

Εκφώνηση εργασίας Αναγνώρισης Προτύπων

Χειμερινό Εξάμηνο, Ακαδημαϊκό έτος 2017-2018

Τμήμα Πληροφορικής

Πανεπιστήμιο Πειραιώς

Δίνεται το σύνολο δεδομένων *MovieLens 100K Dataset*, που περιέχει 100000 βαθμολογήσεις, 943 χρηστών, σε 1682 ταινίες. Το σύνολο αυτό περιγράφεται στη σελίδα <https://grouplens.org/datasets/movielens/100k/>. Χάριν ευκολίας, είναι επίσης διαθέσιμο, στον φάκελο εγγράφων της εργασίας στην πλατφόρμα e-class.

- 1) Αφού αποσυμπιέσετε το σχετικό αρχείο, ακολουθήστε τις οδηγίες που βρίσκονται στο README, για να κατανοήσετε το ρόλο κάθε αρχείου που προκύπτει από την αποσυμπίεση.
- 2) Εφαρμόστε το βασικό σχήμα ακολουθιακής ομαδοποίησης για να εκτιμήσετε το πλήθος των ομάδων των χρηστών, ως προς τις προτιμήσεις τους.
- 3) Με βάση την εκτίμηση του Βήματος 2, εφαρμόστε τον αλγόριθμο k-means και τον αλγόριθμο ιεραρχικής ομαδοποίησης. Θα χρειαστεί να αποφασίσετε για τις λεπτομέρειες των αλγορίθμων και να τεκμηριώσετε τις υποθέσεις σας. Έχουν κάποια αξία/σημασία οι ομάδες που προέκυψαν;
- 4) Χρησιμοποιώντας το 5-fold σχήμα που παρέχεται, σχεδιάστε, υλοποιήστε και αποτιμήστε δύο ταξινομητές, που λύνουν το εξής πρόβλημα: αν δοθεί ένας χρήστης και μία ταινία, ο ταξινομητής αποφασίζει αν ο χρήστης είδε την ταινία (κλάση 1) ή όχι (κλάση 2). Ο ένας ταξινομητής θα είναι νευρωνικό δίκτυο και ο άλλος ελαχίστων τετραγώνων.

Είναι απαραίτητο να τεκμηριώσετε τις λύσεις σας, ως προς τις υποθέσεις, τα στάδια σχεδίασης και τα αποτελέσματα αποτίμησης.

Γενικές οδηγίες

A) Η αντιγραφή οδηγεί σε μηδενισμό.

B) Μπορείτε να χρησιμοποιήσετε βιβλιοθήκες συναρτήσεων από το Διαδίκτυο, αλλά όχι τυχόν έτοιμο κώδικα.

Γ) Καταληκτική ημερομηνία παράδοσης: 31/1/2018, 23:59. Δεν γίνονται δεκτές εργασίες μετά από αυτήν την προθεσμία.

Δ) Η υλοποίηση μπορεί να γίνει μόνο σε Python ή Matlab/Octave

Ε) Για τον ακριβή τρόπο παράδοσης θα ενημερωθείτε με ανακοίνωση.

ΣΤ) Η εργασία εκπονείται σε ομάδες 1-3 φοιτητών.

Καλή επιτυχία