CS 498 – Applied Machine Learning

Assignment 2

bmgriff2 – ramiaa2 – felipea2

Problem 1:

* Accuracy of Model Using Different Regularization Constants

A Support Vector Machine classifier was trained on the data using Stochastic Gradient Descent. In order to get better looking plots and assess our models more accurately, we used 400 epochs of 300 steps each. We also held out 200 items rather than the recommended 50 to get less variance in our accuracies and chose a step of size 5/(5000 + e) (where e is which epoch the training process is in) based on how the training plots looked. Several values of regularization factors were used and a plot showing the different resulting accuracies on the held-out data is shown below in figure 1.

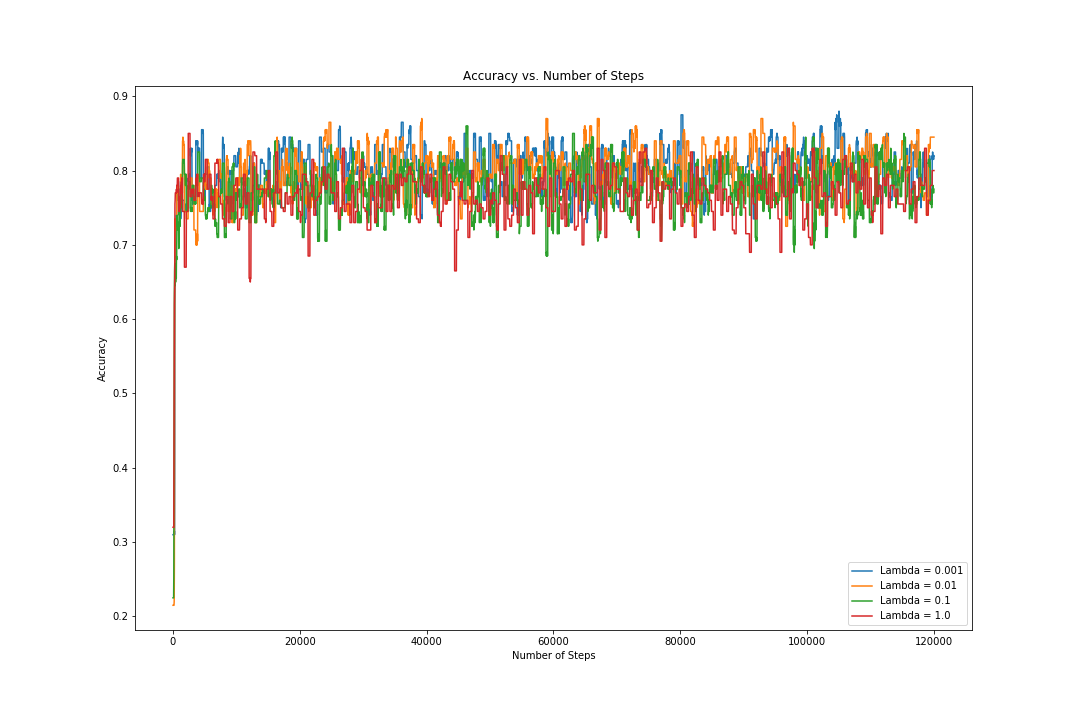


Figure : Accuracy of Held-Out Data vs Number of Steps for Different Regularization Constants

