**Music genre classification**

**Name:Chirag Chandnani**

**ID:201501145**

**Introduction and motivation:**

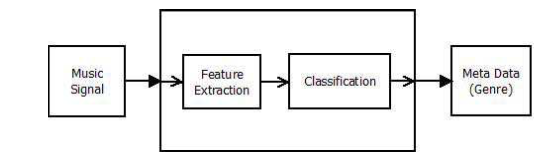
Automatic musical genre classification is very useful for music indexing and retrieval.Wikipedia states that “music genre is a conventional category that identifies pieces of music as belonging to a shared tradition or set of conventions.” Digital music , is one of the important data types distributed on the internet.Genre is an important parameter by which music can be classified apart from artists or album.Genre classification, till now, had been done manually by appending it to metadata of audio files or including it in album info. It can be useful to Recommend songs to user based on the genre he likes (pop,rock,jazz,metal,classical etc).

The aim is : given a piece of music file classify it into a music genre , either pop,rock,blues,metal,hiphop etc.

**Literature search:**

The most influential work on genre classification using machine learning techniques was pioneered by Tzanetakis and Cook. The GTZAN dataset was created by them and is to date considered as a standard for genre classification.. Changsheng Xu et al have shown how to use support vector machines for this task and it has been shown to outperform other algorithms.

**System architecture:**

****

As shown in the figure , given a music signal relevant features are extracted and passed on to a classification algorithm which predicts the genre.

For training the classification algorithm,I have used the GTZAN dataset from the MARYSAS website. It contains 10 music genres, each genre has 100 audio clips in .au format. The genres are - blues, classical, country, disco, pop, jazz, reggae, rock, metal. Each audio clips has a length 30 seconds, are 22050Hz Mono 16-bit files. The dataset incorporates samples from variety of sources like CDs, radios, microphone recordings etc. I split the dataset in 0.9 : 0.1 ratio and used 10-fold cross validation for reporting the results.

**Feature selection:**

Audio sampled at 22000hz . Got 11000 samples .

Audio is processed into 216 frames.So each frame has approx 512 samples which is called hop count.For each frame a value is calculated

After a series of some trials from the features like mfcc,spectral\_centroid,spectral\_bandwidth,spectral\_rolloff,chromatic features,beat features,zero crossing rate, mfcc\_delta**.**

Final feature vector is of length 34.

The 5 following features were found to be useful:

1. MFCC:Mel-Frequency Cepstral Coefficients MFCC represents a set of short term power spectrum characteristics of the sound and have been used in the state-of-the-art recognition and sound categorisation techniques. It models the characterics of human voice.

This features is a large part of the final feature vector (13 coefficients). The method to implement this feature is below :

– Dividing the signal into several short frames. The aim of this step is to keep an audio signal constant.

– For each frame, we calculated the periodogram estimate of the power spectrum. This is to know frequencies present in the short frames.

– Pushing the power spectra into the mel filterbank and collecting the energy in each filter to sum it. We will then know the number of energy existing in the various frequency regions.



Formula to work with Mel Scale

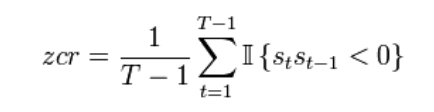
– Calculating the logarithm of the filterbank energies in the previous It enables humans to have our features closer to what humans can hear.

– Calculating the Discrete Cosine Transform (DCT) of the result. It decorrelates the filterbank energies with each others

– Keep first 13 DCT coefficients. We remove the higher DCT coefficients which can introduce errors by representing changing in the filterbank energies

From PracticalCryptography.com tutorial

2.Zero crossing rate: It represents the number of times the waveform crosses 0. It usually has higher values for highly percussive sounds like those in metal and rock



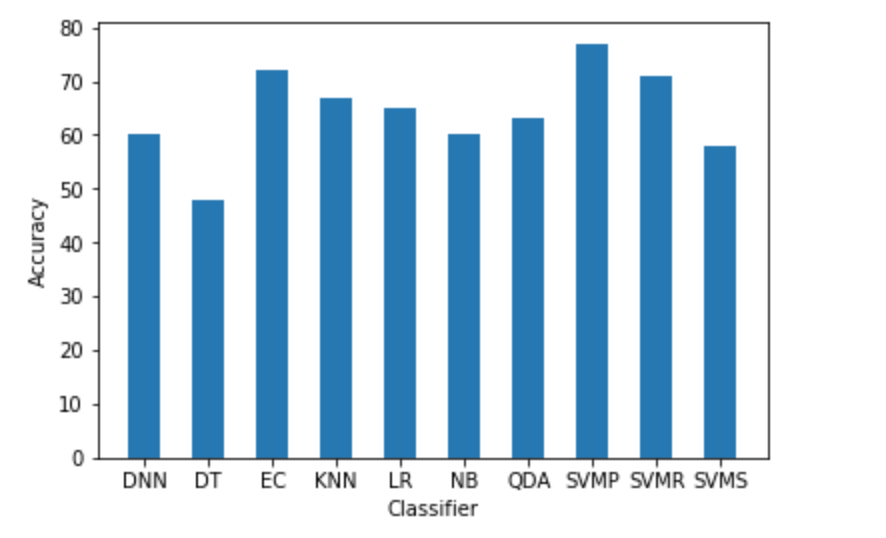
3.Spectral centroid:It describes where the ”centre of mass” for sound is. It essentially is the weighted mean of the frequencies present in the sound

