Music Feature

Music Genre Classification

Yi-Hsin Lu

Statistical Machine Learning Final Presentation

National Dong Hwa University

2022 June 19

Contents

Abstract

Music genre can be predict by using Fourier transform and waveform analysis from audio to numerical data. For the model I used like SVM and Random Forest, the result was better than the others in Kaggle.

1 Introduction

1.1 Motivating Question

When I was a child, I started learning chinese music instrument. That was the beginning that having fun for the music, and I realized the music is also a part of science. For an example, A440 is the musical pitch corresponding to an audio frequency of 440 Hz(from Wikipedia), and every instruments have their own waveform. So I decided to have my master degree that I want to use my statistics knowledge on the music I learned before. The music topic is the work direction for statistical machine learning project.

1.2 Focus Problem

A dataset in Kaggle interests me, 1000 audio tracks each 30 seconds long, it contains 10 genres, each represented by 100 tracks. So my problem is:
1. Is the music genre classified by those numerical data?
2. If it can be classified, which model is the best

1.3 Data

filename	tempo	beats	chroma_stft	rmse	spectral_centroid	spectral_bandwidth	rolloff	zero_crossing_rate
blues.00081.au	103.35938	50	0.3802602	0.2482623	2116.9430	1956.6111	4196.1080	0.1272725
blues.00022.au	95.70312	44	0.3064509	0.1134754	1156.0705	1497.6682	2170.0535	0.0586134
blues.00031.au	151.99908	75	0.2534871	0.1515708	1331.0740	1973.6434	2900.1741	0.0429672
blues.00012.au	184.57031	91	0.2693200	0.1190717	1361.0455	1567.8046	2739.6251	0.0691239
blues.00056.au	161.49902	74	0.3910586	0.1377283	1811.0761	2052.3326	3927.8096	0.0754795

The features in this dataset are extracted from the dataset provided it which consists of 1000 audio tracks each 30 seconds long. It contains 10 genres, each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .wav format. The code used to extract features is at this GitHub repo. Features are extracted using libROSA library.(Content on kaggle)

1.4 Variables/Features

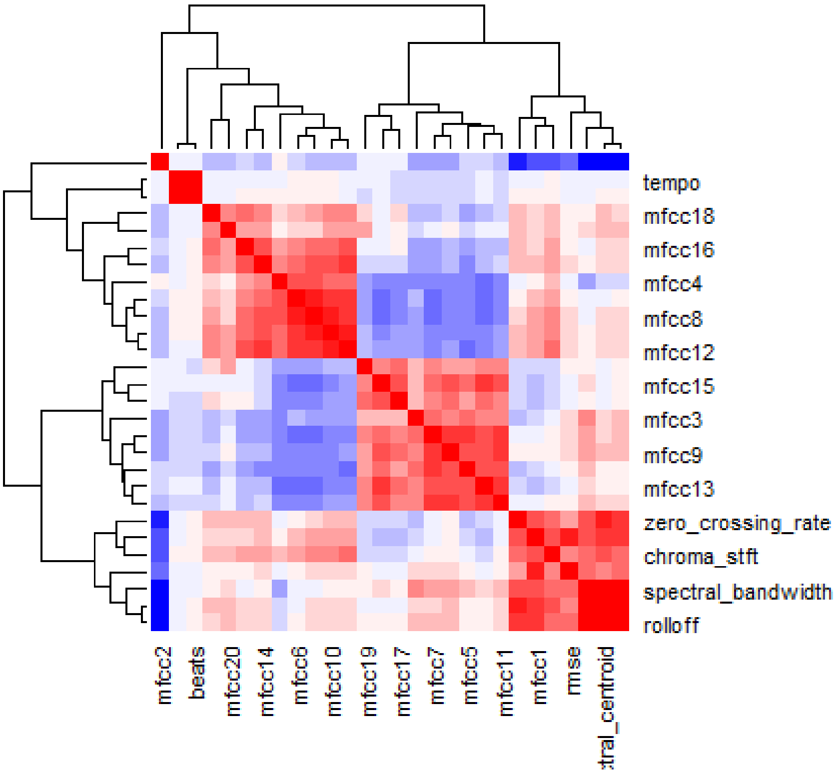
- tempo: The speed at which a passage of music is played
- beats: Rhythmic unit in music
- chroma-stft: Short Time Fourier Transform
- rmse: Root Mean Square Error
- spectral-centroid: Indicates where the "center of mass" of the spectrum is located.
- spectral-bandwidth: It is the Wavelength interval in which a radiated spectral quantity is not less than half its maximum value.
- rolloff: Roll-off is the steepness of a transmission function with frequency.
- zero-crossing-rate: The rate at which the signal changes from positive to negative or back.
- mfcc1-20: Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC.

