

Διευκρινήσεις για τον Naïve Bayes Classifier

$$p(x|\omega_j) = p(x_1|\omega_j)p(x_2|\omega_j)\cdots p(x_d|\omega_j) = \prod_{k=1}^d p(x_k|\omega_j) \quad (1)$$

$$p(x|\omega_j) = \prod_{i=1}^d p(x_i|\omega_j) = \prod_{i=1}^d \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2}} \quad (2)$$

$$g_j(x) = \ln p(x|\omega_j) = \sum_{i=1}^d \ln p(x_i|\omega_j) = -\sum_{i=1}^d \ln \sigma_{ij} - \sum_{i=1}^d \frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2} \quad (3)$$

Για την κατασκευή του *Naïve Bayes classifier* υποθέστε (ανεξάρτητη) κανονική κατανομή για κάθε χαρακτηριστικό. Έτσι, η υπό-συνθήκη συνάρτηση πυκνότητας πιθανότητας για κάθε κατηγορία $p(x|\omega_j)$ θα γράφεται ως γινόμενο των συναρτήσεων πυκνότητας πιθανότητας κάθε συνιστώσας (βλ. σχέση 1). Υποθέτοντας κανονική κατανομή οδηγούμαστε στο γινόμενο της σχέσης (2), ενώ αν θεωρήσουμε το λογάριθμο ως διακρίνουσα συνάρτηση παίρνουμε την μορφή συνάρτησης της σχέσης (3). Αυτή μπορείτε να χρησιμοποιήσετε ως μηχανισμό απόφασης κατηγορίας (κατηγορία με τη μέγιστη τιμή διακρίνουσας).

Μερικές παρατηρήσεις:

- Επειδή οι κατηγορίες είναι συμμετρικές (ίσο πλήθος δεδομένων στο σύνολο εκπαίδευσης) μπορείτε να θεωρήσετε ότι οι εκ των προτέρων πιθανότητες $P(\omega_j)$ είναι ίσες (ίσες με 0.1) και επομένως δεν λαμβάνονται υπόψη στον ταξινομητή κατά Bayes.
- Για τον υπολογισμό των παραμέτρων μ_{ij} και σ_{ij}^2 ακολουθήστε την εξής διαδικασία εκτίμησης: Για κάθε μία κατηγορία j και κάθε μία συνιστώσα i της διανυσματικής έκφρασης των δεδομένων εικόνας, υπολογίστε τα στατιστικά τους, δηλ. το μέσο όρο (*mean*) και την διακύμανσή τους (*variance*), αντίστοιχα. Αυτές θα είναι οι τιμές των παραμέτρων των κανονικών κατανομών και ουσιαστικά αυτή η διαδικασία αποτελεί τη μάθηση στη μέθοδο του Naïve Bayes classifier.
- Προσπαθήστε να υλοποιήσετε τα παραπάνω και μην καταφύγετε στην ευκολία των αυτόματων εργαλείων.
- Τέλος, παρατηρείστε ότι παίρνοντας την απόλυτη τιμή της διακρίνουσας συνάρτησης στον τύπο (3) (δηλ. ορίζοντας την διακρίνουσα ως $g_j(x) = -\ln(p(x|\omega_j))$) στην ουσία είναι σαν να αποφασίζουμε με βάση την **μικρότερη απόσταση** (όχι Ευκλείδεια) από το κέντρο μ_j της κατηγορίας (υπό μία έννοια). Αυτό φυσικά αποτελεί μία απλούστευση της γενικής διαδικασίας ταξινόμησης κατά Bayes.