CS260 HW1 REPORT

Shruti Sharan (UID: 405228029)

February 10, 2019

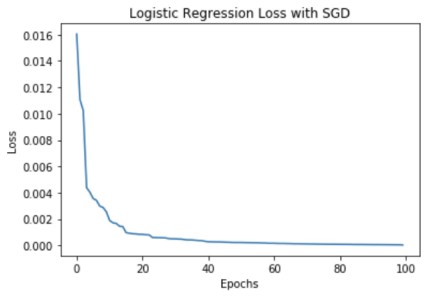
1 Introduction

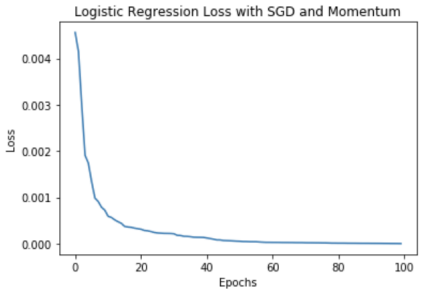
In this homework, we are asked to implement and solve linear Logistic Regression model and Linear SVM model (without regularization term) on the MNIST dataset. The MNIST dataset is a dataset of images of handwritten digits from 0 to 9. In this task, we only need to perform binary classification on digits 0 and 1. We download and load the dataset comprising of 60000 images and split them into training and testing data using the DataLoader in Pytorch. Then we extract the images with only 0 and 1 and save their corresponding labels.

2 Logistic Regression

Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Here we use Logistic Regression to classify the digits into the two classes 0s or 1s.
Learning rates tried: 0.1,0.01,0.001,1.
The best results were obtained using Learning rate=0.1.
The losses were computed over different number of epochs=10,12,15,20,30,50,75,100.
The following table shows the Loss values over 20 epochs, since the values converge within 20 epochs.

```
Epoch [1/20], Loss: 0.0003
Epoch [2/20], Loss: 0.0002
Epoch [3/20], Loss: 0.0005
Epoch [4/20], Loss: 0.0069
Epoch [5/20], Loss: 0.0016
Epoch [6/20], Loss: 0.0005
Epoch [7/20], Loss: 0.0001
Epoch [8/20], Loss: 0.0002
Epoch [9/20], Loss: 0.0001
Epoch [10/20], Loss: 0.0002
Epoch [11/20], Loss: 0.0000
Epoch [12/20], Loss: 0.0081
Epoch [13/20], Loss: 0.0001
Epoch [14/20], Loss: 0.0001
Epoch [15/20], Loss: 0.0010
Epoch [16/20], Loss: 0.0006
Epoch [17/20], Loss: 0.0028
Epoch [18/20], Loss: 0.0018
Epoch [19/20], Loss: 0.0000
Epoch [20/20], Loss: 0.0003
```





Following are the plots for the Logistic Zregression Loss over 100 epochs, with and without momentum. The accuracy is calculated on the model after applying the optimisation technique, Stochastic Gradient Descent, with and without momentum. The loss was optimized with momentum=0.5,0.9 and default (Momentum=True).
The best results were obtained using Momentum=0.5.
The following table shows the Accuracy values for different learning rates, With and Without Momentum.

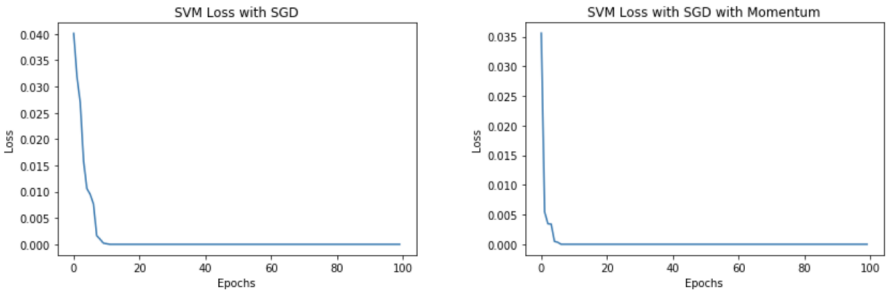
Learning Rate	SGD	SGD with Momentum
0.1	99.905434	99.9527214
0.01	99.905434	99.905434
0.001	99.905434	99.9054344
1	99.905434	nan

We see that the accuracies obtained converge faster with a higher learning rate and perform better without momentum.

3 Support Vector Machines

Support-vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. Learning rates tried: 0.1,0.01,0.001,1.
The best results were obtained using Learning rate=0.01.
The losses were computed over different number of epochs=10,12,15,20,30,50,75,100.
The following table shows the Loss values over 20 epochs, since the values converge within 20 epochs.

```
Epoch [1/20], Loss: 0.0212
Epoch [2/20], Loss: 0.0000
Epoch [3/20], Loss: 0.0000
Epoch [4/20], Loss: 0.0029
Epoch [5/20], Loss: 0.0000
Epoch [6/20], Loss: 0.0000
Epoch [7/20], Loss: 0.0182
Epoch [8/20], Loss: 0.0000
Epoch [9/20], Loss: 0.0000
Epoch [10/20], Loss: 0.0000
Epoch [11/20], Loss: 0.0000
Epoch [12/20], Loss: 0.0058
Epoch [13/20], Loss: 0.0080
Epoch [14/20], Loss: 0.0000
Epoch [15/20], Loss: 0.0000
Epoch [16/20], Loss: 0.0000
Epoch [17/20], Loss: 0.0000
Epoch [18/20], Loss: 0.0000
Epoch [19/20], Loss: 0.0020
Epoch [20/20], Loss: 0.0051
```



Following are the SVM Losses plotted over 100 epochs with and without momentum respectively. The accuracy is calculated on the model after applying the optimisation technique, Stochastic Gradient Descent, with and without momentum. The loss was optimized with momentum=0.5,0.9 and default (Momentum=True).
The best results were obtained using Momentum=0.5.
The following table shows the Accuracy values for different learning rates, with

and without Momentum.

Learning Rate	SGD	SGD with Momentum
0.1	99.952721	99.905434
0.01	99.952721	99.905434
0.001	99.952721	99.9054344
1	100.000000	99.905434

We see that the accuracies obtained are converge faster with a higher learning rate and perform better without momentum.

4 Conclusion

Thus we find that both Logistic Regression and Linear Support Vector Machines both perform extremely well for the binary classification of the MNIST dataset giving us very high testing accuracies of our dataset. It is able to classify the 0s and 1s almost perfectly.