1.5 Genre(label)

- blues
- classical
- country
- disco
- hiphop
- jazz
- metal
- pop
- reggae
- rock

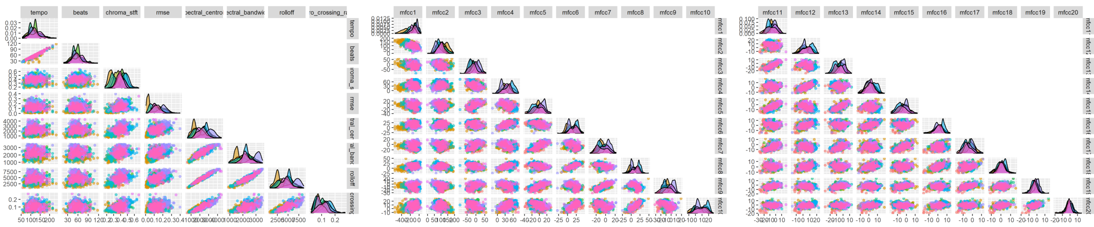
2 EDA

2.1 Correlation Matrix



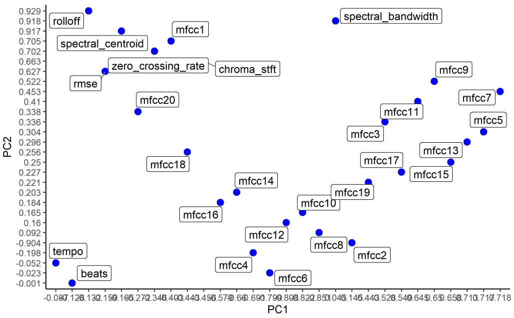
It could be split by three parts in the red area that is the positive high correlation. Then we made the figure of pairs plot fir these three high correlation area.

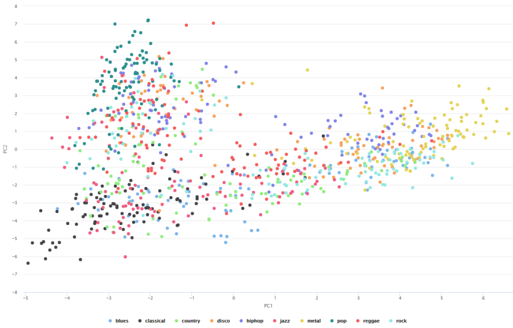
2.2 Pairs plot



There is a high correlation between tempo and beats about 2 times. And spectral-centroid, spectral-bandwidth and rolloff are the results of spectral analysis, they must have correlation for each variables.

