Homework 2

Heena Thakker

CS 498 Applied Machine Learning

SVM using Stochastic Gradient Descent

1. Plot of Accuracy every 30 steps, for each value of the regularization constant

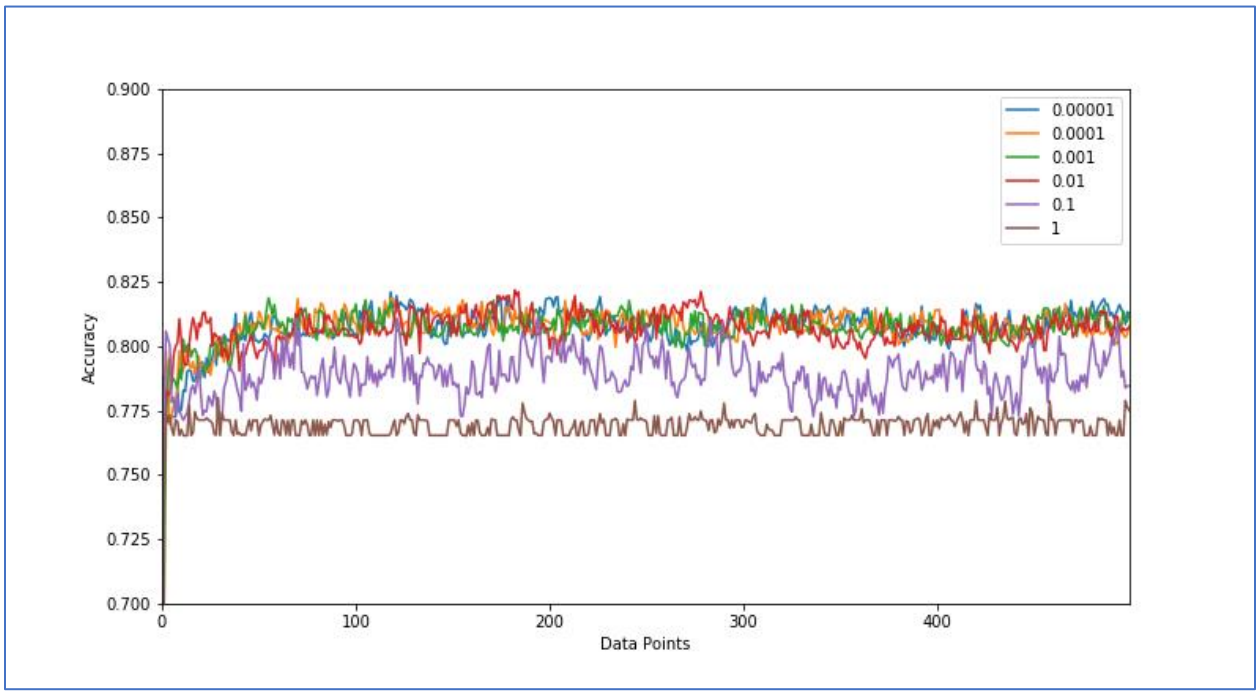

Figure 1

This plot has been produced using the source code in file SGDParameterEvaluation. The legend indicates the values of the regularization constant (lambda)

Accuracies were computed on a 10% validation data set, every 30 steps. The x-axis ranges from 0 to 500 because the accuracies have been computed every 30 steps in 50 epochs of 300 steps each. Hence there are 50*300/30 data points. Progression of the data points also indicate the progression of steps and epochs.

At every transition from one epoch the another, the step size was changed to the value 1/(0.01*epoch +50).

