EE 250 Project Writeup

Description

This project is meant to be a tool that can provide recommendations for songs that would transition well (in terms of melody and tempo) from the song that the user is currently playing on Spotify, which are picked from the user's liked songs. It does this by dividing the user's liked songs into clusters based on the song's ratings for "acousticness," "'danceability," 'energy," and "valence" ratings as provided by the Spotify API. The idea behind this was to let the machine find groupings for the songs, which would provide a smaller subset of songs related to the current song to be considered for recommendations. The candidates for the recommendations are pulled from the cluster that the currently playing song best fits in, and the final recommendations are the tracks in that group that are in the same key and have a tempo within 5 bpm of the tempo of the currently playing song.

Block Diagram Spotify Servers sends authorization via spotipy, a informations python wrapper for the Spotify features of currently played song, user's sends back access liked songs token makes HTTP request to an endpoint set up on the VM with a Local Machine VM Flask app when a url is loaded hosts a webpage that shows the separates liked songs into clusters, current song and the places current song into a cluster, sends information about the recommendations, takes care of the makes final recommendations based current song and spotify authentication on key and bpm similarity recommendations in the response

Frameworks, Protocols and Processing Techiniques

- HTTP used in the authorization flow and for communication between the local machine and the VM
- Flask used to set up the endpoint on the VM and to display the webpage on the local machine
- Spotipy used to make calls to the Spotify API using Python

- K-means clustering was used to divide user's liked songs into clusters, the scikit-learn library was used for this
- The scikit-learn's preprocessing library was used to standardize the liked songs data
- Pandas dataframes were used to organize the data before being fed into the the model

Limitations

The main limitation was how unreliable the key information from the Spotify API was. When I cross-checked with other methods for determining a song's key, like the tools provided in Traktor DJ software, the Spotify API's information about the key a given song was in wasn't consistent, and this led to a lot of less satisfying transitions. Another limitation was the opaqueness of the clustering algorithm's workings - the way it was set up made it more of a black box, so it's possible that the groupings caused some good candidates for transitions from the current song to be overlooked.