4. Spectral rolloff:It is a measure of the shape of the signal. It represents the frequency at which high frequencies decline to 0. To obtain it, we have to calculate the fraction of bins in the power spectrum where 85% of its power is at lower frequencies.

5.Spectral bandwidth:It is the wavelength interval in which a radiated spectral quantity is not less than half its maximum value.It is a measure of the extent of the spectrum

**Results and analysis:**

Various classification algorithms were applied on the features extracted on the dataset.It was observed that any single classifier did not classify all the genres well. For example in the SVM with polynomial kernel worked well for most genres except disco and hiphop.



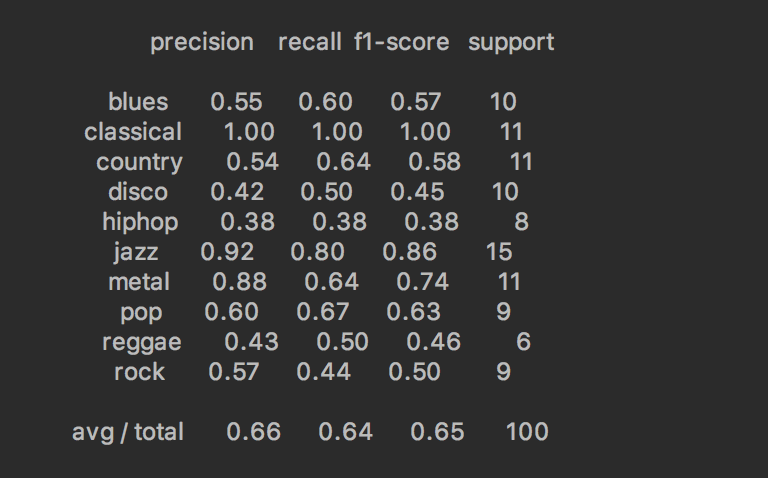
DNN-Deep neural networks,DT-Decision trees , EC- Ensemble classifier consisting of SVM and KNN . KNN-K Nearest neighbours , LR-Logistic regression

QDA-Quadratic discriminant analysis, SVM - P - Support vector machine with polynomial kernel SVM-R - Support vector machine with radial basis kernel

SVM- S - support vector machine with sigmoid kernel.

Decision trees performed the worst with around 50% accuracy.SVM with polynomial kernel performed the best with around 77% accuracy.All other classifiers gave accuracy in the range of 60-70%.Highest verified accuracy on the GTZAN dataset is 84% uptil now

Support vector machine with polynomial kernel:



