Ταξινομητές βασισμένοι στη θεωρία αποφάσεων κατά Bayes

ECE-TEL830 ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

Αθανάσιος Κούτρας

Αναπληρωτής Καθηγητής

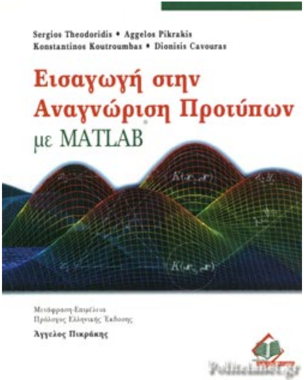
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών,
Παν. Πελοποννήσου

13 Μαρτίου 2023

Περίγραμμα διάλεξης

Υλικό μελέτης

Theodoridis S., Pikrakis, A., Koutroumbas K., Cavouras, D., **"Εισαγωγη στην αναγνώριση προτύπων με MATLAB"**,**ΚΕΦΑΛΑΙΟ** 1



Εισαγωγή

- Οι τεχνικές που θα παρουσιαστούν εδώ έχουν εμπνευστεί από τη θεωρία αποφάσεων του Bayes.
- Στα προβλήματα ταξινόμησης θέλουμε να ταξινομήσουμε ένα πρότυπο (pattern) σε μία από περισσότερες δυνατές κλάσεις c.
- για αρχή, θεωρούμε ότι ο αριθμός των κλάσεων είναι γνωστός εκ των προτέρων σε εμάς.
- Το κάθε πρότυπο το αναπαραστούμε με ένα σύνολο τιμών $x(i), i = 1, 2, , \ldots, l$ οι οποίες δημιουργούν ένα διάνυσμα l διαστάσεων το οποίο καλείται διάνυσμα χαρακτηριστικών

$$x = [x(1), x(2), \ldots, x(l)]^T$$

- υποθέτουμε ότι κάθε πρότυπο αντιπροσωπεύεται με μοναδικό τρόπο από ένα και μόνο διάνυσμα χαρακτηριστικών και μπορεί να ανήκει μόνο σε μία κλάση.

Θεώρημα του Bayes

- δοθέντος ενός διανύσματος x και ενός συνόλου c κλάσεων $\omega_i, i = 1, 2, \ldots, c$, σύμφωνα με το θεώρημα του Bayes ισχύει ότι:

$$P(\omega_i|x)p(x) = p(x|\omega_i)P(\omega_i)$$

και

$$p(x) = \sum_{i=1}^{c} p(x|\omega_i)P(\omega_i)$$

όπου:

- $P(\omega_i)$ είναι η εκ των προτέρων πιθανότητα της κλάσης $\omega_i, i = 1, 2, \ldots, c$
- $P(\omega_i|x)$ είναι η εκ των υστέρων πιθανότητα της κλάσης ω_i δοθέντος του x,
- $p(x)$ είναι η συνάρτηση πυκνότητας πιθανότητας του x
- $p(x|\omega_i)$ είναι η συνάρτηση της υπό συνθήκη πυκνότητας πιθανότητας του x δοθείσης της κλάσης ω_i (καλείται και πιθανοφάνεια της ω_i ως προς το x).

Θεωρία αποφάσεων του Bayes

- Έστω ότι δίνεται ένα πρότυπο αγνωστης κλάσης και έστω $x = [x(1), x(2), \ldots, x(l)]^T$ το αντίστοιχο διάνυσμα χαρακτηριστικών.
- το διάνυσμα χαρακτηριστικών προκύπτει από μετρήσεις που μπορούμε να πραγματοποιήσουμε
- θεωρούμε για το πρόβλημα μας ότι υπάρχουν c δυνατές κλάσεις στις οποίες μπορεί να ταξινομηθεί το πρότυπο άγνωστης κλάσης, οι $\omega_1, \omega_2, \ldots, \omega_c$
- Σύμφωνα με την θεωρία του Bayes, η απόφαση που μπορούμε να πάρουμε για την ταξινόμηση του άγνωστου προτύπου στην κλάση ω_i προκύπτει μετά την εξέταση της παρακάτω σχέσης:

$$P(\omega_i|x) > P(\omega_j|x) \forall j \neq i$$

- από την θεωρία πιθανοτήτων προκύπτει ότι ένας ταξινομητής Bayes είναι βέλτιστος υπό την έννοια ότι ελαχιστοποιεί την πιθανότητα σφάλματος.

