DD2424 Deep Learning in Data Science - Assignment 1 Bonus

Bas Straathof — btstr@kth.se

April 5, 2019

1 Bonus Points

For the first part of the bonus questions, it was chosen to (1) train on all data, (2) use learning rate decay, and (3) use Xavier intialization. For the second part of the bonus questions, the multi-class SVM loss, the same evaluation parameters and settings were used as in the regular cross-entropy loss case.

1.1 Optimize the performance of the network

(a) - Train on all data When trained on all data and a validation set of 1,000 images (i.e. a training set of 49,000 images), the following cost plots (see Figures 1-4 for the cost plots with `np.random.seed(0)`) and mean accuracies (with standard deviation) over 5 runs were obtained:

1 Settings: **lambda=0, n epochs=40, n batch=100, eta=.1**

- The accuracy on the training set is: 0.2960 ± 0.0480
- The accuracy on the validation set is: 0.2762 ± 0.0487
- The accuracy on the testing set is: 0.2786 ± 0.0430

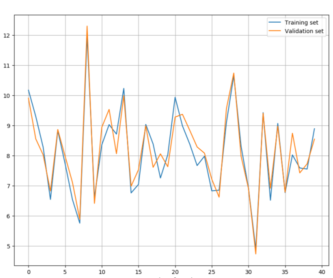

Figure 1: Graph of the total cost on the training and validation data for the parameter settings `lambda=0, n epochs=40, n batch=100, eta=.1`.

2 Settings: **lambda=0, n epochs=40, n batch=100, eta=.01**

- The accuracy on the training set is: 0.4572 ± 0.0030

- The accuracy on the validation set is: 0.3860 ± 0.0019
- The accuracy on the testing set is: 0.3869 ± 0.0012

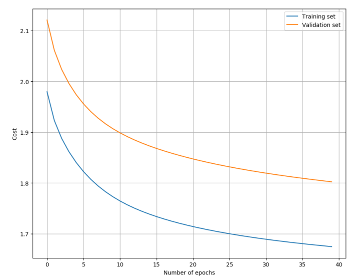

Figure 2: Graph of the total cost on the training and validation data for the parameter settings `lambda=0, n epochs=40, n batch=100, eta=.01`.

3 Settings: **lambda=.1, n epochs=40, n batch=100, eta=.01**

- The accuracy on the training set is: 0.3671 ± 0.0014
- The accuracy on the validation set is: 0.3762 ± 0.0013
- The accuracy on the testing set is: 0.36596 ± 0.0007

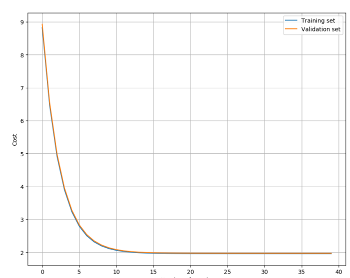

Figure 3: Graph of the total cost on the training and validation data for the parameter settings `lambda=.1, n epochs=40, n batch=100, eta=.01`.

4 Settings: **lambda=1, n epochs=40, n batch=100, eta=.01**

- The accuracy on the training set is: 0.2962 ± 0.0009
- The accuracy on the validation set is: 0.3026 ± 0.0046
- The accuracy on the testing set is: 0.3019 ± 0.0007

We see that even though all accuracies became a bit better, a learning rate η that is too big, or too much regularization still leads to the same problem.

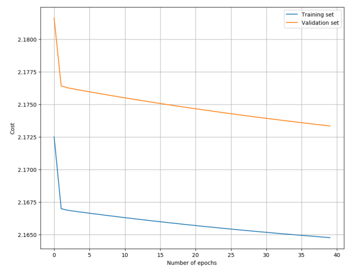

Figure 4: Graph of the total cost on the training and validation data for the parameter settings `lambda=1, n epochs=40, n batch=100, eta=.01`.

(d) - Learning rate decay

When applying a learning rate decay of 0.9 on the parameter settings where $\eta = 0.1$ and $\lambda = 0$, the accuracies obtained are way better (see Figure 5 for the cost plot). This model has obtained the highest accuracies:

1 Settings: **lambda=0, n epochs=40, n batch=100, eta=.1** with 0.9 learning rate decay

- The accuracy on the training set is: 0.4018 ± 0.0133
- The accuracy on the validation set is: 0.4010 ± 0.0072
- The accuracy on the testing set is: 0.3933 ± 0.0091

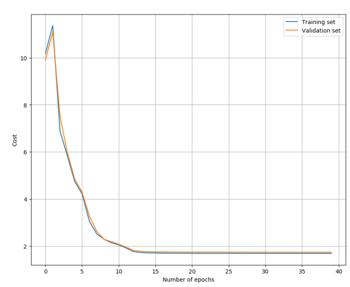

Figure 5: Graph of the total cost on the training and validation data for the parameter settings `lambda=0, n epochs=40, n batch=100, eta=.01` with 0.9 learning rate decay.