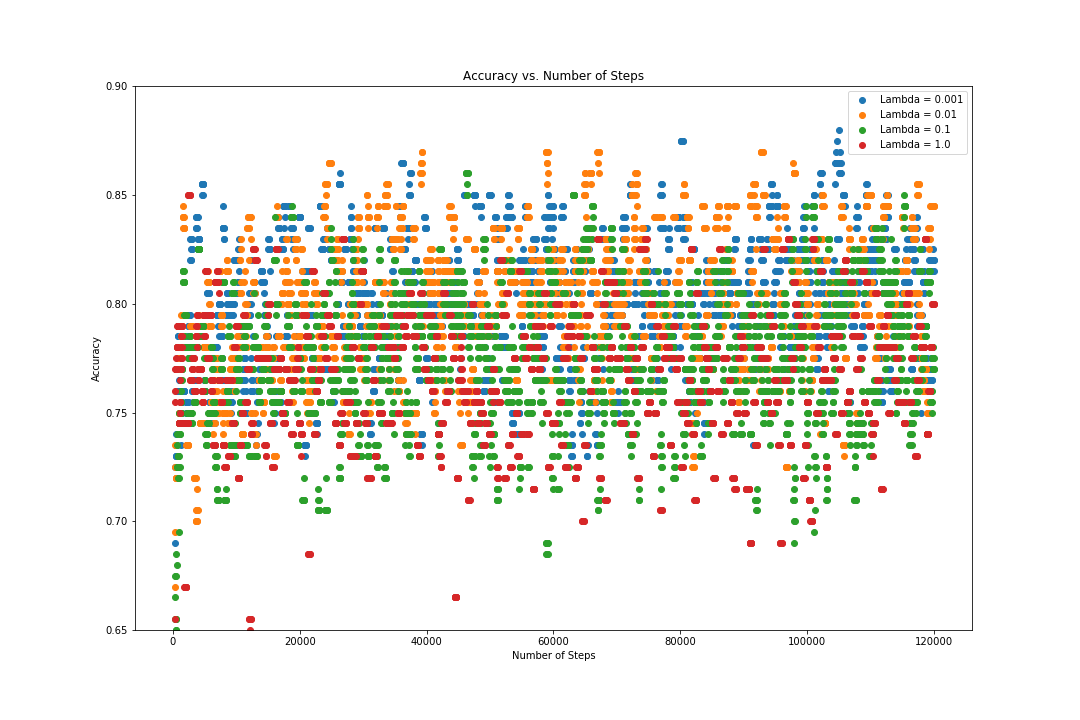


Figure 2: A Closer Look at Accuracy of Held-Out Data vs Number of Steps for Different Regularization Constants

* Magnitude of Coefficient Vector for the Different Regularization Constants

Figure 3 below illustrates the different magnitudes of the coefficient vector achieved with the 4 regularization factors.

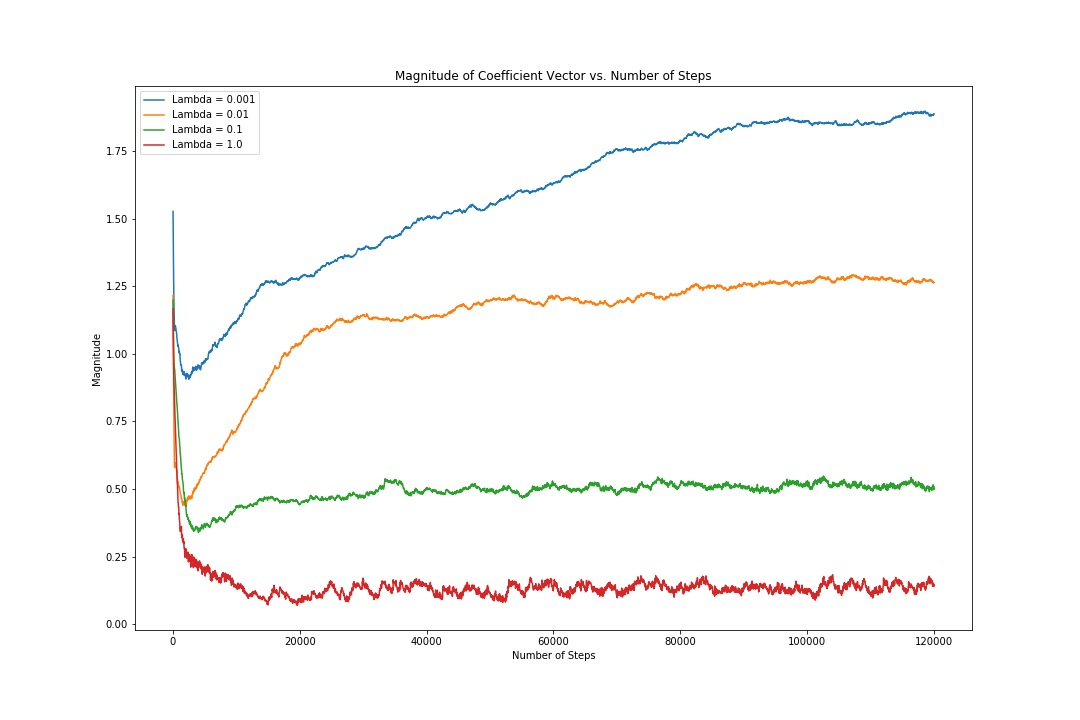


Figure 3: Magnitude of Coefficient Vector vs Number of Steps for Different Regularization Constants

* Estimate of Best Regularization Constant Value

We choose our regularization constant based on how each model performed on the validation set. We found the regularization constant values of 0.001 and 0.01 to perform almost identically and better than the rest. However, we chose 0.01 as our best regularization constant since it results in a smaller magnitude vector and a simpler model that would generalize better to the testing data.

* Estimate of the Accuracy of the Best Classifier on the Remaining 10% Test Dataset

After choosing the best regularization constant, the model was trained on the original 90% training dataset and then tested on the 10% test dataset.

Accuracy of Model = 0.80757