Automatic Music Genre Classification

Indian Institute of Technology, Kanpur
Department of Computer Science and Engineering

Archit Rathore (12152)
Margaux Dorido (EXY1420)

Introduction

- Wikipedia defines music genre as a conventional category that identifies pieces of music as belonging to a shared tradition or set of conventions. Classifying songs according to genre is something that has been till now done by human tagging. The necessity for such a tagged dataset arises for any music search engine that aims to suggest similar songs to the user (e.g. Last.fm or Youtube suggestions).
- Characteristics that define a song to be in a particular genre are usually abstract and often a song may have overlapping genres.
- To automate the task of genre classification one first needs a suitable feature vector to represent the song. We aim to classify genre independent of the metadata (artist information, lyrics etc).

Dataset

- 9 genres x 100 audio clips = 900 audio clips from radio, CDs,…
- 30 secondes audio clips, 22050Hz Mono 16-bit in .au format converted into .wav format to be implemented in python
- Used by G.Tzanetakis and P. Cook in IEEE Transactions on Audio and Speech Processing 2002 for the well known paper in genre classification " Musical genre classification of audio signals "[2]
- 810 audio clips = training – 90 for validation
- 5 fold cross-validation

Approach

Approach 1:

- 6 genres (country, pop, rock, classical, metal and disco)
- MFCC features
- Simple SVM

Approach 2:

- 9 genres (country, pop, rock, classical, metal, disco, hiphop, blues, jazz and raggae)
- Multiple Features
- Ensemble Classifier

Feature Selection and Extraction

- 5 features vectors chosen[4] -> To increase our accuracy
- All features are appended to give a 28 length feature vector

- **Mel-Frequency Cepstral Coefficients**: short term spectral-based features which model amplitude across the spectrum
  - Frame the signal into short frames : *The audio signal will not change too much*
  - Calculate, for each frame, the periodogram estimate of the power spectrum : *To identify which frequencies are present in the frame*
  - Apply the mel filterbank to the power spectra + sum the energy in each filter : *To get an idea of how much energy exists in various frequency regions*

$$M(f) = 1125\ln(1 + f/700)$$

Formula to work with Mel scale

  - Take the logarithm of all filterbank energies : *To make our features match more closely what humans actually hear*
  - Take the Discrete Cosine Transform (DCT) of the log filterbank energies : *To decorrelate the filterbank energies with each others*
  - Keep DCT 13 coefficients , discard the rest : *To remove the higher DCT coefficients which represent fast changes in the filterbank energies*

- **Spectral Centroid**: where the spectral "center of mass" of a sound is. It is calculated as the weighted mean of the frequencies present in the signal with their magnitudes as the weights by using a Fourier Transform.

$$Centroid = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

x(n) : the weighted frequency value of bin number n

f(n) : the center frequency of that bin

Formula to calculate Spectral Centroid

- **Zero Crossing Rate**: the number of times the waveform crosses 0. It is the rate at which the signal changes from positive to negative or back.

