MUSIC GENRE CLASSIFICATION VIA MACHINE LEARNING

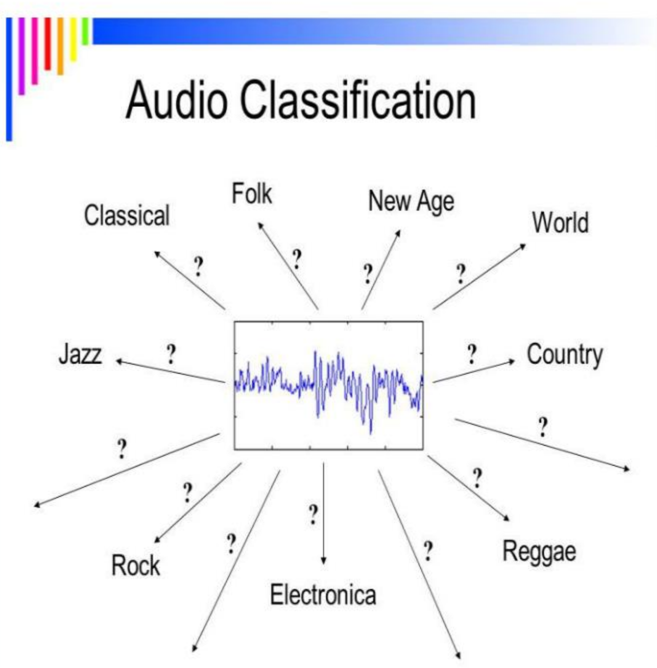
FAİK YUSUF AYAN [040150239]

ADVISOR: Assoc. Prof. ENDER METE EKŞİOĞLU
ISTANBUL TECHNICAL UNIVERSITY, ELECTRONICS AND COMMUNICATION ENGINEEERING DEPARTMENT

INTRODUCTION

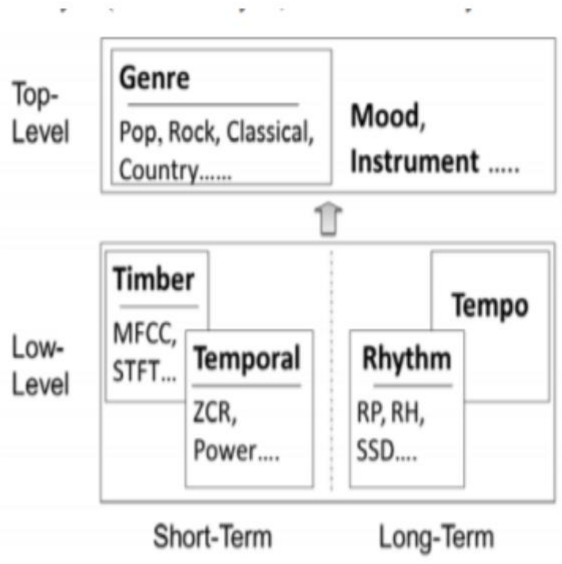
Music genre prediction is one of the topics that digital music processing is interested in. Music styles are classified by a variety of parameters such as beat and timbre. To overcome this challenge, a discipline called MIR(Music Information Retrieval) was created. MIR includes research on the extraction of features from music signals to decide the music genre.

The aim of this work is to predict the genres of songs by using machine learning techniques. For this purpose, feature extraction is done by using signal processing techniques, then machine learning algorithms are applied with those features to do a multiclass classification for music genres.



FEATURE EXTRACTION

Audio features can be mainly divided into two levels as top-level and low-level according to perspective of music understanding . The top level labels provide information on how listeners interpret and understand music using different genres, moods, instruments, etc. Low-level audio features can also be categorized into short-term and long-term features on the basis of their time scale. This figure characterizes audio features from different levels and perspectives.



Most of the features that have been proposed in the literatures are short-time timbre features, which only consider the immediate frequencies and extract the characteristics of the audio signal in a 10-30ms duration small sized window. Long-term features such as rhythm and beat features contain the structural information and normally extracted from the local windows on the large time-scale full song or a sound clip.

