Genre Classifier

EECS 352-Machine Perception of Music
Professor Bryan Pardo
www.WhatGenre.Me

Goal

To take in an audio file as input and return the **Genre** of the inputted audio from a pre-determined list of 10 genres

Methodology

- Use GTZAN Genre Collection Dataset
  - Contains 10 genres with 100 labeled songs for each genre
- Run all 1000 songs on an extraction script based on Librosa to extract 37 features for each song and normalize the data.
- Run a k-nearest neighbors algorithm (KNN) to train the classification model.
  - 80 songs from each genre used for training the model
  - 20 songs from each genre used for testing the model
- Use the trained model to predict the classification on the samples selected.

Analysis

We have an overall hit rate of 54.5% with each genre's hit rate ranging from 5% to 90%.

Blues songs are most often classified incorrectly into disco and rock.

Hiphop songs are most often classified incorrectly into disco and pop.

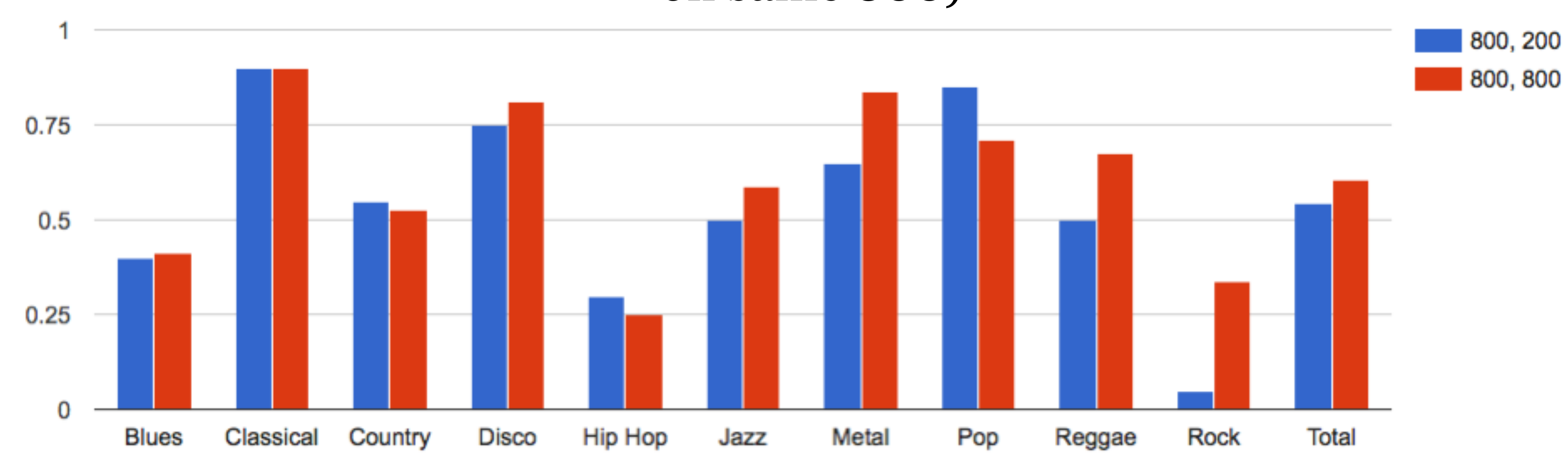
Jazz songs are most often classified incorrectly into country, disco and reggae.

Reggae songs are most often classified incorrectly into disco.

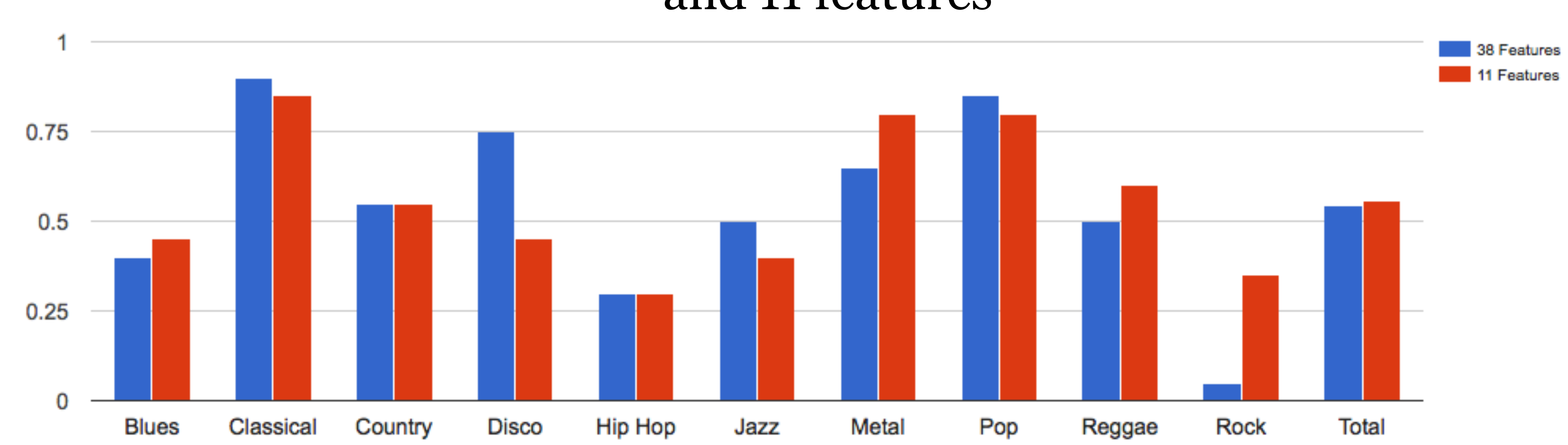
Rock songs are most often classified incorrectly into country, disco and reggae

Results

KNN with 37 Features (800 songs to train, 200 to test, evenly split by genre vs 800 to train, and testing on same 800)



KNN with 800 to train, 200 to test with 37 features and 11 features



Subsequent Testing

We run the 800 songs and test on the 800 songs to see if there is a discrepancy in the setup of the feature vector used.

The result shows that the overall hit rate rises to 60%.

Hypothesis: we have too many features that potentially caused interference and a dampening effect when KNN is running the classification.

We pick 11 features instead of 37 based on its correlation with the standard deviation and mean by genre. It shows us that there is a marginal improvement in overall hit rate.

Conclusion

We use the 11-feature vector as the final model.

- "Fewer Features". We can perform more statistical analysis and find features that correlated well with each other.
- "More Features". We can get more features that correspond to more high-level feature groups such as rhythm and pitch. A more even weight on the distribution on the feature vector will give us better results.
- Capacity of the training data set. If we use a bigger library to train our algorithm, it might be able to achieve better results.
- The defined genres. We could make better definition of the genres. Some genres are outdated, some are sub-genres of another.
- Comparison to Humans. Test our data set on humans, and compare the accuracies of human identification vs computer identification.

Nick David
Yichun Li
Zilun Yu

nicholasdavid2019@u.northwestern.edu
yichunli2019@u.northwestern.edu
zilunyu2019@u.northwestern.edu

Northwestern | McCORMICK SCHOOL OF ENGINEERING