Μη Γραμμικοί Ταξινομητές - Το δίκτυο Ακτινωτής Συνάρτησης Βάσης (RBF)

ECE-TEL830 ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

Αθανάσιος Κούτρας

Αναπληρωτής Καθηγητής

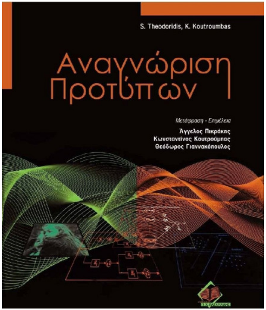
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Παν. Πελοποννήσου

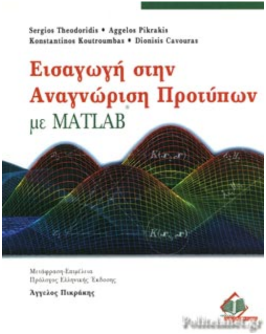
14 Μαΐου 2023

Περίγραμμα διάλεξης

Υλικό μελέτης

- Theodoridis S., Pikrakis, A., Koutroumbas K., Cavouras, D., **"Εισαγωγη στην αναγνώριση προτύπων με MATLAB",ΚΕΦΑΛΑΙΟ** 2
- Theodoridis S., Koutroumbas K., **"Αναγνώριση προτύπων",ΚΕΦΑΛΑΙΟ** 4





Εισαγωγή

- Στη διάλεξη αυτή θα παρουσιαστεί μια διαφορετική τοπολογία δικτύων τα οποία μπορούν να εφαρμοστούν και να λειτουργήσουν καλύτερα σε προβλήματα μη γραμμικά διαχωρίσιμων κλάσεων, όπως και τα πολυεπίπεδα δίκτυα perceptron που παρουσιάστηκαν σε προηγούμενη ενότητα.
- Συγκεκριμένα θα παρουσιάσουμε την περίπτωση των ταξινομητών που χρησιμοποιούν **δίκτυα ακτινωτής συνάρτησης βάσης** (Radial Basis Functions).

Γενικευμένοι Γραμμικοί Ταξινομητές

- στην προηγούμενη ενότητα ασχοληθήκαμε με την επίλυση του προβλήματος XOR το οποίο είναι μη-γραμμικά διαχωρίσιμο πρόβλημα.
- με τη χρήση του πολυεπίπεδου δικτύου (δύο επιπέδων) perceptron, καταφέραμε να μετατρεψουμε το πρόβλημα από μη-γραμμικά διαχωρίσιμο, σε γραμμικά διαχωρίσιμο.
- αυτό το καταφέραμε μέσω μιας **απεικόνισης** που δημιούργησαν οι νευρώνες του κρυφού επιπέδου των παραμέτρων της εισόδου $\mathbf{x}$ σε κάποιες μετασχηματισμένες παραμέτρους $\mathbf{y}$.
- ο μετασχηματισμός αυτός προέκυψε μέσω της **συνάρτησης ενεργοποίησης** των νευρώνων του κρυφού επιπέδου και κατόπιν του γραμμικού συνδυασμού των εισόδων που εφαρμόζονται σε κάθε νευρώνα.
- στην περίπτωση των γενικευμένων γραμμικών ταξινομητών, θέλουμε να μετασχηματίσουμε τα δεδομένα μας από έναν αρχικό χώρο R^l σε έναν χώρο διαφορετικής διάστασης R^k μέσω ενός συνόλου k μη γραμμικών συναρτήσεων $f_1(\cdot), f_2(\cdot), \ldots, f_k(\cdot)$ ώστε να έχουμε την απεικόνιση

$$y \equiv \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_k(\mathbf{x}) \end{bmatrix}$$