2. Plot of magnitude of the coefficient vector every 30 steps, for each value of the regularization constant

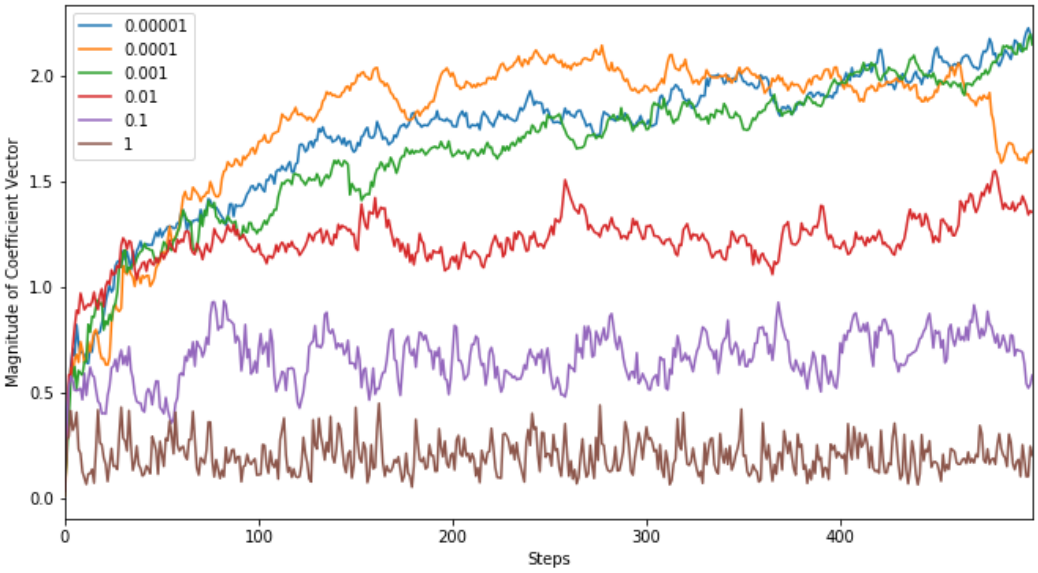

Figure 2

Values plotted for 'Magnitude of Coefficient Vector' have been computed using the same values of epoch count, steps, step size and regularization constants as used in the 1st graph.

The legend indicates different values for the regularization constant.

The magnitude of the coefficient vector was computed as the L2 norm = sqrt($\mathbf{a}^T\mathbf{a}$).

3. Estimate of the best value of regularization constant

I chose **0.01** as the best value for the regularization constant and used it to train 90% of the data (training + validation).

It can be clearly observed from figure 1 that the accuracy values are higher for lambda = 0.1 compared to lambda = 1. However there is not a marked difference in accuracies for lambda values 0.01, 0.001 and higher.

This can also be observed from the table below which averages out the accuracy on validation data set for each of the lambda values. These values were obtained by running the program SGDParameterEvaluation 5 times.

Lambda	1.00E-05	1.00E-04	1.00E-03	1.00E-02	1.00E-01	1.00E+00
Accuracy	0.807187	0.806967	0.806129	0.806264	0.789279	0.768881
Accuracy	0.803157	0.804647	0.802438	0.801853	0.780656	0.75757
Accuracy	0.805648	0.804981	0.80812	0.804918	0.790454	0.769314
Accuracy	0.811399	0.812221	0.808948	0.808964	0.789414	0.769537
Accuracy	0.809659	0.809669	0.809297	0.806403	0.787308	0.765438
Average	0.80741	0.807697	0.806986	0.80568	0.787422	0.766148

If we look at figure 2, it can be observed that the magnitude of the vector coefficients is increasing with lambda. Amongst 1e-5, 1e-4, 1e-3 and 1e-2, 1e-2 has the lowest values for the coefficient vector magnitudes. Hence 1e-2 is chosen as the optimal value for lambda because it strikes a balance between high accuracy values and lower magnitude of the coefficient vector. A lower |a| is preferred because it will help minimize the hinge loss on future unseen examples when misclassified.

4. Estimate of the accuracy of the best classifier on the 10% test dataset data

0.804912998976

This value can be obtained by running program SGDTraining. The program trains the classifier in 100 epochs of 500 steps each to ensure that the entire dataset is seen at least once. At the onset of each epoch, 50 samples are selected at random from the 90% training dataset for evaluating the classifier accuracy. The reported number is obtained by running the classifier on 10% test data which was not used for training.

References:-

- Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. https://archive.ics.uci.edu/ml/datasets/Adult

- Applied Machine Learning, D.A. Forsyth, (approximate 9'th draft) http://luthuli.cs.uiuc.edu/~daf/courses/AML-18/learning-book-15-Jan.pdf