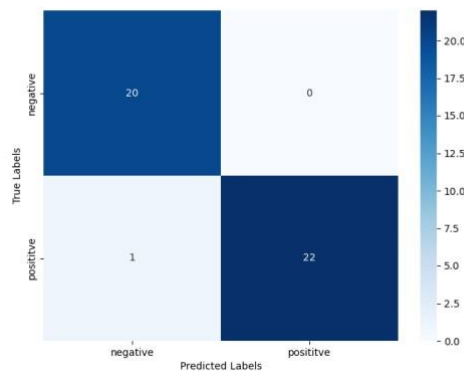


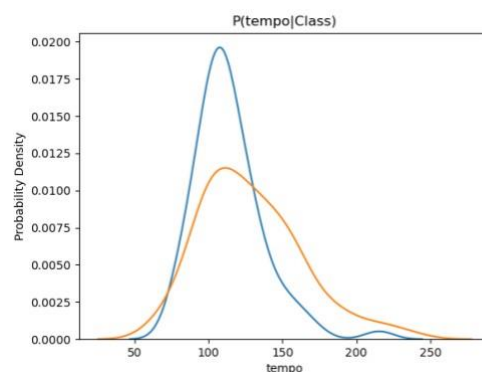
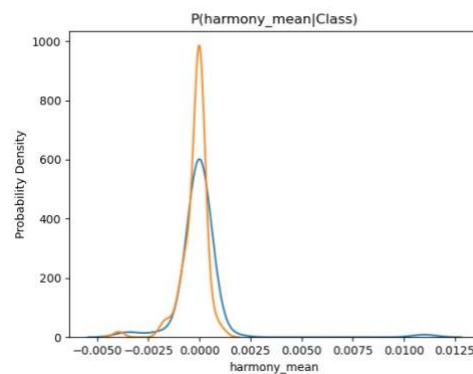
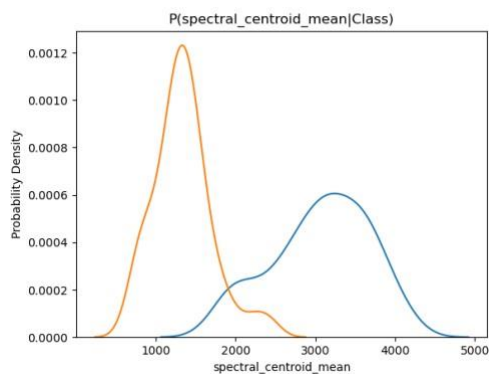
Machine Learning Assignment 1

Task 1 Pop vs. classical music classification

1. The model in general predicts the classes with accuracy of 0.977, precision of 1.0 and recall of 0.957. Meaning the model is a good fit for the given dataset and accurately classifies the data into the appropriate categories. The model has high level of confidence when it predicts positive instances (i.e., classical genre). For this particular dataset, the model does not have any false positive predictions.



2. Below are the respective probability density functions, $P(X|Class = pop)$ and $P(X|Class = classical)$; If I had to choose one of these features to classify data.



I would choose spectral centroid mean to classify the data because after observing these graphs, spectral centroid mean graph has the most separation between the two probability density functions.

As more separation between the two implies that this feature can more accurately distinguish between the two classes. Also, this feature has a strong correlation with the class labels. So, using this feature would result in a more accurate model.

Task 2. 10-way music genre classification

6. Overall accuracy of the model is 0.29. Enclosed a confusion matrix of the dataset which allows us to analyze model's performance. Modified my Naïve Bayes' model to handle missing values. Whenever the model observes a 0/Null value it would just continue the loop and calculate posterior probability of the attribute without the missing values. To test how robust the model is, I created an array which contained number of missing values in the dataset. After that, every time it randomly chooses the column and indexes in that column and the replaces the values with 0/Null. Then pass that modified test set into the model which then outputs the accuracy of the model. The model's performance decreases as number of missing values increases in the test set. The model behaves as we expect it to behave. The plot below shows the evidence to the claim but shape of the graph changes due to randomness.

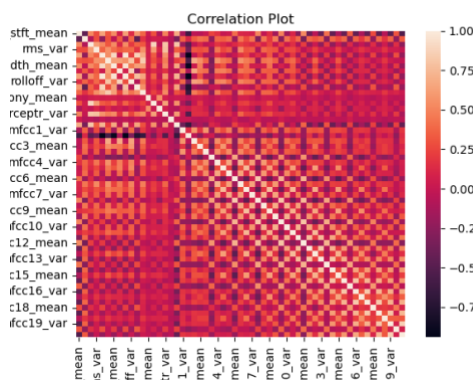
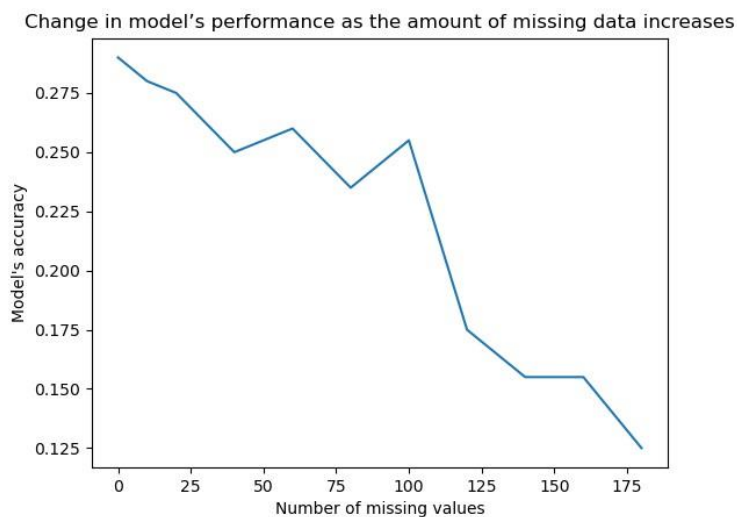


Figure1: correlation plot of all features

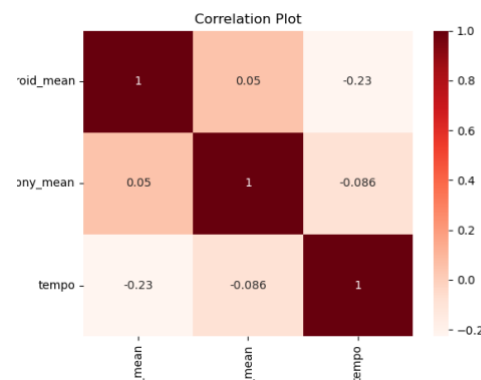


Figure2: correlation plot of 3 features

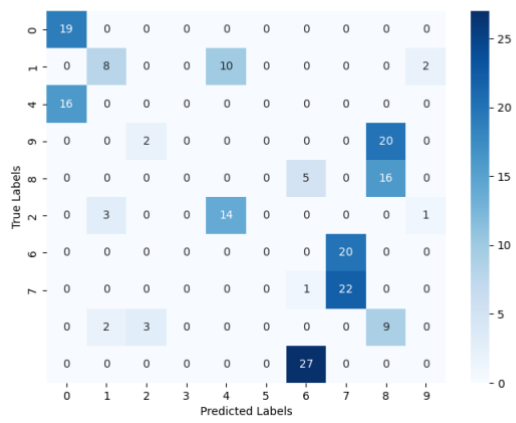


Figure3: Confusion matrix of the dataset