FEATURE TYPES

Audio data can be decoded and transformed into series of digital samples to represent the waveform. But this data cannot be used directly by machine learning algorithms because pattern matching algorithms cannot deal with such an amount of information. So, it is necessary to extract some features that describe the audio wave using a compact representation.

1-Zero Crossing Rate (ZCR)

$$Z(i)=\frac{1}{2N}\sum_{n=0}^{N-1}|sgn[x_i(n)]-sgn[x_i(n-1)]|$$

2-Root Mean Square Energy (RMSE)

$$\sqrt{\frac{1}{N}\sum_{n=1}^{N}|x(n)|^2}$$

3-Spectral Centroid

$$C(i)=\frac{\sum_{N-1}^{k=0} k|Xi(k)|}{\sum_{N-1}^{k=0}|Xi(k)|}$$

4-Spectral Roll-off

$$\sum_{n=1}^{R_t} M_t[n]=0.85*\sum_{n=1}^{N} M_t[n]$$

5-*Mel-Frequency Cepstral Coefficients:Besides from being the most important audio feature MFCC are compact, short time descriptors of the spectral envelope audio feature set and typically computed for audio segments of 10-100ms.

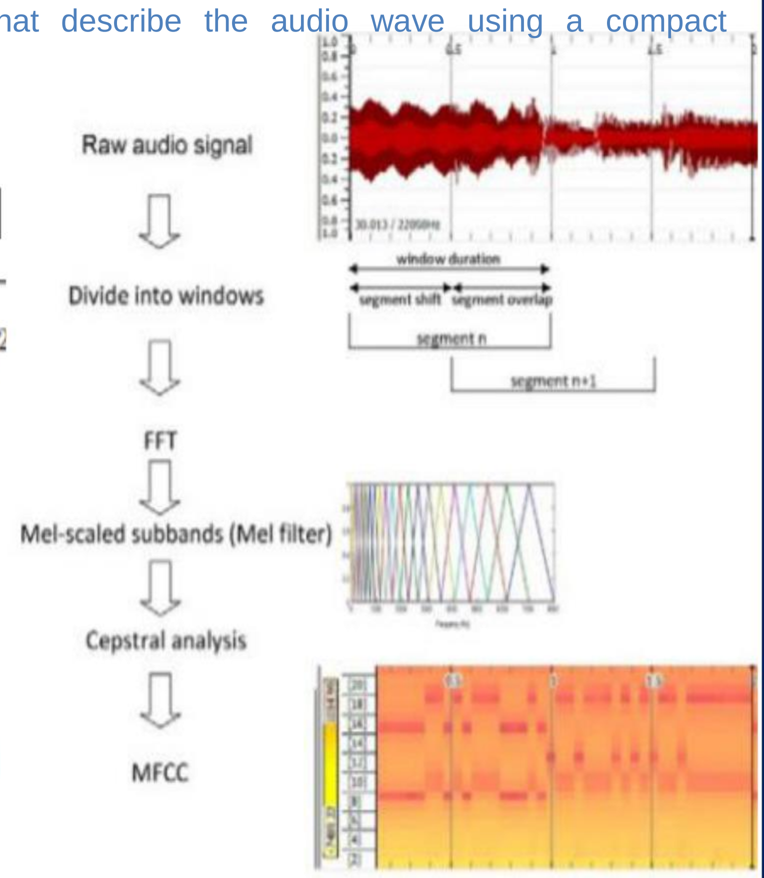
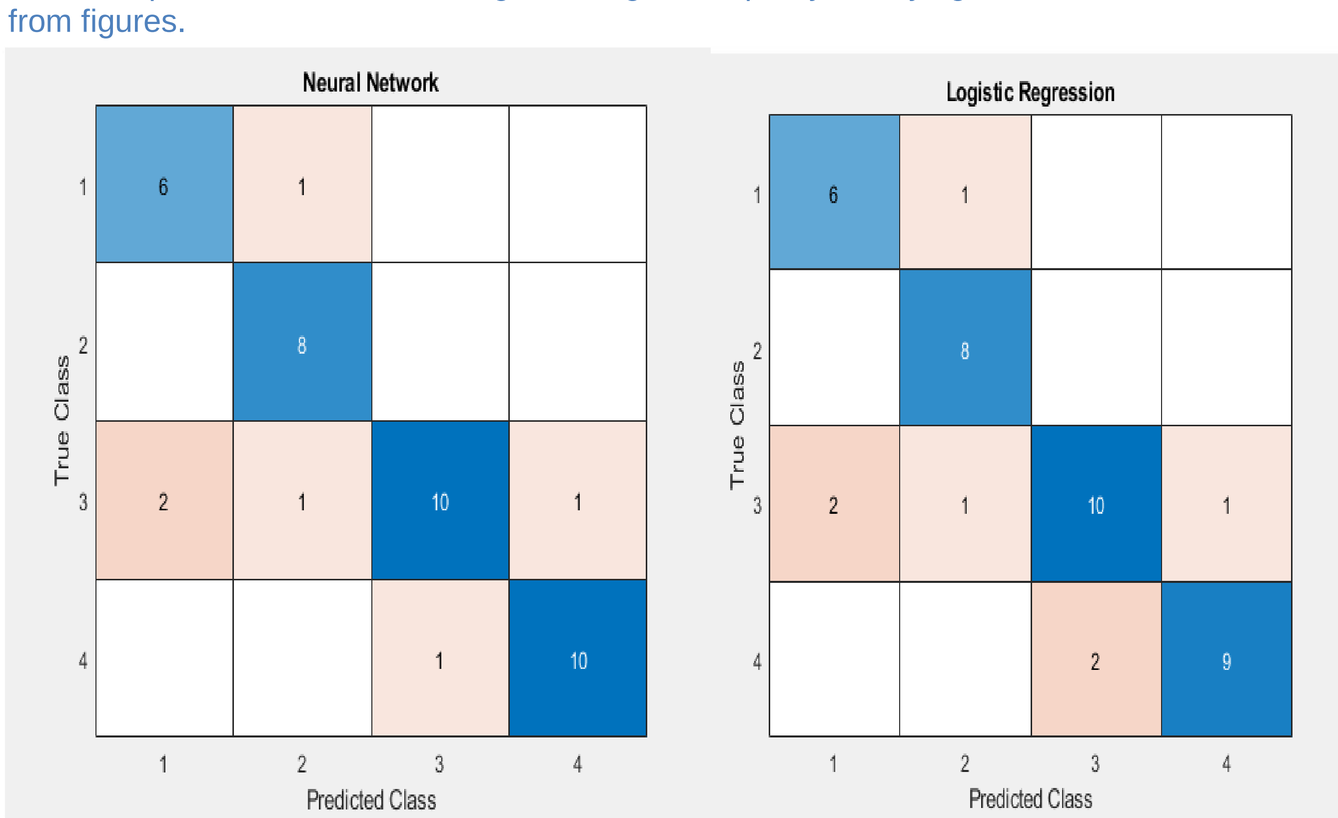


