Assignment 1 - BONUS - DD2424

Alexandra Hotti

June 12, 2019

1 Introduction

This report contains the results from an assignment were the feed forward neural network from the basic part of the assignment was used to classify the images of the Cifar-10 data set.

2 Part 1

The performance of the network was attempted to be improved by implementing suggestions a, d and g from the assignment description (see page 11 in the Assignment description). All of these attempts at improving performance where achieved using the following parameter setting:

$$lambda = 0.1, epochs = 40, batchsize = 100, eta = .01$$

These parameters where chosen since they were the parameters of one of the best performing network in the first basic part of the assignment (see parameter setting 3). I picked this setting as it gave a relatively high final test accuracy of 33.37 %, but also as it produced stable cost and loss plots which did not diverge.

2.1 Shuffling training data

By shuffling the training data at the beginning of every epoch a slightly lower overall loss and cost, as well as a higher final test set accuracy of 34.79 % was achieved.

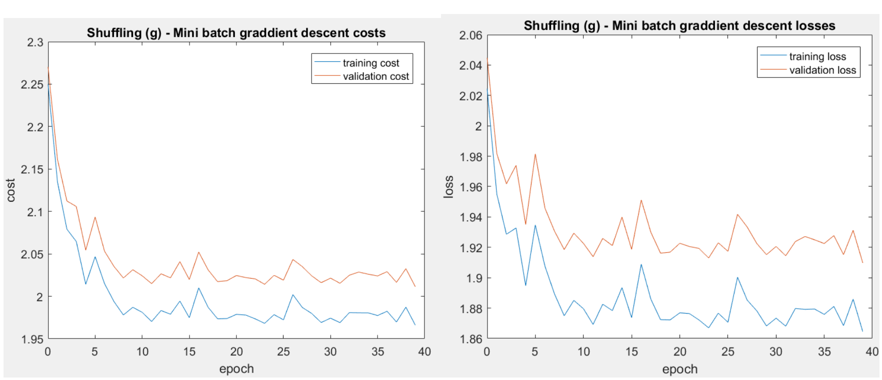

Figure 1: Costs and losses after shuffling the training data during training.

2.2 Using all available training data

By using all available data except for a 1000 data points for training the following results where achieved. The plotted losses and costs stabilized faster compared to in the basic case. Also the final test accuracy of 36.32 % was slightly higher.

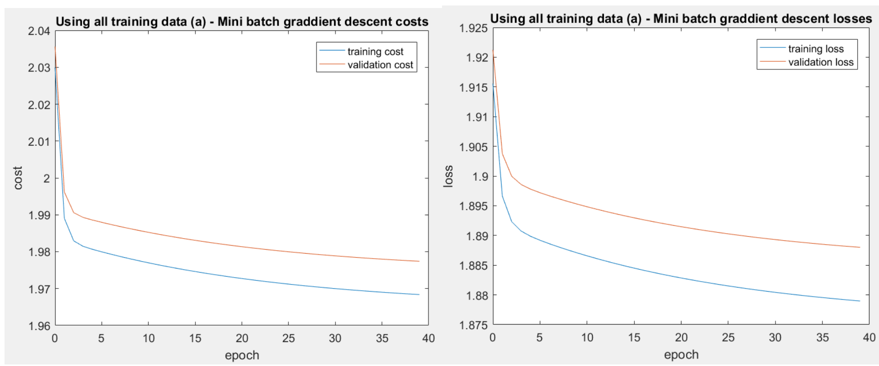

Figure 2: Costs and losses after increasing the amount of training data.

2.3 Decaying learning rate

Decaying the learning rate by a factor 0.9 after each epoch resulted in a slightly higher test accuracy of 36.43 % on the final test dataset.

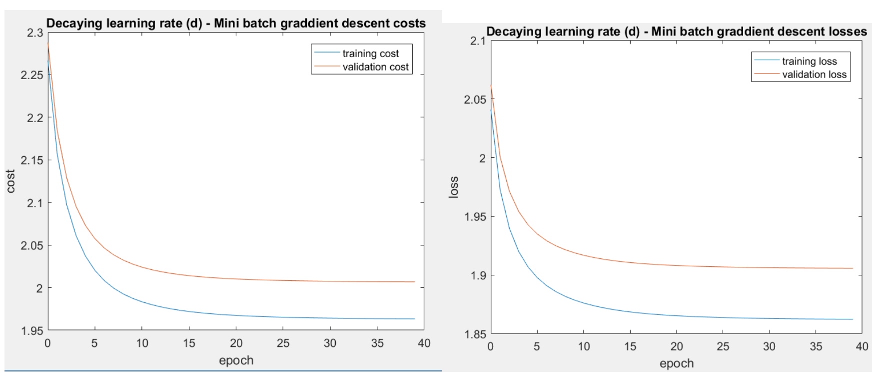

Figure 3: Costs and losses after decaying the learning rate after every epoch of training.

