Kungliga Tekniska Högskolan

DD2424 Deep Learning in Data Science

Kévin YERAMIAN

June 6, 2018

1 Introduction

In this assignment I will train and test a one layer network with multi- ple outputs to classify images from the CIFAR-10 dataset. You will train the network using mini-batch gradient descent applied to a cost function that computes the cross-entropy loss of the classifier applied to the labelled training data and an L2 regularization term on the weight matrix.

2 Gradient

In order to be sure about the gradient, I compared the value with value of the others gradient function. I limited the dataset size at 100, to first see if it is working. There is my result for the gradient:

- gradient of W1:

Respectively 2.2800898156053947e-05 and 2.6422640045174324e-08 for finite difference method and centred difference method.

- gradient of b1:
Respectively 2.707199023141846e-07 and 7.635144507809246e-08 for finite difference method and centred difference method.

I confirmed my result by setting the size at 1000, There is my result for the gradient:

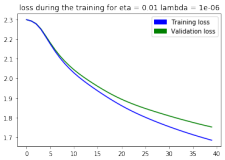
- gradient of W2:
Respectively 1.660935868583902e-05 and 2.3941229365317155e-09 for finite difference method and centred difference method.

- gradient of b2:
Respectively 5.59956958681861e-07 and 2.414321369873949e-08 for finite difference method and centred difference method.

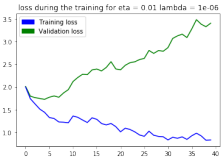
The result is good, we have close similarity. I assumed that the gradient is correct for the rest of the assignment.

3 Momentum term

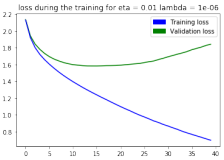
We tried 3 different momentum : 0.5, 0.9, 0.99. We can see below the graphs of the total loss and the cost function on the training data and the validation data after each epoch of the mini-batch gradient descent algorithm and images representing the learnt weight matrix after the completion of training. The learning rate is equal to 0.01, we use the full training set with a batch equal to 100 and 40 epochs.



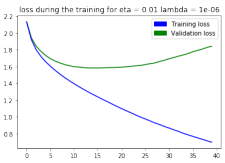
(a) momentum 0



(b) momentum 0.5



(c) momentum 0.9



(d) momentum 0.9

Figure 1: lambda=0.000001, n epochs=40, n batch=100, eta=.01

We see that the momentum have a direct impact on the result. We good momentum that we can choose is 0.9. The momentum have positive impact on the result if it is well fixed.

4 The 3 best performing networks during the coarse search.

I took 9000 of the 10000 images of the trainset to train each model. I search lot of parameters of eta and lambda, I set batch=100,decay-rate=0.95, rho=0.9 and epochs=5:

- lambda=0.01, eta = 0.01 and accuracy = 39.83.
- lambda=0.0001, eta = 0.02 and accuracy = 42.4.
- lambda=0.000001, eta = 0.1 and accuracy = 40.44.

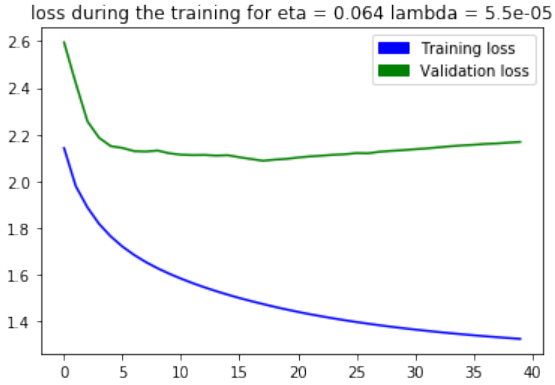
5 Grid Search

I took 9000 of the 10000 images of the trainset to train each model. I set batch=100,decay-rate=0.95, rho=0.9 and epochs=5. I look between 0.01 and 0.1 for eta and between 0.00001 and 0.0001 for lambda with 10 different eta and 20 different lambda so 200 combinations. For each combination I run 5 epochs and compute the accuracy on the testset. This is the 3 bests parameters I found:

- lambda=1e-05, eta = 0.055 and accuracy = 44.3.
- lambda=4.6e-05, eta = 0.064 and accuracy = 44.5.
- lambda=5.5e-05, eta = 0.064 and accuracy = 45.1.

6 Best Model

I took the all images of the trainset to train each model and the all images of the validationset. I set batch=100,decay-rate=0.95, rho=0.9, lambda=5.5e-05 and eta = 0.064. The accuracy is 45.05 on the testset.



(a) Best model

Figure 2: batch=100,decay-rate=0.95, rho=0.9, lambda=5.5e-05 and eta = 0.064, epochs = 40

7 Conclusion

In this lab we saw the influence of many parameters for a neural network with one layers. Indeed, the influence of the learning rate, momentum, decay-rate and lambda. We also focus on the fact that the only way to find the best parameters is to do a grid search on a interval we think it's good.