Assignment 2 in DD2424 Deep Learning

C. Chou

Kungliga Tekniska Hogskolan

chchou@kth.se

April 20, 2019

Gradient Computation

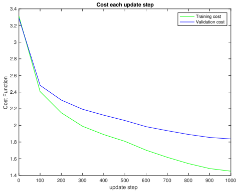
I successfully managed to compute the gradient analytically by using following code:

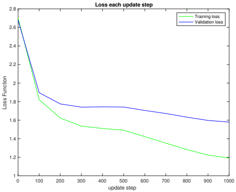
```
gradW = cell(1,2);
gradb = cell(1,2);
g = - (Y - P);
gradW2 = g * h'./n + 2 * lambda * W2 ;
gradb2 = sum(g,2)./n;

gbatch = W2'* g;
indh = h > 0;
gbatch = gbatch .* ind_h;
gradW1 = gbatch * X'./n + 2 * lambda * W1;
gradb1 = sum(gbatch,2)./n;
```

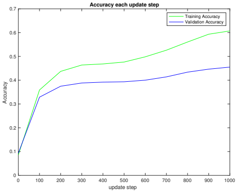
And the result was tested by computing the absolute difference between this result and the result from centered difference formula. The difference is less than 1e-6.

Cylinical Learning Rate - 1

The result figures are showed below:



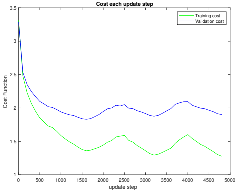


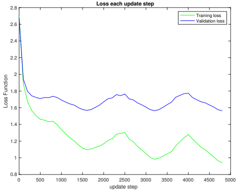


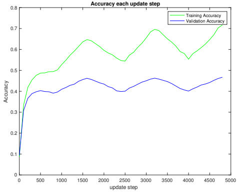
The accuracy reaches 46.00%.

Cylinical Learning Rate - 2

The result figures are showed below:



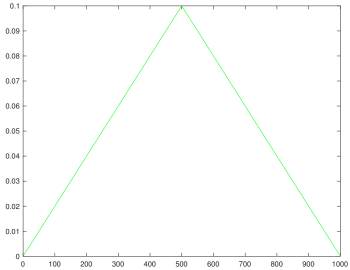


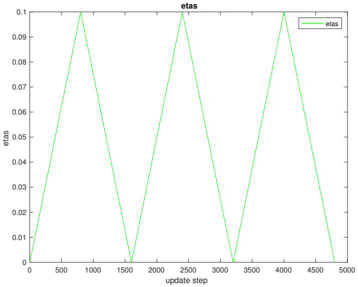


The accuracy reaches 46.83%. After three cycles of training set, my result didn't reach as what professor got when testing, but in the above plots we notice that how the loss and accuracy vary as the learning rate varies, and the learning rate is designed exactly the same as the instruction required.(see the next slide). so I convinced myself that I have a bug free implementation of the cyclic scheduled learning rates.

Cylinical Learning rate - 3

The following figures show the Cylinical Learning rate required in the instruction:





The left plot is the learning rate for one cycle of training ,following by figure 3 in the instruction. The right plot is the learning rate for three cycles of training following by figure 4 in the instruction.

Coarse Search

The following table shows the first three search with best performance with their *lambda* :

l	λ	Accuracy(%)
-4.8507	1.4103e-5	46.72%
-4.3972	4.0071e-5	46.62%
-2,5395	2.9e-3	46.44%

with l_{min} = -5 ; l_{max} = -1 ; 2 cycles ; n_{epoch} = 10 ; n_s = 75 ; n_{batch} = 1500 ,using 45000 training data with 5000 validation data.

Fine Search

After the Coarse search, I found that model performed better with λ in the range between e-5 to e-3 with 2 cycles, so I narrow down the λ and increased the number of update steps. After several time training with 49000 training data and 1000 validation data, The best results with their parameters are shown below:

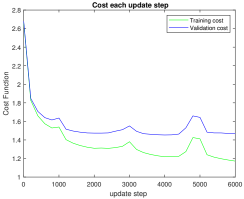
i $\lambda = 0.0014$, n_{epoch}=40 , n_{batch}= 490, ns = 2000, 1 cycle (51.19 %)

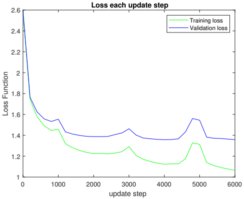
ii $\lambda = 0.00076$, n_{epoch}=40 , n_{batch}= 490, ns = 2000, 1 cycle (50.90 %)

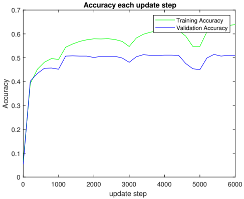
iii $\lambda = 0.000538$, n_{epoch}=40 , n_{batch}= 490, ns = 1000, 2 cycles (50.87 %)

best λ

From the previous slide, the best λ is found as 0.0014. Now, using λ as 0.0014, I set $n_{epoch}=60$, $n_{batch}=490$, ns $=1000$ for 3 cycles, and the following figures are the result:







The accuracy is 50.78 %