Homework Report 5

Prompt:

- Explain how you used gradient descent with regularization to learn the weights.
- Think back to when you implemented Logistic Regression on MNIST. How would it have been different if you applied Tikhonov regularization?
- Produce a model selection curve, that is, plot training and validation error with respect to the value of lambda. It may be useful to graph the log values of lambda rather than the actual values of lambda here. Then, conclude what the best value of lambda is and explain why. NOTE: please set your default lambda in the constructor to your optimal lambda you discovered for TA testing purposes.

Response: I used exactly the same process of gradient descent as our previous assignment except for the inclusion of the regularization term. We updated the weights similarly to last time except for the fact that the regularization term was added to the updates. This addition allowed me to determine the balance between fitting to the training data and still refraining for building to complex of a model. If I had included regularization on the Logistic Regression on MNIST we probably would have had a slightly different model that was less accurate but also less complex. You can see my graph below! The best value of lambda is 10. You can see that is exactly where the validation error is minimized in my graph. It also indicates that is where the error on non-training upon sets will be smallest. You can see that as we continue, the training erro continues to decrease. However, the separation of the validation set from the training set allows us to see where our most beneficial model lies.