχου της, δηλαδή της μορφής της, προκειμένου η αεροτομή να βελτιστοποιηθεί ως προς την αεροδυναμική της απόδοση.

Για τη βέλτιστη προσέγγιση μιας δοθείσης αεροτομής από καμπύλες Bezier, εφαρμόστηκε ένας ΕΑ με ελεύθερες παραμέτρους τις συντεταγμένες των σημείων ελέγχου των καμπυλών Bezier, με στόχο να βρεθούν οι βέλτιστες θέσεις τους, δηλαδή οι θέσεις αυτές που ελαχιστοποιούν την απόκλιση της παραγόμενης γεωμετρίας της αεροτομής από την πραγματική. Στην περίπτωση αυτή δηλαδή, η αντικειμενική συνάρτηση που καλείται να ελαχιστοποιήσει ο ΕΑ είναι η απόκλιση που εμφανίζει η γεωμετρία της δοθείσης αεροτομής, από τη γεωμετρία της αεροτομής που παράγεται με τις καμπύλες Bezier. Υπάρχουν πολλές εναλλακτικές μαθηματικές εκφράσεις αυτής της απόκλισης, όπως για παράδειγμα:

$$E = \int_{s_{\min}}^{s_{\max}} (x_{given}(s) - x_{bez}(s))^2 \, ds$$

όπου $x_{given}(s), x_{bez}(s)$ τα τόξα της δοθείσης και της παραγόμενης αεροτομής αντίστοιχα, παραμετροποιημένα με την παράμετρο s. Περαιτέρω αναφορά στον ακριβή μαθηματικό τύπο της γεωμετρικής απόκλισης των δύο αεροτομών δεν έχει ουσιαστική σημασία.

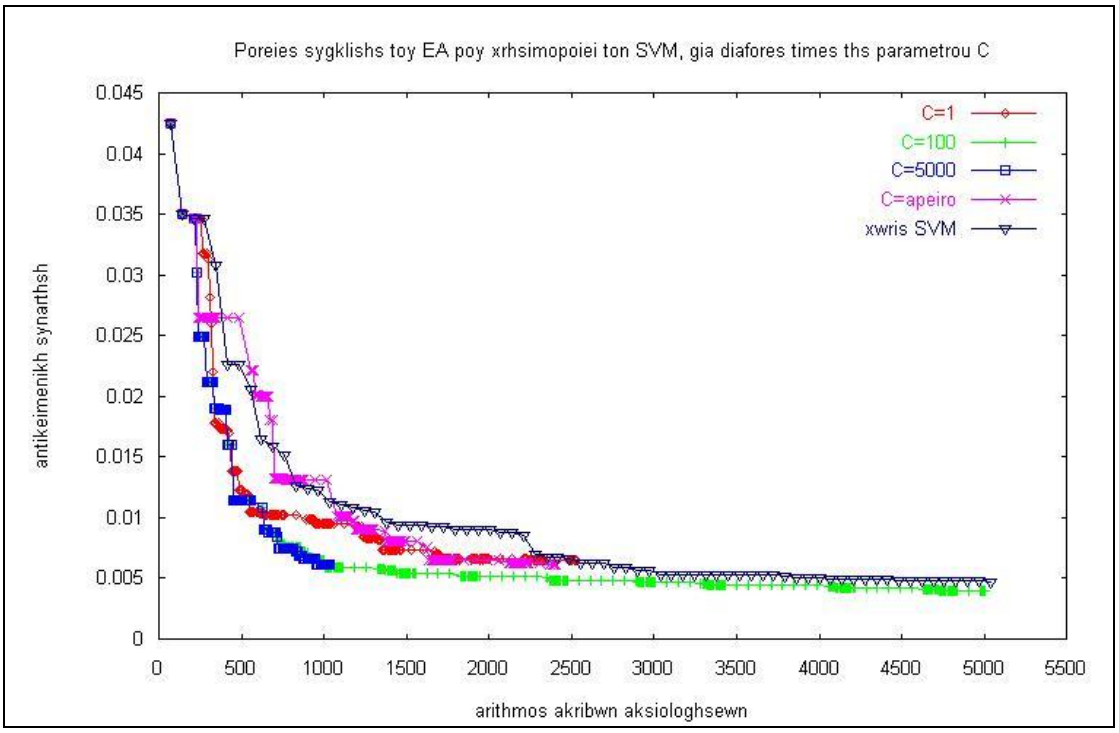
Αυτό που έχει σημασία είναι ότι στα συγκριτικά τρεξίματα του ΕΑ με και χωρίς την εφαρμογή του *SVM*, τα αντίστοιχα αποτελέσματα καταδεικνύουν την αποτελεσματικότητα της μεθόδου *SVM* στο να μειώσει τον απαιτούμενο αριθμό ακριβών αξιολογήσεων. Στο σχήμα 22 φαίνεται το αποτέλεσμα της σύγκλισης του ΕΑ βελτιστοποίησης, δηλαδή η βέλτιστη προσέγγιση της δοθείσης αεροτομής με καμπύλες Bezier.



Σχήμα 22: Βέλτιστη προσέγγιση της δοθείσης αεροτομής από καμπύλες Bezier, ως αποτέλεσμα της σύγκλισης του ΕΑ βελτιστοποίησης

Και στην περίπτωση αυτή, οι παράμετροι ***Κ,Λ,Μ,Ν*** διερευνήθηκαν ως προς την επίδρασή τους στην αποτελεσματικότητα του αλγορίθμου.

