

Music Genre Classification

Chunya Hua, 2017

Introduction

Understanding the genre of music is the first step in music recommendation. The goal of this project is to explore music genre classification using various machine learning techniques.

Data Source

- 106,574 processed tracks from [Free Music Archive](#) (fma) [1].
- 100778 tracks are labeled to 16 top genres.

Data Understanding

- Frequency level: Pitch features (chroma) and Timbre features(mfcc).
- Time level: Rhythm features (temporal).
- In fma dataset, the features are generated using librosa, and stored as statics, including kurtosis, max, min, mean, median, std and skew, for each feature.

K-Nearest-Neighbors

KNN algorithm fits this problem well because the music genres has no clear borders. Trained a KNN on full dataset with different feature sets. The accuracy is as follows:

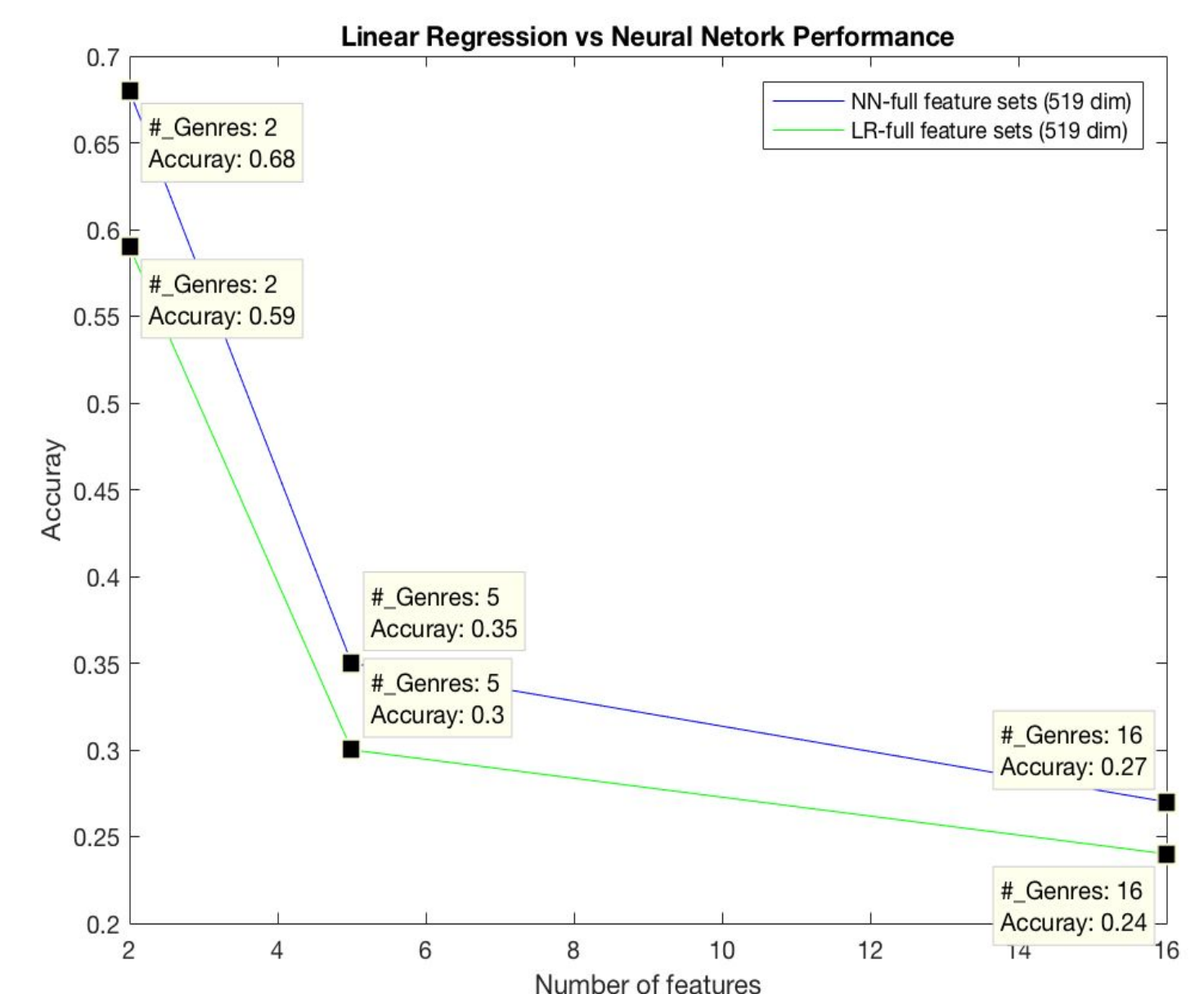
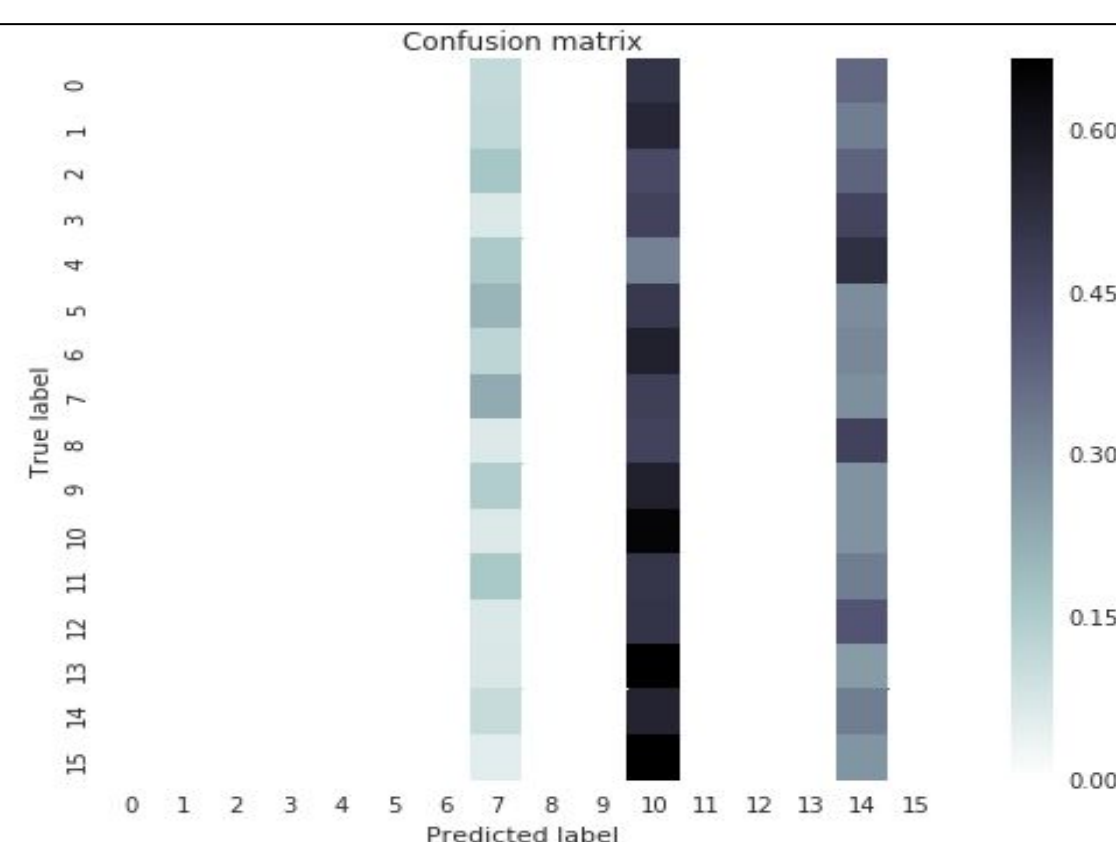
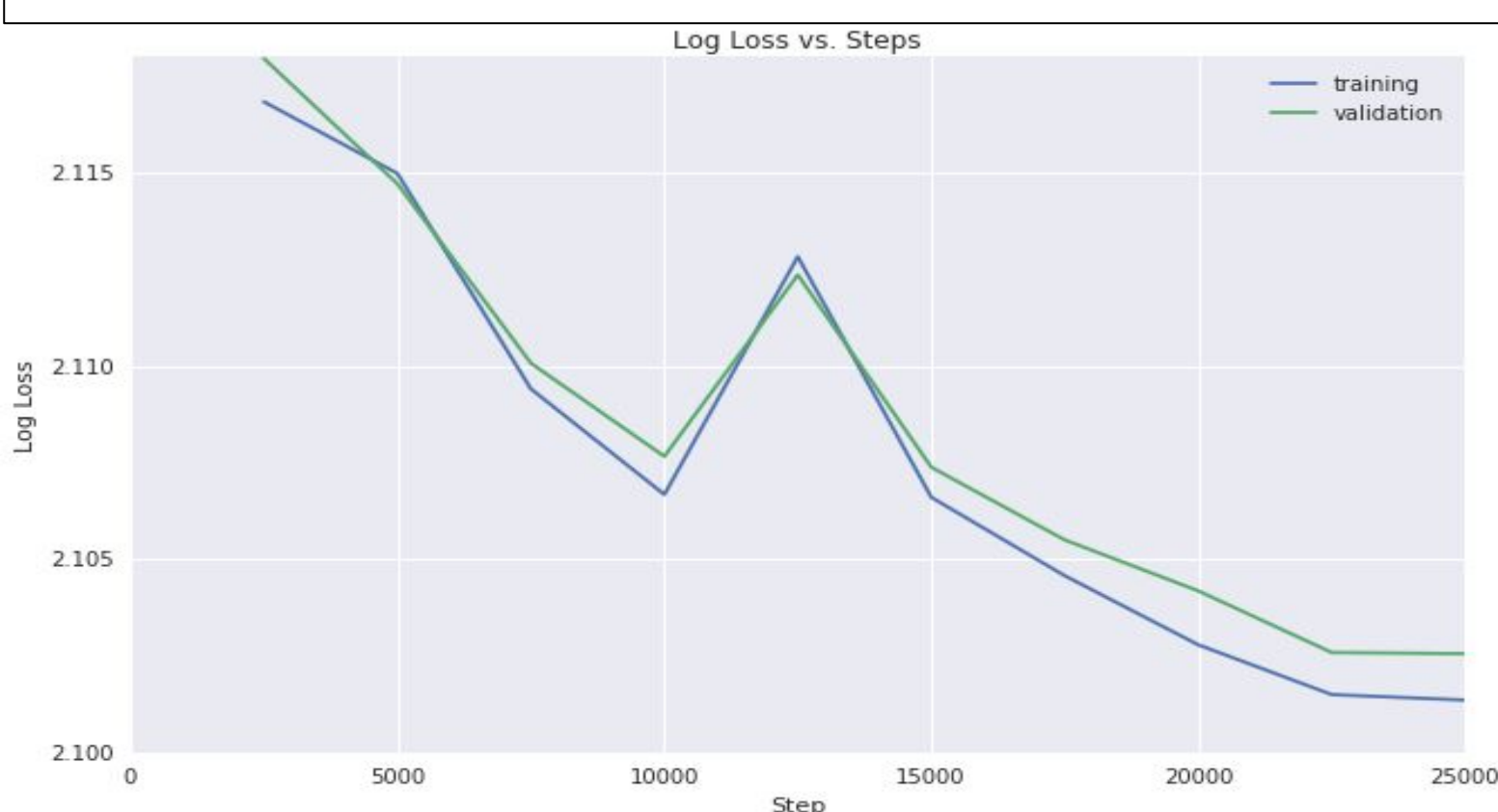
- Full training examples, full features with PCA, 101 neighbors. Accuracy is 46.19%
- Full training examples, chroma feature, 101 neighbors. Accuracy is 30.72%.
- Full training examples, chrom, mfcc, and tonnetz features. Accuracy is 32.06%.