2.3 PCA(Principle Component Analysis)



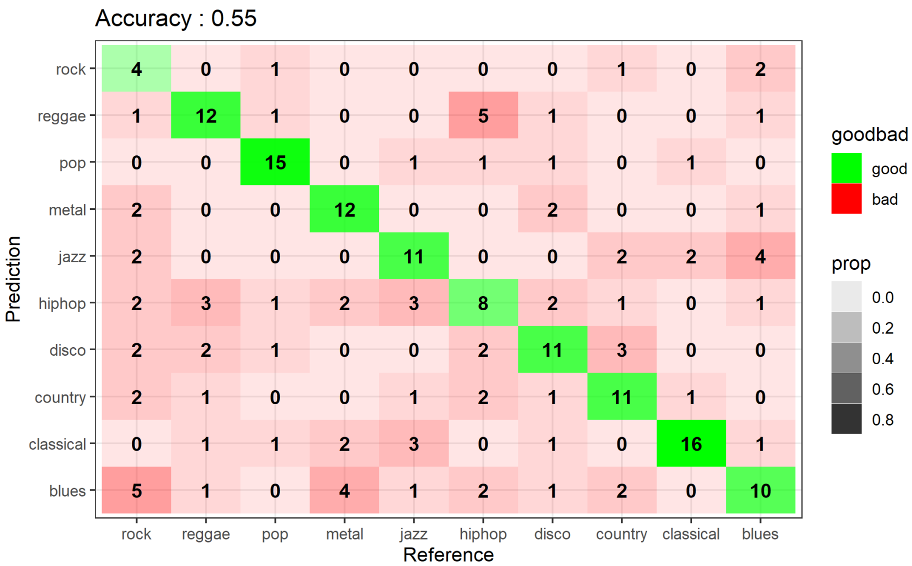


3 Models

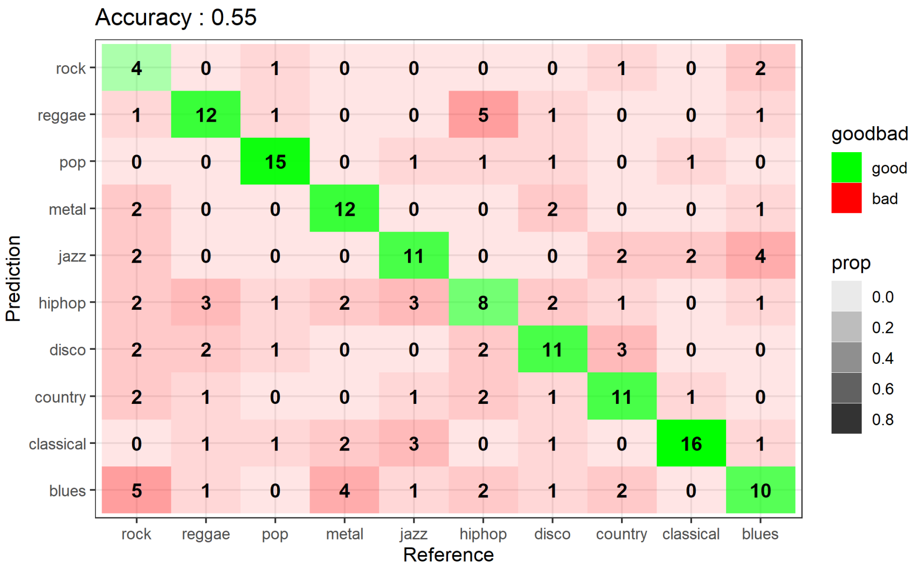
3.1 Training and Testing data

For the balance in this data, taking 80% as training data and 20 % as testing data from each genres, so training data contains 10 genres, each represented by 80 tracks. And so does testing data.

