Multi-class Music Genre classification using Machine Learning techniques

Name:	**Sarthak Mishra**
Registration No./Roll No.:	18388
Institute/University Name:	IISER Bhopal
Program/Stream:	BS-MS in Physics
Problem Release date:	February 02, 2022
Date of Submission:	April 24, 2022

1 Introduction

Music genre is a key feature of any song that can guide users to their preferred category. In this work, I apply several machine learning techniques on the given data set to classify 10 music genres based on various input features. The goal of this work is that this genre classifier can be used to correctly classify a new music track given its associated features into a plausible music genre. A curated data set has been used for building the music genre classification system and implementing further calibration on the system such as error analysis. Then, several models such as Gaussian Naive Bayes (GNB), k-Nearest Neighbors (kNN), and Support Vector Classifier (SVC) with optimized parameters have been used to train the data set.

2 Preliminary Experiments

2.1 Gathering Data and Data Visualization

The data set includes 900 tracks each trimmed to 30s. They are arranged into 10 genres, as shown in the figure 1. Each track is described by 58 attributes including the length of the audio track.

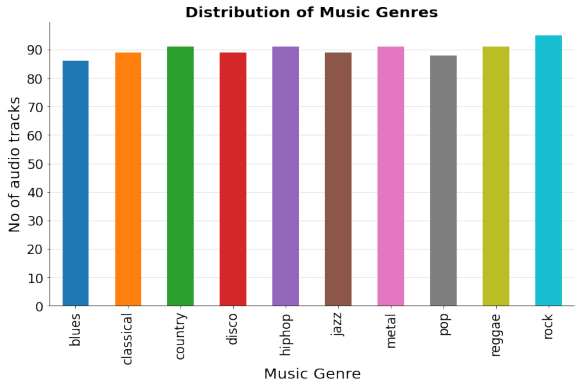

Figure 1: Distribution of Music Genres

2.2 Data Preprocessing

The raw data was very much imbalanced and skewed with standard deviation ranging from as low as 10^{-5} to as high as 10^{5} in some features, employing a `MinMaxScaler` and scaling the features between 0 to 1 solves the issue. And then the 900 tracks were splitted into training set and test set with

Table 1: Performance Of Different Feature Selection Algorithms

Algorithm	GNB	kNN	SVC
Chi2	0.533	0.652	0.752
PCA	0.607	0.667	0.748
DT	0.433	0.593	0.615

sizes of 630, 270 rows respectively proportional to the distribution of labels with the use of `stratify` parameter. Therefore, training examples were represented as a matrix of 630 rows and 58 columns, and a label of the genre.

2.2.1 Feature Selection

In order to overcome complexity and overfitting, The following feature selection algorithms were used: chi-square test (Chi2), Dimensionality Reduction using Principal Component Analysis (PCA), and Decision Tree (DT) for Feature Selection algorithm. The accuracy of various models are reported as per the given table 1. Chi2 returned 35 selected features out of 57 initials, whereas PCA reduced it to 14.

A chi-square test is a statistical test used to compare observed results with expected results. The values are scaled and it is a categorical data of frequencies of occurrences of various genres in the labels, the chi-square test gives the best possible result. Thus, the 35 features selected is carried forward to model evaluation explained in further steps.

3 Methods

I have broadly explored three baseline models: Gaussian Naive Bayes, k Nearest Neighbors and Support Vector Classifier with linear kernel [1]. The initial results were tabulated as follows 1. The C parameter found for the linear kernel was much smaller than for the other two kernels, indicating that the linear kernel is less able to perfectly separate the classes. Model selection produced desirable results where only a subset of all features were required to minimize classification errors.

3.1 Support Vector Classifier

Support Vector Machines constructs a decision boundary using the input data set so that the minimal distance from data points to the decision boundary is maximized. In this project I have implemented SVM with linear kernel with initial setting to be regularization parameter, C=10, kernel coefficient, gamma=0.08.

3.1.1 Hyperparameter Tuning

Upon this I have run a `GridSearchCV` to find the best setting to give best possible results. I have used `SelectKBest(chi2, k='all'` pipeline using all selected features. The parameters being fed into it were `'clf__C': (0.1,1,10,100),` and `'clf__kernel': ('linear','poly','rbf','sigmoid'),`. It takes all these ensemble of parameters into account and runs through each setting the `GridSearchCV`. The `scoring` was done based on 'f1_micro' and 'accuracy' as well as. `scoring='f1_micro'` gives the best result. The optimal regularization parameter was found to be SVC(C=100, class_weight='balanced', probability=True) with a Medium Probability of Confidence which is a good result.

3.2 k-Nearest Neighbors

K-nearest neighbors is a simple non-parametric classification technique. The number of neighbors, k, controls model flexibility and adjusts the bias variance trade off. [2]

The performance of various neighbors and their accuracy have been shown in the figure 2.