Neural Network

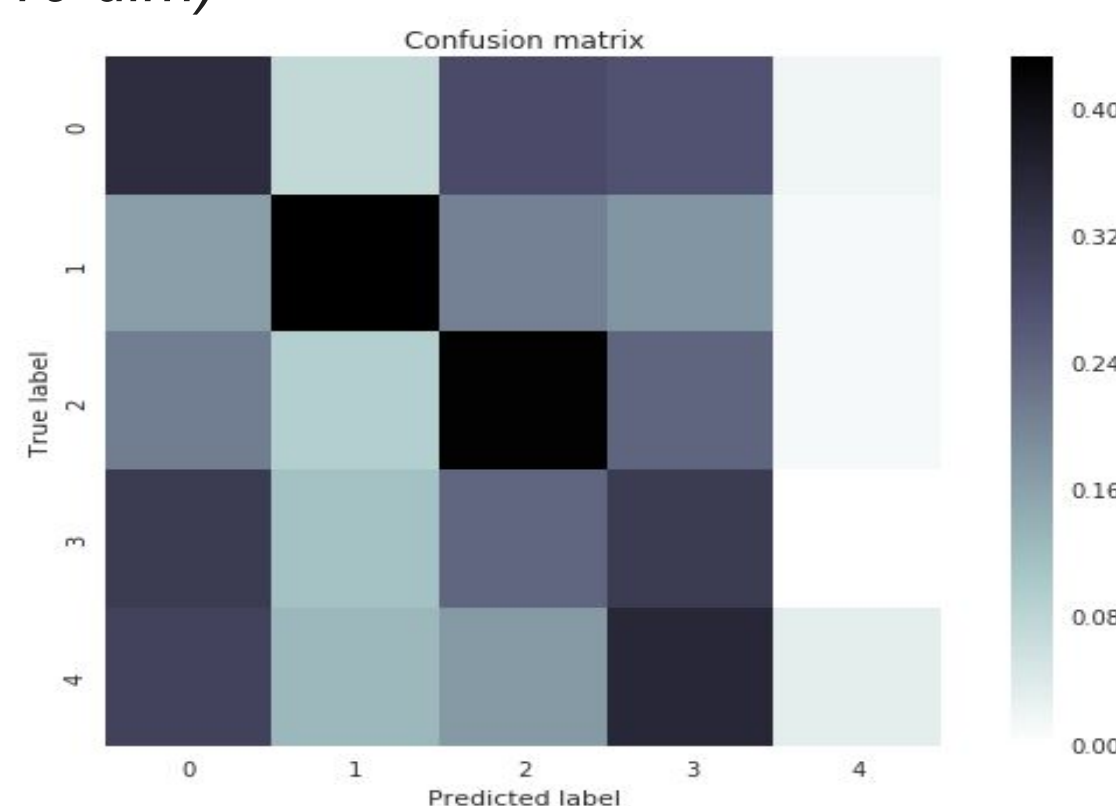
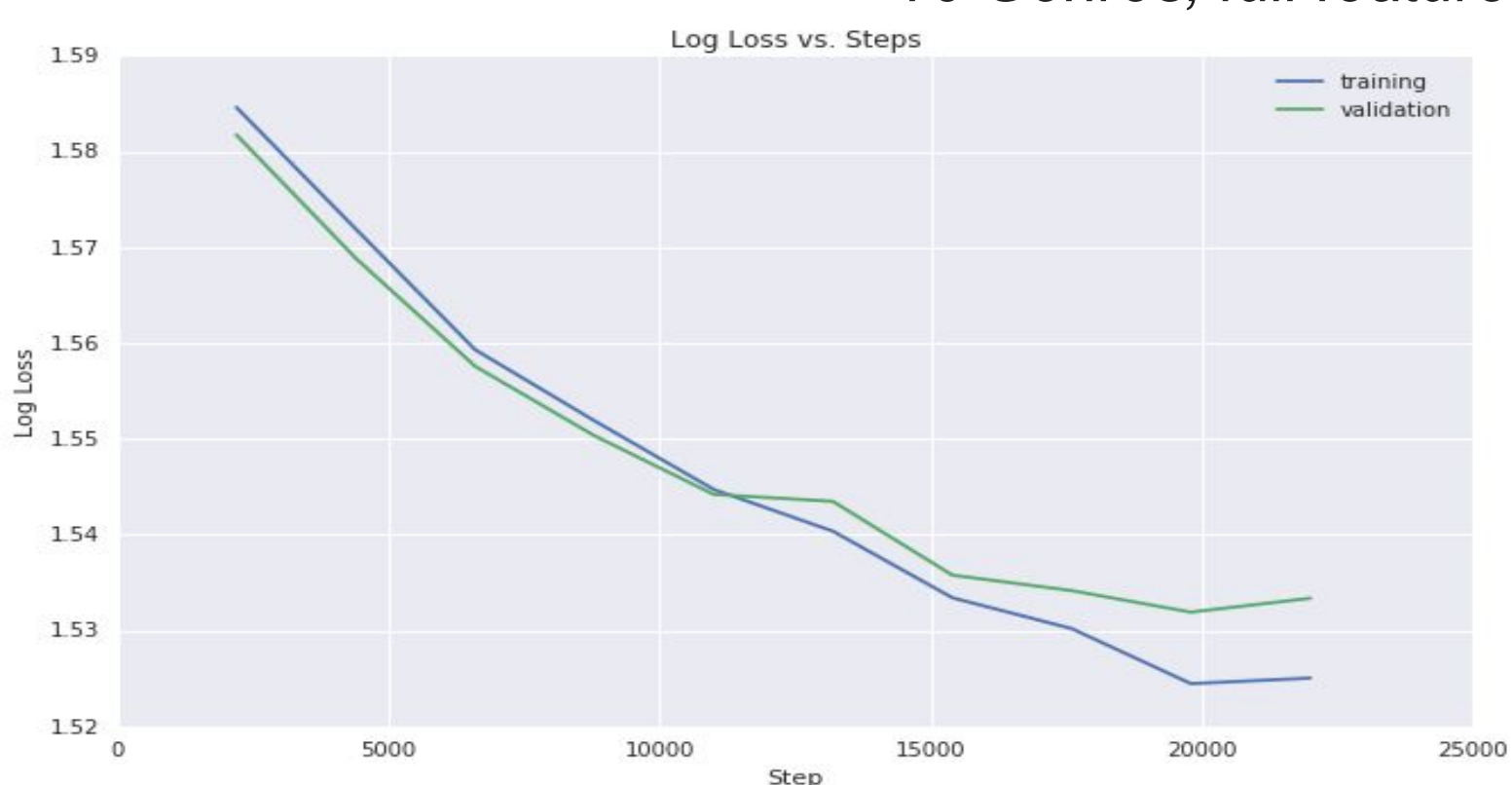
We experimented with various batch size, learning rate, hidden units architecture and regularization methods to train the full data set into 16 genres. But the best accuracy is only 27%. When we focus on train tracks in 5 genres, the accuracy increases to 35%. When we only train tracks in 2 genres, the accuracy increases to 68%

