STAT30270/STAT40590 – Statistical Machine Learning

Assignment 1

Deadline - Thursday 27 February at 16:00

The data

Dataset `data_spotify_songs.rda` contains data about audio features for a collection of songs extracted from the music streaming platform Spotify. The songs belong to three genres: `acoustic`, `pop`, `rock`. For each song, measurements about audio features and popularity are recorded; more information about the audio features is available here:

`https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/`

Assignment

Complete a cluster analysis of the Spotify audio features data using k-means and write a report about your findings. In the report, in particular:

- Motivate the selection of the number of clusters.

- Evaluate the quality of the clustering results and provide an interpretation of the clusters found.

- **STAT40590 students** - Perform a cluster analysis also using k-medoids. Compare the results of k-medoids with the clustering results obtained with the k-means algorithm.

Submission rules

- **Deadline for submission** is **Thursday 27 February at 16:00**.

- Write a short report and submit it as a single pdf file. Indicatively, for STAT30270 students max 4 pages, for STAT40590 students max 5 pages.

- Include the R code used for analysis in the appendix. **The code must be working and the analysis must be reproducible in all parts**.

- Multiple submissions before deadline are allowed and only the latest one will be considered for marking.

- Submission after deadline will incur in penalization as UCD rules.

- It is students' responsibility taking care of uploading all homework parts before the deadline. **Pending or incomplete submissions after the deadline will be considered as they are at submission**.

- **Plagiarism is strictly prohibited**.