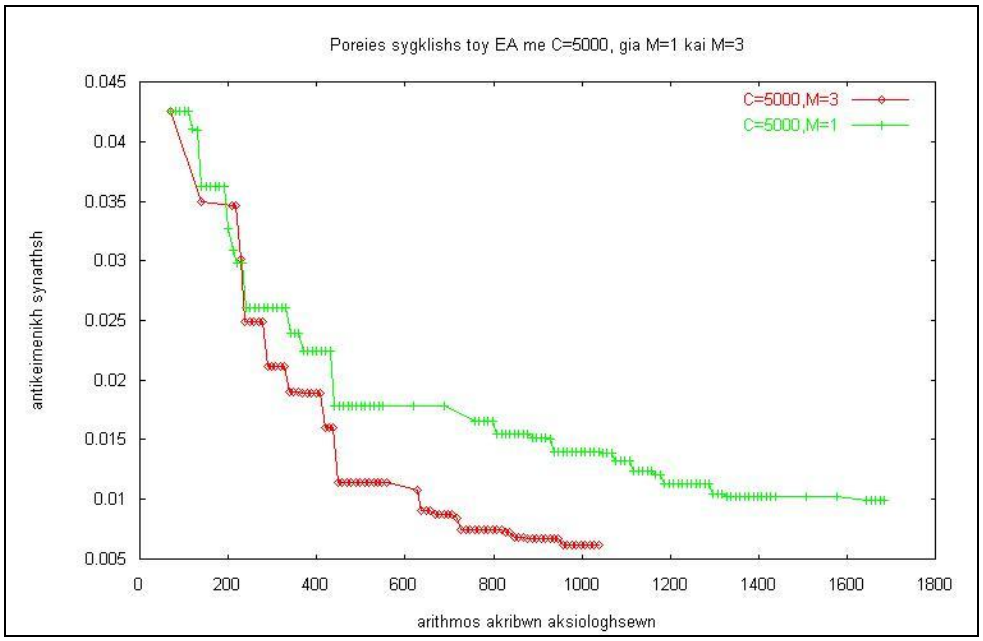
<u>*Περίπτωση (Α):*</u> Στο σχήμα 23 φαίνονται οι πορείες σύγκλισης του ΕΑ που χρησιμοποιεί *SVM*, για διάφορες τιμές της παραμέτρου C, συγκρινόμενες με την πορεία σύγκλισης του ΕΑ που δεν χρησιμοποιεί *SVM*. Οι τιμές των υπολοίπων παραμέτρων που επιλέχτηκαν αρχικά είναι ***Μ=3,Κ=10,Λ=10,Ν=20.***



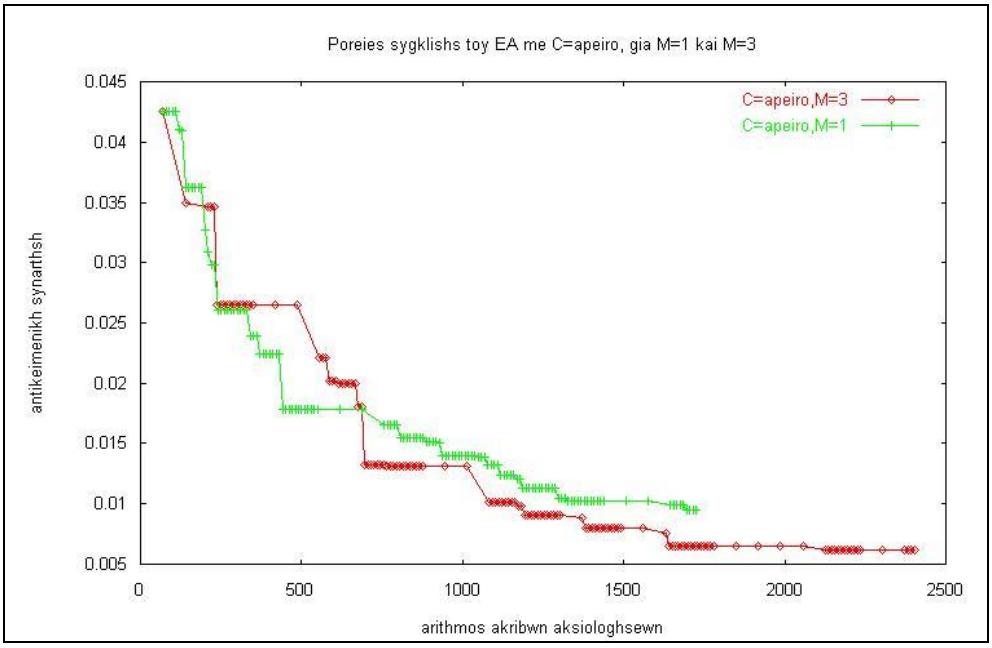
Σχήμα 23: Πορείες σύγκλισης του ΕΑ που χρησιμοποιεί τον *SVM*, για διάφορες τιμές της παραμέτρου C

Παρατηρούμε ότι την ταχύτερη σύγκλιση επιτυγχάνει ο ΕΑ που χρησιμοποιεί *SVM* με C=5000. Στα παρακάτω συγκρίνουμε (όπως και στην περίπτωση της ελαχιστοποίησης της συνάρτησης του Rastrigin) τη συμπεριφορά αυτού του διαχωριστή (C=5000) με τον «αυστηρό» διαχωριστή $C=\infty$, για διάφορες τιμές των παραμέτρων ***Μ,Ν,Κ,Λ.***

Περίπτωση (Β): Μειώνουμε τον αριθμό ***M*** των γενεών που υφίστανται αρχικά ακριβή αξιολόγηση πριν την εφαρμογή του *SVM*. Τα αντίστοιχα αποτελέσματα σύγκλισης φαίνονται για τους δύο διαχωριστές στα σχήματα 24 και 25. Όπως παρατηρούμε από τα σχήματα, η απόδοση των δύο διαχωριστών (και συνεπώς και οι ρυθμοί σύγκλισης που επιτυγχάνονται για τους αντίστοιχους ΕΑ) είναι χειρότερη από πριν. Αυτό σημαίνει ότι η αρχική βάση δεδομένων δεν είναι τώρα επαρκής για την ικανοποιητική εκπαίδευση του *SVM* και συνεπώς δεν μπορεί να εξασφαλιστεί καλή γενίκευση.



