Team GW-Music

Learning Algorithms for Music Classification

Aziz Burak Gülen, Sacha L'Heveder, Afsin Ozdemir, Ming Zhang

Why classify music with AI?

There’s too much!

Classifying into genres allows stakeholders to narrow their search for music, saving on time.

Such stakeholders could be those creating and maintaining:

- Music databases for advertisers and content creators
- Sample libraries for music production

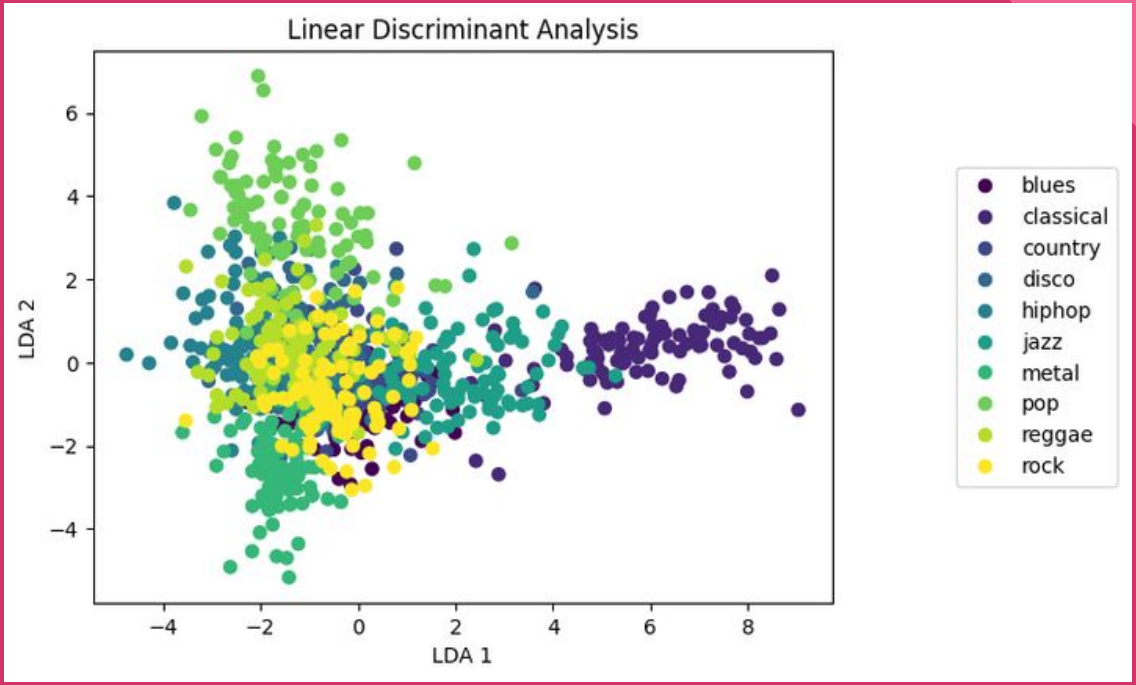
Our Goal

Leverage the power of state of the art learning algorithms to classify large quantities of music data.

We experimented with:

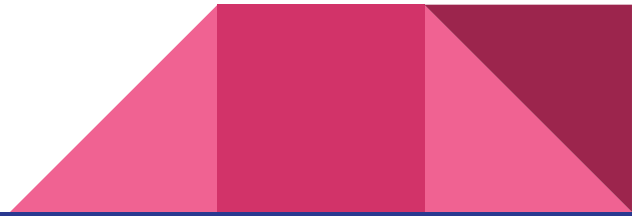
- Network architecture
- Distance function
- Data augmentation
- Feature extraction

The Data


Linear Discriminant Analysis
LDA 2
LDA 1
6
4
2
0
−2
−4
−4
−2
0
2
4
6
8
blues
classical
country
disco
hiphop
jazz
metal
pop
reggae
rock

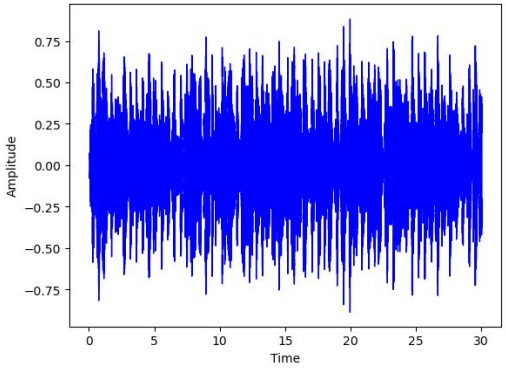
We used the publicly available GTZAN dataset which provides 30 second samples of music from a variety of genres (blues, pop, country, etc.)

