Homework Project Definition

Yazid Mouline – Guillaume Requena (Group 9)

Context and problem definition

Music streaming platforms such as Spotify or Deezer have totally modified the way we listen to music today. Instead of always listening to albums we tend to listen to playlists of songs grouped by genre.
Also, these applications allow us to discover many songs easily. Indeed, these platforms have developed techniques (Flow for Deezer and Mix for Spotify) that suggest songs or tracks that are similar to one song or an artist etc.
It could be interesting for us to implement Machine Learning techniques in order to classify music by genre.

Methodology

We will choose wisely our features in the dataset below. Then we will use a classification method such as logistic regression in order to classify songs by genre. We will perhaps try to increase the accuracy of our results by using a neuronal networks.

Data

We will use the Free Music Archive (FMA) dataset (https://github.com/mdeff/fma). FMA is an interactive library of high quality, legal audio downloads. The github link provides these csv files :

- **tracks.csv**: per track metadata such as ID, title, artist, genres, tags and play counts, for all 106,574 tracks.
- **genres.csv**: all 163 genre IDs with their name and parent (used to infer the genre hierarchy and top-level genres).
- **features.csv**: common features extracted with librosa.
- **echonest.csv**: audio features provided by Echonest (now Spotify) for a subset of 13,129 tracks.