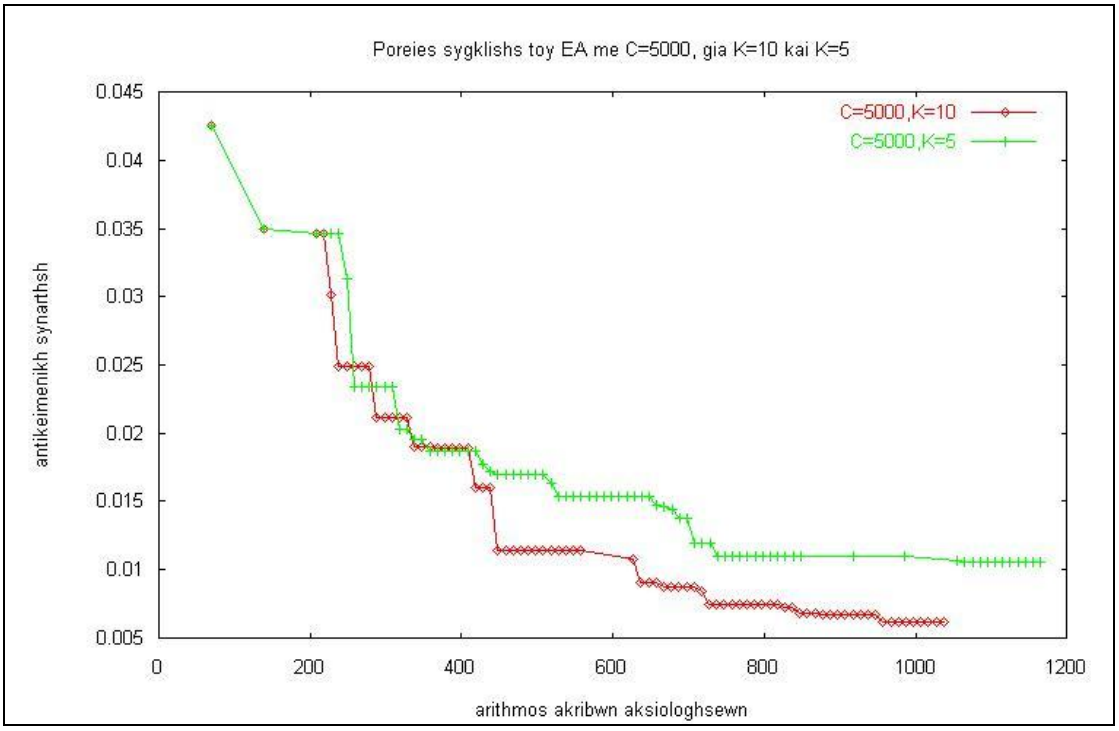
Σχήμα 24: Πορείες σύγκλισης του ΕΑ C=5000, για M=1 και M=3.



Σχήμα 25: Πορείες σύγκλισης του ΕΑ με $C=\infty$, για M=1 και M=3.

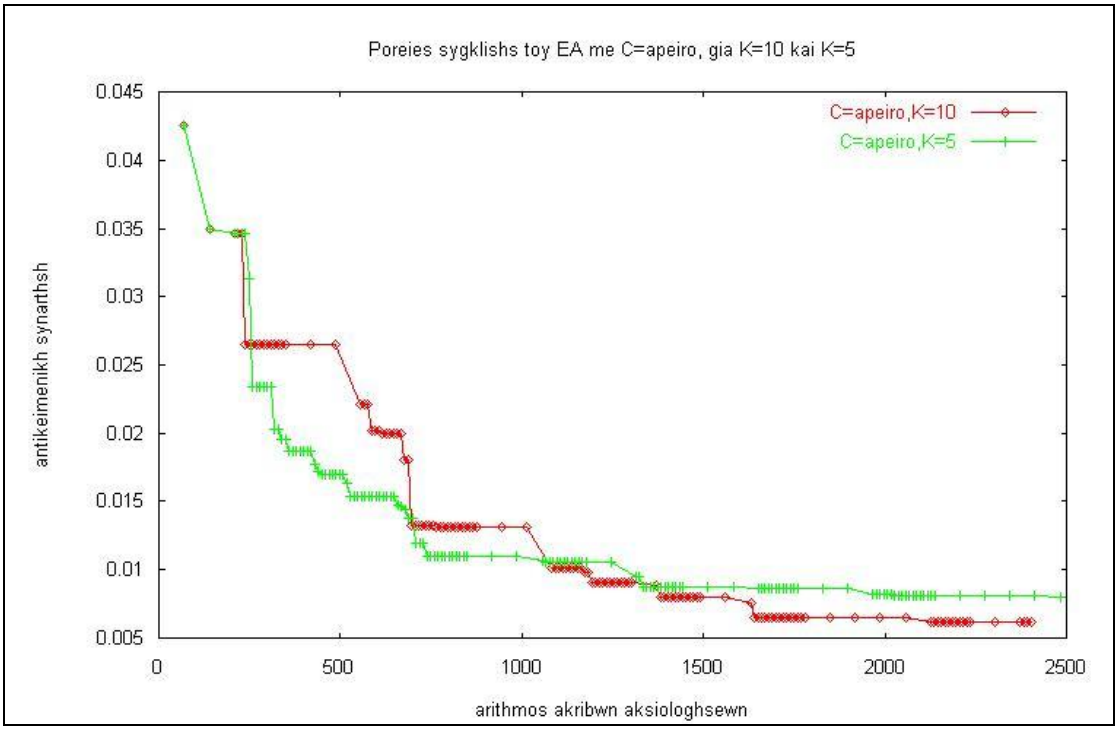
Το συγκεκριμένο πρόβλημα ελαχιστοποίησης είναι πιο σύνθετο από ότι η ελαχιστοποίηση της συνάρτησης του Rastrigin που εξετάσαμε πριν, καθώς εμπλέκονται περισσότερες ελεύθερες παράμετροι σε σχέση με πριν (οι συντεταγμένες των σημείων ελέγχου των καμπυλών Bezier). Συνεπώς η μείωση της αρχικής πληροφορίας που έχει ο SVM κάτω από ένα όριο επιφέρει χειρότερα και όχι καλύτερα αποτελέσματα.

