# Assignment 1 Bonus Points

DD2424 Deep Learning in Data Science

Anna Canal Garcia
annacg@kth.se

April 11, 2018

## 1 Introduction

This is an extension of the assignment 1 that aims to optimize the performance of the network and to compare network trained in Exercise 1 with a trained network by minimizing the SVM multi-class loss.

## 2 Method

First of all some tricks are applied to the network in order to optimize its performance. The first change is to use a learning rate with a decay factor of 0.9 after each epoch. The second is to use more data for the training set and less data for the validation set. The third step is to extend the time of training, so using 50 epochs. And the last change is to do a grid-search of the best parameters to then finally put together all the changes and achieve the maximum test accuracy.

Secondly, SVM multi-class loss function is implemented in order to compare the network performance with this loss function to the one implemented in Exercise 1 (cross-entropy loss function).

$$l_{svm} = \sum_{j=1, j\neq y} max(0, s_j - s_y + 1) = \sum_{j=1, j\neq y} max(0, f(\mathbf{x}_i; \mathbf{W})_j - f(\mathbf{x}_i; \mathbf{W})_y + 1) \tag{1}$$

$$J(D, \lambda, \boldsymbol{W}) = \frac{1}{D} \sum_{i} \sum_{j=1, j\neq y} max(0, f(\mathbf{x}_i; \mathbf{W})_j - f(\mathbf{x}_i; \mathbf{W})_y + 1) + \lambda \sum_{x,l} \mathbf{W}_{xl}^2 \tag{2}$$

# 3 Results

## 3.1 Exercise 2.1

1. The decay of learning rate

   Taking the best configuration from Exercise 1 ($\lambda = 0$ and $\eta = 0.01$) and applying a decay factor of 0.9 to the learning rate ($\eta$) after each epoch the test accuracy is 37.42% that means a gain= 0.59% since the test accuracy without using a decay learning rate is 36.83%. Moreover, when tested with $\lambda = 0.1$ and $\eta = 0.01$ the test accuracy is equal to 36.47% which represents a gain = 3.08%.

2. More training data, less validation data

   Training data was increased using another batch file, so I take data from data_batch_1 and data_batch_3. Validation dataset is the first 1000 images from data_batch_2. It was not possible to use a bigger training set with Matlab. So with this new amount of data, the test accuracy of the network for $\lambda = 0.0$ and $\eta = 0.01$ is 38.56%, which represents an gain of 1.73% in respect to exercise 1 (so bigger gain than when using the decay learning rate). With $\lambda = 0.1$ and $\eta = 0.01$ the test accuracy is 36.32%, a gain of 2.93% (slightly lower than the gain with the decay learning rate).

3. Longer training time

   With 50 epochs the test accuracy of the network for $\lambda = 0.1$ and $\eta = 0.01$ is increased by 3.6% in respect to exercise 1, but with $\lambda = 0.0$ and $\eta = 0.01$ the test accuracy is lower, so we don't have gain.

4. Grid Search

   Finally a grid search is used to find good values for the amount of regularization ($\lambda$) and the learning rate ($\eta$). The values that give the maximum accuracy are $\lambda = 0.1$ and $\eta = 0.001$.

   If we use these values without the decay of the learning rate and a longer training time (60 epochs), we obtain an accuracy of 36.01%. If we add the decay factor to the learning rate equal to 0.9 and then we start it from $\eta$=0.01 we end with an accuracy of 36.45%. But if we finally test with lambda = 0 and the test accuracy is 37.4%. It would be higher using more training data as in the test 2, but it takes more time.

## 3.2 Exercise 2.2

In this section are the results from using a svm-multi class loss classifier with different parameters:

1. **Configuration 1: $\lambda = 0$, n_epochs = 60, n_batch = 100, $\eta = 0.1$, decay_$\eta = 0.9$**

The test accuracy obtained is 34.58%. Fig.1 shows the svm loss function of this network.

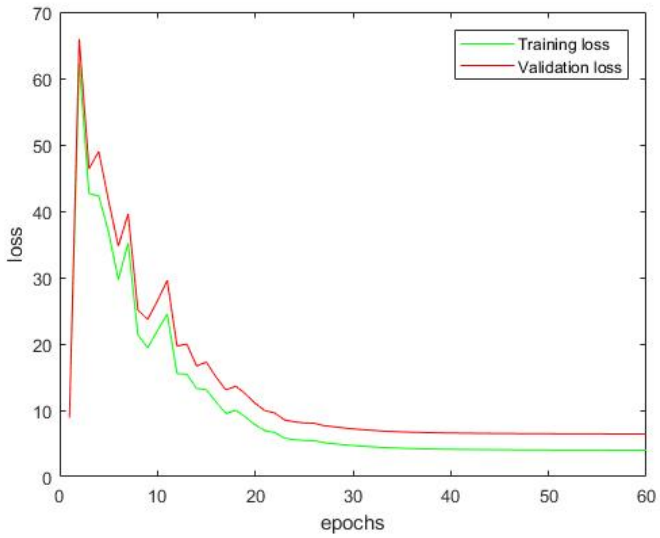


Figure 1: SVM Loss function with $\lambda = 0$, $\eta = 0.1$, epochs = 60 and n_batch = 100.