We processed the data using **Librosa**, a python library designed for processing audio signals



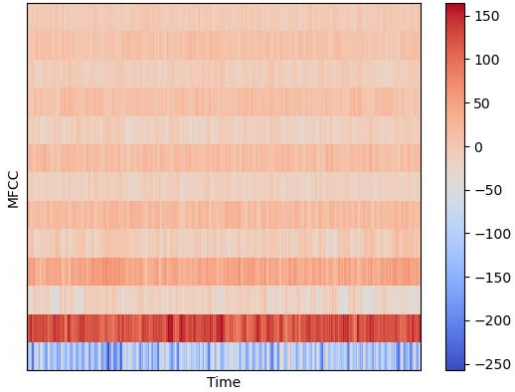
Processing the data:

Audio file

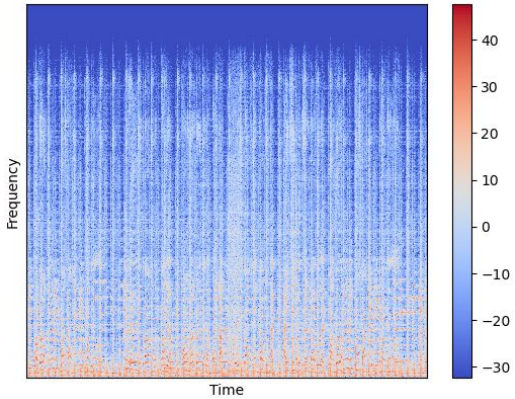


feature extraction

Mel-Frequency Cepstral Coefficients
MFCCs: A “picture of the song”



fast fourier transform

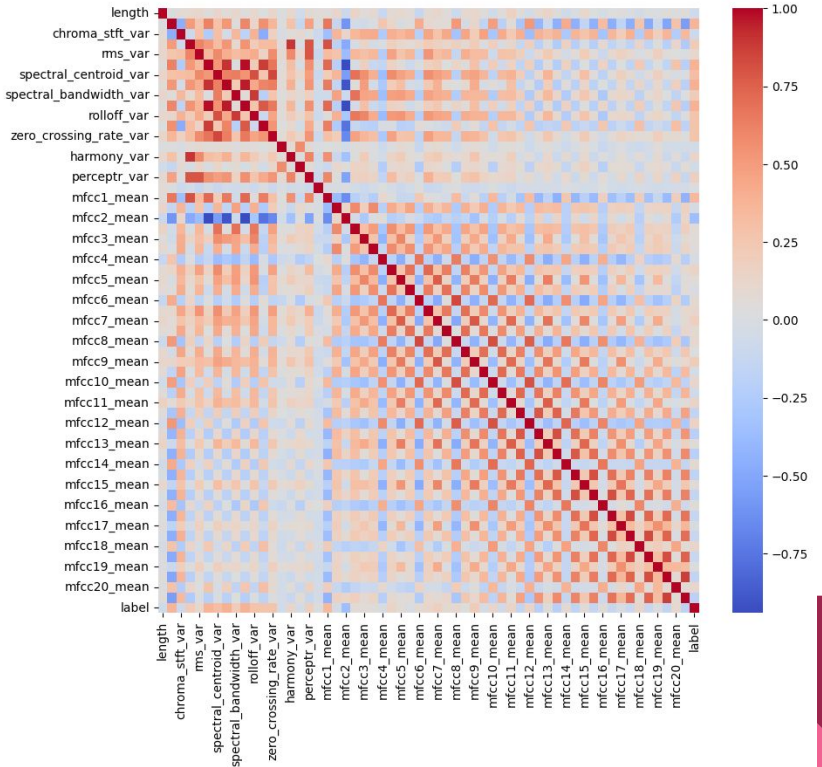


Spectrogram

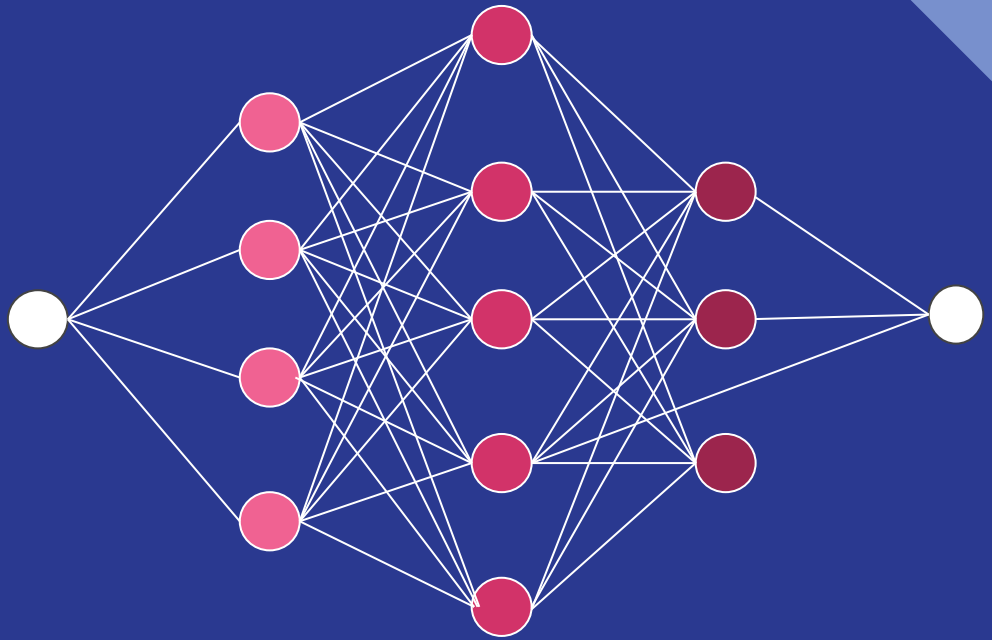
Extracting the most important features from the data.

