LIU, Dingdong
20583503

Programming Assignment 1

COMP5212
October 3, 2022

Report

Training Loss and Accuracy of Each Model

The training loss of two models over each epochs is shown as follows in Figure 1. After 30 epochs, both model were evaluated on test set. Results are:

1. For Logistic Regression: 99.85%. Hyper-parameters are: 1.0 for step size. 0 for momentum (i.e., use standard SGD optimiser without momentum). Batch size=64.

2. For SVM: 99.91%. Hyper-parameters are: 10.0 for step size. 0 for momentum (i.e., use standard SGD optimiser without momentum). Batch size=64.

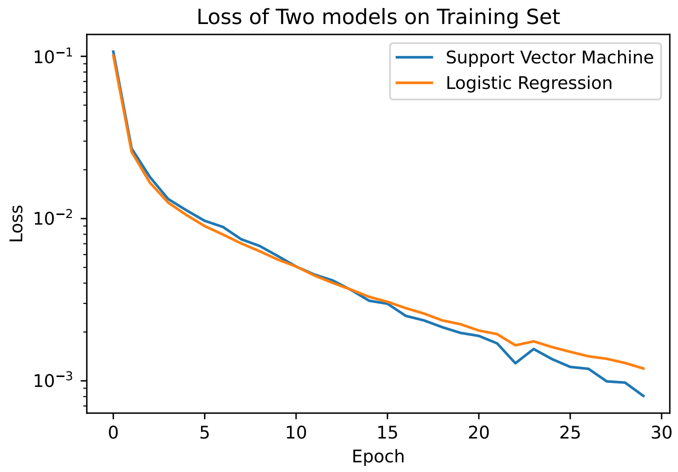

Figure 1: Training loss of SVM and Logistic Regression model over 30 epochs.

Results for Two Optimisers

I experimented with different momentum value for SGD optimiser. When $momentum = 0$, original SGD optimiser is adopted. Results are shown below in Figure 2.

We can find from the figure 2 that when value of momentum increase, the loss become more unstable though still decreasing. From the equation of SGD-momentum we can know that when value of momentum increase, in gradient, more information from previous samples are kept. Under some cases this is helpful for reducing training loss, but may negatively influence the stability of loss when momentum is large.

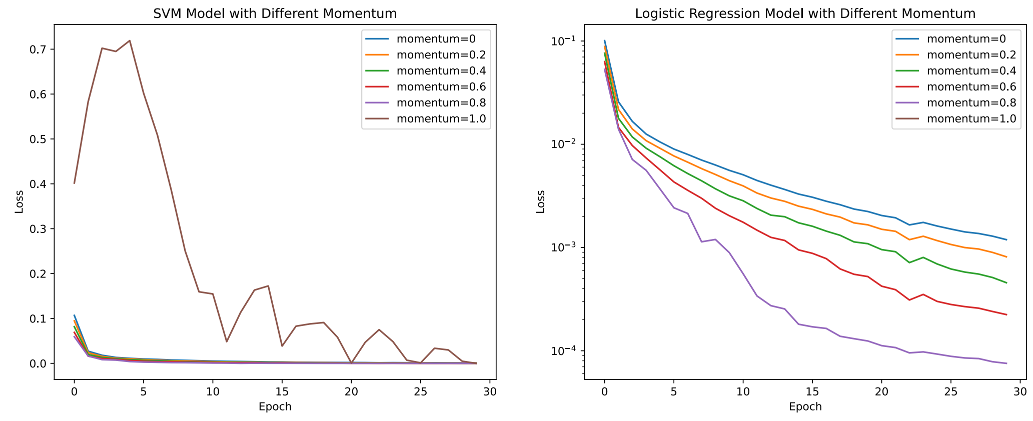

Figure 2: Different momentum value for SVM and Logistic Regression model under training set. In this experiment, step size was fixed to 0.1 for both models

Training with Different Step Size

I experimented with step size on log scale, i.e., $lr \in \{10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1, 10\}$. The results are shown in Figure 3.

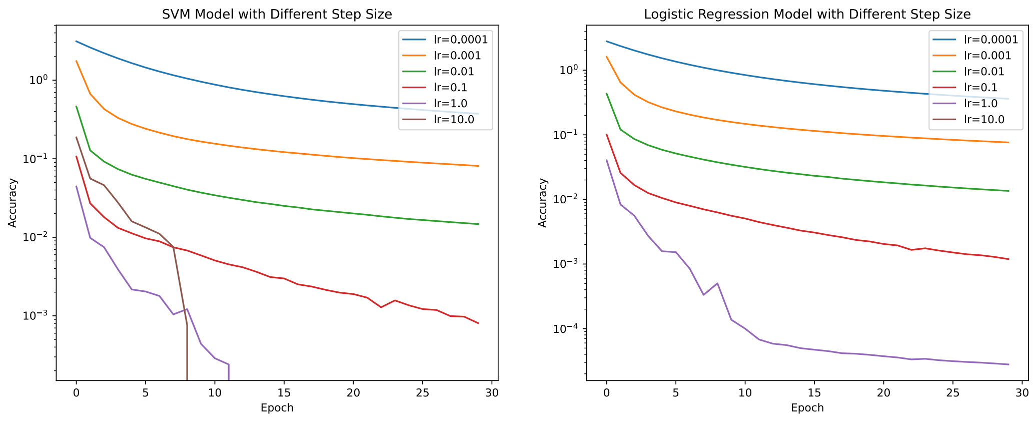

Figure 3: Train model with different step size.

Generally speaking, a larger step size helps model to converge faster. However, when step size is too large, the loss may explode and result in “nan” in Python. This is observed when step size is 10 for logistic model.

Besides, a large step size may add to instability during training. As shown in the figure 3, with a large step size, the curve of loss may not be as smooth as the loss curve with smaller step size.