

Αναγνώριση Προτύπων – Μηχανική Μάθηση

Προαιρετική Εργασία

Ημερομηνία παράδοσης: 28/6/2019

Στόχος: Στο πλαίσιο της εργασίας αυτής, θα υλοποιήσετε και θα αξιολογήσετε βασικούς αλγόριθμους μηχανικής μάθησης χρησιμοποιώντας το περιβάλλον MATLAB.

Περιγραφή: Με τη χρήση του περιβάλλοντος προγραμματισμού MATLAB:

- Αφού μεταφορτώσετε το σύνολο δεδομένων Wisconsin Breast Cancer από τον ιστότοπο του UCI Repository (<https://archive.ics.uci.edu/ml/index.php>) σε ένα αρχείο excel, να συμπληρώσετε τα ελλιπή δεδομένα με την επικρατέστερη τιμή για κάθε χαρακτηριστικό. Προσοχή, η πρώτη στήλη των δεδομένων δεν είναι χρήσιμη για την ταξινόμηση των στιγμιοτύπων, καθώς αποτελεί ένα id για κάθε περιστατικό, οπότε μπορείτε να την αγνοήσετε.
- Να κανονικοποιήσετε τα χαρακτηριστικά της εισόδου στο εύρος [0,1].
- Να εφαρμόσετε τη μέθοδο διασταυρούμενης επικύρωσης (k-fold cross validation) με $k=10$ για να πραγματοποιήσετε ταξινόμηση των στιγμιοτύπων σε καλοήγη (benign) και κακοήγη (malignant) περιστατικά με τη χρήση των ακόλουθων ταξινομητών:
 - Πολυεπίπεδα Νευρωνικά Δίκτυα Πρόσθιας Τροφοδότησης (Multilayer Perceptron)
 - Ταξινομητής K-Κοντινότερων Γειτόνων (K-Nearest Neighbor)
 - Ταξινομητής Bayes
 - Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)
 - Δένδρα Απόφασης (Decision Tree)
- Για κάθε ταξινομητή να πειραματιστείτε με τουλάχιστον δύο διαφορετικούς συνδυασμούς παραμέτρων σχεδίασής του.
- Να αξιολογήσετε τους ταξινομητές που υλοποιήσατε με βάση τα κριτήρια της Ακρίβειας (Accuracy), Ευαισθησίας (Sensitivity) και Ειδικότητας (Specificity). Ποιος είναι κατά τη γνώμη σας ο καταλληλότερος ταξινομητής για το συγκεκριμένο πρόβλημα και γιατί;

Για την υλοποίηση των παραπάνω ερωτημάτων μπορείτε να χρησιμοποιήσετε έτοιμες συναρτήσεις του MATLAB, όπως είναι οι xlsread, fitcsvm κ.α.

Παραδοτέο: Το παραδοτέο της συγκεκριμένης εργασίας θα είναι ένας φάκελος, τον οποίο θα ανεβάσετε στο e-class, ο οποίος θα περιέχει τα ακόλουθα αρχεία:

- Μία γραπτή αναφορά στην οποία θα περιγράψετε συνοπτικά το σύνολο δεδομένων που χρησιμοποιήσατε και τα βήματα υλοποίησης που ακολουθήσατε. Θα παραθέσετε, επίσης, σε έναν πίνακα τις τιμές των παραμέτρων και τις τιμές των κριτηρίων αξιολόγησης (Ακρίβεια, Ευαισθησία και Ειδικότητα) για κάθε ταξινομητή που υλοποιήσατε και θα αιτιολογήσετε την επιλογή του καλύτερου ταξινομητή από αυτούς που υλοποιήθηκαν με βάση αυτά τα αποτελέσματα.
- Αρχεία με τον κώδικα του MATLAB για την υλοποίηση των ερωτημάτων της εργασίας. Στη συγγραφή του κώδικα καλό θα ήταν να συμπεριλάβετε σχόλια με λατινικούς χαρακτήρες, τα οποία να εξηγούν τη λειτουργία των βασικών του μερών.

Καλή επιτυχία!