COMP30027 Machine Learning

Project 1: Music genre classification with Naïve Bayes

Student Code: 1079860

Task 1. Pop vs. classical music classification

Question 1.

Table 1:

The classification report / evaluation generated for the Naïve Bayes Model trained on pop_vs_classical_train.csv and tested on pop_vs_classical_test.csv

	classical	pop	accuracy	macro avg	weighted avg
precision	0.9524	1.0000	0.9767	0.9762	0.9779
recall	1.0000	0.9565	0.9767	0.9783	0.9767
f1-score	0.9756	0.9778	0.9767	0.9767	0.9768
support	20	23	0.9767	43	43

Accuracy = 97.67% ≈ 98%

Precision (Classical as positive class) = 95.24% ≈ 95%

Recall (Classical as positive class) = 100%

Question 2.

If we had to classify pop or classical music using just one of these features, then using the spectral centroid mean would be the most logical choice. As seen in Figure 1, this feature has the least amount of overlap present between the probability density functions for both classes and therefore we can predict the class based on this attribute with a high degree of certainty. If the spectral centroid mean for an instance is below approximately 2000 then there is a larger probability that the instance is an example of classical music and simultaneously a small probability that the instance is pop music, and vice versa if the spectral centroid mean is above 2000. Due to the lack of overlap between the probability densities, there is a very small chance of an error when a prediction is made purely using the probability density in relation to this single feature.

If the other 2 features were to be used to make predictions, predicting the class label based on which class has the higher probability density would likely result in more errors in our predictions due to the greater overlap between the probability densities for those attributes, as seen in Figures 2 and 3.

