Practical Write-Up

Rodney Lafuente
lafuentemercado@college.harvard.edu

Roger Brockett
rbrockett@college.harvard.edu

April 1, 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	Class %
Overall	0.198	0.252	100%
Class 0	0.197	0.160	12.6%
Class 1	0.026	0.462	3.54%
Class 2	0.592	0.020	12.53%
Class 3	0.096	0.183	9.42%
Class 4	0.072	0.227	10.93%
Class 5	0.174	0.303	12.98%
Class 6	0.033	0.433	1.49%
Class 7	0.119	0.102	11.85%
Class 8	0.140	0.945	12.03%
Class 9	0.160	0.133	12.61%

Table 1: Accuracies of Logistic Regression models on Raw Amplitude and 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?

Accuracy	Raw Amplitude	Mel Spectogram	Class %
Overall	0.248	0.334	100%
Class 0	0.230	0.280	12.6%
Class 1	0.000	0.256	3.54%
Class 2	0.746	0.211	12.53%
Class 3	0.035	0.424	9.42%
Class 4	0.110	0.216	10.93%
Class 5	0.390	0.375	12.98%
Class 6	0.033	0.167	1.49%
Class 7	0.102	0.356	11.85%
Class 8	0.212	0.441	12.03%
Class 9	0.127	0.437	12.61%

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

2.1.3 Discussion

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

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.

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.

3 Optional Exploration, Part C: Explore some more!

3.1 Approach

What did you do? Credit will be given for:

- Diving deeply into all of the model classes and/or pre-processing algorithms that you tried (rather than just trying off-the-shelf tools with default settings). When relevant, provide mathematical descriptions or pseudocode to help us understand how the models you tried make predictions and are trained.

3.2 Results

Describe your results in a way that is appropriate for the experiments that you ran.

3.3 Discussion

Credit will be given for:

- Explaining the your reasoning for why you sequentially chose to try the approaches you did (i.e. what was it about your initial approach that made you try the next change?).
- Explaining the results. Did the adaptations you tried improve the results? **Why or why not?** Did you do additional tests to determine if your reasoning was correct?