<u>*Περίπτωση (Γ):*</u> Μειώνουμε την παράμετρο ***K*** σε ***K=5***, δηλαδή τον αριθμό των γειτόνων ενός μέλους που μαρκάρονται ως «καλοί». Οι πορείες σύγκλισης των ΕΑ που χρησιμοποιούν τους δύο διαχωριστές φαίνονται στα σχήματα 26,27.



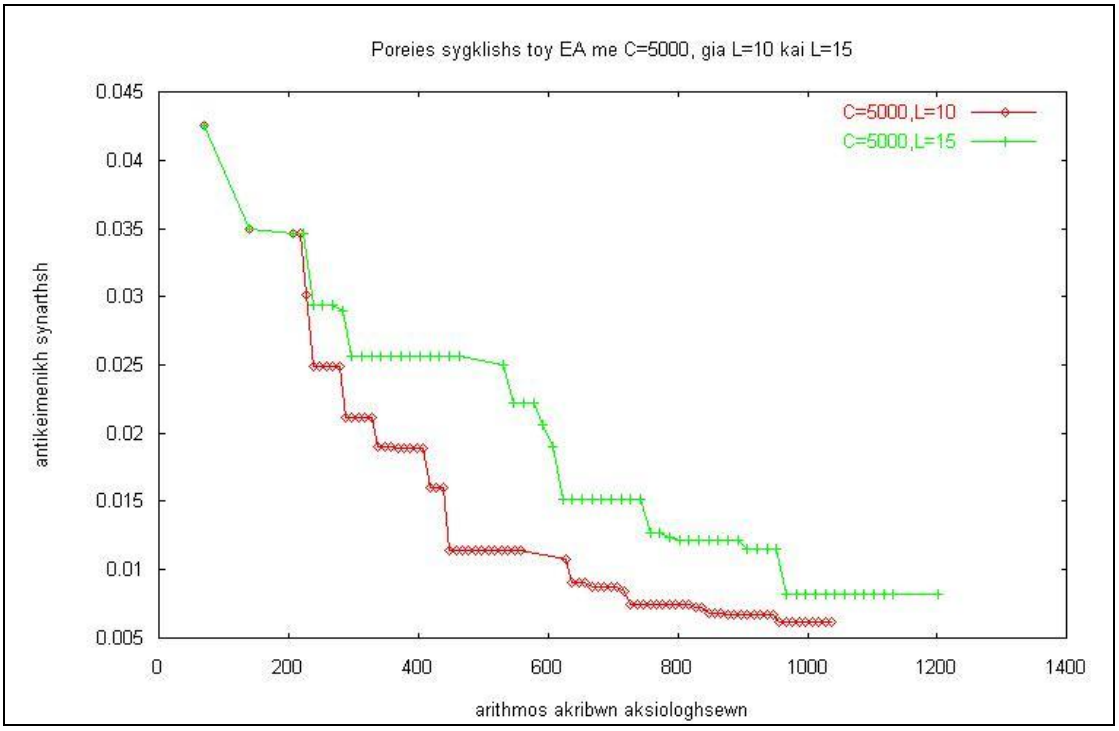
Σχήμα 26: Πορείες σύγκλισης του ΕΑ με C=5000, για K=10 και K=5.

Όπως και στην περίπτωση της συνάρτησης του Rastrigin, η μείωση του αριθμού ***K*** , δηλαδή της πληροφορίας που έχει ο SVM για τα «καλά» μέλη του συνόλου εκπαίδευσης, επιδρά αρνητικά στον «αυστηρό» διαχωριστή $C=\infty$ (βλ. σχήμα 27). Ωστόσο, στην προκειμένη περίπτωση, το ίδιο συμβαίνει και για τον πιο «ελαστικό» διαχωριστή (C=5000). Αυτό πιθανόν οφείλεται στο ότι το εν λόγω πρόβλημα ταξινόμησης είναι αρκετά σύνθετο (πολλές εμπλεκόμενες παράμετροι, άρα διαχωρισμός σε πολυδιάστατο χώρο) και η πληροφορία που παρέχει η τιμή ***K=5*** δεν είναι επαρκής.

