

**Αναγνώριση Προτύπων**

*Εργασία 2018-2019*

Π16036 – Ιωαννίδης Παναγιώτης

Π16112 – Παραβάντης Αθανάσιος

Περιεχόμενα

[1 Εισαγωγή 2](#_Toc822012)

[2 Ερώτημα 1 3](#_Toc822013)

[3 Ερώτημα 2 4](#_Toc822014)

[3.1 Ζητούμενο 4](#_Toc822015)

[3.2 Τεκμηρίωση 4](#_Toc822016)

[3.2.1 Βασικό ακολουθιακό αλγοριθμικό σχήμα (BSAS) 4](#_Toc822017)

[3.2.2 Εκτίμηση αριθμού ομάδων 4](#_Toc822018)

[3.2.3 Εκτέλεση 4](#_Toc822019)

# Εισαγωγή

Η εργασία μας είναι υλοποιημένη σε python και συγκεκριμένα την έκδοση 3.7.2.  
Επίσης έχουμε χρησιμοποιήσει τις ακόλουθες βιβλιοθήκες:

* Pandas
* Scikit Learn
* Numpy
* SciPy
* Matplotlib

Στην παρούσα εργασία έχουν απαντηθεί όλα τα ερωτήματα.

# Ερώτημα 1

Αφού διαβάσαμε το αρχείο README και κατανοήσαμε τη δομή των αρχείων αποφασίσαμε να δημιουργήσουμε τη κλάση MovielensData στο αρχείο movielens\_data.py ώστε να διαχειριζόμαστε τα αρχεία αυτά και να λαμβάνουμε, κάθε φορά, τα δεδομένα που μας είναι απαραίτητα.

Η επεξεργασία των δεδομένων γίνετε με τη χρήση της βιβλιοθήκης pandas.

# Ερώτημα 2

## Ζητούμενο

Εφαρμόστε το βασικό σχήμα ακολουθιακής ομαδοποίησης για να εκτιμήσετε το  
πλήθος των ομάδων των χρηστών, ως προς τις προτιμήσεις τους.

## Τεκμηρίωση

### Βασικό ακολουθιακό αλγοριθμικό σχήμα (BSAS)

Στο ερώτημα αυτό, υλοποιήσαμε το βασικό ακολουθιακό αλγοριθμικό σχήμα (BSAS) όπως και τον αλγόριθμο εκτίμησης αριθμού ομάδων.

Στην περίπτωση μας ο αριθμός των ομάδων δεν είναι γνωστός. Το ζητούμενο μας είναι να βρούμε τον αριθμό αυτόν. Εφόσον δεν γνωρίζουμε τον αριθμό των ομάδων, η παράμετρος η οποία εμπλέκεται στον αλγόριθμο και η οποία ορίζεται από τον χρήστη, είναι το κατώφλι ανομοιότητας Θ.

Κάθε νέο διάνυσμα νέο διάνυσμα εξετάζεται από τον αλγόριθμο και καταχωρείται είτε σε μια από τις υπάρχουσες ομάδες είτε σε μια νέα ομάδα που δημιουργείται, ανάλογα με την απόσταση από τις άλλες υπάρχουσες ομάδες.

Για τον υπολογισμό της απόστασης θα χρησιμοποιήσουμε το τετράγωνο της ευκλείδειας απόστασης.

Η διάταξη με την οποία παρουσιάζονται τα διανύσματα στον BSAS παίζει σημαντικό ρόλο στη διαδικασία ομαδοποίησης, καθώς μπορεί να οδηγήσουν σε διαφορετικές ομαδοποιήσεις. Επίσης σημαντικός παράγοντας είναι και το Θ.

Κάθε ομάδα αντιπροσωπεύεται από ένα διάνυσμα. Ο αντιπρόσωπος τις ομάδας μετρά από κάθε προσθήκη ενός νέου διανύσματος στην ομάδα θα πρέπει να ενημερώνεται.

### Εκτίμηση αριθμού ομάδων

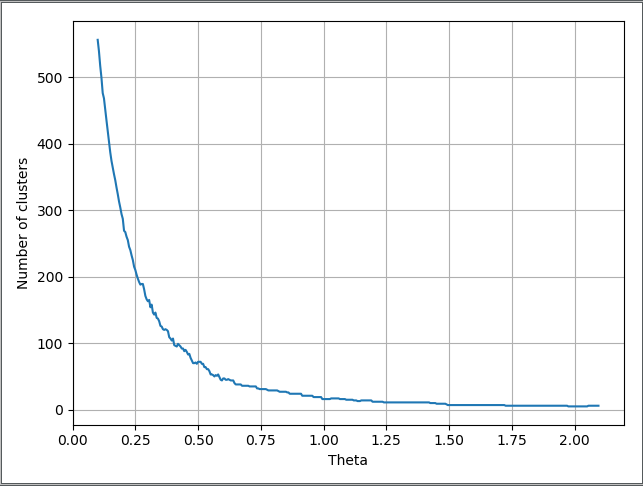
Για να μπορέσουμε να εκτιμήσουμε τον αριθμό των ομάδων, αρκεί να εκτελέσουμε επαναληπτικά τον αλγόριθμο BSAS, για Θ από a έως b με βήμα c. Όσο περισσότερες φόρες εκτελέσουμε τον αλγόριθμο BSAS τόσο μεγαλύτερη θα είναι και η ακρίβεια των αποτελεσμάτων.

### Εκτέλεση

Ο αλγόριθμος BSAS και αλγόριθμος εκτίμησης αριθμού ομάδων βρίσκονται στο αρχείο bsas.py. Ως ελάχιστο Θ δίνουμε την τιμή 0.1. ως μέγιστο Θ την τιμή 2.1 και ως βήμα την τιμή 0.005. Τα δεδομένα που δίνονται ως είσοδος στο αλγόριθμο είναι το άθροισμα των κατηγοριών των ταινιών με βαθμολογία από 4 και πάνω. Στη συνέχεια κανονικοποιούμε τα διανύσματα που προκύπτουν ώστε οι τιμές να είναι μεταξύ 0 και 1.

Αρκεί να εκτελέσουμε το αρχείο bass.py και στο αρχείο result.txt που θα δημιουργηθεί θα δούμε τον αριθμό των ομάδων που δημιουργούνται για κάθε Θ.

Μετά την εκτέλεση του αλγόριθμου προκύπτει η ακόλουθη γραφική απεικόνιση του αριθμού των ομάδων ως προς Θ, εικόνα 3.1.



Εικόνα .

Με βάση τις παραμέτρους που έχουμε θέσει ο αριθμός των ομάδων είναι n = 7.