Η Gaussian συνάρτηση πυκνότητας πιθανότητας

- η συνάρτηση αυτή χρησιμοποιείται εκτενώς στην αναγνώριση προτύπων λόγω του μαθηματικού υπόβαθρου που την συνοδεύει αλλά και λόγω του κεντρικού οριακού θεωρήματος.
- σύμφωνα με το θεώρημα αυτό, η συνάρτηση πυκνότητας πιθανότητας του αθροίσματος ενός πλήθους στατιστικώς ανεξάρτητων τυχαίων μεταβλητών, τείνει στην Gausian συνάρτηση πυκνότητας πιθανότητας, όταν το πλήθος των όρων του αθροίσματος τείνει στο άπειρο.
- στην πράξη αυτό ισχύει κατά προσέγγιση για έναν αρκετά μεγάλο αριθμό όρων αθροίσματος.
- η μορφή της πολυμεταβλητής συνάρτησης Gauss είναι η:

$$p(x) = \frac{1}{(2\pi)^{\frac{1}{2}} |S|^{\frac{1}{2}}} exp\left(-\frac{1}{2}(x-m)^T S^{-1} (x-m) \right)$$

  όπου $m = E[x]$ είναι το μέσο διάνυσμα, S είναι ο πίνακας συνδιασποράς ο οποίος ορίζεται $S = E[(x-m)(x-m)^T]$ και $|S|$ είναι η ορίζουσα του πίνακα S.
- συχνά η Gaussian συνάρτηση καλείται και κανονική και χρησιμοποιείται ο συμβολισμός $N(m, S)$
- για την μονοδιάστατη περίπτωση, προκύπτει από τον παραπάνω τύπο η πιο γνώριμη μορφή:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{(x-m)^2}{2\sigma^2} \right)$$

Ταξινομητές ελάχιστης απόστασης: Η περίπτωση της Ευκλείδειας απόστασης

- ο βέλτιστος Bayesian ταξινομητής απλοποιείται σημαντικά υιοθετώντας τις παρακάτω παραδοχές:
  - οι κλάσεις είναι ισοπίθανες
  - τα δεδομένα σε όλες τις κλάσεις ακολουθούν την κανονική κατανομή
  - το μητρώο (πίνακας) συνδιασποράς είναι ίδιο για όλες τις κλάσεις
  - το μητρώο συνδιασποράς είναι διαγώνιο και όλα τα στοιχεία της διαγωνίου είναι μεταξύ τους ίσα
- υπό αυτές τις προϋποθέσεις προκύπτει ότι ο βέλτιστος Bayesian ταξινομητής είναι ισοδύναμος με τον ταξινομητή ελάχιστης Ευκλείδειας απόστασης.
- το άγνωστο διάνυσμα x ταξινομείται στην κλάση ω_i αν

$$||x - m_i|| = \sqrt{(x - m_i)^T (x - m_i)} < ||x - m_j||, \forall i \neq j$$

- επειδή ο ταξινομητής αυτός είναι αρκετά απλός, χρησιμοποιείται αρκετά συχνά ακόμα και αν γνωρίζουμε ότι οι παραπάνω προϋποθέσεις δεν ισχύουν.
- αυτό που καταφέρνει είναι να ταξινομήσει ένα πρότυπο στην κλάση της οποίας μέση τιμή είναι πιο κοντά στο πρότυπο με βάση την Ευκλείδεια νόρμα.

Ταξινομητές ελάχιστης απόστασης: Η περίπτωση της απόστασης Mahalanobis

- ο ταξινομητής αυτός προκύπτει όταν χαλαρώσουμε τα παραπάνω κριτήρια και συγκεκριμένα αφαιρέσουμε την τελευταία υπόθεση (διαγώνιο μητρώο συνδιασποράς και τα στοιχεία της διαγωνίου να είναι ίσα μεταξύ τους)
- Σε αυτή την περίπτωση το διάνυσμα x ταξινομείται στην κλάση ω_i αν

$$\sqrt{(x-m_i)^T S^{-1}(x-m_i)} < \sqrt{(x-m_j)^T S^{-1}(x-m_j)}, \forall i \neq j$$

