Music genre classification

March 17, 2022

In this project, the task is to classify the music genre of an audio track that is 30 seconds long. In the data set, there are a total of ten music genres; *pop*, *metal*, *disco*, *blues*, *reggae*, *classical*, *rock*, *hiphop*, *country*, and *jazz*. The data is based on the **GTZAN data set**[1], which consist of 1000 audio tracks, with 100 audio tracks from each music genre.

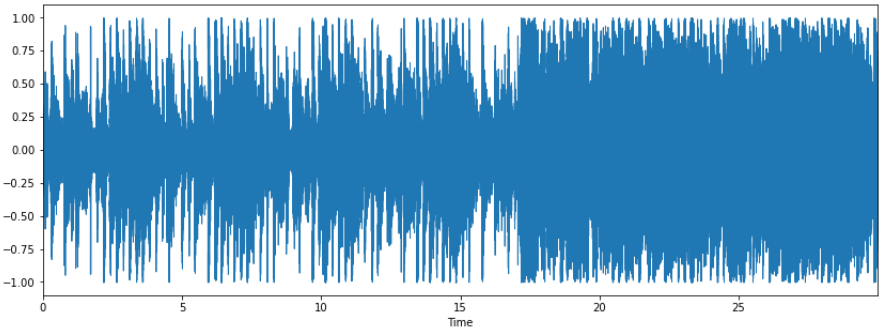

Figure 1: Wave plot of an audio track.

The model features, i.e., the input variables for the classifier, are commonly used audio features that are extracted from the audio track using the LIBROSA[2] package in Python (with default hyperparameter settings). Each feature represent either the mean and standard deviation for an audio feature, indicated by the suffix in the feature name, e.g., *zero_cross_rate_mean*. The majority of the audio features are calculated from the spectrogram of the audio track, see Figure 2.

In the three data files *GenreClassData_5s.txt*, *GenreClassData_10s.txt*, and *GenreClassData_30s.txt*, respectively, you find features that have been calculated by splitting each audio track into non-overlapping segments of either 5, 10 or 30 seconds, before calculating the audio features, see Figure 3. If two rows have the same Track ID, they have been derived from the same audio track. In

[1] http://marsyas.info/downloads/datasets.html
[2] https://librosa.org/doc/main/feature.html

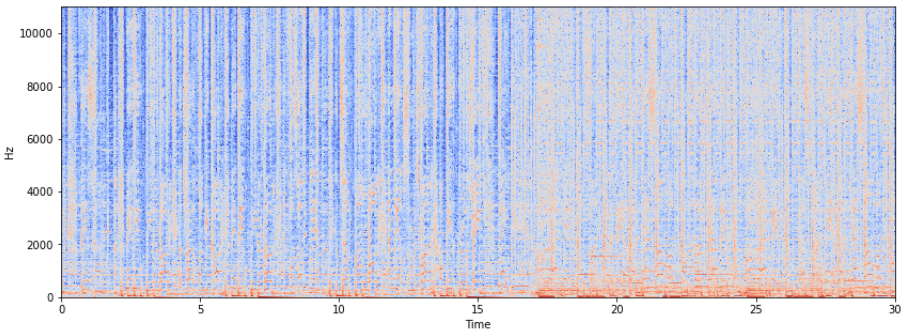

Figure 2: Spectrogram of an audio track.

Metadata_GenreClass.pdf you find a detailed description of what the columns in the data matrix represent.

The training set is represented by 800 audio tracks, with 80 audio tracks from each of the ten music genres. The test set is represented by 200 audio tracks, with 20 audio tracks from each of the ten music genres. The same audio tracks are placed in the same train and test set for all of the data files *GenreClassData_5s.txt*, *GenreClassData_10s.txt* and *GenreClassData_30s.txt*.

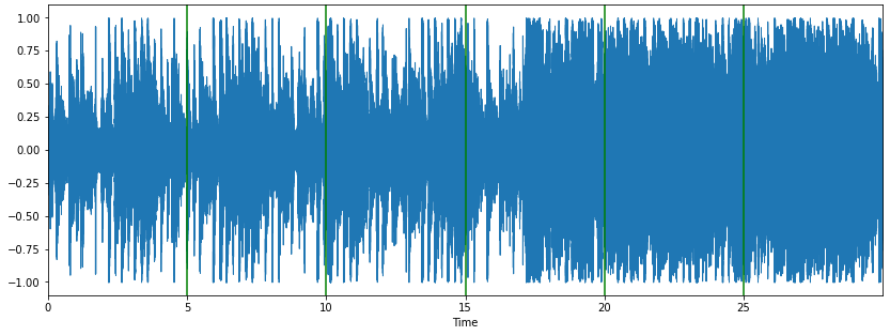

Figure 3: Audio track split into 5 second segments (green vertical lines) before calculating the features.

Tasks

For tasks 1-3 use only the data set *GenreClassData_30s.txt*.

1. Design a k-NN classifier (k=5) for all ten genres using only the following four features; *spectral_rolloff_mean*, *mfcc_1_mean*, *spectral_centroid_mean* and *tempo*. Evaluate the performance of the classification model.

2. For each of the four features; *spectral_rolloff_mean*, *mfcc_1_mean*, *spectral_centroid_mean* and *tempo*, compare the feature distribution for the four classes; **pop**, **disco**, **metal** and **classical**. Analyze how the feature distribution relates to the performance of your classifier.

3. Design a k-NN classifier (k=5) for all ten genres using only four features with at least three features being; *spectral_rolloff_mean*, *mfcc_1_mean*, *spectral_centroid_mean* or *tempo*. Motivate why you selected the particular four features.

4. Design a classifier for all ten genres that classifies the audio tracks, each represented by a *Track ID*. You are allowed to use any classifier, as many features as you like and all of the available data sets *GenreClassData_5s.txt*, *GenreClassData_10s.txt*, and *GenreClassData_30s.txt* as input data.