

3^η Εργασία

Υλοποιήστε σε Java ή C++ (ή άλλη γλώσσα που θα σας επιτρέψουν οι υπεύθυνοι των εργαστηρίων) δύο ή τρεις (ανάλογα με το αν η ομάδα σας έχει δύο ή τρία μέλη) από τους ακόλουθους αλγορίθμους μάθησης:

- **Αφελής ταξινομητής Bayes** (πολυμεταβλητή μορφή Bernoulli ή πολυωνυμική μορφή).
- **ID3** (προαιρετικά με πριόνισμα ή πρόωρο τερματισμό της επέκτασης του δέντρου).
- **AdaBoost** (με δέντρα απόφασης βάθους 1 ως βασικό ταξινομητή).
- **Λογιστική παλινδρόμηση** (με στοχαστική ανάβαση κλίσης, προσθέτοντας όρο κανονικοποίησης στην αντικειμενική συνάρτηση).

Προαιρετικά μπορείτε να προσθέσετε αυτόματη επιλογή ιδιοτήτων (π.χ. μέσω υπολογισμού κέρδους πληροφορίας) στον αφελή ταξινομητή Bayes ή/και τη λογιστική παλινδρόμηση, κάτι που θα προσμετρηθεί θετικά στο βαθμό σας.

Επιδείξτε τις δυνατότητες μάθησης των υλοποιήσεών σας χρησιμοποιώντας υπάρχοντα σύνολα δεδομένων (datasets). Μπορείτε, για παράδειγμα, να χρησιμοποιήσετε σύνολα δεδομένων του UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>) ή της Ομάδας Επεξεργασίας Φυσικής Γλώσσας του ΟΠΑ (βλ. <http://nlp.cs.aueb.gr/software.html>). Θα πρέπει να περιλάβετε στην αναφορά σας **αποτελέσματα των πειραμάτων** που θα εκτελέσετε, δείχνοντας π.χ. καμπύλες μάθησης (ποσοστό σφάλματος σε δεδομένα εκπαίδευσης και ελέγχου συναρτήσει του πλήθους των παραδειγμάτων εκπαίδευσης), αποτελέσματα ακρίβειας (precision) και ανάκλησης (recall) κλπ.¹ Θα πρέπει να αναφέρετε επίσης τις τιμές των παραμέτρων που χρησιμοποιήσατε (π.χ. βάρος λ του όρου κανονικοποίησης στη λογιστική παλινδρόμηση) και πώς τις επιλέξατε (π.χ. με δοκιμές σε ξεχωριστά δεδομένα επικύρωσης).

Μπορείτε προαιρετικά να συγκρίνετε τις επιδόσεις των υλοποιήσεών σας με τις επιδόσεις άλλων διαθέσιμων υλοποιήσεων (π.χ. του Weka) ή άλλων ομάδων, κάτι που θα προσμετρηθεί θετικά στο βαθμό σας. Δεν επιτρέπεται να χρησιμοποιήσετε έτοιμες υλοποιήσεις αλγορίθμων μηχανικής μάθησης. Περαιτέρω διευκρινίσεις θα δοθούν από τους υπευθύνους των εργαστηρίων.

Η προθεσμία παράδοσης της εργασίας θα ανακοινωθεί στο e-class. **Διαβάστε προσεκτικά και το έγγραφο με τις γενικές οδηγίες των εργασιών του μαθήματος** (βλ. e-class). Αν οι κανόνες εκείνου του εγγράφου σας επιτρέπουν να υποβάλετε την εργασία ατομικά, αρκεί να υλοποιήσετε έναν από τους παραπάνω αλγορίθμους.

¹ Βλ. https://en.wikipedia.org/wiki/Precision_and_recall.