Εκτίμηση μέγιστης πιθανοφάνειας των παραμέτρων της Gaussian σ.π.π

- το πρόβλημα που αντιμετωπίζουμε σε κάποιες περιπτώσεις, είναι ότι μπορεί να μην γνωρίζουμε τις συναρτήσεις πυκνότητας πιθανότητας που περιγράφουν τις στατιστικές κατανομές των δεδομένων στις κλάσεις
- η εκτίμηση αυτών των παραμέτρων θα πρέπει να γίνει με χρήση του συνόλου των δεδομένων εκπαίδευσης.
- η λύση είναι να υποθέσουμε ότι η εκάστοτε συνάρτηση πυκνότητας πιθανότητας υπακούουν σε μια συγκεκριμένη συνάρτηση, αλλά δεν γνωρίζουμε τις τιμές των παραμέτρων που καθορίζουν πλήρως την συνάρτηση αυτή.
- για παράδειγμα μπορεί να γνωρίζουμε ότι η κατανομή ακολουθεί την Gaussian κατανομή, αλλά δεν γνωρίζουμε την μέση τιμή και τα στοιχεία του μητρώου συνδιασποράς.
- προς αυτή την κατεύθυνση, ο εκτιμητής μέγιστης πιθανοφάνειας προσφέρει μια μέθοδο παραμετρικής εκτίμησης μιας άγνωστης συνάρτησης πυκνότητας πιθανότητας.

- στην περίπτωση που δίνονται N σημεία $x_i, i = 1, 2, \ldots, N$ που είναι γνωστό ότι είναι κατανεμημένα σύμφωνα με την κανονική κατανομή, οι εκτιμήσεις της άγνωστης μέσης τιμής και του μητρώου συνδιασποράς δίνονται από τις σχέσεις:

$$m_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\sigma_{ML} = \frac{1}{N} \sum_{i-1}^{N} (x_i - m_{ML})(x_i - m_{ML})^T$$

- συχνά, αντί για την μεταβλητή N, χρησιμοποιούμε στο άθροισμα για τον υπολογισμο του πίνακα συνδιασπορών το $N - 1$

Μοντέλα Μίξης

- Όταν δεν είναι γνωστή η συνάρτηση πυκνότητας πιθανότητας που περιγράφει τα δεδομένα μιας κλάσης, τότε πρέπει να γίνει εκτίμηση τους πριν χρησιμοποιηθεί ο Bayesian ταξινομητής.
- μια από τις πιο γνωστές μεθόδους μοντελοποίησης άγνωστων συναρτήσεων πυκνότητας πιθανότητας, ειναι η Μοντελοποίηση Μίξης.
- σύμφωνα με αυτή, κάθε συνάρτηση πυκνότητας πιθανότητας μπορεί να μοντελοποιηθεί ως γραμμικός συνδυασμός από J συναρτήσεις πυκνότητας ως

$$p(x) = \sum_{j=1}^{J} P_j p(x|j)$$

όπου

$$\sum_{j=1}^{J} P_j = 1, \int p(x|j)dx = 1$$

για μεγάλη τιμή του J.

- στην πράξη στις περισσότερες των περιπτώσεων, οι συναρτήσεις $p(x|j)$ επιλέγονται να είναι Gaussian.

Ο αλγόριθμος Expectation-Maximization (EM)

- θεωρούμε ότι διαθέτουμε ένα σύνολο από N σημεία, των οποίων οι στατιστικές ιδιότητες περιγράφονται από μια συνάρτηση πυκνότητας πιθανότητας όπως πριν.
- αυτό που θέλουμε είναι να χρησιμοποιήσουμε τα δεδομένα για να μπορέσουμε να εκτιμήσουμε τις παραμέτρους που συμμετέχουν στο προηγούμενο ανάπτυγμα:
- αυτές είναι οι πιθανότητες $P_j, j = 1, 2, \ldots, J$ καθώς και οι παράμετροι που σχετίζονται με κάθε έναν από τους όρους $p(x|j), j = 1, 2, \ldots, J$
- για παράδειγμα αν υποθέσουμε ότι κάθε μια από τις σ.π.π. του αθροίσματος είναι Gaussian με μητρώο συνδιασποράς $\sigma_j^2 I$:

$$p(x|j) = \frac{1}{(2\pi)^{\frac{l}{2}}\sigma_j^l} exp\left(-\frac{(x-m_j)^T(x-m_j)}{2\sigma_j^2}\right), j = 1, 2, \ldots, J$$

  τότε οι άγνωστες παράμετροι είναι οι μέσες τιμές $m_j, j = 1, 2, \ldots, J$ (συνολικά lJ παράμετροι) και οι J διασπορές $\sigma_j^2, j = 1, 2, \ldots, J$ (συνολικά J παράμετροι)
- ο EM αλγόριθμος υπολογίζει επαναληπτικά τις ζητούμενες παραμέτρους, ξεκινώντας από αρχικές τιμές που καθορίζονται από τον χρήστη.

