

CMSC510 - Assignment 3

Steven M. Hernandez, V00528395

October 2020

In this assignment, we perform gradient descent with logistic loss, with L_1 regularization and proximal step soft thresholding. Logistic loss ensures we minimize our error while L_1 regularization penalizes weights as they grow larger. Finally, proximal soft-thresholding says that we should check each element in our weight vector to identify which elements are contributing with high weights and which are low. If the weights are too low (very little innovation), then we should ignore this by replacing it with zero.

1 Results

For experiments, I use the MNIST dataset classes 5 and 9 which are transformed into labels -1 and $+1$ respectively.

First, we look at selecting values for our hyperparameters C and L . We can see in Fig. 1 (a) three experiments where we change the value for C . We can see that the accuracy for $C = 10$ and $C = 100$ are similar, but $C = 1$ has a lower accuracy. In Fig. 1 (b) we change the value for L . $L = 1e6$ and above seem to work similarly well so we select $L = 1e6$.

Now we want to see how our two regularization method change our model's performance. The two regularization methods used are 1.) L_1 regularization and 2.) proximal gradient descent. In Fig. 2 (a) we can see that applying regularization reduces our accuracy to a small degree. This might be expected because regularization such as proximal soft thresholding attempts to essentially reduce the learning rate of different weights by reducing their values to 0. We can see in Fig. 2 (b) that this is true. For both cases where Proximal gradient descent is used, a higher number of weights are set to $w_i = 0$ while without proximal, more weights have a value $abs(w_i) > 0$. So this shows that our soft thresholding is working as expected.

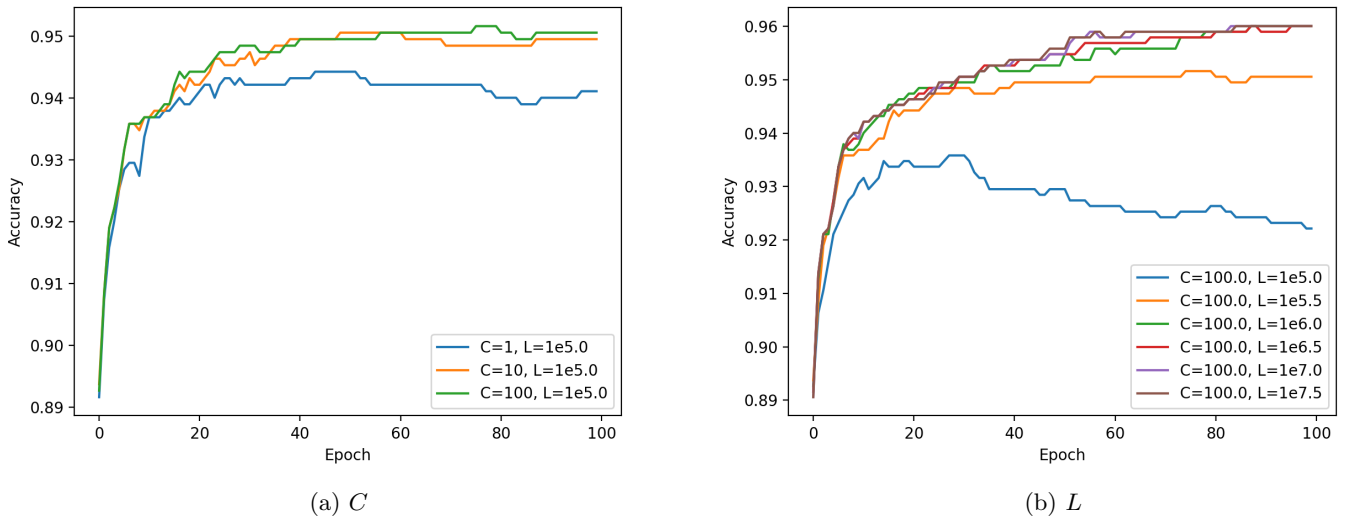
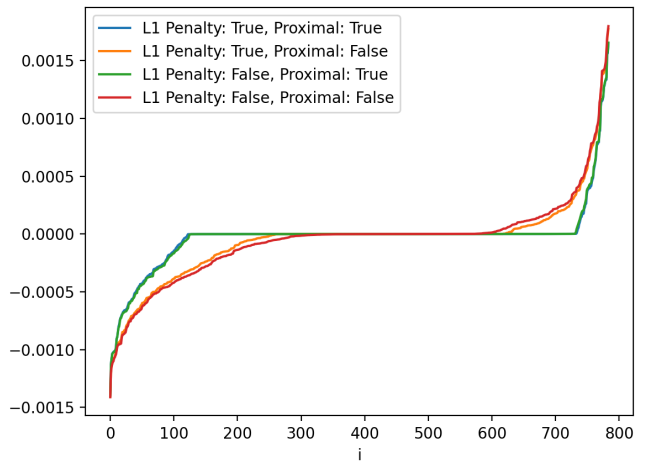
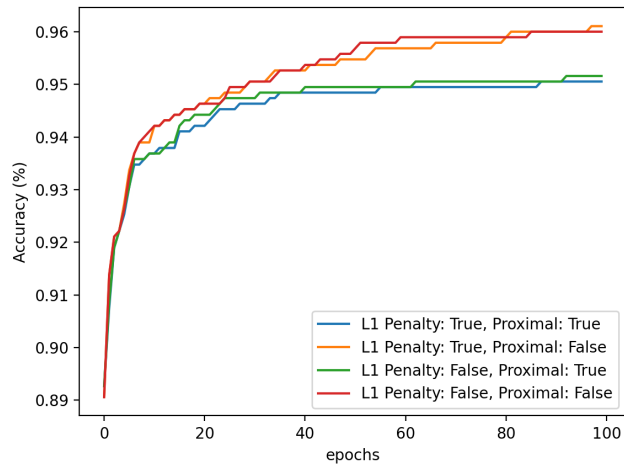


Figure 1: Simple hyperparameter search for C and L



(a) Accuracy with and without applying regularization techniques. (b) Weights with and without applying regularization techniques.

Figure 2: Comparing the use of combinations of L_1 regularization and Proximal gradient descent