Figure illustrates the series of steps in the calculation of MFCC features from raw audio signal. 13 MFCC features is selected for the solution of genre specification.

RESULTS

At the beginning there was 4 different music classes (blues, metal, country, and pop) with 100 examples each. Both of the algorithms gave us pretty satisfying results as it can be seen from figures.

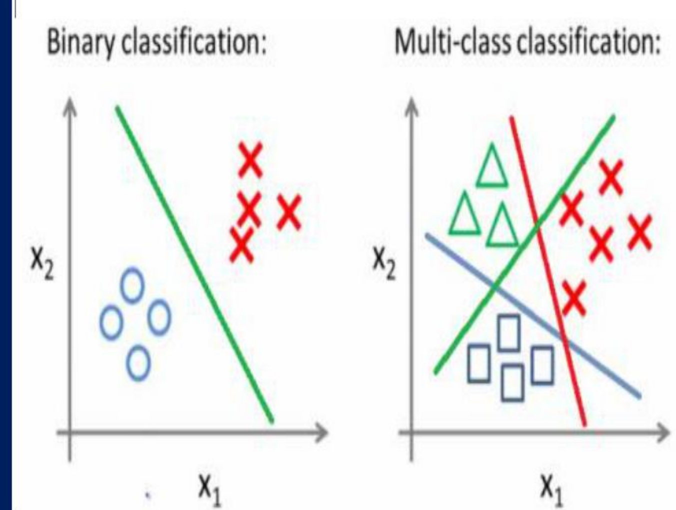



```
Iteration  4989 | Cost: 1.006330e-01
Iteration  4990 | Cost: 1.006330e-01
Iteration  4991 | Cost: 1.006330e-01
Iteration  4992 | Cost: 1.006330e-01
Iteration  4993 | Cost: 1.006330e-01
Iteration  4994 | Cost: 1.006329e-01
Iteration  4995 | Cost: 1.006329e-01
Iteration  4996 | Cost: 1.006329e-01
Iteration  4997 | Cost: 1.006329e-01
Iteration  4998 | Cost: 1.006329e-01
Iteration  4999 | Cost: 1.006329e-01
Iteration  5000 | Cost: 1.006329e-01

Test Accuracy: 82.500000
```