$$zcr = \frac{1}{T-1}\sum_{t=1}^{T-1}\mathbb{I}\{s_t s_{t-1} < 0\}$$

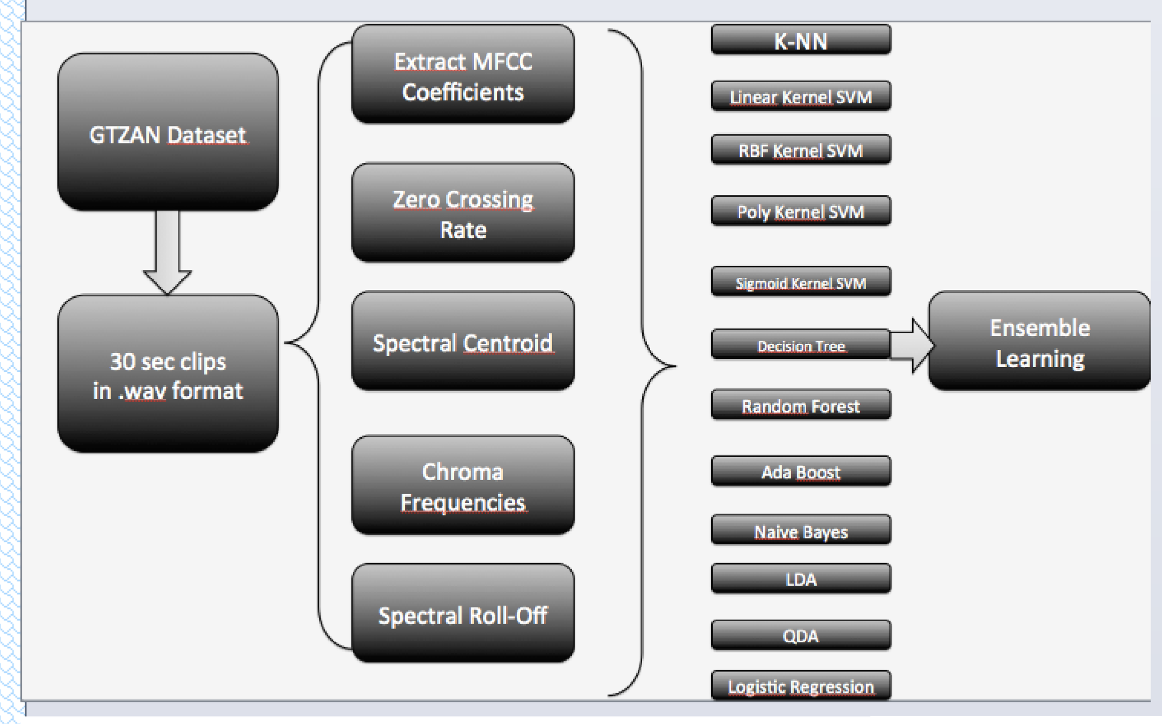
s_t : signal of length t
|| { X } : indicator function (=1 if X true, else =0)