2.4 Combining the above optimization techniques

Finally, by combining all of the above improvements, the following results where achieved. Again, the final test accuracy of 36.81 % was higher than in the basic case.

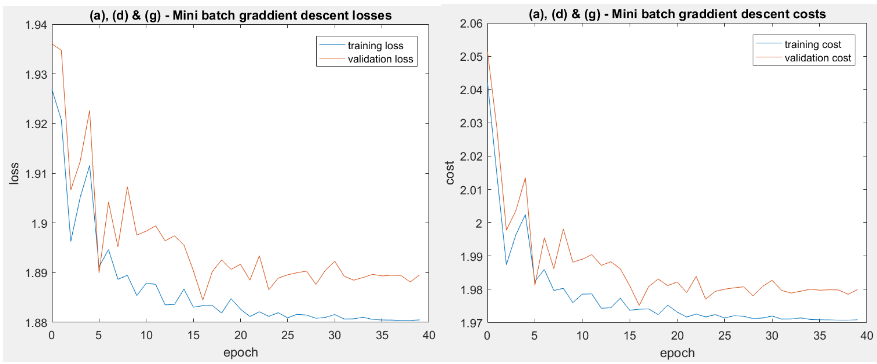

Figure 4: Costs and losses after increasing the amount of training data.

3 Conclusion

The highest test accuracy was achieved by combining all of the above improvements. However, this final accuracy was only slightly higher than when we used the other improvements in isolation. Overall this final test accuracy was 3,44 percentage points higher than the accuracy achieved in the basic version of the network.

The optimization that in isolation resulted in the best test accuracy was decaying the learning rate by a factor of 0.9. This method increased the final test accuracy by 3.06 percentage points.

4 Part 2

Below the final test accuracies for 17 different parameter settings have been computed with a SVM and a cross entropy loss respectively (See Table 1). These tests were carried out by varying one parameter at the time within a sensible range.

Overall the cross-entropy achieved higher test accuracies compared to the SVM loss. Also, the SVM loss resulted in accuracies as low as 20.05 %, while the cross-entropy loss at its lowest was at 25.79 %. Furthermore, both networks performed better when the learning rate decreased, however the cross-entropy performance dropped considerably when the learning rate became at its smallest (0.0005) while the SVM network performed at its best here. Lastly, both network's performance dropped when lambda increased to much (0.2).

Finally, two networks were trained using the best performing parameters based on test accuracy for the individual networks. These parameters are marked in bold text in Table 1. The SVM loss achieved a test accuracy of 34.99 % (see Table 2) while the cross-entropy loss resulted in a higher test accuracy of 38.75 % (see Table 3).

Table 1: Comparison of test accuracies for different parameters for the SVM and the cross-entropy loss.

λ	batch size	eta	epochs	SVM Accuracy	Cross Entropy Accuracy
0	100	0.1	40	24.07 %	25.79 %
0	100	0.01	40	30.61 %	**36.78 %**
0	100	0.005	40	32.6 %	35.89 %
0	100	0.001	40	35.51 %	31.41 %
0	100	0.0005	40	**35.59 %**	27.92 %
0.01	100	0.005	40	27.63 %	**35.94%**
0.1	100	0.005	40	**29.56 %**	35.61%
0.2	100	0.005	40	28.8 %	34.92 %
0.1	100	0.005	10	28.5 %	33.59 %
0.1	100	0.005	20	**28.91 %**	**35.66 %**
0.1	100	0.005	50	27.98 %	35.6 %
0.1	100	0.005	70	28.28 %	35.61 %
0.1	20	0.005	20	23.23 %	35.60 %
0.1	50	0.005	20	**31.65 %**	**36.58 %**
0.1	70	0.005	20	22.83 %	36.30 %
0.1	120	0.005	20	20.05 %	35.17 %
0.1	150	0.005	20	31.00 %	34.88 %

The combination of the parameters which achieved the highest test set accuracy for the SVM loss can be found below. These obtained a test set accuracy of 34.99 %.

$$\lambda = 0.1, batchsize = 50, eta = 0.0005, epochs = 20$$

Next, the combination of the parameters which achieved the highest test set accuracy for the cross-entropy loss can be found below. These obtained a test set accuracy of 38.75 %.

$$\lambda = 0, batchsize = 100, eta = 0.01, epochs = 40$$