3.2 Logistic Regression



3.3 SVM(Support Vector Machine)

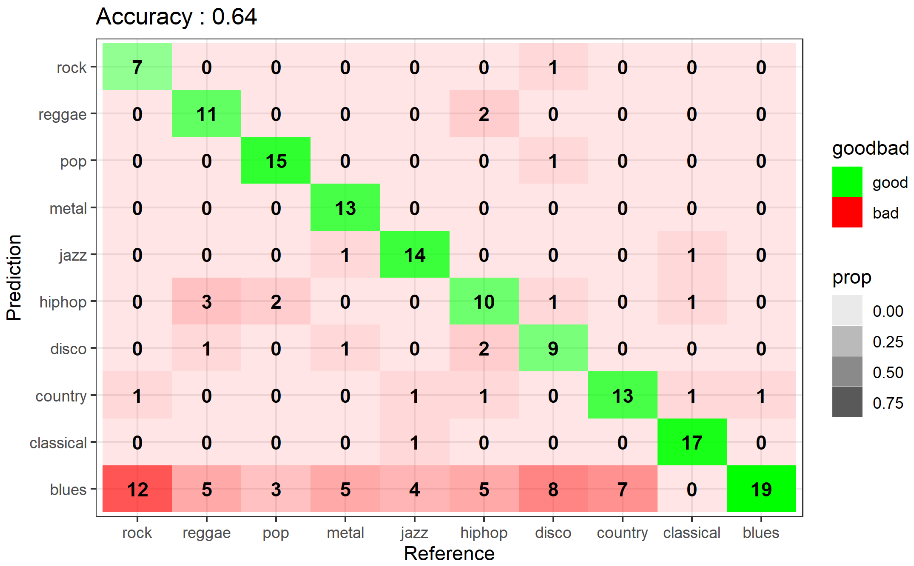


3.4 SVM(one-hot encoding)

Turning this 10 class prediction into a simple version, We do the one-hot encoding for the genres as the figure down below.

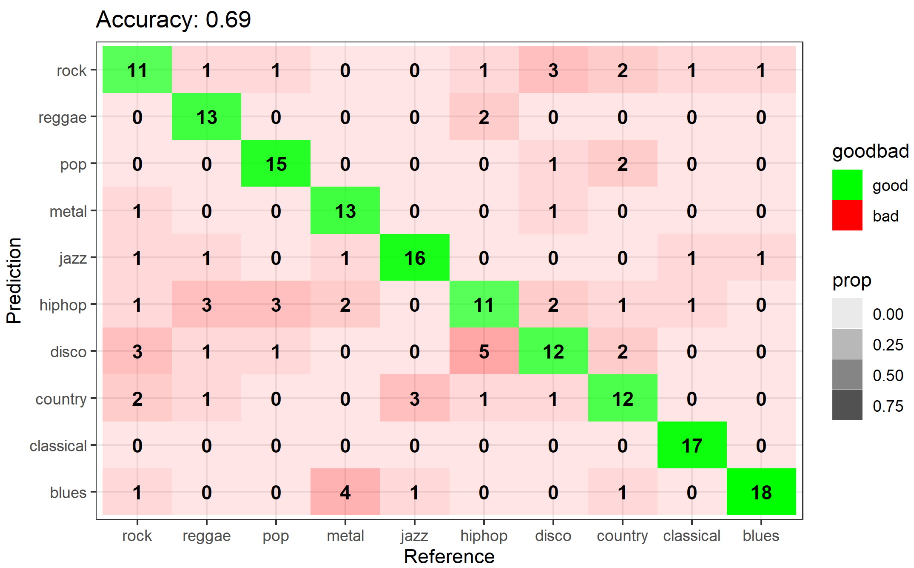
label	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0
blues	1	0	0	0	0	0	0	0	0	0

Predict for each genres, and take the highest accuracy to be the predictor.