Formula to calculate Spectral Centroid

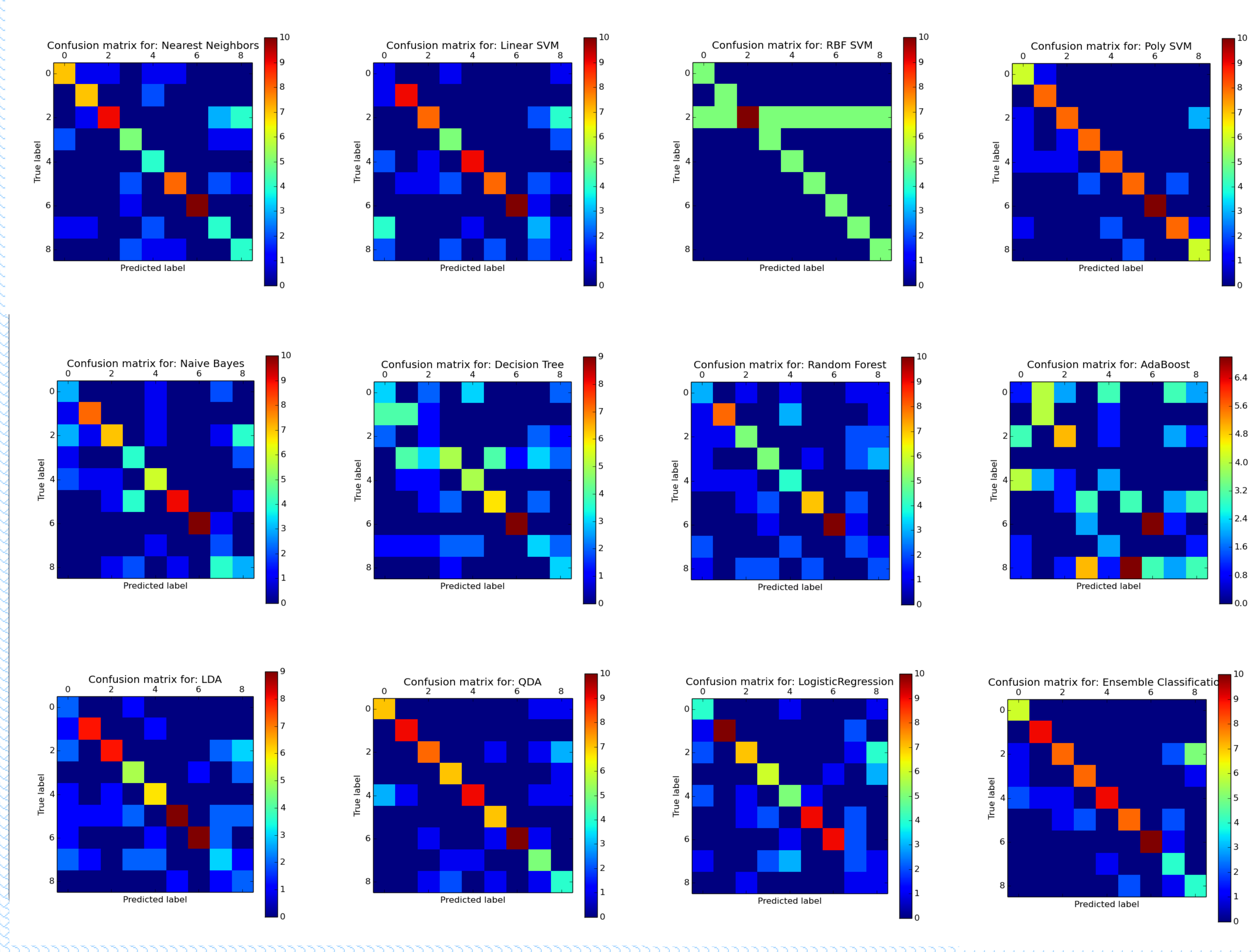
- **Chroma Frequencies**: discretizes the spectrum into chromatic keys, and represents the presence of each key. Creates a histogram of present notes on a 12-note scale.

- **Spectral Roll-Off**: the frequency at which high frequencies decline to 0. It is the fraction of bins in the power spectrum at which 85% of the power is at lower frequencies.

Classification



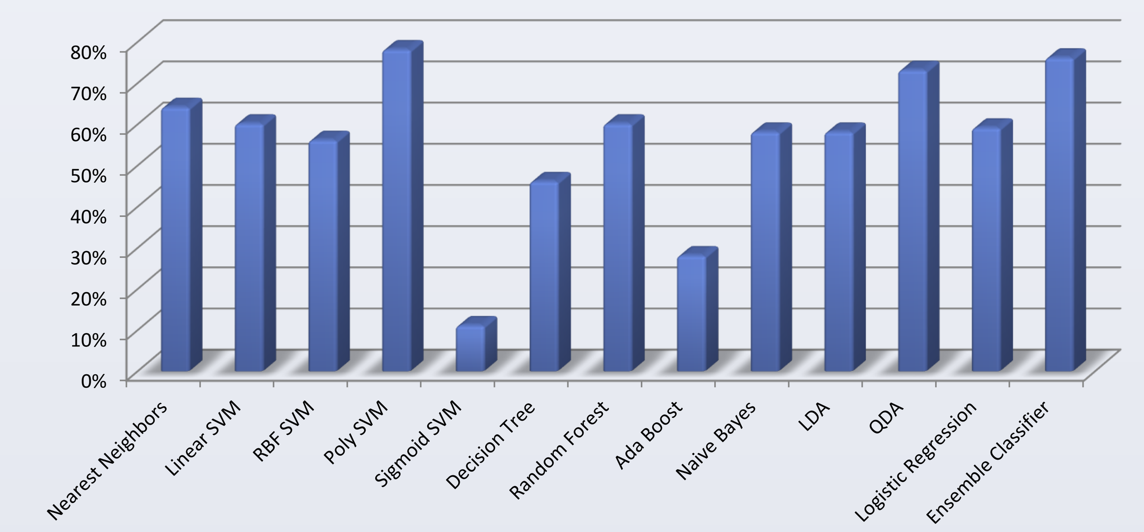
- Using any single classifier often tends to either provide poor results on its own or largely misclassify a subset of the genres.
- Of all the classifiers, SVM with the polynomial kernel provides the highest accuracy and moderate precision except for a few genres.
- We train an ensemble classifier which uses a weighted "hard" voting heuristic to combine the predictions of various classifiers (12 of them). The weights are assigned propotional to the accuracy of each classifier.