2. **Configuration 2: $\lambda$ = 0.1, n_epochs = 60, n_batch = 100, $\eta$ = 0.01, decay_$\eta$ = 0.9**

   The test accuracy obtained is 37.13%. Fig.2 shows the svm loss function of this network.

3. **Configuration 3: $\lambda$ = 0.1, n_epochs = 60, n_batch = 100, $\eta$ = 0.001, no decay_$\eta$**

   The test accuracy obtained is 34.9%. Fig.3 shows the svm loss function of this network.

4. **Configuration 4: $\lambda$ = 1, n_epochs = 60, n_batch = 100, $\eta$ = 0.01, decay_$\eta$ = 0.9**

   The test accuracy obtained is 35.5%. Fig.4 shows the svm loss function of this network.

Finally, we can see than the network that uses the svm-multi class loss function has a slightly worst performance than the one using the cross-entropy loss function, so lower test accuracy.

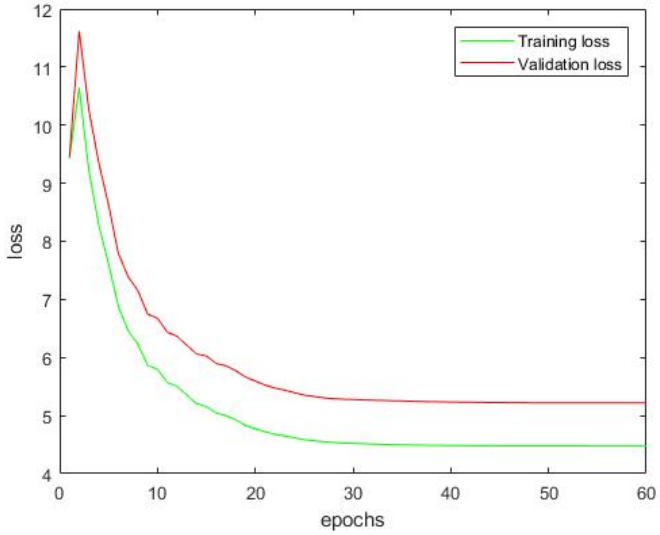


Figure 2: SVM Loss function with $\lambda = 0.1$, $\eta = 0.01$, epochs = 60 and n_batch = 100.

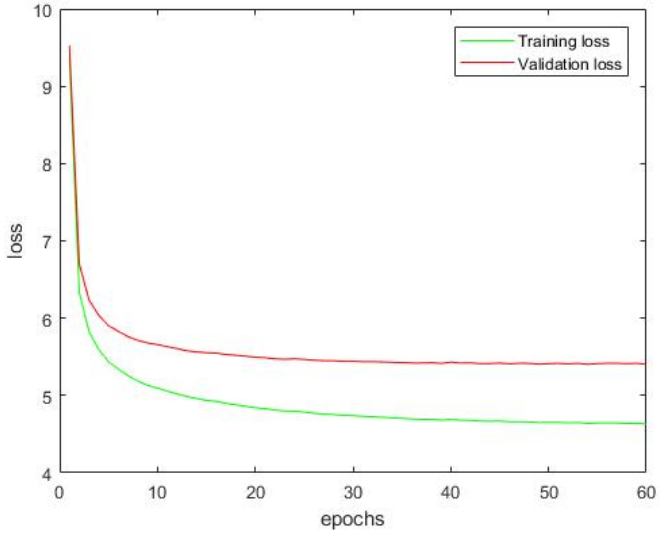


Figure 3: SVM Loss function with $\lambda = 0.1$, $\eta = 0.001$, epochs = 60 and n_batch = 100.

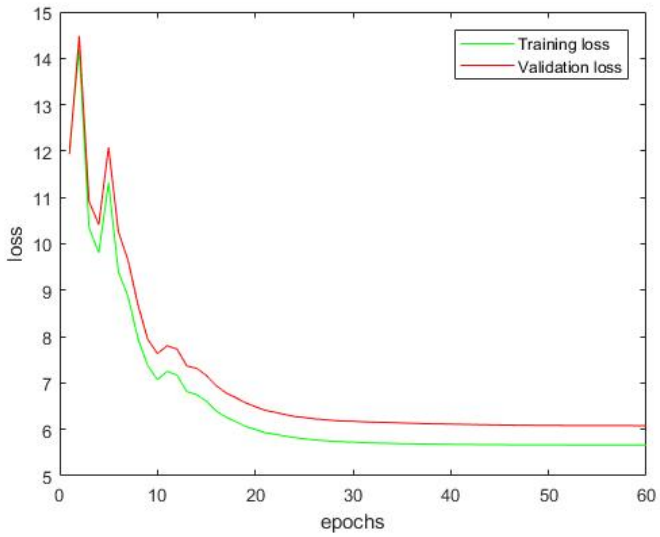


Figure 4: SVM Loss function with $\lambda = 1$, $\eta = 0.01$, epochs = 60 and n_batch = 100.