Γραμμικοί Ταξινομητές - Μηχανές Διανυσματικής Στήριξης

ECE-TEL830 ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

Αθανάσιος Κούτρας
Αναπληρωτής Καθηγητής

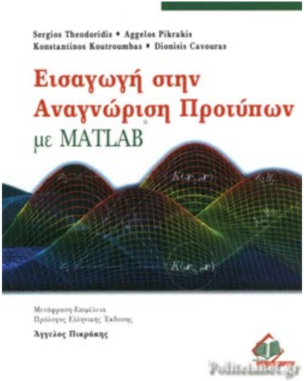
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών,
Παν. Πελοποννήσου

9 Απριλίου 2023

Περίγραμμα διάλεξης

Υλικό μελέτης

Theodoridis S., Pikrakis, A., Koutroumbas K., Cavouras, D., **"Εισαγωγη στην αναγνώριση προτύπων με MATLAB",ΚΕΦΑΛΑΙΟ** 2



Εισαγωγή

Στη διάλεξη αυτή θα παρουσιαστεί μια διαφορετική λογική για το σχεδιασμό γραμμικών ταξινομητών ξεκινώντας από τον διαχωρισμό δύο κλάσεων και προχωρώντας σε περισσότερες.

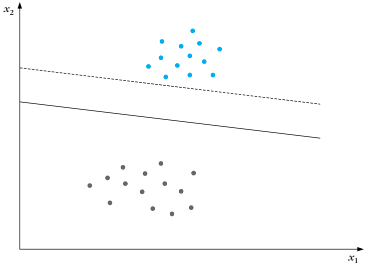
Μηχανές Διανυσματικής Στήριξης (Support Vector Machines - SVM)

- Για τον γραμμικό ταξινομητή θέλουμε να σχεδιάσουμε ένα υπερεπίπεδο:

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$$

  το οποίο θα ταξινομεί σωστά τα διανύσματα εκπαίδευσης.
- όπως είδαμε στην προηγούμενη διάλεξη, αυτό το επίπεδο δεν είναι μοναδικό.
- στην προσέγγιση που θα ακολουθήσουμε, θα επιβάλουμε μεγαλύτερους περιορισμούς.

- στο σχήμα αυτό παρουσιάζεται το πρόβλημα ταξινόμησης με δύο δυνατά υπερεπίπεδα τα οποία δουλεύουν σωστά στο σύνολο των δεδομένων μας.
- Ποιο όμως από τα δύο θα επιλέγαμε στην περίπτωση που θα είχαμε και δεδομένα εκτός του συνόλου εκπαίδευσης;
- η προφανής λύση θα ήταν αυτό με τη συμπαγή γραμμή. Γιατί;
- διότι αφήνει περισσότερο χώρο σε κάθε πλευρά του για να κινηθούν πιο ελεύθερα τα δεδομένα και από τις δύο κατηγορίες.



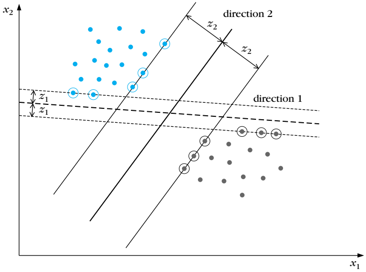
Σχήμα: Παράδειγμα δύο γραμμικά διαχωρίσιμων κλάσεων με δύο πιθανούς γραμμικούς ταξινομητές

Η απόδοση γενίκευσης

- η επιλογή που παρουσιάσαμε θίγει ένα σημαντικό πρόβλημα στην αναγνώριση προτύπων, αυτό της απόδοσης γενίκευσης του ταξινομητή (generalization performance).
- αυτή αναφέρεται στην ικανότητα του ταξινομητή ο οποίος έχει σχεδιαστεί χρησιμοποιώντας το σύνολο των δεδομένων εκπαίδευσης, να λειτουργεί ικανοποιητικά με δεδομένα και εκτός συνόλου.
- το ζητούμενο λοιπόν είναι να μεγιστοποιήσουμε το περιθώριο που αφήνει ο ταξινομητής και από τις δύο κλάσεις.