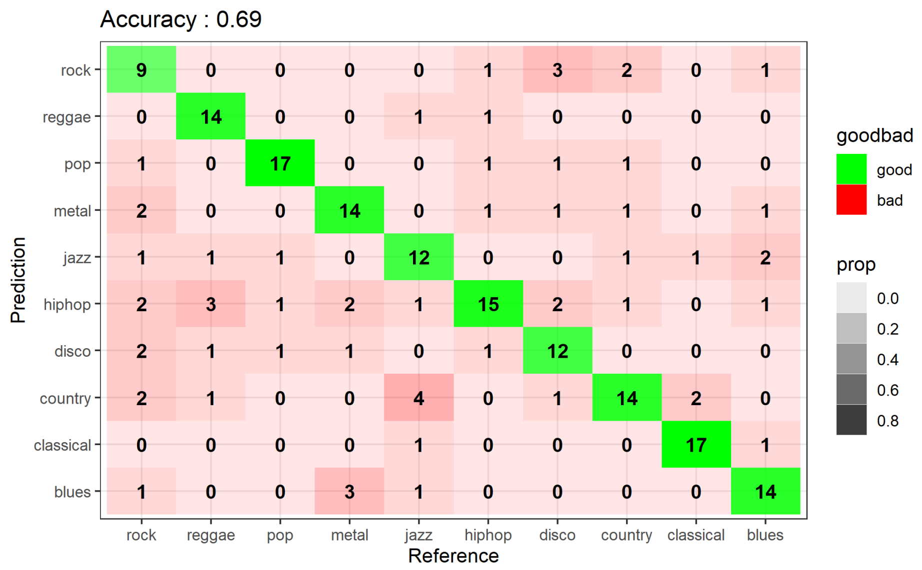


3.5 SVM(scaling)

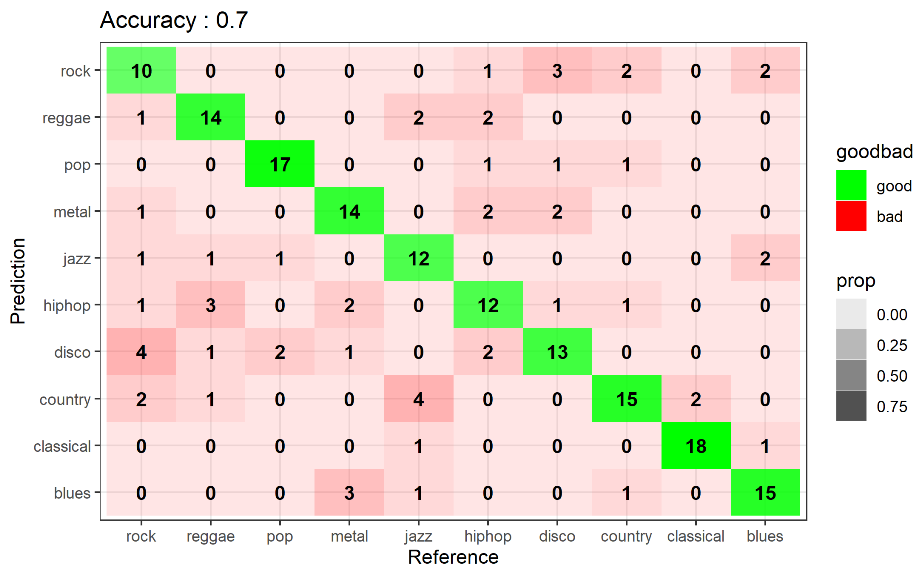
After reading the paper of tuning of SVM, scaling is very essential.



3.6 Random Forest



3.7 Random Forest(scaling)



4 Conclusion

- The music could be classified by those signal transform.
- SVM and Random Forest are nice classifier for this dataset.

4.1 Results

logistic regression	0.55
svm(one-hot encoding)	0.64
svm	0.67
svm(scaling)	0.69
random forest	0.69
random forest(scaling)	0.7

Reference

- (TextBook)Hastie, Tibshirani and Friedman (2009). The Elements of Statistical Learning: Data Mining, Inference and Prediction. 2nd Edition.
- (TextBook)Hardle and Simar (2015). Applied Multivariate Statistical Analysis, 4th Edition.
- (Paper)Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin (2016). A Practical Guide to Support Vector Classification.
- Meinard Muller, Stefan Balke (2015). Short-Time Fourier Transform and Chroma Features.
- (Website)Librosa
- (Website)Tempo vs Rhythm