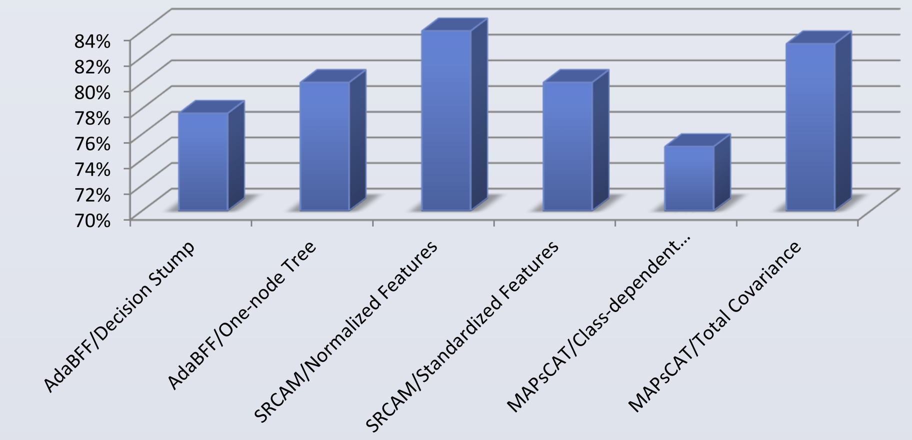
Results and Issues

- ✓ Approach 1 - obtained 82% accuracy on restricted set of genres. But the accuracy reduces on inclusion of more genres.
- ✓ Approach 2 - 9 genres (country, pop, rock, classical, metal, disco, blues, jazz and raggae). Obtained 78% accuracy with Poly SVM and 76% with ensemble classifier. The ensemble method had better genre-wise precision and recall

Accuracy obtained by different Classifiers



Highest verified Accuracy on GTZAN



Conclusion

- ✓ The classifier that works best for our data is SVM with the Polynomial Kernel. Some classifier are very efficient for some specific genres (like Ada Boost which give good results for "pop"). An ensemble classifier is implemented to achieve overall better results.
- ✓ [4] reported an 85% accuracy for 8 genres. Our Results: 78% for 9 genres.
- ✓ The highest verified accuracy on the GTZAN dataset is reported at 84%.
- ✓ The ensemble method improves upon Poly SVM classifier by reducing large errors in classifying specific genres.

References

[1] Changsheng Xu, N.C. Maddage, Xi Shao, Fang Cao, and Qi Tian. Musical genre classification using support vector machines. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)., 2003.

[2] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5):293-302, 2002.

[3] Fang Zheng, Guoliang Zhang, and Zhanjiang Song. Comparison of different implementations of mfcc. J. Comput. Sci. Technol., 16(6):582-589, 2001

[4] Diab, O., Mainero, A. and Watson, R. (2012). 1 Musical Genre Tag Classification With Curated and Crowdsourced Dataset. [online] Available at: http://cs229.stanford.edu/proj2012/DiabMaineroWatson-MusicalGenreTagClassificationWithCuratedAndCrowdsourcedDatasets.pdf