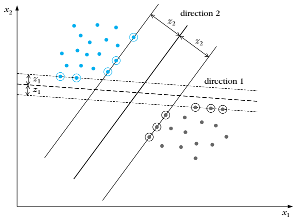
Η έννοια του περιθωρίου

- για τον σχεδιασμό του βέλτιστου υπερ-επιπέδου, θα πρέπει να ποσοτικοποιήσουμε το περιθώριο (margin) που αφήνει αυτό από κάθε κλάση.
- γνωρίζουμε ότι το υπερεπίπεδο ορίζεται από την διεύθυνση του ($\mathbf{w}$) και την θέση του στον χώρο (w_0)
- για να μην δείξουμε προτίμηση σε κάποια από τις δύο κλάσεις, υποθέτουμε ότι το υπερεπίπεδο που ψάχνουμε ισαπέχει από τα πλησιέστερα σε αυτό σημεία των ω_1 και ω_2



Σχήμα: Παράδειγμα δύο γραμμικά διαχωρίσιμων κλάσεων με δύο δυνατούς γραμμικούς ταξινομητές

- τα υπερεπίπεδα που έχουν σχεδιαστεί με σκούρες γραμμές είναι εκείνα που έχουν επιλεγεί στην αντίστοιχη διεύθυνση
- το περιθώριο για την "διεύθυνση 1", είναι το $2 \cdot z_1$ ενώ για την "διεύθυνση 2", είναι το $2 \cdot z_2$
- το ζητούμενο είναι να ψάξουμε για την διεύθυνση που μας δίνει το μέγιστο δυνατό περιθώριο.
- επειδή κάθε υπερεπίπεδο καθορίζεται από έναν συντελεστή κλίμακας, προηγείται μια κλιμάκωση ώστε όλα να είναι συγκρίσιμα μεταξύ τους.
- η κλιμάκωση των $\mathbf{w}$ και w_0 έχει ως αποτέλεσμα η συνάρτηση $g(\mathbf{x})$ στα πλησιέστερα σημεία των ω_1, ω_2 να είναι ίση με $+1$ για την ω_1 και -1 για την ω_2



Σχήμα: Παράδειγμα δύο γραμμικά διαχωρίσιμων κλάσεων με δύο δυνατούς γραμμικούς ταξινομητές

- τα παραπάνω ισοδυναμούν με τις ακόλουθες σχέσεις που πρέπει να ικανοποιούνται:
- θέλουμε να έχουμε ένα περιθώριο για το οποίο:

$$\frac{\mathbf{1}}{||\mathbf{w}||} + \frac{\mathbf{1}}{||\mathbf{w}||} = \frac{\mathbf{2}}{||\mathbf{w}||}$$

  και να απαιτούμε να είναι:

$$\mathbf{w}^T\mathbf{x} + w_0 \geq 1, \forall \mathbf{x} \in \omega_1$$

$$\mathbf{w}^T\mathbf{x} + w_0 \leq -1, \forall \mathbf{x} \in \omega_2$$

- ο γραμμικός SVM ταξινομητής δίνει την λύση για το παραπάνω επιλύοντας το πρόβλημα ελαχιστοποίησης υπό συνθήκη:
  - να ελαχιστοποιείται η συνάρτηση $J(\mathbf{w}, w_0) = \frac{1}{2}||\mathbf{w}||^2$
  - υπό τους περιορισμούς $y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1, i = 1, 2, \ldots, N$
- η λύση του παραπάνω προβλήματος, μας δίνει τις παραμέτρους $\mathbf{w}$ και w_0 του υπερεπιπέδου.

