Practical Write-Up

Rodney Lafuente
lafuentemercado@college.harvard.edu

Roger Brockett
rbrockett@college.harvard.edu

April 15, 2022

1 Part A: Feature Engineering, Baseline Models

1.1 Approach

What did you do? When relevant, provide mathematical descriptions or pseudocode. Credit will be given for:

- PCA: Describe what the top 500 principal components represent, and how you computed them.
- Logistic regression: Describe how the model you trained predicts output probabilities for each class.

1.2 Results

This section should report on the following questions:

- What is the **overall** and **per-class** classification accuracy of the models that you implemented?

Accuracy	Raw Amplitude	Mel Spectogram
Overall	0.198	0.252
Class 0	0.197	0.160
Class 1	0.026	0.462
Class 2	0.592	0.020
Class 3	0.096	0.183
Class 4	0.072	0.227
Class 5	0.174	0.303
Class 6	0.033	0.433
Class 7	0.119	0.102
Class 8	0.140	0.945
Class 9	0.160	0.133

Table 1: Accuracies of Logistic Regression models on Raw Amplitude and Mel Spectogram Data.

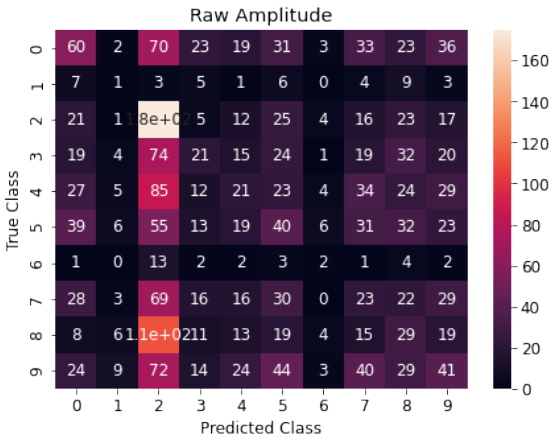

Figure 1: Confusion Matrix of Logistic Regression on Raw Amplitude Data

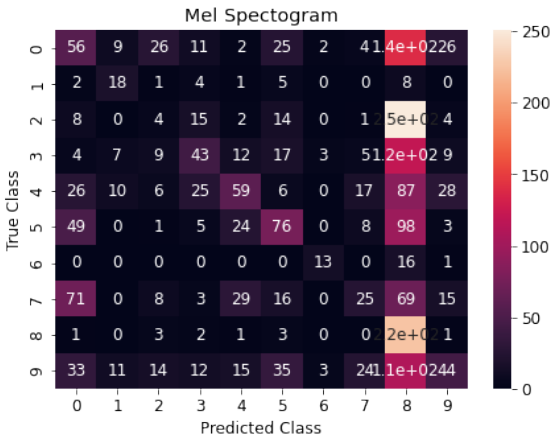

Figure 2: Confusion Matrix of Logistic Regression on Mel Spectogram Data

1.3 Discussion

This section should report on the following questions:

- Why do you hypothesize one feature representation performed better than the other?
- Why might have asked you to perform PCA first, and what is the impact of that choice?

2 Part B: More Modeling

2.1 First Step

2.1.1 Approach

What did you do? Credit will be given for:

- Provide mathematical descriptions or pseudocode to help us understand how the models you tried make predictions and are trained.

2.1.2 Results

This section should report on the following questions:

- What is the overall and per-class classification accuracy of the models that you implemented?

2.1.3 Discussion

Compare your results to the logistic regression models in Part A and discuss what your results imply about the task.

Accuracy	Raw Amplitude	Mel Spectogram
Overall	0.248	0.334
Class 0	0.230	0.280
Class 1	0.000	0.256
Class 2	0.746	0.211
Class 3	0.035	0.424
Class 4	0.110	0.216
Class 5	0.390	0.375
Class 6	0.033	0.167
Class 7	0.102	0.356
Class 8	0.212	0.441
Class 9	0.127	0.437

Table 2: Accuracies of Random Forest Classifier models on Raw Amplitude and Mel Spectogram Data.

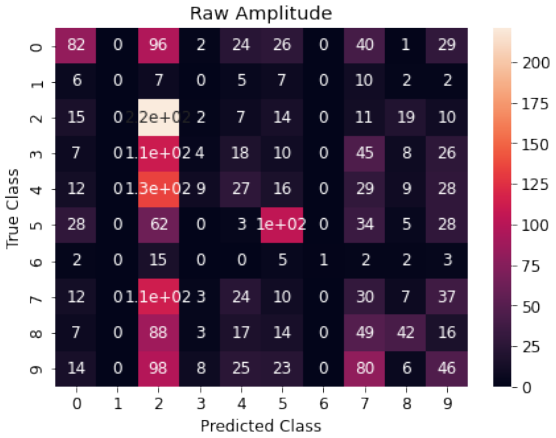

Figure 3: Confusion Matrix of Random Forest Classifier on Raw Amplitude Data

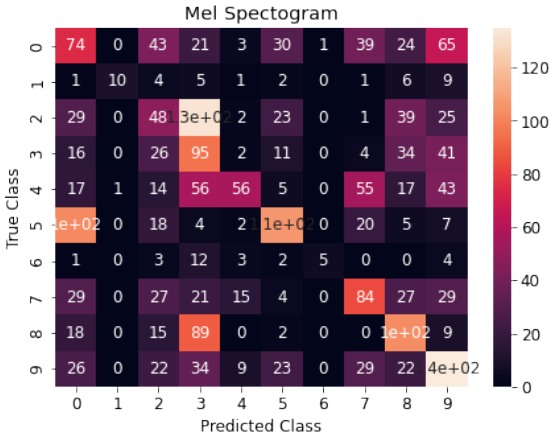

Figure 4: Confusion Matrix of Random Forest Classifier on Mel Spectogram Data

2.2 Hyperparameter Tuning and Validation

2.2.1 Approach

What did you do? Credit will be given for:

- Making tuning and configuration decisions using thoughtful experimentation. How did you perform your hyperparameter search, and what hyperparameters did you search over?

2.2.2 Results

Present your results of your hyperparameter search in a way that best reflects how to communicate your conclusions.

- RFC: The results from the hyperparameter grid search on Random Forest Classifiers showed...

# Estimators	Raw Amplitude	Mel Spectogram	Overall Rank
360	0.235	0.423	5
400	0.235	0.428	3
440	0.238	0.424	4
480	0.240	0.426	1
520	0.242	0.424	1

Table 3: Accuracies of Random Forest Models of Varying Estimator Counts

- SVM: The results on ...

C Value	Raw Amplitude	Mel Spectogram	Overall Rank
0.001	0.123	0.123	5
0.01	0.141	0.123	4
0.1	0.160	0.185	3
1	0.175	0.267	2
10	0.188	0.322	1

Table 4: Accuracies of Support Vector Machine models of Varying C Values

2.2.3 Discussion

Why do you expect the tuned models to perform better than the baseline models and the model used in First Step? Discuss your validation strategy and your conclusions.