In addition to the MFCCs, there are attributes which greatly impacted the distance computed between data points. We found that those with the greatest impact were:

- RMS amplitude (how loud is the song?)
- Chroma (captures harmonic information)



The Models



The process

1. Prepare the data and implement an 80-20 train test split.

2. Build/adjust the model using **Tensorflow**.

3. Train the model

4. Test the accuracy of the model

Parameters of the model

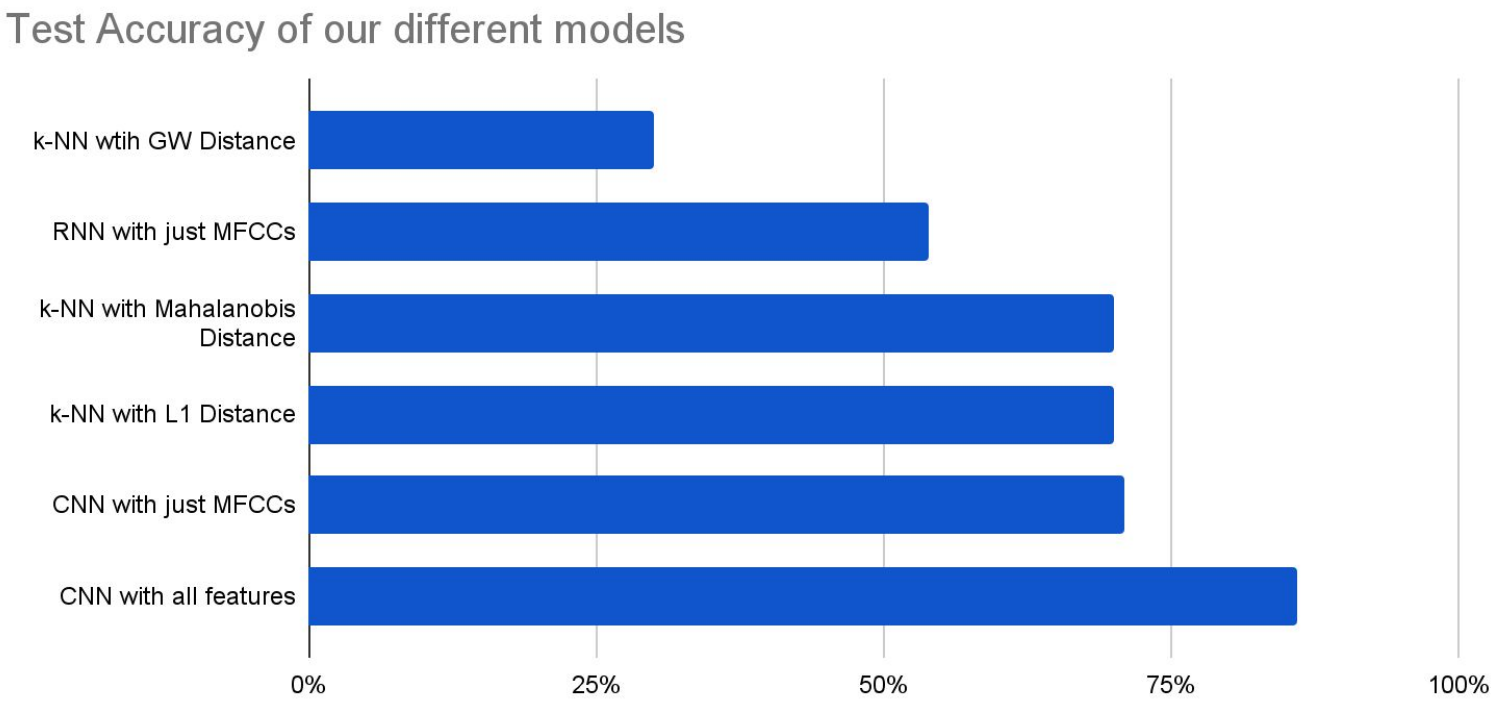
Model

- k-Nearest Neighbors
- Neural Network
  - Convolutional
  - Recurrent

Distance Function

- L1 Distance
- Mahalanobis Distance
  - Takes into account the covariance in the genres
- Gromov-Wasserstein (GW) Distance
  - We use a certain type of GW distance developed for studying time-series

Model Comparison



After testing, the model which provided the best result was CNN with a test accuracy of

86%

Summary

We were able to, given a large and unwieldy set of music samples, accurately classify them (up to 86%) according to pre-prescribed genres.