Οικονομικό Πανεπιστήμιο Αθηνών, Τμήμα Πληροφορικής
Μάθημα: Τεχνητή Νοημοσύνη
Ακαδημαϊκό έτος: 2019–20
Διδάσκων: Ι. Ανδρουτσόπουλος

2η Εργασία

Υλοποιήστε σε Java ή C++ ή Python (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 στον βαθμό σας.

Επιδείξτε τις δυνατότητες μάθησης των υλοποιήσεών σας χρησιμοποιώντας **τουλάχιστον ένα** από τα σύνολα δεδομένων **Enron-Spam**, **Ling-Spam**, **PU**. (βλ. http://nlp.cs.aueb.gr/software.html). Θα πρέπει να περιλάβετε στην αναφορά σας **αποτελέσματα των πειραμάτων** που θα εκτελέσετε, δείχνοντας (τουλάχιστον) **καμπύλες** (και αντίστοιχους πίνακες) με ποσοστό **σφάλματος** (accuracy) σε **δεδομένα εκπαίδευσης** (training) και **ελέγχου** (test) συναρτήσει του πλήθους των παραδειγμάτων εκπαίδευσης, καθώς και αντίστοιχες καμπύλες (και πίνακες) με αποτελέσματα **ακρίβειας** (precision), **ανάκλησης** (recall) και **F1** συναρτήσει του πλήθους των παραδειγμάτων εκπαίδευσης.[1] Θα πρέπει να αναφέρετε επίσης στην αναφορά σας τις **τιμές** των **υπερ-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).

Δεν επιτρέπεται να χρησιμοποιήσετε έτοιμες υλοποιήσεις αλγορίθμων μηχανικής μάθησης. Μπορείτε, όμως, να συγκρίνετε προαιρετικά τις επιδόσεις των υλοποιήσεών σας με τις επιδόσεις άλλων διαθέσιμων υλοποιήσεων (π.χ. του Weka ή του Scikit-learn) ή άλλων ομάδων, κάτι που θα προσμετρηθεί θετικά στον βαθμό σας. Επιτρέπεται, επίσης, να χρησιμοποιήσετε έτοιμες βιβλιοθήκες για την κατασκευή πινάκων και διαγραμμάτων με καμπύλες. 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

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

σας επιτρέπουν να υποβάλετε την εργασία ατομικά, αρκεί να υλοποιήσετε έναν από τους παραπάνω αλγορίθμους.