Εκτίμηση πυκνότητας πιθανότητας με βάση τους k-πλησιέστερους γείτονες

- αν θεωρήσουμε ένα σύνολο N σημείων $x_1, x_2, \ldots, x_N$ που προέρχονται από μια άγνωστη στατιστική κατανομή. Ο στόχος είναι να εκτιμήσουμε την τιμή της άγνωστης σ.π.π. για κάποιο δεδομένο σημείο x
- ο εκτιμητής πυκνότητας πιθανότητας που βασίζεται στους k-πλησιέστερους γείτονες ακολουθεί τα παρακάτω βήματα:
  1. επιλέγουμε μια τιμή για το k (αριθμός γειτόνων)
  2. βρίσκουμε την απόσταση του x από κάθε σημείο του συνόλου της εκπαίδευσης $x_i, i = 1, 2, \ldots, N$. Ως μέτρο απόστασης μπορεί να χρησιμοποιηθεί οποιοδήποτε όπως Ευκλείδεια, Mahalanobis κλπ.
  3. βρίσκουμε τους k πλησιέστερους γείτονες του x
  4. υπολογίζουμε το μέγεθος (όγκο) $V(x)$ της περιοχής στην οποία βρίσκονται οι k-πλησιέστεροι γείτονες
  5. υπολογίζουμε την εκτίμηση σύμφωνα με την

$$p(x) \approx \frac{k}{NV(x)}$$

Εκτίμηση πυκνότητας πιθανότητας με βάση τους k-πλησιέστερους γείτονες

Για την περίπτωση της Ευκελείδειας απόστασης ισχύει:

- $V(x) = 2\rho$ στον μονοδιάστατο χώρο
- $V(x) = \pi\rho^2$ στον δισδιάστατο χώρο
- $V(x) = \frac{4}{3}\pi\rho^3$ στον τρισδιάστατο χώρο

Ο Naive Bayes ταξινομητής

- σύμφωνα με τον ταξινομητή Naive Bayes, η εκτίμηση της σ.π.π. για κάποιο σημείο $x = [x(1), x(2), \ldots, x(N)]^T$ δίνεται από τη σχέση

$$p(x) = \prod_{j=1}^{l} p(x(j))$$

- υποθέτουμε δηλαδή όται τα στοιχεία του διανύσματος χαρακτηριστικών x είναι στατιστικώς ανεξάρτητα μεταξύ τους.
- μας είναι χρήσιμο σε περιπτώσεις μεγάλης διάστατης των διανυσμάτων χαρακτηριστικών καθώς αυτό οδηγεί στην απαίτηση για μεγάλο αριθμό και δεδομένων εκπαίδευσης για να έχουμε αξιόπιστη εκτίμηση της πολυδιάστατης σ.π.π.
- ακόμα και αν δεν ισχύει η υπόθεση της στατιστικής ανεξαρτησίας, στην περίπτωση του naive Bayes η απόδοση μπορεί να είναι καλή καθώς αξιόπιστες εκτιμήσεις για μονοδιάστατες σ.π.π. είναι εφικτές με σχετικά λίγα μόνο διαθέσιμα σημεία.

Ταξινομητής πλησιέστερων γειτόνων

- Είναι ένας από τους δημοφιλέστερους αν και παλιός ταξινομητής.
- Δίνονται c κλάσεις $\omega_i, i = 1, 2, \ldots, c$ και ένα σημείο x και N σημεία εκπαίδευσης $x_i, i = 1, 2, \ldots, N$ στον l-διάστατο χώρο μαζί με τις αντίστοιχες ετικέτες κλάσης.
- με δεδομένο το σημείο x του οποίου δεν γνωρίζουμε την κλάση, θέλουμε να το ταξινομησουμε σε μια από τις παραπάνω κλάση.
- τα βήματα του ταξινομητή είναι τα ακόλουθα:
    1. ψάξτε ανάμεσα στα N σημεία εκπαίδευσης τους k κοντινότερους γείτονες του x χρησιμοποιώντας ένα μέτρο απόστασης (π.χ. Ευκλείδεια, Mahalanobis κ.λ.π.). Η παράμετρος k καθορίζεται από τον χρήστη και δεν πρέπει να είναι πολλαπλάσιο του c
    2. προσδιορίστε ποιοι από τους k πλησιέστερους γείτονες ανήκουν στην κλάση ω_i. Προφανώς $\sum_{i=1}^{c} k_i = k$
    3. το x ταξινομείται σε εκείνη την κλάση ω_i για την οποία $k_i > k_j, j \neq i$
- Για μεγάλη τιμή του N, όσο μεγαλύτερο είναι το k, τόσο θα πλησοάσει σε απόδοση αυτή του Naive Bayes.
- για μικρές τιμές όμως, μεγάλες τιμές του k δεν είναι απαραίτητο ότι θα οδηγήσουν σε βελτίωση της απόδοσης αναγκαστικά.

Signal & Image Processing, Pattern Recognition Group (SIPPRE)

www.sippre-group.com