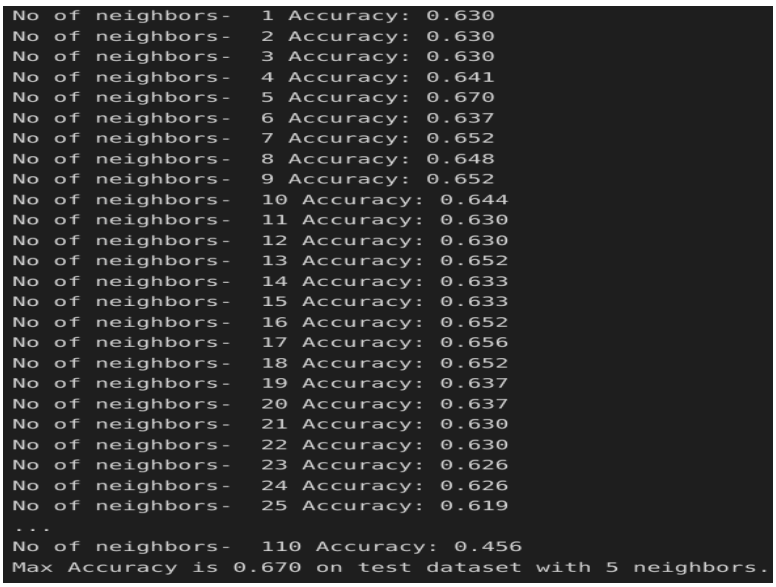

Figure 2: Accuracy of various neighbors in kNN

Table 2: Performance Of Support Vector Classifier Using All Terms

Classes	Precision	Recall	F-measure	support
country	0.67	0.77	0.71	26
classical	1.00	1.00	1.00	27
pop	0.71	0.74	0.73	27
metal	0.52	0.59	0.55	27
jazz	0.76	0.48	0.59	27
reggae	0.89	0.93	0.91	27
disco	0.89	0.93	0.91	27
blues	0.77	0.77	0.77	26
rock	0.65	0.74	0.69	27
hip hop	0.62	0.52	0.57	29

3.3 Gaussian Naive Bayes

Gaussian Naive Bayes inputs continuous valued features and models each as conforming to a Gaussian (normal) distribution. I have used this model along with Decision Tree Classifier for Feature Selection and then Gaussian Naive Bayes model evaluates the model performance. However, the results were not good because there were multiple correlated values and the tree greedily chooses only one out of them reducing the feature dimension to as small as 14 from large feature columns of 57. [3]

4 Evaluation Criteria

In the case of multi-class classification, we adopt averaging methods for F1 score calculation, resulting in a set of different average scores (macro, weighted, micro) in the classification report. Precision is a metric that quantifies the number of correct positive predictions made whereas Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made. F1Score (or simply F-measure) is the harmonic mean of precision and recall for a more balanced summarization of model performance. The performance of the best performing classifier has been tabulated in the following table 2.

Table 3: Genre Classification Accuracy

Model	Train(%)	Test(%)
GNB	74.2	54.1
kNN	78.1	67.0
SVC	99.7	77.4

5 Analysis of Results

Each of the models was evaluated using the same training set of 630 examples, and test set of 270 examples. Feature selection and regularization are conducted for each model. The training accuracy and test accuracy are shown in following table 3. SVM model with linear kernel has the highest test accuracy. We can also see that KNN classification technique gives a low test accuracy, compared to other models. This is because with increasing dimensionality, the volume of feature space increases and training examples become sparse. The confusion matrix of the SVC-linear displayed below shows test prediction accuracy for each genre.

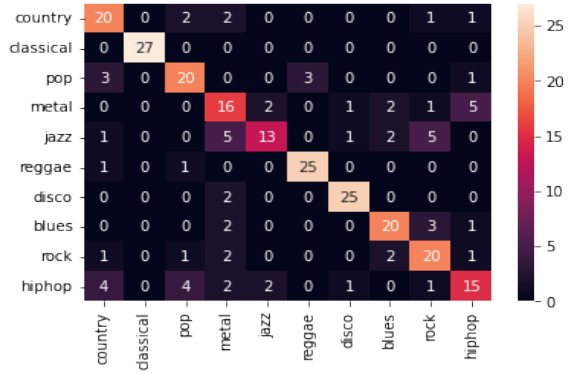

Figure 3: Confusion Matrix

6 Discussions and Conclusion

As a first step to a real world machine learning problem, I have implemented several machine learning techniques to classify music genres. However, no single model is the best. There is always a scope for further improvements. There are a number of extensions that can be done. 1) Try and explore the SVC-RBF kernel along with other models such Random Forest Classifier, Logistic Regression, which I could not because of time constraint. 2) Gathering more data for genres with less data currently to balance the data distribution. Extending the range of labels to more than just ten as of current data set.

References

[1] Hareesh Bahuleyan. Music genre classification using machine learning techniques. *arXiv preprint arXiv:1804.01149*, 2018.

[2] Li Guo, Zhiwei Gu, and Tianchi Liu. Music genre classification via machine learning.

[3] A.Parveenbanu Mrs.Jetlin CP M.D.Nevetha, A.Nithyasree. Music genre classification using machine learning. 9:155.