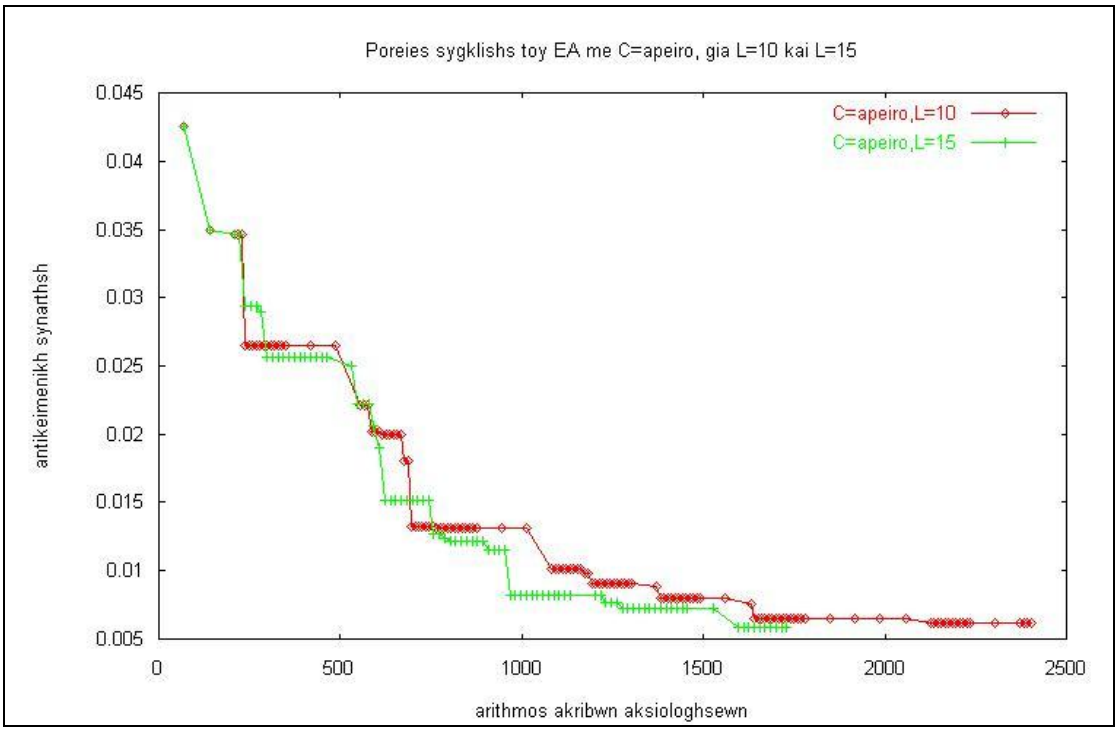


Σχήμα 27: Πορείες σύγκλισης του ΕΑ με $C=\infty$, για Κ=10 και Κ=5.

<u>*Περίπτωση (Δ):*</u> Κρατάμε τις υπόλοιπες παραμέτρους ίδιες, και αυξάνουμε τον αριθμό ***Λ*** των μελών κάθε γενιάς που υφίστανται ακριβή αξιολόγηση. Όπως φαίνεται και από τα σχήματα 28,29 ο «αυστηρός» διαχωριστής ευνοείται από τον εμπλουτισμό σε κάθε βήμα της βάσης δεδομένων με περισσότερη πληροφορία, ενώ το αντίθετο συμβαίνει με τον «ελαστικό» διαχωριστή. Η ίδια συμπεριφορά στους δύο διαχωριστές παρατηρήθηκε και στην πρώτη εφαρμογή που εξετάσαμε.



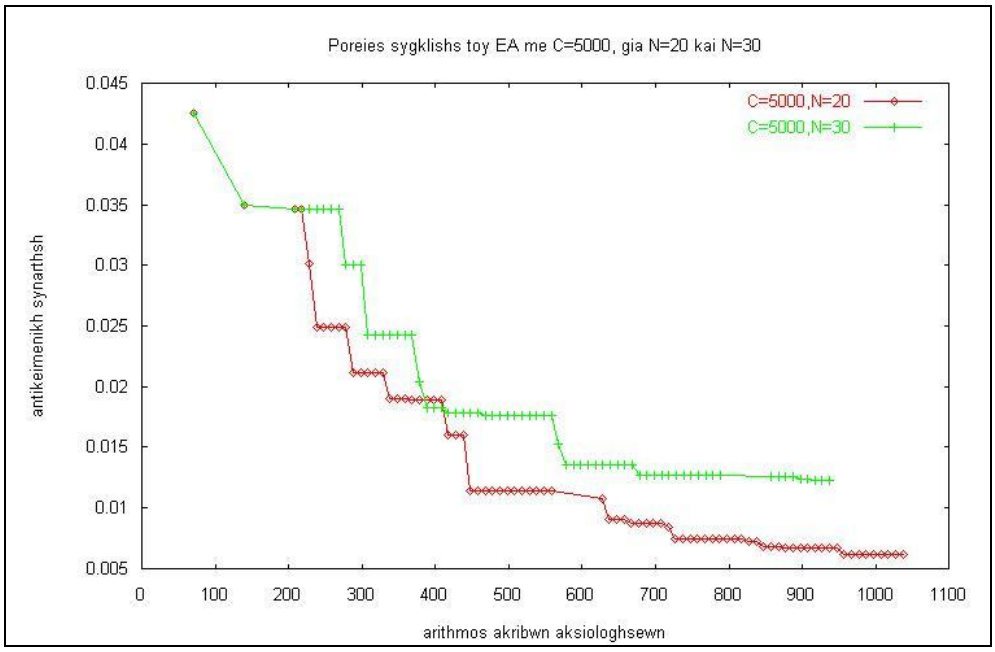
Σχήμα 28: Πορείες σύγκλισης του ΕΑ με C=5000, για Λ=10 και Λ=15.



