

# Εργασία 2

---

## Εισαγωγή

Η εργασία αυτή έχει ως στόχο την εξοικείωση με την μάθηση από δεδομένα πολλαπλών ετικετών (Μάθημα 7), την μάθηση από σάκους περιπτώσεων (Μάθημα 8) και την ενεργή μάθηση (Μάθημα 9) καθώς και των λύσεων που υπάρχουν σε αυτά τα προβλήματα. Στην εργασία θα δουλέψετε με το σύνολο δεδομένων [DeliciousMIL](#).

## Μέρος Α

Στο πρώτο μέρος της εργασίας θα δουλέψετε στο αντικείμενο της μάθησης από δεδομένα πολλαπλών ετικετών. Εφαρμόστε μία μέθοδο μάθησης από δεδομένα πολλαπλών ετικετών στα παραπάνω δεδομένα, όπου η αναπαράσταση κάθε εγγράφου αφορά στις λέξεις που περιλαμβάνει (bag of words). Παρουσιάστε και σχολιάστε τα αποτελέσματα των πειραμάτων σας με βάση τα δοθέντα train και test splits.

## Μέρος Β

Στο δεύτερο μέρος της εργασίας θα δουλέψετε στο πρόβλημα της μάθησης από σάκους περιπτώσεων. Εστιάστε στην συχνότερη από τις 20 κλάσεις, ώστε να αντιμετωπίσετε ένα δυαδικό πρόβλημα ταξινόμησης. Θεωρείστε ότι κάθε έγγραφο είναι ένας σάκος από προτάσεις. Εφαρμόστε μία μέθοδο μάθησης από σάκους περιπτώσεων. Παρουσιάστε και σχολιάστε τα αποτελέσματα των πειραμάτων σας με βάση τα δοθέντα train και test splits.

## Μέρος Γ

Στο τρίτο μέρος της εργασίας θα δουλέψετε στο πρόβλημα της ενεργούς μάθησης. Εστιάστε στην σπανιότερη από τις 20 κλάσεις αυτήν την φορά. Χωρίστε το test set σε 2 υποσύνολα, ένα unlabeled pool set και ένα test set σε αναλογία 50%. Εφαρμόστε 10 επαναλήψεις του αλγόριθμου uncertainty sampling. Παρουσιάστε και σχολιάστε το αντίστοιχο learning curve.

## Λογιστικά

Οι εργασίες προτείνεται να γίνουν σε ομάδες 2-3 ατόμων. Θα πρέπει να παραδώσετε μέσω του elearning μια αναφορά (αρχείο τύπου pdf) που θα παρουσιάζει και θα σχολιάζει τα αποτελέσματα των πειραμάτων σας. Η αναφορά θα πρέπει να περιέχει σύνδεσμο προς τον κώδικα ή/και το λογισμικό που χρησιμοποιήσατε. Προθεσμία υποβολής: 7/6/2018. Εκπρόθεσμη υποβολή μέχρι και 2 μέρες μετά την προθεσμία θα έχει ως αποτέλεσμα τη μείωση του βαθμού κατά 10% την ημέρα.