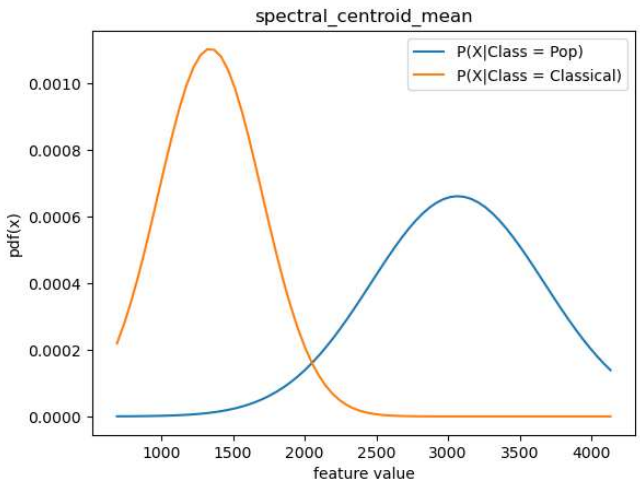

Figure 1: pdf of spectral centroid mean given class

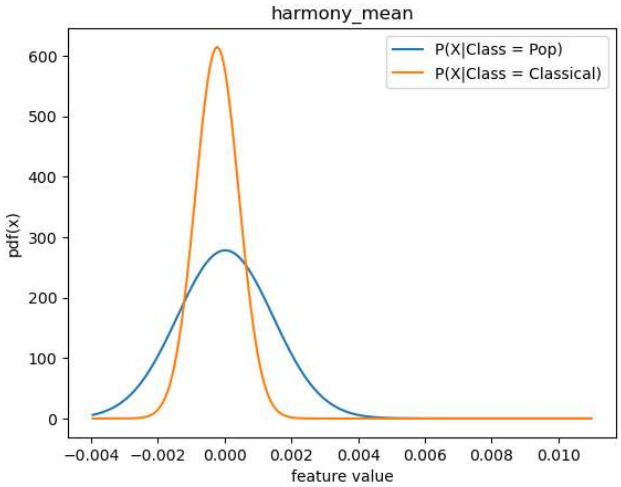

Figure 2: pdf of harmony mean given class

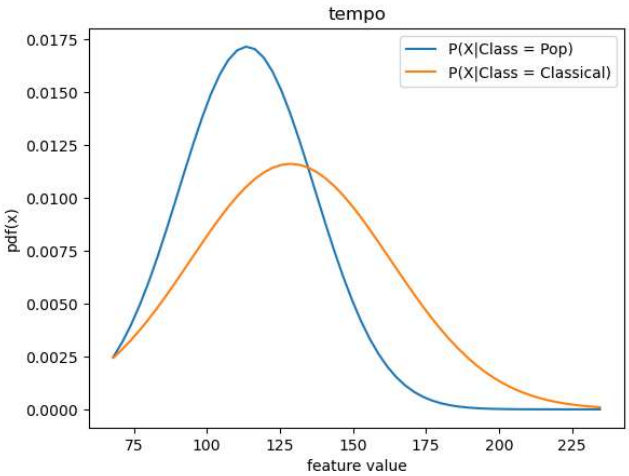

Figure 3: pdf of tempo given class

Task 2. 10-way music genre classification

Question 4.

Table 2: Averaged Evaluation Metrics for 2-fold cross validation (50% training, 50% test)

	blues	classical	country	disco	hip hop	jazz	metal	pop	reggae	rock	accuracy	macro avg	weighted avg
f1-score	0.280	0.824	0.407	0.253	0.321	0.445	0.614	0.671	0.404	0.162	0.456	0.438	0.438
precision	0.406	0.836	0.350	0.347	0.335	0.484	0.503	0.628	0.401	0.225	0.456	0.451	0.451
recall	0.220	0.820	0.490	0.200	0.310	0.420	0.830	0.730	0.410	0.130	0.456	0.456	0.456
support	50	50	50	50	50	50	50	50	50	50	0.456	500	500

Table 3: Averaged Evaluation Metrics for 5-fold cross validation (80% training, 20% test)

	blues	classical	country	disco	hip hop	jazz	metal	pop	reggae	rock	accuracy	macro avg	weighted avg
f1-score	0.263	0.854	0.464	0.309	0.323	0.445	0.618	0.659	0.436	0.127	0.47	0.45	0.45
precision	0.392	0.860	0.401	0.324	0.504	0.584	0.515	0.614	0.465	0.141	0.47	0.48	0.48
recall	0.200	0.880	0.570	0.300	0.240	0.380	0.840	0.740	0.420	0.130	0.47	0.47	0.47
support	20	20	20	20	20	20	20	20	20	20	0.47	200	200

Table 4: Averaged Evaluation Metrics for 10-fold cross validation (90% training, 10% test)

	blues	classical	country	disco	hip hop	jazz	metal	pop	reggae	rock	accuracy	macro avg	weighted avg
f1-score	0.235	0.848	0.491	0.329	0.388	0.441	0.645	0.684	0.471	0.157	0.495	0.469	0.469
precision	0.321	0.850	0.415	0.342	0.599	0.521	0.566	0.641	0.506	0.171	0.495	0.493	0.493
recall	0.200	0.890	0.620	0.330	0.300	0.390	0.850	0.760	0.450	0.160	0.495	0.495	0.495
support	10	10	10	10	10	10	10	10	10	10	0.495	100	100

As training set size increases, it is quite clear from Tables 2-4 that the average accuracy of the model increases steadily from 45.6% to 49.5%, which intuitively makes sense as more data is being used to inform the Naïve Bayes model.

The performance of the model doesn't seem to increase significantly in relation to classes which it is able predict relatively well with a smaller amount of training data such as classical, pop, or metal. Whereas, with genres such as reggae, country, hip hop and disco, the model seems to perform notably better in terms of predicting these genres as the training set size increases (i.e. the average f1-score increases and thus the combined metrics of precision and recall are increasing). However, there are also genres such as blues, jazz and rock that aren't predicted accurately by the model with a smaller training set and also don't appear to be predicted more accurately as training set size increases. In fact, the model tends to make worse predictions about blues music as the training set becomes larger.

From these results, we can conclude that while this model's ability to predict certain music genres improves and its overall average accuracy increases as the training set becomes larger, the model still doesn't improve or becomes worse at predicting other music genres. This is likely because the features present in the data used for training don't correlate well with these particular music genres or reveal unique information about them. Therefore, even if more samples of these features are provided from songs belonging to those genres, the model isn't learning anything useful and still won't be able to identify them accurately.