(e) - Xavier initialization

In this case, Xavier initialization does not make a big difference. Before, W was implemented as a $(10, 3072)$ matrix with entries sampled from a normal distribution with mean 0 and standard deviation 0.01. In the case of Xavier initialization, we still sample from a normal distribution but the standard deviation is set to $\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{3072}} \approx 0.018$. These weight initialization are sufficiently similar to not have a big impact on training stability. For concision, the plots have therefore been omitted.

1.2 Train network by minimizing the SVM multi-class loss

See Figures 6-9 for the cost plots of the training and validation sets for the different parameter settings with multi-class SVM loss. From these plots we can conclude that this type of classifier also suffers from unstable training when the learning rate is to high and that not much can be learned when the regularization term is high. There are some discrepancies in the performance of the multi-clss SVM classifier in comparison to the softmax classifier. These are most likely due to suboptimal parameter settings. When training on all data, applying learning rate decay and Xavier optimalization and optimizing the regularization parameter λ, the multi-class SVM classifier should perform about as well as the best softmax classifier. The following mean classification (with standard deviation) results over 10 runs were obtained:

1 Settings: **lambda=0, n epochs=40, n batch=100, eta=.1**

- The accuracy on the training set is: 0.3063 ± 0.0604
- The accuracy on the validation set is: 0.2461 ± 0.0403
- The accuracy on the testing set is: 0.2476 ± 0.0391

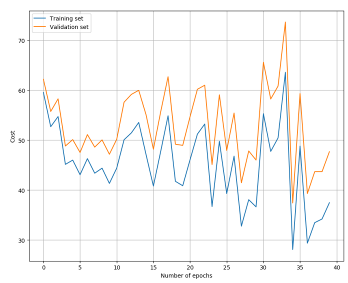

Figure 6: Graph of the total cost on the training and validation data for the parameter settings `lambda=0, n epochs=40, n batch=100, eta=.1`.

2 Settings: **lambda=0, n epochs=40, n batch=100, eta=.01**

- The accuracy on the training set is: 0.5036 ± 0.0053
- The accuracy on the validation set is: 0.2971 ± 0.0036
- The accuracy on the testing set is: 0.3013 ± 0.0043

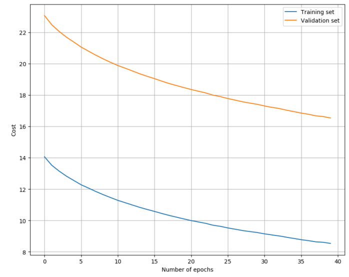

Figure 7: Graph of the total cost on the training and validation data for the parameter settings `lambda=0, n epochs=40, n batch=100, eta=.01`.

3 Settings: **lambda=.1, n epochs=40, n batch=100, eta=.01**

- The accuracy on the training set is: 0.2926 $\pm$ 0.0154
- The accuracy on the validation set is: 0.2671 $\pm$ 0.0130
- The accuracy on the testing set is: 0.2703 $\pm$ 0.0122

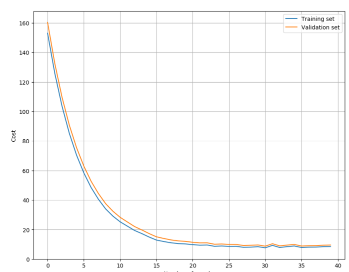

Figure 8: Graph of the total cost on the training and validation data for the parameter settings `lambda=.1, n epochs=40, n batch=100, eta=.01`.

4 Settings: **lambda=1, n epochs=40, n batch=100, eta=.01**

- The accuracy on the training set is: 0.2344 $\pm$ 0.0270
- The accuracy on the validation set is: 0.2247 $\pm$ 0.0259
- The accuracy on the testing set is: 0.2272 $\pm$ 0.0254

Settings : **lambda=0, n epochs=40, n batch=100, eta=.01** on the full dataset with learning rate decay and Xavier initialization.

- The accuracy on the training set is: 0.2699 $\pm$
- The accuracy on the validation set is: 0.2680 $\pm$
- The accuracy on the testing set is: 0.2623 $\pm$

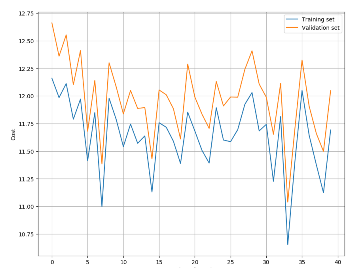

Figure 9: Graph of the total cost on the training and validation data for the parameter settings `lambda=1, n epochs=40, n batch=100, eta=.01`.