Επίλυση του προβλήματος βελτιστοποίησης

- η ελαχιστοποίηση της συνάρτησης με τους περιορισμούς που εξετάστηκαν, οδηγεί σε ένα μη γραμμικό (τετραγωνικό) πρόβλημα βελτιστοποίησης.
- η λύση του προβλήματος είναι η:

$$\mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i$$

$$\sum_{i=1}^{N} \lambda_i y_i = 0$$

όπου λ_i είναι οι πολλαπλασιαστές Lagrange (μπορεί να είναι είτε θετικοί, είτε μηδέν)

- το διάνυσμα παραμέτρων $\mathbf{w}$ είναι ένας γραμμικός συνδυασμός από $N_s \leq N$ διανύσματα χαρακτηριστικών τα οποία σχετίζονται με μη μηδενικούς συντελεστές λ_i
- τα διανύσματα αυτά καλούναι διανύσματα στήριξης (support vectors) και ο βέλτιστος ταξινομητής (υπερεπίπεδο) είναι γνωστός ως μηχανή διανυσματικής στήριξης (support vector machine - SVM)

Μηχανές Διανυσματικής Στήριξης - SVM

- από τους περιορισμούς του προβλήματος προκύπτει ότι τα διανύσματα στήριξης βρίσκονται σε ένα από τα δύο υπερεπίπεδα

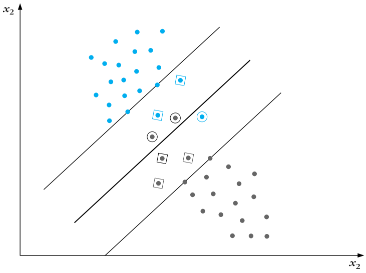
$$\mathbf{w}^T\mathbf{x} + w_0 = \pm 1$$

- είναι τα διανύσματα εκπαίδευσης τα οποία βρίσκονται πιο κοντά στον γραμμικό ταξινομητή και αποτελούν τα κρίσιμα στοιχεία του συνόλου εκπαίδευσης
- το βέλτιστο υπερεπίπεδο που υπολογίζεται είναι **μοναδικό**
- αν και η λύση που προκύπτει είναι μοναδική, εντούτοις οι πολλαπλασιαστές Lagrange που υπολογίζονται δεν είναι

Μη γραμμικά διαχωρίσιμες κλάσεις

- όταν οι κλάσεις δεν είναι γραμμικά διαχωρίσιμες, τα προηγούμενα δεν ισχύουν.
- στο παρακάτω σχήμα, ότι και να κάνουμε, δεν μπορούμε να βρούμε μια ζώνη διαχωρισμού κλάσεων χωρίς σημεία στο εσωτερικό της
- γνωρίζουμε ότι το περιθώριο ορίστηκε ως η απόσταση ανάμεσα στο ζεύγος των παράλληλων υπερεπιπέδων τα οποία περιγράφονται από τις

$$\mathbf{w}^T\mathbf{x} + w_0 = \pm 1$$



Σχήμα: περίπτωση μη διαχωρίσιμων κλάσεων

Μη διαχωρίσιμες κλάσεις

- τα παραδείγματα εκπαίδευσης ανήκουν σε μία από τις τρεις περιπτώσεις:
  - διανύσματα που βρίσκονται εκτός ζώνης και είναι σωστά ταξινομημένα

$$y_i(\mathbf{w}^T\mathbf{x} + w_0) > 1$$

  - διανύσματα που βρίσκονται εντός ζώνης και είναι σωστά ταξινομημένα

$$0 \leq y_i(\mathbf{w}^T\mathbf{x} + w_0) < 1$$

  - διανύσματα που ταξινομούνται εσφαλμένα

$$y_i(\mathbf{w}^T\mathbf{x} + w_0) < 0$$

- οι παραπάνω τρεις περιπτώσεις μπορούν να αναπαρασταθούν από κοινού ως

$$y_i(w^Tx + w_0) \geq \xi_i$$

$$\xi_i = 0$$

$$0 < \xi_i \leq 1$$

$$\xi_i > 1$$

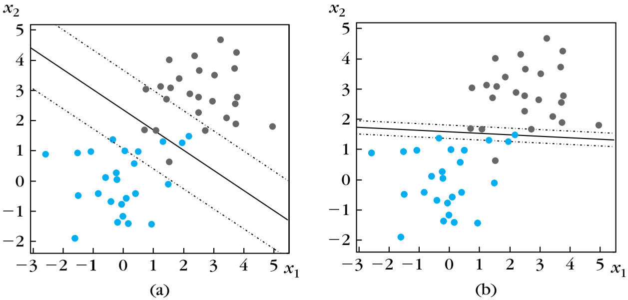
με την παράμετρο ξ_i γνωστή ως **μεταβλητή χαλαρότητας** (slack variable)

