Project: Music genre classification with neural networks

Name: Pinak Mandal

Problem Statement: Music genre classification is a problem in the field of music information retrieval which has found many popular applications in modern times. Fluidity of modern music and the subjective nature of genres make it a difficult task.

Data set: This work aims to analyze the GTZAN Genre Collection which was used in the paper "Musical genre classification of audio signals" by G. Tzanetakis and P. Cook (2002) which has since become a standard dataset in music analysis.

Challenges: 1. Data needs to be processed for easier manipulation.
2. Huge dimensionality of data needs to be reduced in a way that retains useful features before analysis.
3. Implementation of a neural network in Python.

Workflow

1. First we convert the tracks to .wav format for easier manipulation (implemented in convert.py).
2. We extract MFCC features for each track and tabulate them in csv files (implemented in process.py).
3. We do principal component analysis on the processed data and try to infer some characteristics of individual genres.
4. We divide the data into training and test sets (implemented in process.py).
4. We code a feedforward artificial neural network, feed it the training data and create a model for genre prediction (implemented in main.py).
5. We test our test data with the computed model and calculate accuracy of the model (implemented in main.py and test.py).
Code for this project is available at the github repository pinakm9/musis (link: https://github.com/pinakm9/musis).

Observations

1. Not all genres are equally distinguishable.
2. Classification problem gets harder as number of genres increase.
3. Inclusion of more distinguishable genres make classification problem somewhat easier.
4. MFCCs can be used to encode human perception of speech/music.
5. Neural networks can be effective tools in pattern recognition but they lack illuminating interpretation.
6. Increasing number of hidden layers doesn't drastically change the prediction capability of the neural network used in our experiment.