Γενικευμένοι Γραμμικοί Ταξινομητές

- ο σκοπός μας είναι να διερευνήσουμε αν υπάρχει μια κατάλληλη τιμή για το k και τις συναρτήσεις f_i έτσι ώστε οι κλάσεις που θέλουμε να ταξινομήσουμε, να γίνουν γραμμικά διαχωρίσιμες στον $k-$διάστατο χώρο.
- στον χώρο αυτό μετά, μπορούμε να κατασκευάσουμε ένα υπερεπίπεδο $\mathbf{w} \in R^k$ ώστε

$$w_0 + \mathbf{w}\mathbf{y}^T > 0, \mathbf{x} \in A$$

$$w_0 + \mathbf{w}\mathbf{y}^T < 0, \mathbf{x} \in B$$

- αν στον αρχικό χώρο οι δύο κλάσεις ήταν διαχωρίσιμες από μια (μη γραμμική) υπερεπιφάνεια $g(\mathbf{x}) = 0$, τότε μπορούμε να γράψουμε την προσέγγιση της μη γραμμικής $g(\mathbf{x})$ από έναν γραμμικό συνδυασμό των $f_i(\mathbf{x})$:

$$\boxed{g(x) = w_o + \sum_{i=1}^{k} w_i f_i(\mathbf{x})}$$

- το παραπάνω είναι ένα τυπικό πρόβλημα προσέγγισης συνάρτησης από μια προεπιλεγμένη κλάση **συναρτήσεων παρεμβολής** $f_i(\cdot)$

Δίκτυα Ακτινωτής Συνάρτησης Βάσης

- οι **συναρτήσεις παρεμβολής** (πυρήνες - kernels) που θα χρησιμοποιηθούν εδώ είναι της γενικής μορφής

$$f(||\mathbf{x} - \mathbf{c}_i||)$$

- το όρισμα της συνάρτησης είναι η Ευκλείδεια απόσταση του διανύσματος εισόδου από ένα κέντρο $\mathbf{c}_i$ κάτι που δικαιολογεί και το όνομα *συνάρτηση ακτινωτής βάσης*
- η συνάρτηση αυτή μπορεί να πάρει διάφορες τιμές:

$$\boxed{f(\mathbf{x}) = exp\left(-\frac{1}{2\sigma^2}||\mathbf{x} - \mathbf{c}_i||^2\right)}$$

$$\boxed{f(\mathbf{x}) = \frac{\sigma^2}{\sigma^2 + ||\mathbf{x} - \mathbf{c}_i||^2}}$$

Προσέγγιση μέσω αθροίσματος συναρτήσεων ακτινωτής βάσης

- από τις προηγούμενες συναρτήσεις, η πιο συχνά χρησιμοποιούμενη είναι η Gaussian. Για μεγάλες τιμές του k η προσέγγιση της μη-γραμμικής συνάρτησης μπορεί να θεωρηθεί ίση με:

$$\boxed{g(\mathbf{x}) = w_o + \sum_{i=1}^{k} w_i exp\left(-\frac{(\mathbf{x}-\mathbf{c}_i)^T(\mathbf{x}-\mathbf{c}_i)}{2\sigma_i^2}\right)}$$

- η παραπάνω εξίσωση μπορεί να ερμηνευτεί ως η έξοδος ενός δικτύου με ένα κρυφό επίπεδο που αποτελείται από κόμβους με RBF συναρτήσεις ενεργοποίησης και έναν γραμμικό κόμβο εξόδου.

Διαφορές πολυεπίπεδων δικτύων perceptron - RBF δικτύων

- στα δίκτυα **perceptron**, οι **είσοδοι** των συναρτήσεων ενεργοποίησης είναι **γραμμικοί συνδυασμοί** των παραμέτρων των χαρακτηριστικών εισόδου.
- η **έξοδος** κάθε νευρώνα είναι **ίδια** για όλα τα $\mathbf{x}$
- στα δίκτυα **RBF**, η **έξοδος** κάθε κόμβου είναι **ίδια** για όλα τα **σημεία που έχουν την ίδια Ευκλείδεια απόσταση** από το αντίστοιχο κέντρο και μειώνεται εκθετικά ως προς την απόσταση (σε περίπτωση Gaussian kernel)
- για τα **RBF** δίκτυα, οι αποκρίσεις ενεργοποίησης των κόμβων είναι **τοπικού** χαρακτήρα
- για τα πολυεπίπεδα **perceptron**, οι αποκρίσεις ενεργοποίησης των κόμβων είναι **καθολικού** χαρακτήρα.

