



Εργασία 4

(Προθεσμία: Κυριακή 22 Νοεμβρίου 2020)

Άσκηση 4.1. (Θεοδωρίδης 2.36 και 2.37)

Η σ.π.π. μιας τυχαίας μεταβλητής δίνεται από την σχέση $p(x) = \begin{cases} \frac{1}{2} & \text{για } 0 < x < 2 \\ 0 & \text{διαφορετικά} \end{cases}$.

α) Χρησιμοποιείστε την μέθοδο παραθύρων *Parzen* για να την προσεγγίσετε χρησιμοποιώντας ως συνάρτηση πυρήνα την Gaussian $N(0,1)$. Επιλέξτε την παράμετρο λείανσης να είναι $h=0.05$ και $h=0.2$. Για κάθε περίπτωση, σχεδιάστε την εκτίμηση βασιζόμενοι σε $N=32$, $N=256$ και $N=5000$ σημεία, τα οποία παράγονται από μία γεννήτρια ψευδοτυχαίων αριθμών, σύμφωνα με την $p(x)$.

β) Επαναλάβετε το προηγούμενο ερώτημα έχοντας $N=5000$ σημεία και χρησιμοποιώντας την εκτίμηση *k-πλησιέστερων γειτόνων* με $k=32$, 64 και 256 αντιστοίχως.

Ταξινομητές *Parzen* και *k-πλησιέστερων γειτόνων*

Άσκηση 4.2.

Έστω πρόβλημα με 3 κλάσεις με $P(x/\omega_1) = N(2, 0.5)$, $P(x/\omega_2) = N(1, 1)$, $P(x/\omega_3) = N(3, 1.2)$ με Priors: $P(\omega_1) = 0.5, P(\omega_2) = 0.3, P(\omega_3) = 0.2$

A. Δημιουργήστε τυχαίο δείγμα $\{x_i\}$ 100 δειγμάτων (προτύπων) που να ακολουθεί τις παραπάνω κατανομές, σύμφωνα με τις δεδομένες εκ των προτέρων πιθανότητες. Χρησιμοποιήστε αυτό το δείγμα ως «σύνολο εκμάθησης». Δημιουργείστε τυχαίο δείγμα $\{y_i\}$ 1000 προτύπων που να ακολουθεί τις παραπάνω κατανομές, σύμφωνα με τις δεδομένες εκ των προτέρων πιθανότητες. Χρησιμοποιείστε αυτό το σύνολο ως σύνολο δοκιμής.

B. Ταξινομείστε τα δείγματα $\{y_i\}$ χρησιμοποιώντας τον αλγόριθμο *k-nearest-neighbor* και υπολογίστε την πιθανότητα λάθους. Υλοποιήστε τον αλγόριθμο για $k=1,2,3$. Σχολιάστε τα αποτελέσματά σας. Συγκρίνετε την πιθανότητα λάθους με αυτή του Bayesian ταξινομητή. Προαιρετικό: Πως μπορείτε να υπολογίσετε ένα βέλτιστο k για το σύνολο $\{x_i\}$; Ποιό είναι αυτό;

Γ. Χρησιμοποιήστε τα παραπάνω δείγματα για κατηγοριοποίηση με *Parzen Windows*. Χρησιμοποιήστε τουλάχιστον 4 διαφορετικές τιμές της παραμέτρου λείανσης $h_n = \sigma$ (*spread*) και διαλέξτε εκείνη που δίνει τα καλύτερα αποτελέσματα. Σχολιάστε το αποτέλεσμα.

Προαιρετικό: Πως μπορείτε να υπολογίσετε ένα βέλτιστο h_n για το σύνολο $\{x_i\}$; Ποιό είναι αυτό;

Δ. Χρησιμοποιήστε τα παραπάνω δείγματα για κατηγοριοποίηση με Parzen Windows/Probabilistic Neural Networks. Χρησιμοποιήστε 4 διαφορετικές τιμές του $h_n = \sigma$ (spread). Χρησιμοποιήστε την σχετική συνάρτηση του Neural Network Toolbox του Matlab® ή άλλη της επιλογής σας (στο λογισμικό που διανεμήθηκε υπάρχει και ο φάκελος parzenPNN με τα προγράμματα που περιγράφονται στις διαφάνειες 5b). Σχολιάστε τα αποτελέσματα. Συγκρίνετε με τον k-NN.

Ξάνθη, 11/11/2020