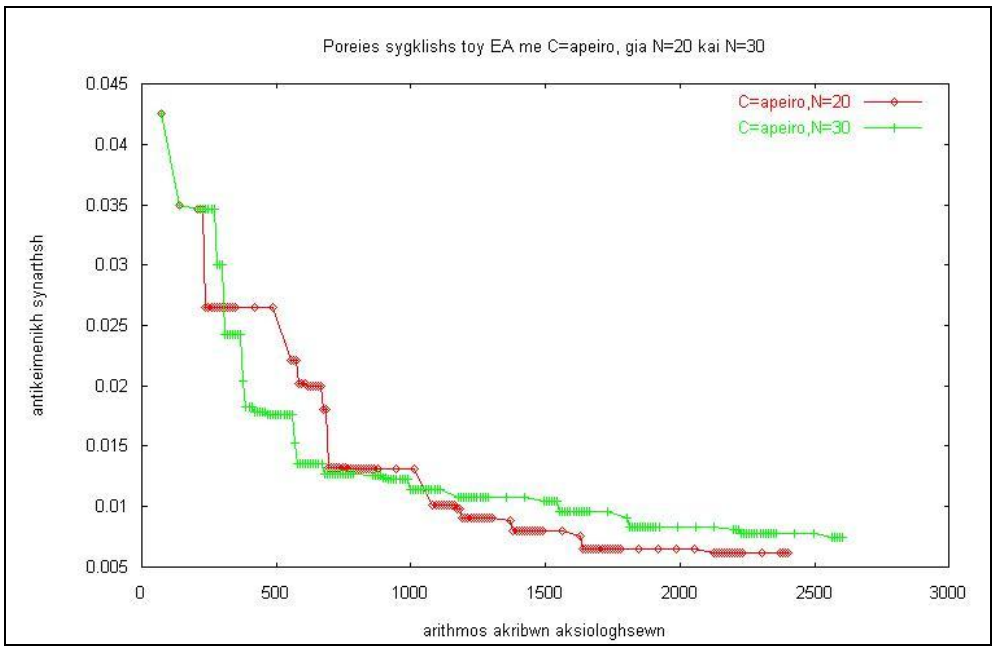
Σχήμα 29: Πορείες σύγκλισης του ΕΑ με $C=\infty$, για Λ=10 και Λ=15.

<u>*Περίπτωση (Ε)*</u>: Τέλος, αυξάνουμε το μέγεθος του συνόλου εκπαίδευσης των τοπικών SVM από ***N=20*** σε ***N=30.*** Οι αντίστοιχες πορείες σύγκλισης

των ΕΑ που χρησιμοποιούν τους δύο διαχωριστές φαίνονται στα σχήματα 30,31.



Σχήμα 30: Πορείες σύγκλισης του ΕΑ με C=5000, για N=20 και N=30



Σχήμα 31: Πορείες σύγκλισης του ΕΑ με $C=\infty$, για N=20 και N=30.

Παρατηρούμε ότι και οι δύο διαχωριστές εμφανίζουν χειρότερη απόδοση. Αυτό πιθανόν οφείλεται στο ότι μειώνεται ο λόγος ***K/N,*** που στην

προκειμένη περίπτωση φαίνεται να επιδρά με τον ίδιο τρόπο τόσο στον «ελαστικό» όσο και στον «αυστηρό» διαχωριστή.

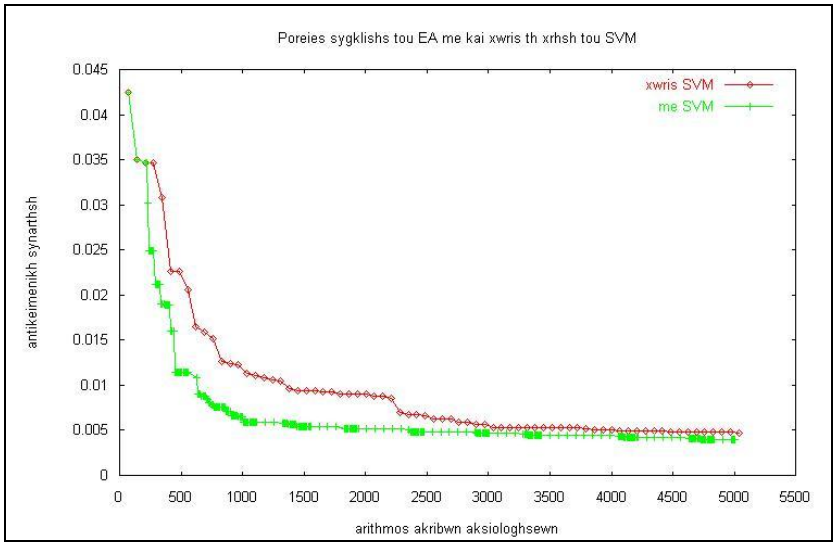
Καταλήγοντας, επιλέγουμε ***M=3*** όπου ο καλύτερος διαχωριστής είναι ο C=5000, και οι υπόλοιπες παράμετροι επιλέγονται με βάση την παραπάνω ανάλυση:

N=20 *(μέγεθος συνόλου εκπαίδευσης – αριθμός γειτόνων κάθε υπό εξέταση μέλους)*

K=10 *(αριθμός γειτόνων του συνόλου εκπαίδευσης που μαρκάρονται ως «καλοί»)*

Λ=10 *(αριθμός «καλών» μελών της τρέχουσας γενιάς που στέλνονται για ακριβή αξιολόγηση)*

Η αντίστοιχη πορεία σύγκλισης του ΕΑ, συγκρινόμενη με την πορεία σύγκλισης του ΕΑ που δεν χρησιμοποιεί SVM, φαίνεται στο σχήμα 32.



Σχήμα 32: Πορείες σύγκλισης του ΕΑ, με και χωρίς τη χρήση SVM, για το «βέλτιστο» συνδυασμό παραμέτρων που βρέθηκε μετά από συγκριτικές δοκιμές.

