Spotify songs clustering

Ali Saleh 97222053

In this project we want to cluster spotify song features and create a song recommender system based on input playlist.

Our input is a playlist containing 47 songs in different genres.

Training dataset is audio features of 42305 songs in 15 genres.

First we combine the training dataset and input dataset to cluster all of our songs and in the end we pick songs that are in the same cluster with our input playlist songs.

Data Preprocessing

- For combining input and training datasets we drop useless features like type, id, uri, track_href, … and make datasets shape equal.
- Change dataset type from float64 and int64 to float16 and int16 to avoid memory limit problem
- Scale dataset using MinMaxScaler

Output dataset has 42352 rows and 13 columns and the values are normal.
Clustering in 13 dimensions is infeasible so we try to reduce dimensions with different methods.

Dimension reduction methods:

- PCA
- SVD
- Feature Selection

PCA

With the PCA dimension reduction method we decrease our dataset features to 5 features and cluster this dataset with DBSCAN.

Results are full of outlier points.

PCA and DBSCAN wont work on this task.

SVD

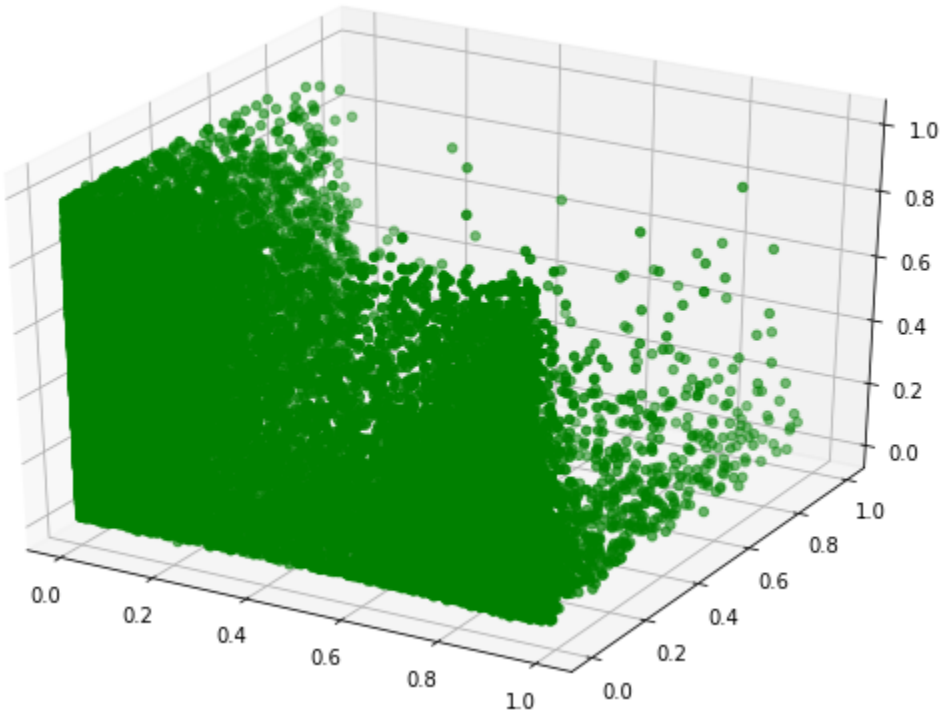
SVD results won't have a notable difference from PCA and results were similar to PCA.

Feature Selection

In the feature selection method we pick random 3 and 4 features from the dataset and compare their clusters.

3 best features were acousticness, instrumentalness, speecheness

This 3 feature plot:



But the DBSCAN method didn't work again.

KMeans

We use the k-means method and cluster data into 10 clusters. And get better results than DBSCAN.

Our input songs cluster into 5 different clusters.
20 song were in cluster 2
10 songs in cluster 5
9 songs in cluster 8
5 songs in cluster 3
3 songs in cluster 9

MIx Playlists

Our first mix consists of 5 songs in cluster 2 which were nearest to the mean of 20 input songs that were in this cluster.

Second mix is from cluster 5
Third mix is from cluster 8
Fourth mix is from cluster 3
Fifth mix is from cluster 9