EECS545(Section001): Homework #4

Due on Mar.16, 2022 at 11:59pm

Instructor:Honglak Lee

Tiejin Chen **tiejin@umich.edu**

Problem 1

(a)

It is easy for us to compute:

$$\frac{\partial Y_{kl}}{\partial b} = 1$$

$$\frac{\partial L}{\partial b} = \sum_{k=1}^{N}\sum_{l=1}^{dout}\frac{\partial L}{\partial Y_{kl}} \times 1 = \sum_{i=1}^{N}\frac{\partial L}{\partial Y_i}$$

Now consider the W and X. For W, we have:

$$\frac{\partial Y_{kl}}{\partial W_{ij}} = \frac{\partial \sum_s (X_{ks}W_{sl})}{\partial W_{ij}} = \frac{\partial X_{ki}W_{il}}{\partial W_{ij}} + 0 = X_{ki}I(l=j)$$

$$\frac{\partial L}{\partial W_{ij}} = \sum_{j,l}\frac{\partial L}{\partial Y_{kl}}X_{ki}I(l=j)$$

Thus we can get:

$$\frac{\partial L}{\partial W} = X^T\frac{\partial L}{\partial Y}$$

Similarly, we can compute $\frac{\partial L}{\partial X}$:

$$\frac{\partial Y_{kl}}{\partial X_{ij}} = \frac{\partial \sum_s (X_{ks}W_{sl})}{\partial X_{ij}} = \frac{\partial X_{kj}W_{jl}}{\partial W_{ij}} = I(i=k)W_{jl}$$

$$\frac{\partial L}{\partial X_{ij}} = \sum_{j,l}\frac{\partial L}{\partial Y_{kl}}W_{jl}I(i=k)$$

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y}W^T$$

(b)

It is easy to see:

$$\frac{\partial L}{\partial X_ij} = \begin{cases} \frac{\partial L}{\partial Y_ij} \times 1, & X_ij \geq 0 \\ 0, & X_ij < 0 \end{cases}$$

(c)

First consider $X_{n,c,i,j}$. Then we have:

$$\frac{\partial Y_{n,f,m,n}}{\partial X_{n,c,i,j}} = \begin{cases} K_{f,c,k,l}, & X_{n,c,i,j}\ contribute\ to Y_{n,f,m,n} \\ 0, & else \end{cases}$$

More accurately, we have:

$$\frac{\partial Y_{n,f,m,o}}{\partial X_{n,c,i,j}} = \sum_{k=1}^{ph}\sum_{l=1}^{pw}K_{f,c,k,l}\frac{\partial x_{n,c,m+k-1,o+l-1}}{\partial X_{n,c,i,j}}$$

where ph and pw are the height and weight of K. Then we have $m+k-1=i, o+l-1=j$ Hence we get:

$$dX_{n,c,i,j} = \sum_{m=i-ph+1}^{i}\sum_{n=j-pw+1}^{j}\frac{\partial L}{\partial Y_{n,f,m,o}}K_{f,c,i-m+1,j-n+1} = (K_{f,c} *_{full} \frac{\partial L}{\partial Y_{n,f}})_{i,j}$$

Then we sum all K to get:

$$\frac{\partial L}{\partial X_{n,c}} = \sum_{i=1}^{F}(K_{f,c} *_{full} \frac{\partial L}{\partial Y_{n,f}})$$

Problem 1 continued on next page...

Now, let us consider $K_{f,c,k,l}$. We can get:

$$\frac{\partial Y_{n,f,m,o}}{\partial K_{f,c,k,l}} = X_{n,c,m+k-1,o+l-1}$$

Then we have:

$$dK_{f,c,k,l} = \sum_{m=1}^{M}\sum_{o=1}^{O} X_{n,c,m+k-1,o+l-1}\frac{\partial L}{\partial Y_{n,f,m,o}} = (X_{n,c} *_{filt} (\frac{\partial L}{\partial Y_{n,f}})_{k,l}$$

where M and O is the the third dimension and fourth dimension of shape of Y. Then to sum all N, we can get:

$$\frac{\partial L}{\partial K_{f,c}} = \sum_{n=1}^{N} X_{n,c} *_{filt} (\frac{\partial L}{\partial Y_{n,f}})$$

And that proves what we need.