Διαφορές πολυεπίπεδων δικτύων perceptron - RBF δικτύων

Ταχύτητα σύγκλισης

- τα δίκτυα **perceptron** μαθαίνουν πιο **αργά** σε σχέση με τα RBF

Ιδιότητες Γενίκευσης

- τα δίκτυα **perceptron** παρουσιάζουν **γενικευμένες** ιδιότητες γενίκευσης, ειδικά για περιοχές που δεν αντιπροσωπεύονται επαρκώς στο σύνολο εκπαίδευσης

Παράδειγμα χρήσης ταξινομητή RBF

- Έστω το γνωστό πρόβλημα ταξινόμησης XOR. Για την επίλυση του επιλέγουμε 2 κέντρα τα $c_1 = [1, 1]$ και $c_2 = [0, 0]$ καθώς και $f(x) = exp(-||x - c_i||^2)$
- εφαρμόζοντας την απεικόνιση προκύπτουν οι μετασχηματισμοί των αρχικών σημείων:

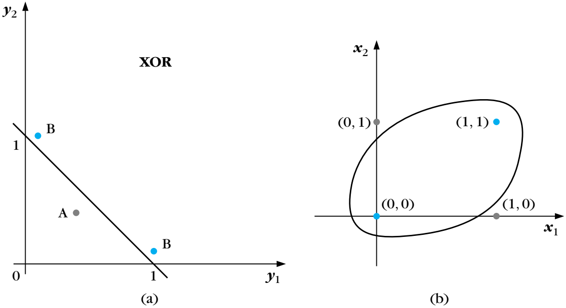
$$(0,0) \rightarrow (0.135, 1), (1,1) \rightarrow (1, 0.135),$$

$$(1,0) \rightarrow (0.368, 0.368), (0,1) \rightarrow (0.368, 0.368)$$

- το γραμμικά διαχωρίσιμο πρόβλημα, μπορεί να επιλυθεί με την $g(y) = y_1 + y_2 - 1$ και την ισοδύναμη καμπύλη απόφασης:

$$\boxed{g(x) = exp(-||x - c_1||^2) + exp(-||x - c_2||^2) - 1 = 0}$$

Το παράδειγμα XOR



Σχήμα: Καμπύλες απόφασης που σχηματίζονται από έναν RBF γενικευμένο γραμμικό ταξινομητή για το πρόβλημα XOR. Η ακμπύλη απόφασης είναι γραμμική στον μετασχηματισμένο χώρο και μη γραμμική στον αρχικό χώρο.

Επιλογή των κέντρων

- η επιλογή των κέντρων μπορεί να γίνει με **τυχαίο τρόπο** από το σύνολο της εκπαίδευσης
- αν το σύνολο της εκπαίδευσης είναι κατανεμημένο με αντιπροσωπευτικό τρόπο σε όλο τον χώρο διανυσμάτων χαρακτηριστικών, η τεχαία επιλογή φαίνεται ένας λογικός τρόπος επιλογής.
- ένας δεύτερος τρόπος είναι **μέσω της εκπαίδευσης** μαζί με τα βάρη w και τις διασπορές σ_i^2
- επειδή το παραπάνω είναι πολύ υπολογιστικά πολύπλοκο, χρησιμοποιείται η **τεχνική της ομαδοποίησης (clustering)** των δεδομένων και την επιλογή ενός αντιπροσώπου για κάθε ομάδα ως κέντρο.

Signal & Image Processing, Pattern Recognition Group (SIPPRE)

www.sippre-group.com