Ανακεφαλαίωση – Συμπεράσματα

Στην παρούσα εργασία διερευνήθηκε για πρώτη φορά μια μέθοδος ταξινόμησης δεδομένων με το όνομα Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – *SVM*). Η μέθοδος βασίζεται στη στατιστική θεωρίας εκμάθησης και εφαρμόζει την Αρχή Ελαχιστοποίησης Κατασκευαστικού Ρίσκου, εξασφαλίζοντας καλύτερη γενίκευση απ' ότι προγενέστερες μέθοδοι εκμάθησης. Ο αλγόριθμος της μεθόδου προγραμματίστηκε και εντάχθηκε στο λογισμικό βελτιστοποίησης του Εργαστηρίου Θερμικών Στροβιλομηχανών ΕΜΠ που βασίζεται σε Εξελικτικούς Αλγορίθμους, όπου χρησιμοποιήθηκε για την οικονομική προαξιολόγηση των υποψήφιων λύσεων - μελών μιας γενιάς του ΕΑ. Έγινε εφαρμογή της μεθόδου σε ένα δύσκολο πρόβλημα ταξινόμησης, όπου χρησιμοποιήθηκε αυτόνομα, καθώς και σε δύο προβλήματα ελαχιστοποίησης στα οποία χρησιμοποιήθηκε υποβοηθητικά στον Εξελικτικό Αλγόριθμο βελτιστοποίησης, σαν εργαλείο προεπιλογής υποψήφιων λύσεων. Τα αποτελέσματα που προέκυψαν και στις τρεις περιπτώσεις κρίνονται ως ιδιαίτερα ικανοποιητικά, δεδομένου ότι αποτελούσαν μια πρώτη προσπάθεια δοκιμής της αποτελεσματικότητας της μεθόδου σε δύσκολα προβλήματα. Από τις δοκιμές φάνηκε η σημασία που έχει η επιλογή κατάλληλων τιμών για τις ρυθμιστικές παραμέτρους της μεθόδου, οι οποίες καθορίζουν την ικανότητα γενίκευσης σε νέα σημεία. Έγινε δε μια πρώτη διερεύνηση των βέλτιστων τιμών των παραμέτρων αυτών καθώς και των παραγόντων που επηρεάζουν την εκλογή τους. Ειδικότερα, στις παραπάνω δοκιμές χρησιμοποιήθηκε κέλυφος ακτινικής βάσης με τυπική απόκλιση σ=0.9, επιλογή που έχει και στο παρελθόν (σε άλλες μεθόδους) δείξει καλά αποτελέσματα σε

αντίστοιχα προβλήματα. Η διερεύνηση εστιάστηκε στην τιμή του άνω ορίου C των πολλαπλασιαστών Lagrange, που επηρεάζει την ελαστικότητα του διαχωριστή σε σημεία παραπλάνησης, καθώς και στις παραμέτρους που καθορίζουν τη συνεργασία του *SVM* με τον ΕΑ βελτιστοποίησης. Στη μεγάλη πλειοψηφία των περιπτώσεων, η εφαρμογή του *SVM* υποβοηθητικά στον ΕΑ βελτιστοποίησης οδήγησε σε μείωση του απαιτούμενου αριθμού ακριβών αξιολογήσεων, επιτυγχάνοντας έτσι εξοικονόμηση υπολογιστικού χρόνου.

Αντικείμενο μελλοντικής έρευνας πρέπει να αποτελέσει η δοκιμή και άλλων τύπων κελύφους πέραν του κελύφους ακτινικής βάσης, καθώς και η εξέταση ενός πιο συστηματικού τρόπου επιλογής των βέλτιστων παραμέτρων της μεθόδου. Ενδιαφέρει επίσης η εφαρμογή του *SVM* σε προβλήματα αεροδυναμικής βελτιστοποίησης, όπου μπορεί να χρησιμοποιηθεί για προβλήματα τόσο ενός όσο και πολλών στόχων. Τέλος, υπάρχει έδαφος για την εφαρμογή της μεθόδου και σε προβλήματα παλινδρόμησης, όπου η τελευταία θα μπορεί να χρησιμοποιηθεί και σαν εργαλείο προ-αξιολόγησης (και όχι μόνο προεπιλογής) υποψήφιων λύσεων στον ΕΑ, υποκαθιστώντας τα μέχρι τώρα χρησιμοποιούμενα μεταπρότυπα παλινδρόμησης (ΤΝΔ, πολυωνυμική παρεμβολή κτλ.).

[BUR] G. Burel and D. Carel, "Detection and Localization of Faces on Digital Images," *Pattern Recognition Letters*,Vol. 15, 1994, pp. 963–967.

[EMM02] M. EMMERICH, A. GIOTIS, M. OZDEMIR, T. BACK, K.C. GIANNAKOGLOU: Metamodel-Assisted Evolution Strategies, 7th Intern. Conf. on Parallel Problem Solving from Nature (PPSN 2002), Granada, Spain, Sept. 7-11, 2002.

[GGP00] A.P. GIOTIS, K.C. GIANNAKOGLOU and J. PERIAUX: A Reduced-Cost Multi-Objective Optimization Method Based on the Pareto Front Technique, Neural Networks and PVM, ECCOMAS 2000, European Congress on Computational Methods in Applied Sciences and Engineering, Barcelona, Sept. 2000.

[GIA01] K.C. GIANNAKOGLOU: Optimization and Inverse Design in Aeronautics: How to Couple Genetic Algorithms with Radial Basis Function Networks, EURODAYS 2000, Innovative Tools for Scientific Computation in Aeronautical Engineering, Paris, March 2001.

[GIA03] Κ.Χ.Γιαννάκογλου. Μέθοδοι Βελτιστοποίησης στην Αεροδυναμική, ΕΜΠ Αθήνα 2003

[GIA99] K.C. GIANNAKOGLOU: Designing Turbomachinery Blades Using Evolutionary Methods, ASME Paper 99-GT-181, 44th ASME Gas Turbine & Aeroengine Congress, Indianapolis, IN, USA, June 7-10, 1999.

[GIO01] A. GIOTIS, M. EMMERICH, B. NAUJOCS, K.C. GIANNAKOGLOU, T.BACK, Low-Cost Stochastic Optimization for Engineering Applications, Int. Conference EUROGEN 2001 Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems, Athens, Greece, Sept. 19-21, 2001.

[GIO99] A.P. GIOTIS and K.C. GIANNAKOGLOU: Single- and Multi-Objective Airfoil Design Using Genetic Algorithms and Artificial Intelligence, EUROGEN 99, Evolutionary Algorithms in Engineering and Computer Science, May 1999

[JOA] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," published in *Proc.10th European Conf. Machine Learning*

[KAMP04] I.C. KAMPOLIS, D.I. PAPADIMITRIOU, K.C. GIANNAKOGLOU: Evolutionary Optimization Using a New

Radial Basis Function Network and the Adjoint Formulation, Inverse Problems, Design and Optimization (IPDO) Symposium, Rio de Janeiro, Brazil, March 17-19, 2004.