Linear-regression

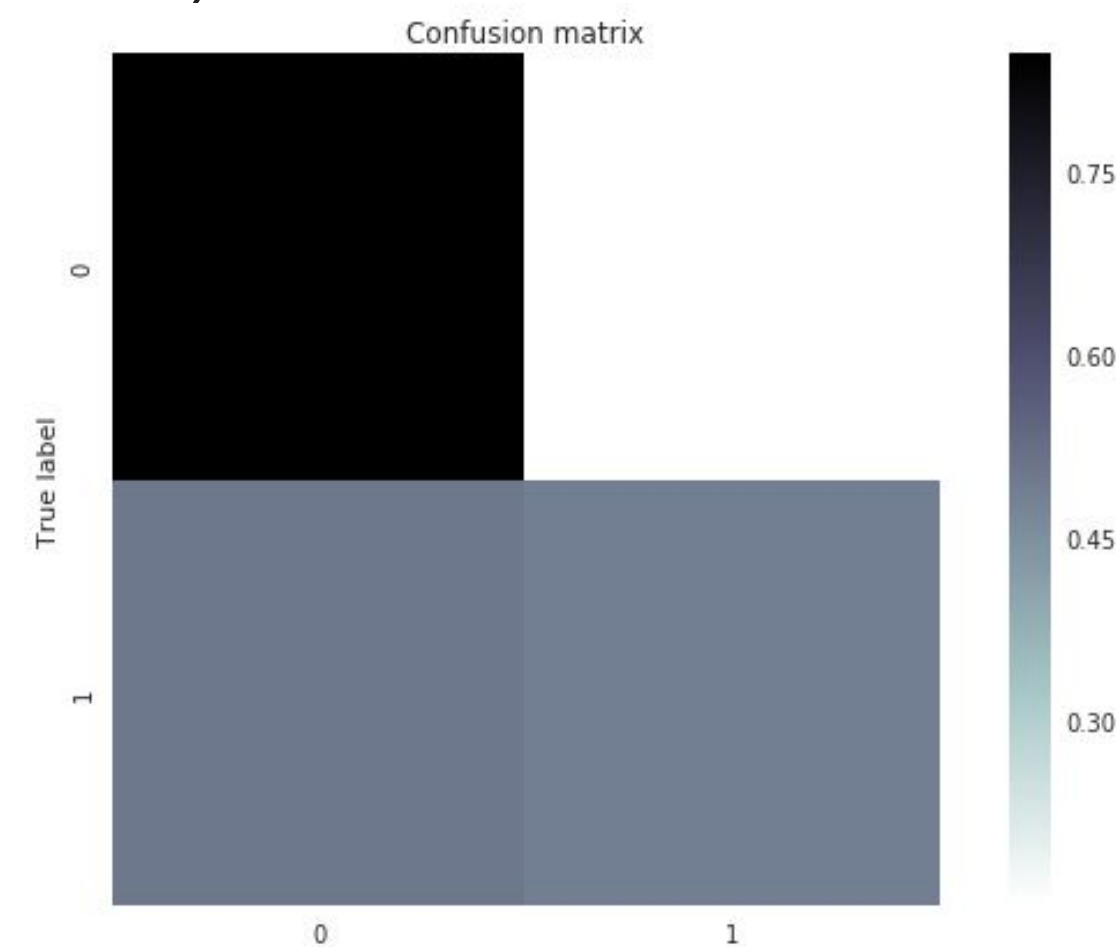
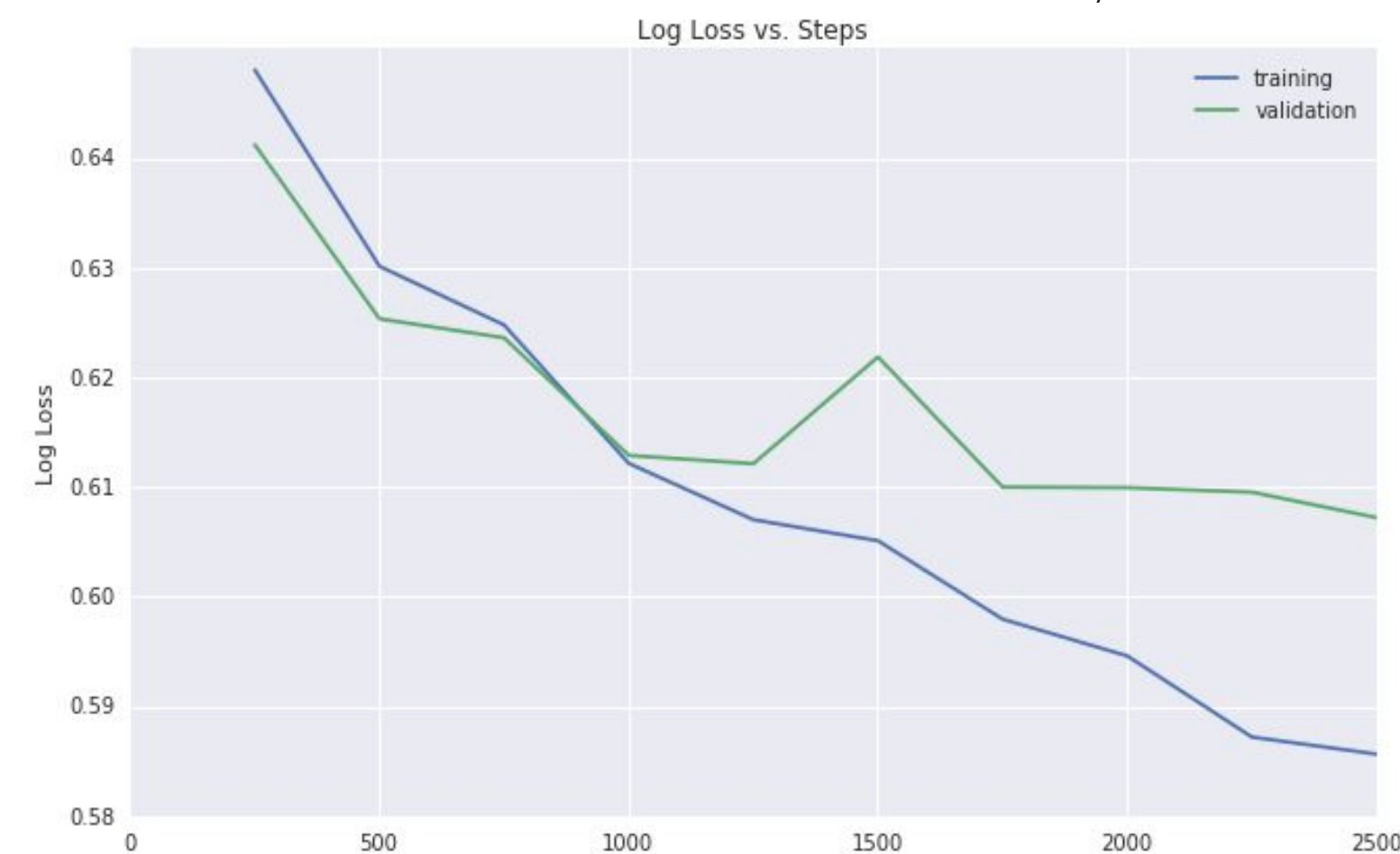
Trained a linear model as a baseline model to compare against Neural Network. The performance is in figure 1.



16 Genres, full feature sets (519 dim)



5 Genres, full feature sets (519 dim)



2 Genres, full feature sets (519 dim)

Conclusions and Future Work

- K-mean performs the best among three algorithms. The testing accuracy of neural network is better than the testing accuracy of the linear regression model. Also, the larger set of the target label is, the better NN outperforms than LR.
- When the target label set is 16, the accuracy of NN model is low (27%). This result might be caused by the fact that the data used are already compressed (statistic data instead of original music frequency). Needs more work to verify. Another guess is because the border of genres are fuzzy. If the overlap part of two genres are large, it is difficult to clarify them.

References

- [1] Defferrard, Michael and Benzi, Kirell and Vandergheynst, Pierre and Bresson, Xavier, FMA: A Dataset for Music Analysis, 18th International Society for Music Information Retrieval Conference, 2017

We further explored different features, chroma, mfcc, and combination of chroma and mfcc, on tracks in 5 Genres (Jazz, classical, hip-hop, rock and country). But none of them beats performance using full feature set.

- The accuracy using chroma only is 24%
- The accuracy using mfcc only is 28%
- The accuracy using chroma and mfcc is 30%. This accuracy is close to the accuracy using full feature set. This result is as expected because chroma and mfcc are two major independent features of music.