Problem 2

(c)

We try 9 different settings for hidden units, And we get the result in validation set:

1. number of hidden units:10, validation acc:0.9496
2. number of hidden units:30, validation acc:0.9718
3. number of hidden units:50, validation acc:0.9764
4. number of hidden units:80, validation acc:0.9792
5. number of hidden units:100, validation acc:0.9788
6. number of hidden units:120, validation acc:0.9790
7. number of hidden units:150, validation acc:0.9810
8. number of hidden units:180, validation acc:0.9798
9. number of hidden units:200, validation acc:0.9794

The best setting for validation set is hidden units equal to 150. And we re-train the network with 150 hidden units and get the testing accuracy is 0.9767.

Problem 3

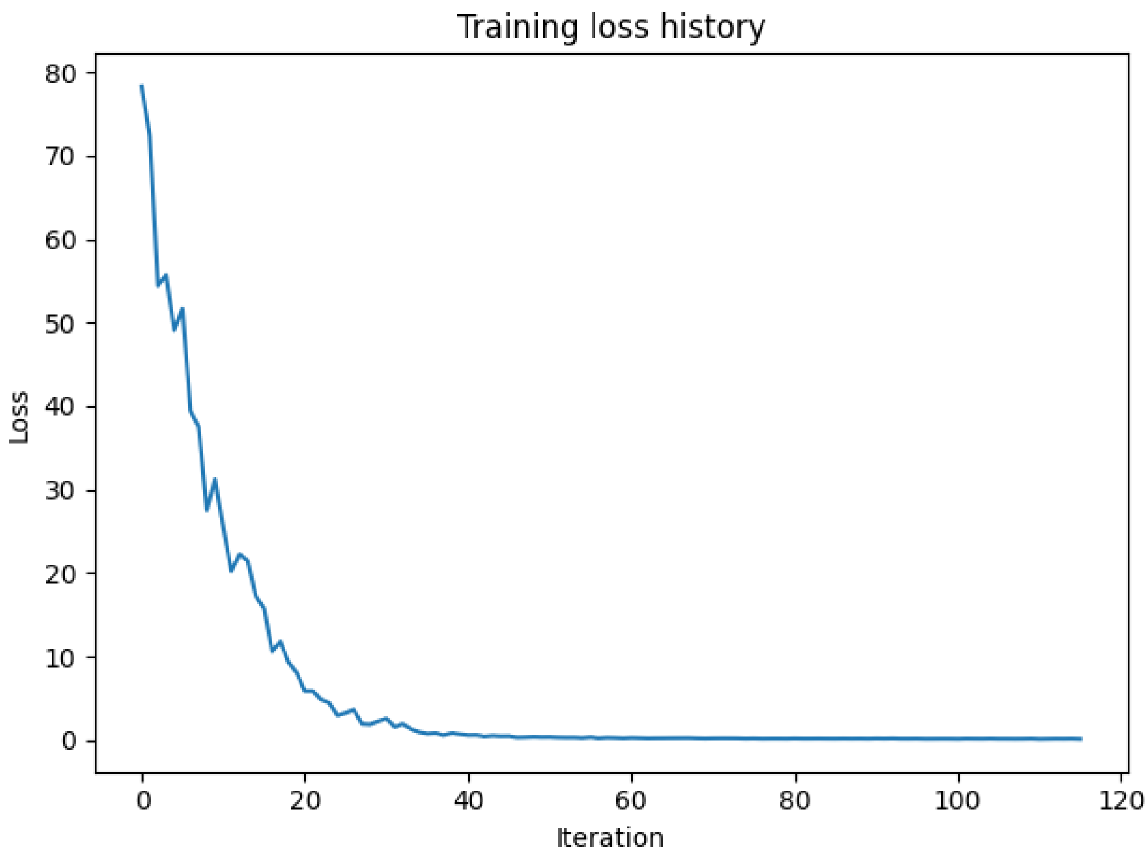
(c)

We run about 35 minutes to finish the training and get the test accuracy is 0.9795.

Problem 4

(c)

We can get the learning curve for training loss



And we also have caption samples:

train
four horses standing in a area that is fenced in <END>
GT:<START> four horses standing in a area that is fenced in <END>

train
a passenger train is parked while a person <UNK> to board it <END>
<START> a passenger train is parked while a person <UNK> to board it <EN

val
a player of a <UNK> with a a <UNK> <END>
RT> a surfer holding a surf board at the edge of the beach watching the surf

val
a round clock with <UNK> <UNK> hanging on the side of a building <END>
GT:<START> a large white clock is on a tower <END>

Problem 5

(e)

For fine tune, After 24 epoches, the validation accuracy is 0.9216. And the best validation accuracy during 24 epoches is 0.934641.

For freeze, after 24 epoches, the validation accuracy is 0.9346, And the best validation accuracy during 24 epoches is 0.954248.