The classifier is more accurate for predicting

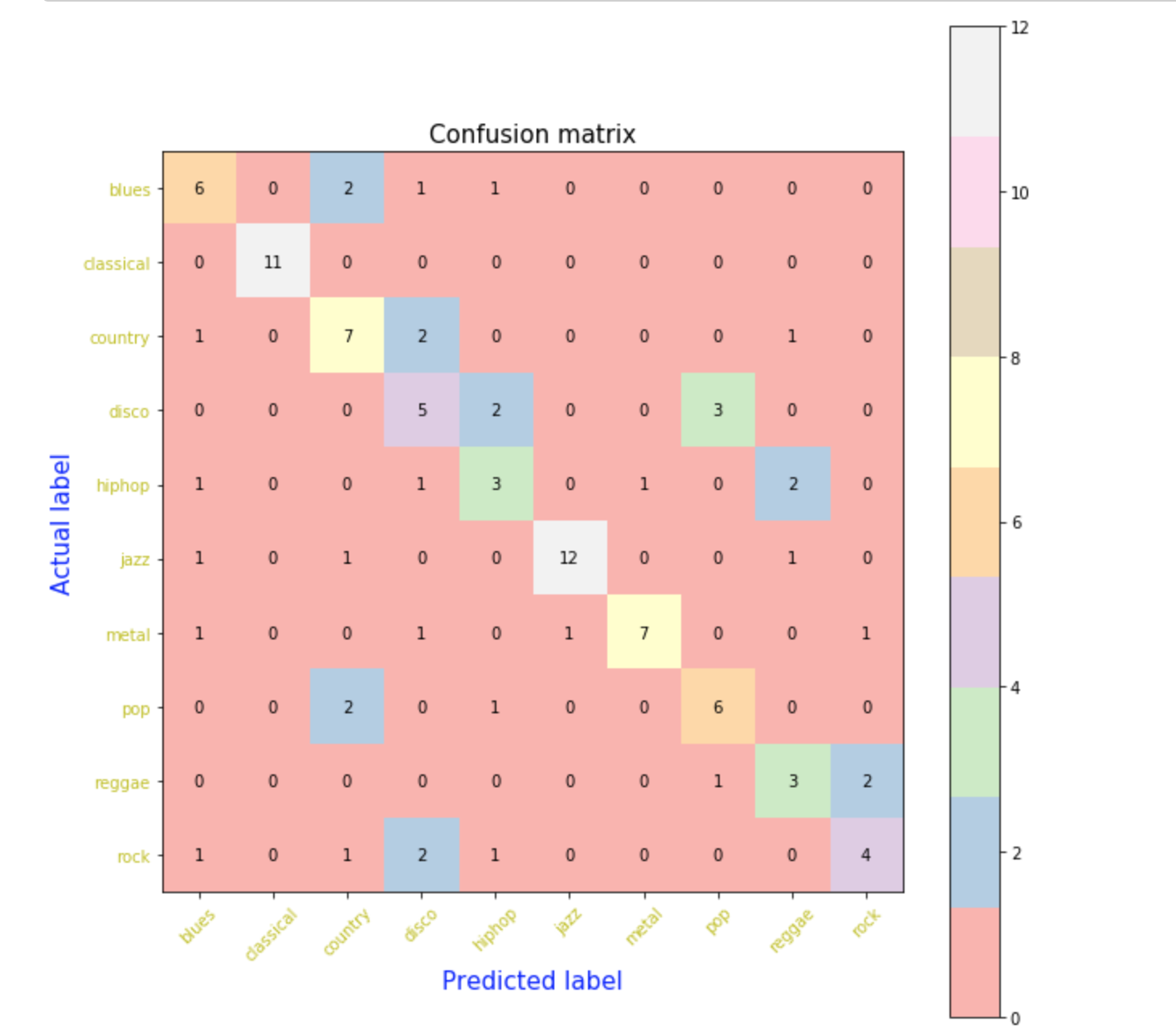
Classical,jazz,metal ,rock and pop while poor for

predicting disco,hiphop . KNN performed well for pop with 0.83 precision.The same is shown through confusion matrix

Precision=TP/TP+FP

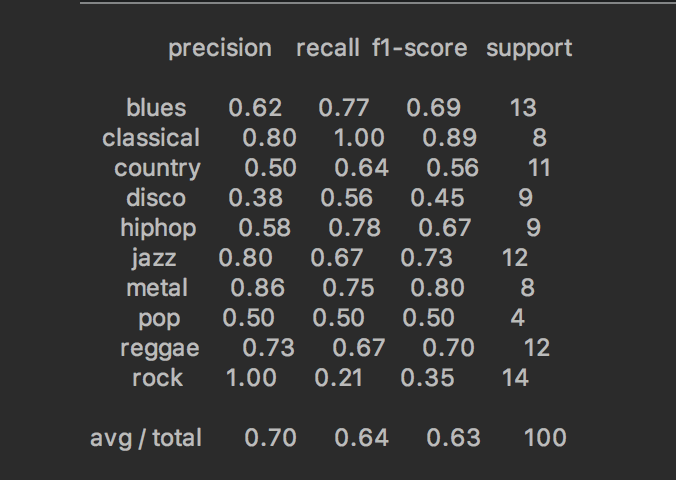
Recall=TP/P . P=TP+FN

F1=2/(1/Precision+1/Recall)



