

DD2424 – Assignment 1 Bonus

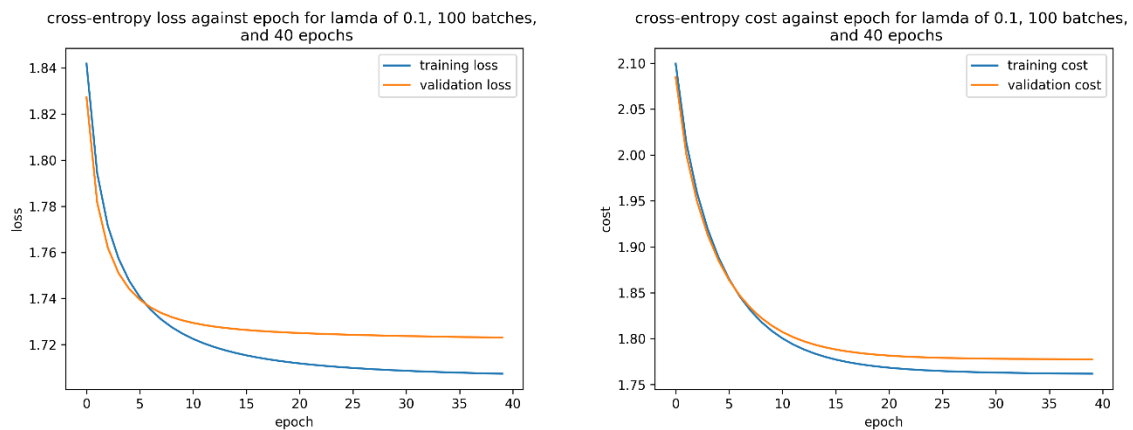
Aloysius Chow

26/04/2023

Training the network

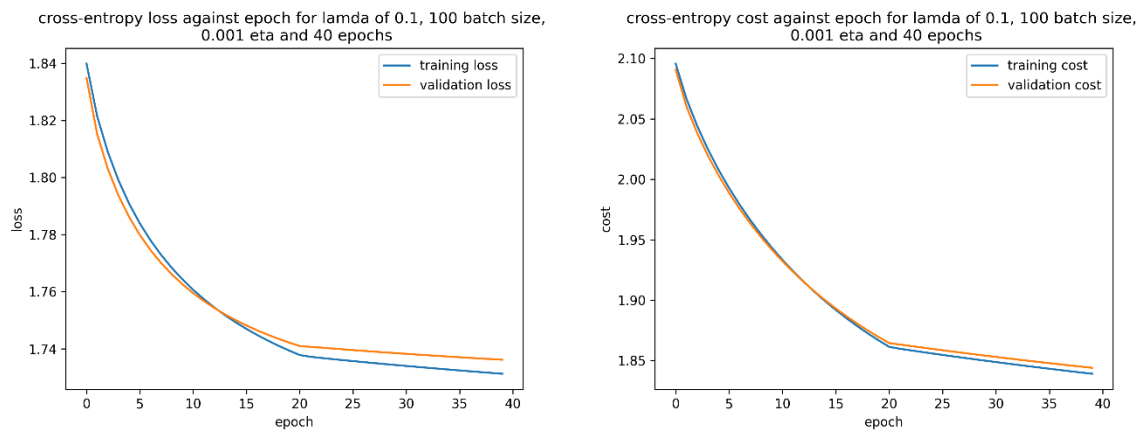
First, a baseline model was trained for 40 epochs with a lambda of 0.1, batch size of 100, and eta of 0.001.

The cross-entropy loss and cost curves are shown below, and the model achieved a final accuracy on the test set of 40.91%.



Next, decaying eta was introduced, where the eta would be divided by 4 every 20 epochs.

The cross-entropy loss and cost curves are shown below, and the model achieved a final accuracy on the test set of 40.48%.



