Machine Learning: Programming HW 2

Challa Priyankar
`USN: ENG20AM0017`

Raghav Nanjappan
`USN: ENG20AM0044`

Punith Pechetti
`USN: ENG21AM3028`

Sai Nishwanth Raj Reddy
`USN: ENG21AM3031`

Due Date: November 7^{th}, 2022

1 Introduction

1.1 Problem Statement

Write a program to train a support vector machine on UCI Adult Dataset. using stochastic gradient descent without using a library package to train the classifier. Do this exercise for two case:

1. Without scaling the variables
2. Scaling the variables so that each has unit variance

Try at least the regularization parameters values $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$ for at least 50 epochs with a minimum step size of 300. Plot the accuracy graph for every 30 steps.

1.2 Values Used

1. *Learning Rate:* 0.1
2. *Epochs:* 100
3. *Regularization Parameter:* $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$
4. *Step Length:* $\eta = 300$

1.3 `.ipynb` Files

Please click on the following text(s) to get to the `.ipynb` files for the SVM classifier with unscaled and scaled data.

1. `Without scaling the variables`
2. `Scaling the variables so that each has unit variance`

The dataset has been taken from the UCI Adult Dataset.

2 Plots

2.1 Without Scaling

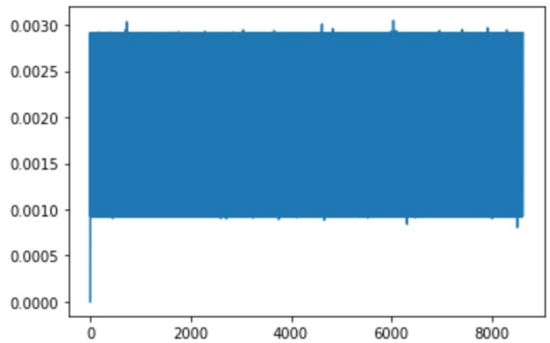

Figure 2.1.1: Regularization Parameter $\lambda = 10^{-3}$

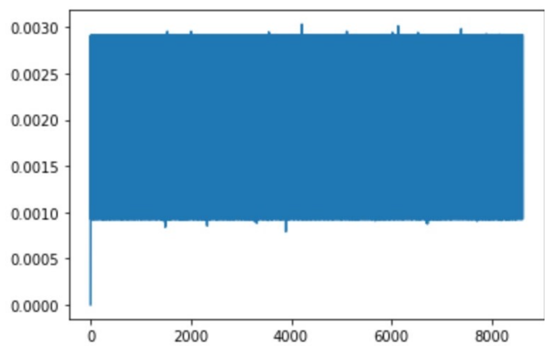

Figure 2.1.2: Regularization Parameter $\lambda = 10^{-2}$

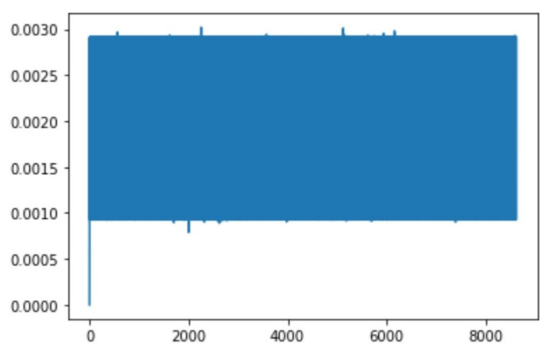

Figure 2.1.3: Regularization Parameter $\lambda = 10^{-1}$

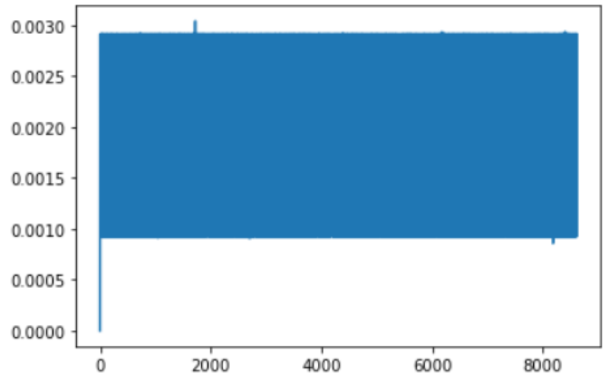

Figure 2.1.4: Regularization Parameter $\lambda = 1$

2.2 With Scaling

