1 Logistic Regression (40pts)

Solution.

1. What is the role of the learning rate (eta) on the efficiency of convergence during training?

2. What is the role of the number of epochs on test accuracy?

2 Feature Engineering (40 points)

1. What custom features did you add/try (other than n-grams)? How did those additional features affect the model performance? Why do you think those additional features helped/hurt the model performance?

2. What are unigrams, bigrams, and n-grams? When you added those features to the FeatureUnion, what happened to the model performance? Why do these features help/hurt?

3 Gradient Descent Learning Rule for Multi-class Logistic Regression (20 pts)

Solution. *1.* Derive the negative log likelihood for multi-class logistic regression.

2. The gradient descent learning rule for optimizing weight vectors generalizes to the following form: $\beta_j^{t+1} = \beta_j^t - \eta \nabla \beta_j^t$ where η is the learning rate. Find the $\nabla_{c,j}$ (the parameter for feature x_j in class c) for a multi-class logistic regression model.