Μη διαχωρίσιμες κλάσεις

- το πρόβλημα που καλούμαστε να λύσουμε τώρα είναι πιο πολύπλοκο αλλά στην ίδια λογική:
  - θέλουμε να κάνουμε το **περιθώριο όσο το δυνατό πιο μεγάλο**
  - την ίδια στιγμή να κρατήσουμε **το πλήθος των σημείων** για τα οποία ισχύει $\xi_i > 0$ όσο το **δυνατό πιο μικρό**
- η μόνη διαφορά που προκύπτει από την επίλυση του προβλήματος βελτιστοποίησης σε σχέση με τις διαχωρίσιμες κλάσεις, είναι η ύπαρξη της παραμέτρου C στους περιορισμούς,

Παράδειγμα μη γραμμικώς διαχωρίσιμων κλάσεων

το πλάτος του περιθωρίου δεν εξαρτάται ολοκληρωτικά από την κατανομή των δεδομένων όπως στις γραμμικά διαχωρίσιμες κλάσεις, αλλά επηρεάζεται από την επιλογή του C



Σχήμα: Παράδειγμα δύο μη γραμμικώς διαχωρίσιμων κλάσεων με τον SVM γραμμικό ταξινομητή (συμπαγής γραμμή) και τα σχετικά περιθώρια (διακεκομμένες γραμμές) για (α) $C = 0.2$ και (β) $C = 1000$. Στη δεύτερη περίπτωση η θέση και η διεύθυνση του ταξινομητή καθώς και το πλάτος του περιθωρίου έχουν αλλάξει προκειμένου να περιέχονται λιγότερα σημεία μέσα στο περιθώριο.

Η Περίπτωση των πολλαπλών κλάσεων

- για να λύσουμε το πρόβλημα των M κλάσεων, μια απλή επέκταση είναι να το θεωρήσουμε ως ένα σύνολο M προβλημάτων δύο κλάσεων (κάθε μια έναντι των υπόλοιπων - one-against-all).
- για κάθε μια από τις κλάσεις θέλουμε να σχεδιάσουμε μια βέλτιστη συνάρτηση διάκρισης $g_i(\mathbf{x}), i = 1, 2, \ldots, M$ ώστε $g_i(\mathbf{x}) > g_j(\mathbf{x}), \forall j \neq i, x \in \omega_i$
- η παραπάνω τεχνική μπορεί να οδηγήσει σε απροσδιόριστες περιοχές
- επιπλέον ο κάθε ταξινομητής ασχολείται με ένα ασύμμετρο πρόβλημα όπου η εκπαίδευση γίνεται με περισσότερα αρνητικά παραδείγματα, παρά θετικά (εντονότερο στην περίπτωση περισσότερων κλάσεων).

Η Περίπτωση πολλών κλάσεων

- για την αντιμετώπιση των παραπάνω, εναλλακτικά χρησιμοποιούμε την τεχνική μία-έναντι-μίας (one-against-one).
- σε αυτή την περίπτωση εκπαιδεύονται $M(M-1)/2$ δυαδικοί ταξινομητές και κάθε ένας από αυτούς διαχωρίζει ένα ζεύγος κλάσεων.
- η τελική απόφαση βγαίνει με πλειοψηφία.
- βασικό μειονέκτημα της μεθόδου αυτής είναι ότι πρέπει να εκπαιδευτεί σε ένα μεγάλο πλήθος δυαδικών ταξινομητών

Signal & Image Processing, Pattern Recognition Group (SIPPRE)
www.sippre-group.com