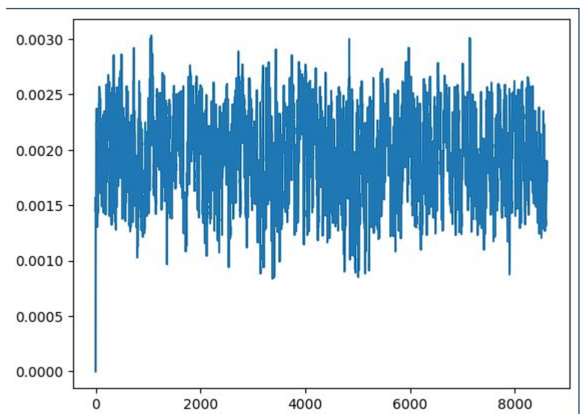

Figure 2.2.1: Regularization Parameter $\lambda = 10^{-3}$

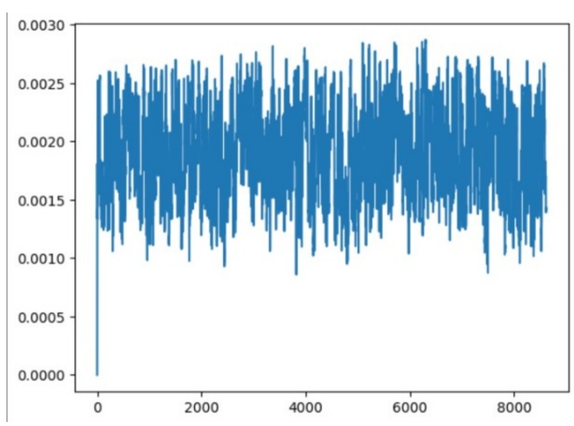

Figure 2.2.2: Regularization Parameter $\lambda = 10^{-2}$

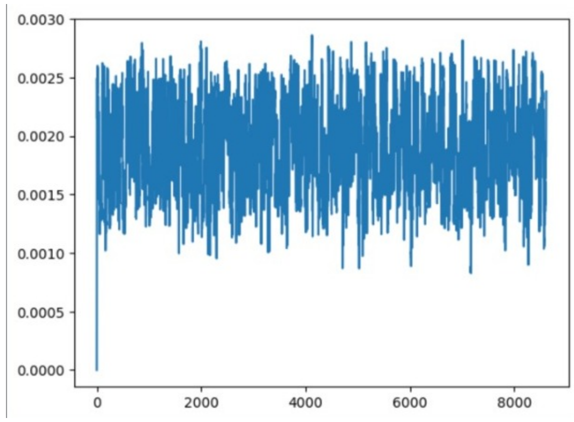

Figure 2.2.3: Regularization Parameter $\lambda = 10^{-1}$

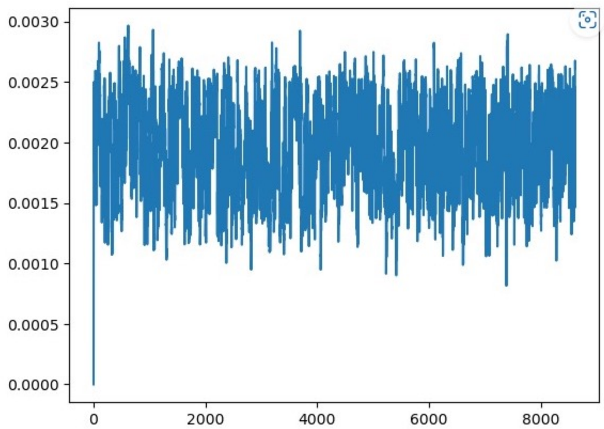

Figure 2.2.4: Regularization Parameter $\lambda = 1$

3 Estimation of Best Regularization Parameter λ

Our best value of the regularization parameter λ is 10^{-3} (for the scaled data) with respect to cost, as for every incremental change in the value of n in 10^n, the cost increases by 10 fold. So, our estimate for the best value of λ is the one where $10^n \approx 0$ with respect to cost.

4 Expression for Modifying the Step Length η per Epoch

From our understanding, the step length η can be modified per epoch in the following way:

$$\eta = \frac{Total\ Number\ of\ Training\ Examples}{Epoch\ Number} \tag{1}$$

5 Estimation of Best Accuracy

The best value for accuracy we had achieved is 67% (for scaled data) where $\lambda = 10^{-3}$. We have observed for every incremental change in the value of n in 10^n, the accuracy increases by a significant amount. So, our estimate for the best value of accuracy is $\approx$ 90%.

6 Conclusion

From the graphs in section 2, we can infer that scaling the data with 0 mean and unit variance produces far more accurate results and an overall lower cost function compared to unscaled data. We can also further conclude that from the reguralization constant λ values that we had selected, $\lambda = 1(10^0)$ produces the best accuracy for the given data set.