DD2424 Deep Learning in Data Science

Assignment 1

Ali Shibli

April 2021

1 Introduction

Here I will report on the bonuses that I tackled form assignment 1 of the Deep Learning Course. From the first part, I solved 1.a, 1.b. and 1.d. As well, I solved bonus question 2.

2 Bonus 1

1.a (Big training set and small validation set)

This exercises aimed to use the entire training data (all batches) for training, with randomly splitting around 1000 sample images for validation.

1.b (Early stopping and more epochs)

This exercise aimed at increasing the number of epochs, as well as performing early stopping of the gradient updating while monitoring the validation loss (when val_loss decreases below a certain threshold, we stop training). Here we use a threshold of 0.4 increased of validation loss.

1.d (Decaying learning rate)

This exercise aimed on performing a decay for the learning rate. We perform this decay as multiplicative factor of 0.9 after each complete epoch.

Results for Bonus 1

The following table highlights to testing accuracies from the different implemented methods.

Case	Big training set	Early Stopping	Learning rate decay	Combined
1	0.3243	0.2895	0.3042	0.3812
2	0.4083	0.3859	0.3508	0.407
3	0.3997	0.3907	0.3595	0.4046
4	0.363	0.3688	0.3591	0.3774

From this table of testing accuracies, we can see that the highest testing accuracies where generally reported with the bigger training set, which makes sense since we are given a lot of training data and thus more capacity to learn.

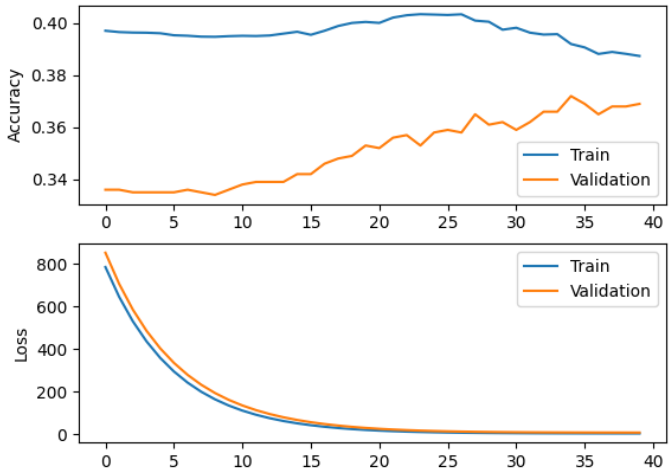
- The best results for using only 1.a were reported by the second case which had learning rate = 0.001 and lambda = 0.0. The test accuracy for this model was 40.82%
- The best results for using only 1.b were reported by the third case as well which had learning rate = 0.001 and lambda = 0.1. The test accuracy for this model was 39.07%.
- The best results for using only 1.d were reported by the third case as well. The test accuracy for this model was 35.95%.
- For the combination of the three methods, the second case got higher testing accuracy of 40.7% .

From these results we conclude that generally the best performing model was the of case 2, with the using the combination of the three implemented methods, since if we notice the column of the combined methods, generally the the values are higher (higher average accuracy).

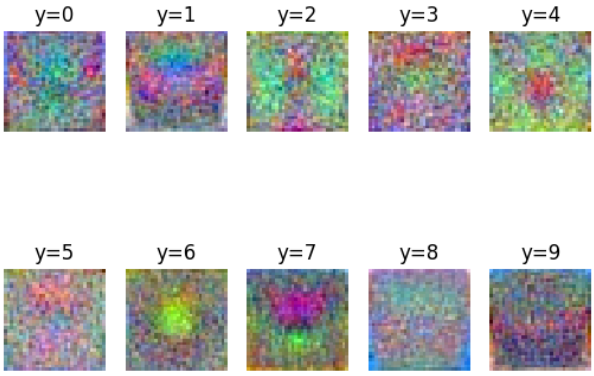
3 Bonus 2

For this part, we implement the SVM loss function. As before, this loss function is divided into SVM loss l_svm and regularization loss $l_regularization$.

Using the best performing network we deduced in the previous section (combination), we train the network again using the SVM loss function instead of the cross entropy loss. The results for the best performing parameters of the previous section with SVM loss instead are shown in the figure below:



(a) Accuracy and loss graphs for train and validation sets



(b) Weight matrix visualization

Figure 1: Case 4 results

The model in this case has $\lambda = 0.1$ and $\mu = 0.001$ (case 3). The test accuracy in this case is 39.93%.