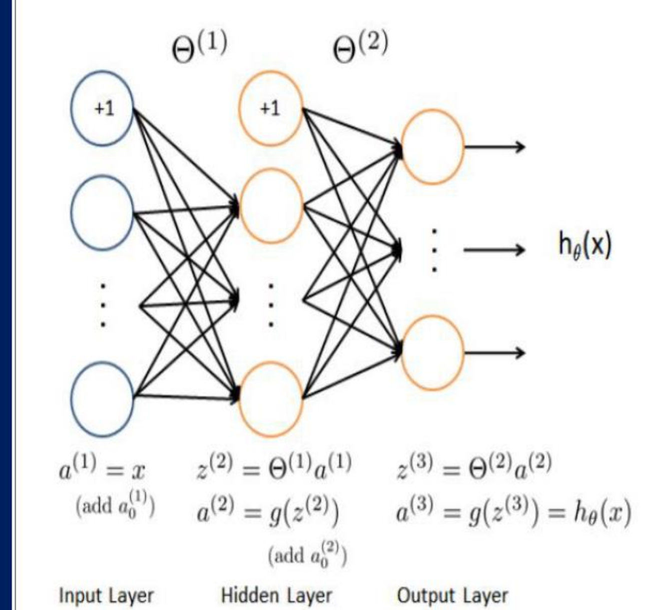
```
Iteration  19988 | Cost: 9.935718e-01
Iteration  19989 | Cost: 9.935717e-01
Iteration  19990 | Cost: 9.935717e-01
Iteration  19991 | Cost: 9.935715e-01
Iteration  19992 | Cost: 9.935714e-01
Iteration  19993 | Cost: 9.935712e-01
Iteration  19994 | Cost: 9.935712e-01
Iteration  19995 | Cost: 9.935703e-01
Iteration  19996 | Cost: 9.935600e-01
Iteration  19997 | Cost: 9.935593e-01
Iteration  19998 | Cost: 9.935575e-01
Iteration  19999 | Cost: 9.935518e-01
Iteration  20000 | Cost: 9.935475e-01

Test Accuracy: 85.000000
```

METHODS

1-One vs All Logistic Regression: is a form of logistic regresssion used to predict a variable have more than 2 classes.



2-Neural Networks: Neural network consists of various neuron layers, each layer receiving inputs from previous layers and transferring outputs to next layers.



CONCLUSION

In summary, objective of the project was to get accuracy rate over 80% for determining the type of four different music genres. If the training data increases, accuracy will increase and if number of classes rise, accuracy may drop. I could have achieved even better results by increasing the iteration, but I left it because I knew that the process would increase and I already had good results.

REFERENCES

[1]https://www.coursera.org/learn/machine-learning
[2]Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches"GSTF International Journal on Computing (JoC), Vol. 3 No.2, July 2013
[3]" Music Genre Classification using Machine Learning Techniques" Hareesh Bahuleyan
[4]https://ch.mathworks.com/help/audio/ref/audiofeatureextractor.html
[5]http://mirlab.org/jang/books/audioSignalProcessing/appNote/musicGenreClassification/html/goTutorial.html
[6]https://www.udemy.com/course/machine-learning-for-datascience-using-matlab/