[KAR01] M.K. KARAKASIS, A.P. GIOTIS and K.C. GIANNAKOGLOU: Efficient Genetic Optimization Using Inexact Information and Sensitivity Analysis. Application in Shape Optimization Problems, ECCOMAS Computational Fluid Dynamics Conference 2001, Swansea, Sept. 2001.

[KAR04] M.K. KARAKASIS, K.C. GIANNAKOGLOU: On the Use of Surrogate Evaluation Models in Multi-Objective Evolutionary Algorithms, ECCOMAS 2004, 4th European Congress on Computational Methods in Applied Sciences and Engineering, Jyvaskyla, Finland, July 24-28, 2004.

[KGK05] E.A. KONTOLEONTOS, K.C. GIANNAKOGLOU, D.G. KOUBOGIANNIS: Robust Design of Compressor Cascade Airfoils Using Evolutionary Algorithms and Surrogate Models, 1rst International Conference on Experiments/Process/System Modelling/Simulation/Optimization, Athens, July 6-9, 2005.

[ROW] H. Rowley, S. Baluja, and T. Kanade, *Human Face Detection in Visual Scenes*, Tech. Report 95–158, Computer Science Dept., Carnegie Mellon Univ., Pittsburgh, 1995.

[YAN] G. Yang and T. Huang, "Human Face Detection in a Complex Background," *Pattern Recognition*,Vol. 27, 1994, pp. 53–63.

Βιβλιογραφία 2ου κεφαλαίου

[AIZ64] M.A.Aizerman, E.M.Braverman,L.I.Rozoner. Theoritical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control, 25:821-837, 1964*

[AKK] S.C.Ahn,G.Kim,M.Kim. A note on applications of Support Vector Machines

[BAR01] P.Bartlett. Statistical Learning and VC Theory, ISCAS, May 2001

[BEN] K.P.Bennett,C.Campbell. Support Vector Machines Hype or Hallelyjah?

[BOUG] S.Boughorbel,J.P.Tarel,N.Boujemaa. The LCCP for Optimizing Kernel Parameters for SVM

[BOZ] B.E.Bozer,I.M.Guyon,V.Vapnik. A training algorithm for optimal margin classifiers. In *Fifth Annual Workshop on Computational Learning Theory,* Pittsburgh, 1992. ACM.

[BUR98] C.J.C.Burges. A Tutorial on Support Vector Machines for Pattern Recognition, 1998

[CAW] G.C.Cawley. Model Selection for Support Vector Machines via Adaptive Step-Size Tabu Search

[CHE98] V.Cherkassky,F.Mulier. Learning From Data, John Wiley&Sons,Inc., 1998

[CHELSC] P.H.Chen,C.J.Lin,B.Scholkopf. A Tutorial on ν-Support Vector Machines

[CHEMA] V.Cherkassky,Y.Ma. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression

[CHR] N.Christianni,U.C.Davis. Kernel Methods for Pattern Analysis

[COR] C.Cortes,V.Vapnik. Support Vector Networks. Machine Learning, 20:273-297, 1995

[COU] R.Courant,D.Hilbert. Methods of Mathematical Physics. Interscience, 1953.

[EAD] D.Eads,D.Hill,S.Davis,S.Perkins,J.Ma,R.Porter,J.Theiler. Genetic Algorithms and Support Vector Machines for Time Series Classification

[FAR04] A.Farag,R.M.Mohamed. Regression using Support Vector Machines: Basic Foundations, Technical Report,December 2004

[FLE87] R.Fletcher. Practical Methods of Optimization, John Wiley & SonsInc., 2nd edition, 1987

[GUN98] S.R.Gunn. Support Vector Machines for Classification and Regression, Technical Report, May 1998

[HEA98] M.A.Hearst,S.T.Dumais,E.Osuna,J.Platt,B.Scholkopf. Support Vector Machines, Trends & Controversies July/August 1998

[HSU] C.W.Hsu,C.C.Chang,C.J.Lin. A Practical Guide to Support Vector Classification

[KEE] S.S.Keerthi,C.-J.Lin. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation 15(7), 1667-1689*

[MAR] M.Martin. On-line Support Vector Machines for Function Approximation

[MOO] A.W.Moore. Cross-Validation for detecting and preventing overfitting

[MOO01] A.W.Moore. VC-dimension for characterizing classifiers,November 2001

[PLAa] J.C.Platt. Fast Training of Support Vector Machines using Sequential Minimal Optimization

[PLAb] J.C.Platt. Using Analytic QP and Sparseness to Speed Training of Support Vector Machines

[SBS98] B.Scholkopf,C.J.Burges,A.J.Smola. Introduction to Support Vector Machines,Berlin,Holmdel,July 1998

[SCH00] B.Scholkopf. Statistical Learning and Kernel Methods, February 2000

[SCHI] K.Schittkowski. Optimal Parameter Selection in Support Vector Machines

[SMO96] A.Smola et al. Regression Estimation with Support Vector Machines,Version 1.01,December 1996

[SMO98] A.J.Smola,B.Scholkopf. A Tutorial on Support Vector Regression, NeuroCOLT2 Technical Report Series, October 1998

[VAGSM] V.Vapnik,S.E.Golowich,A.Smola. Support Vector Method for Function Approximation, Regression Estimation and Signal Processing

[VAP95] V.N.Vapnik. The Nature of Statistical Learning Theory, John Wiley&Sons,Inc.,1995

[VAP98] V.N.Vapnik. Statistical Learning Theory, John Wiley&Sons,Inc., 1998