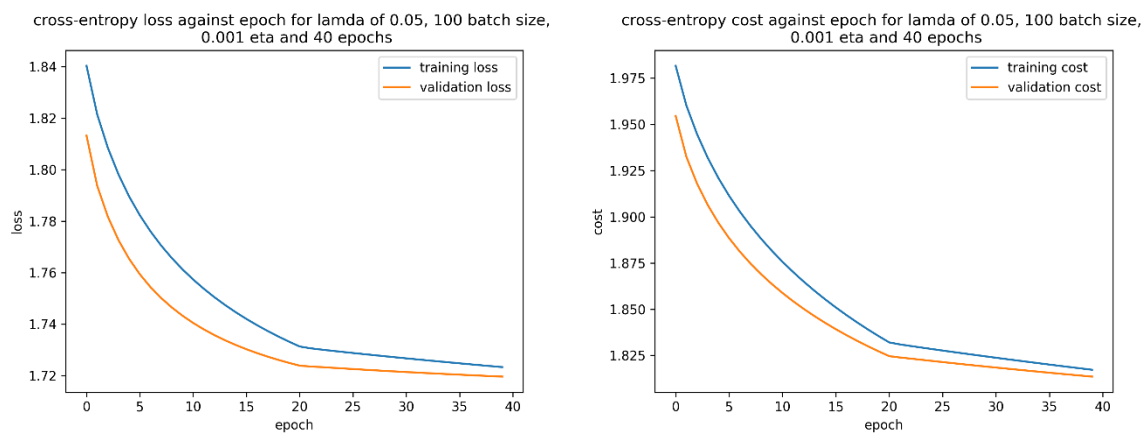
Finally, the last enhancement added was image augmentation, done by randomly mirroring a proportion of images about the vertical axis.



The function was tested by running on the first 10 images of the first batch, with a probability of 1 to ensure all images were flipped. The results can be seen above.

Then, training was performed with a probability of 0.1 (i.e, 10% of images would be flipped), as well as decaying eta. Lambda was reduced as recommended, to 0.05.

The cross-entropy loss and cost curves are shown below, and the model achieved a final accuracy on the test set of 40.45%.



Parameter search

A grid search was performed with the following ranges:

$\lambda = 10^i$ for 5 values of i , linearly spaced from $i = -3$ to $i = -2$

$\eta = 10^i$ for 5 values of i , linearly spaced from $i = -3$ to $i = -1$

batch size = i for 5 values of i , linearly spaced from $i = 20$ to $i = 120$

Image augmentation and decaying eta were used.

Validation accuracy for each permutation of parameters was obtained, and the top 10 permutations by accuracy are shown below:

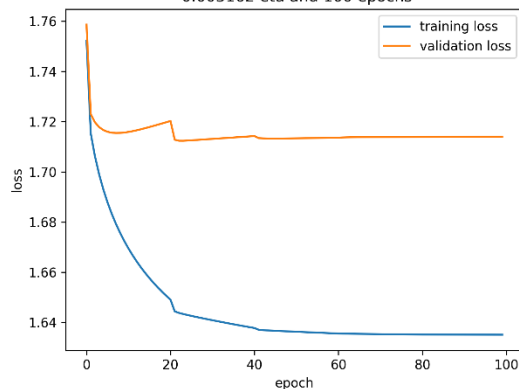
lambda	eta	batch size	average validation accuracy
0.003162	0.003162	45	0.432
0.003162	0.01	95	0.43
0.001	0.01	120	0.427
0.01	0.003162	45	0.426
0.001778	0.003162	45	0.425
0.005623	0.003162	70	0.425
0.003162	0.003162	70	0.424
0.003162	0.01	120	0.424
0.005623	0.01	95	0.424
0.005623	0.001	20	0.423

A final model was trained for 100 epochs with the best performing parameters of:

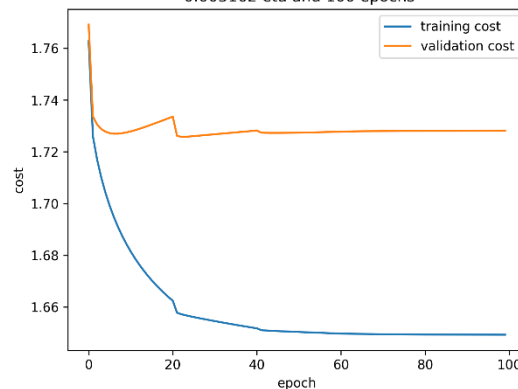
lambda = 0.003162, eta = 0.003162, n_batch = 45, as well as image augmentation and decaying eta.

The cross-entropy loss and cost curves are shown below, and the model achieved a final accuracy on the test set of 41.53%.

cross-entropy loss against epoch for lamda of 0.003162, 45 batch size, 0.003162 eta and 100 epochs



cross-entropy cost against epoch for lamda of 0.003162, 45 batch size, 0.003162 eta and 100 epochs



Interestingly, we notice a spike in the validation loss and cost at the first reduction of eta, which we did not observe in the earlier tests.

Conclusion

This served as a means of better understanding how the performance of neural networks can be improved through various means, like dynamically changing the learning rate (which we explore more in assignment 2), increasing the amount of data available, and by using data augmentation. Furthermore, it also allowed me to explore how grid search can be performed manually, even though I may have used it in simple means through other packages like Scikit-learn in the past.