Project: Music Genre Classification

Name: Pranav Kandwal
Roll no.: 20208
Name: Saksham
Roll no.: 20240
Institute name: IISER-B
Department: EECS
Instructor: Dr. Tanmay Basu
Problem/Release Date: 12/1/2023
Date of submission:15/4/2023

Introduction

We are provided with a labelled training data set X(matrix of order $900 * 59$) where (X_i,y_i) be an instance in it. X has 900 data points(X_i). Where X_i,∀ i ∈ $\{1, 2, ..., 900\}$ be a m-dimemsional feature vector given as $[x_{i1}\ x_{i2}\ x_{i3}\ .\ .\ .\ .\ x_{im}]$ where m be the total no. of features. Here m=58. Here $x_{ij} \forall\ j \in \{1, 2, ..., 58\}$ be individual features and in the given data set the features are numeric. And $y_i, \forall\ i \in \{1, 2, ..., 900\}$ be the target variables which are categorical.

The given problem falls under the classification problem.Here we need to classify/predict the target variable of the data point, where the target variables are discrete. Hence our task is to build a classifier ,i.e., a mathematical model and train its parameters through the given labelled training data set and thereby producing an infered function so as to map a new data point or a test point to the most suitable discrete class.

Methods

Before fitting our model, we performed feature selection based on covariance criterion with the assumed threshold of 0.85, thereby removing the redundant features and making our data set cleaner. We also scaled our data set in the range(-1,1) to ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent(η) are updated at the same rate for all the features. Having features on a similar scale will help the gradient descent converge more quickly towards the minima.

We have worked with the classifiers which we studied during the coursework and chose the relevant ones which best suit our classification problem. We have used the following classifiers for our problem:

A) **Decision tree Classifier**.
B) **Random Forest Classifier**.
C) **Support Vector Machine**.
D) **k-Nearest Neighbour Classifier(KNN)**.
E) **Multinomial Logistic regression**.
F) **Ada Boost Classifier**.

We have run each of these models on our data set which is divided broadly into 4 major parts:

1) Without feature selection and without scaling.
2) Without feature selection but with scaling.
3) With feature selection but without scaling.
4) With feature selection and with scaling.

We first dry run the model on our data set with the default parameter values and then improved their peeformances by hypertuning them using **GridSearchCV()** and **RandomizedSearchCV()**. With the learnings in the coursework, we fine tuned our models by giving a set of important hyperparameter values to the model and mofifying the set untill we get a satisfactory result. We have, however not run the SVM on data set without scaling as it was taking more than 5 hrs and the runtime was also crashing out. But we have included it in the code within the comments.

We have found that the data set when performed feature selection and scaling(StandardScalar()) has shown the best metric statistics in kNN, Random Forest Classifier and SVM while also performing second best in other classifiers.

Github link of the project

Experimental Analysis

Classifier	Macro-Average Precision	Macro-Average Recall	f1-Score	Accuracy
Decision Tree	0.58	0.58	0.58	0.5786
Random Forest	0.76	0.75	0.75	0.7528
SVM	0.70	0.70	0.70	0.6966
Multinomial Logistic Regression	0.70	0.70	0.69	0.6969
kNN	0.69	0.67	0.67	0.6685
Ada Boost	0.79	0.79	0.78	0.7865

During the evaluation of the model and choosing the set of hyperparameter values we have used **f1 macro** as the scoring parameter in the grid and randomized search algorithms. The reason for choosing the same is that since we have a multiclass classification problem we have chosen macro average statistics as also stated in the report guidelines. We rely on f1 score due its equity for precision and recall as it is the harmonic mean of the two. Also it is a better evaluating metric than accuracy as we have discussed in the class by citing the example of classification problem of Malignant and Benign Tumour.

We found the best accuracy and f1 score for **Ada Boost Classifier** with Random Forest Classifier as its base estimator which came around 0.79 and 0.78 respectively. Here the data set used is without feature selection but with scaling. Moreover we hypertuned our model for fitting this data set and got $n\_estimators$=10 and $learning\_rate$=0.3 and thereby chose this model as the best model for test data set.

Discussions

Looking at the merits of the proposed method, we can say that the methods used are simple and easily interpretable. We have used simple yet rigorous machine learning algorithms and dived a pretty deep into them by not just using them blindly but also tuning the hyperparamaters of each carefully and appropriately. **Application:** With the rapid advancement of technology espicially AI(Artificial Intelligence) in the music industry worldwide, one can find the very need of this work. Where by looking at the features of the music genre such as spectral statistics, one can report the corresponding form of music such as hip-hop or rock and tune the audio devices accordingly as per the requirement of the artist.

Since there is always a scope of improvement, here also there's a room of further evolving our methods used. We're thinking to explore more ensembler classifiers like XGM and CatBoost and also deep learning models like ANN(Artificial Neural Network) or Multilayer Perceptron algorithm. We'll also be likely to explore more about feature selection which plays a key role specially when we are dealing with such a large number of features as here and just can't rely on the covariance between them as the sole indicator of selecting of features.

We have also found out some significant observations by looking at the spatial distribution of features. For example for chroma statistics, one of the important aspects in music blending, one can infer that metal music is more likely to have the lower statistics. We also report that the f-measure for "classical" genre in the given data set is the highest whilest that of "rock" genre is the least. Which means that our model fits the data points belonging to "classical" class really well but fails to give the similar convincing results for the "rock" class. Hence we'll be carefully exploring the feature vectors belonging to the low f-score classes as further improvement to it.

References

Machine Learning by Tom M. Mitchell.
Class notes(DSML Spring 2023) by Dr.Tanmay Basu.
https://towardsdatascience.com
https://www.kaggle.com/
https://www.analyticsvidhya.com
https://machinelearningmastery.com
https://scikit-learn.org