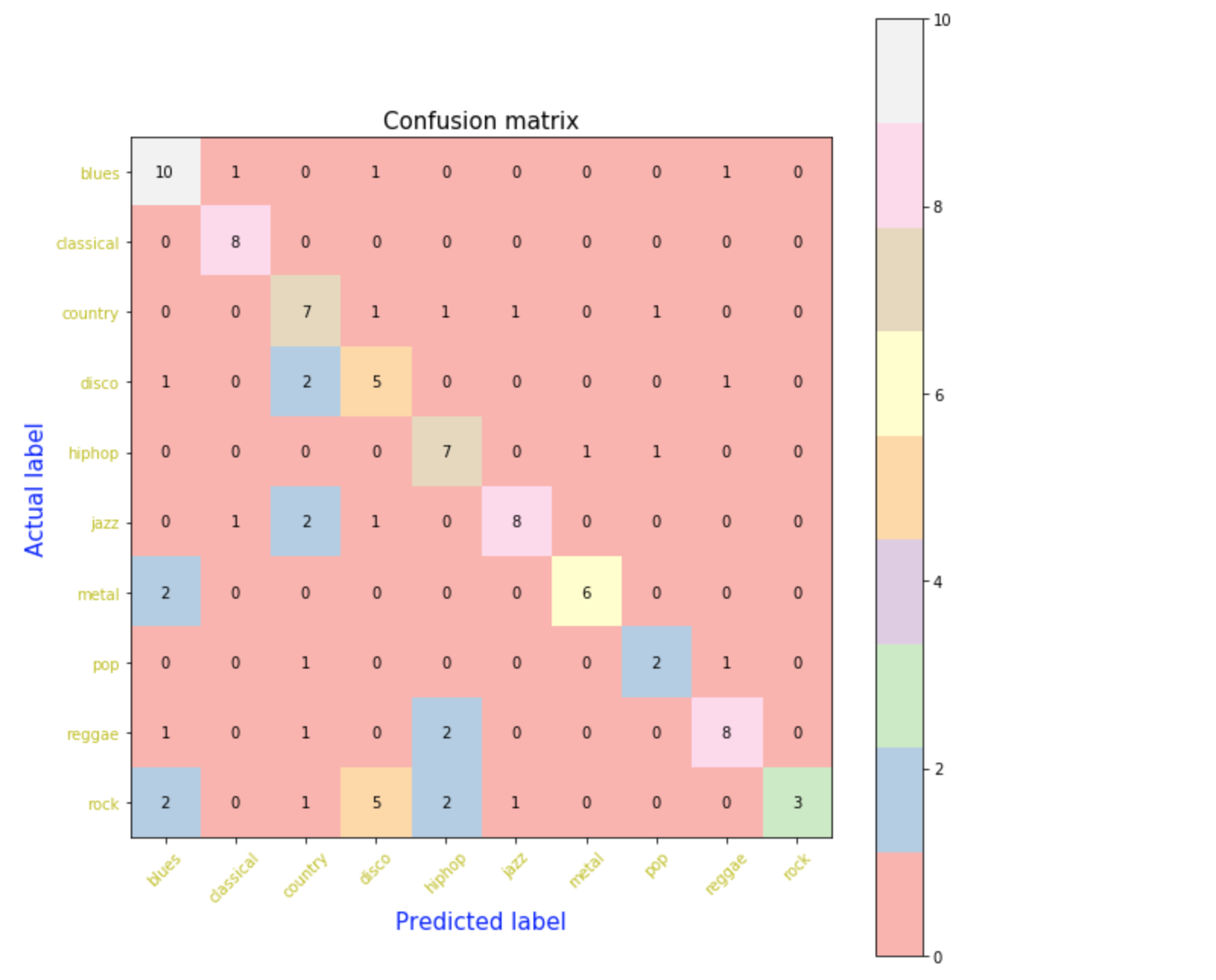
Ensemble classifier:

This ensemble classifier used the prediction of each classifier and run a majority voting heuristic to obtain the optimal class label for given test input.The classifiers used in ensemble are SVM with polynomial kernel,SVM with radial basis kernel,K-NN.



Ensemble performs well for classical,jazz,metal,rock,hiphop but not for disco,pop

Precision=TP/TP+FPRecall=TP/P . P=TP+FNF1=2/(1/Precision+1/Recall)



**Improvements:**

Ensemble classifier improves upon the svm classifier in some specific genre and overall performs well

The ensemble seemed to give a better overall precision but the accuracy dropped by a small amount

**Limitations:**

For some genres the classifiers doesn’t work well.For music with some noise the classifier fails as observed in the research paper.For music consisting of two genres we can classify only one genre..Finding out the proper set of features which works is difficult

**Further improvements:**

The classifier that works best is SVM with Polynomial Kernel. Some classifier are very efficient for some specific genres**.**Use of CRNN as proposed in the another research paper.Use of a larger dataset for training .Trying some other features which may improve the accuracy.Classifying live music.To classify songs consisting of a combination of genres

**References:**

[1] Xu, Changsheng & Maddage, Namunu & Shao, Xi & Cao, Fang & Tian, Qi. (2003). Musical genre classification using support vector machines. 5. V - 429. 10.1109/ICASSP.2003.1199998. <http://ieeexplore.ieee.org/document/1199998/>

[2]Choi, Keunwoo et al. “Convolutional recurrent neural networks for music classification.” *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2017): 2392-2396.

[3]<http://marsyasweb.appspot.com/download/data_sets/>