Deep Learning Assignment2 Bonus

Mengdi Xue, mengdix@kth.se

April 2018

1 Exercise 5.1

I tried the following methods to help bump up performance:

1. Use more hidden nodes improves the final classification rate and increase the amount of regularization a little bit. I've tried several settings and found that **100 hidden nodes** are good enough for this problem. Too much hidden nodes, such as 150 nodes may cause the model to overfit so that the test accuracy would be decrease.

2. Use different approaches to anneal the learning rate. I tried to decay η by 0.1 after 10 epoches instead of decay it by 0.9 every epoch. It seems not very helpful and the loss becomes stepped. Then I tried to decay η by different factors and found that **decay it by 0.8** every epoch is better for the current parameter.

3. I switched it to **He initialization** and it performs well. This method may bring the largest gains.

With the above three improvements, the final test accuracy was improved to 51.88% as in Figure 1 and the loss is as in Figure 2

```
>> acc = ComputeAccuracy(X_test, y_test, W, b);
disp(acc);
    0.5188
```

Figure 1: Best test accuracy

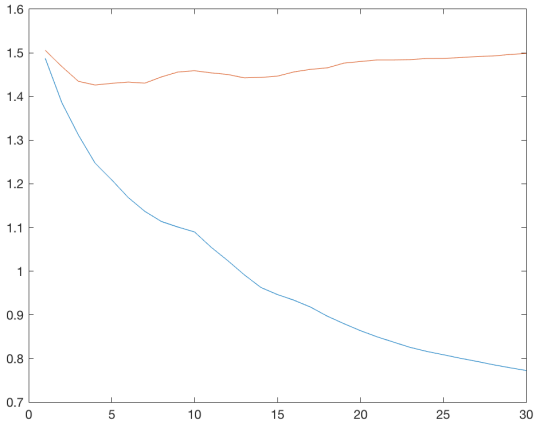

Figure 2: Loss

2 Exercise 5.2

The test accuracy of the network trained with **tanh** activation function compared to the same network with a ReLu activation function is as in Table 1 and the loss is as in Figure 3. Here I used 1000 samples with settings of $n\_epochs = 10, n\_batch = 100$ and random initialization.

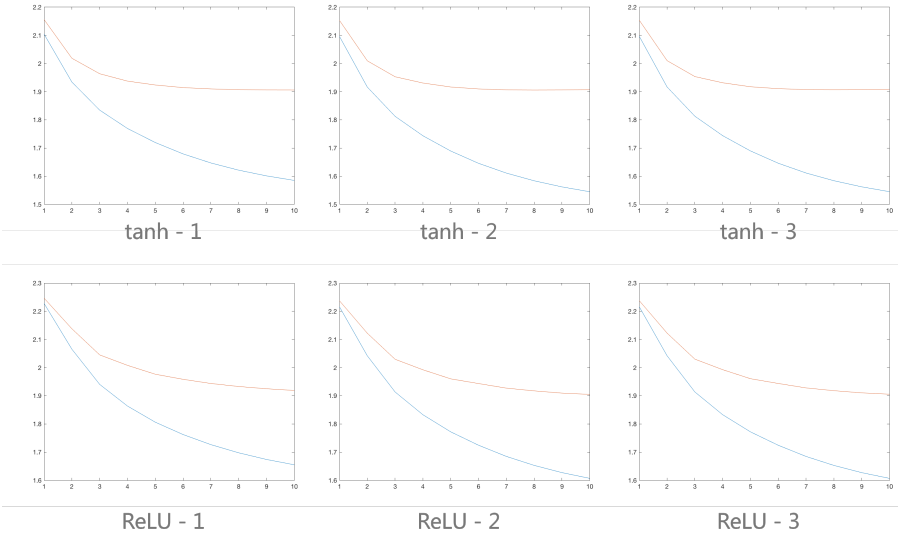

Figure 3: Loss of the three settings

Table 1: 3 Best parameters for find search

	setting 1	**setting 2**	**setting 3**
λ	2.24e-05	1.24e-05	3.91e-05
η	0.108993	0.118811	0.119100
ReLU accuracy	32.84%	33.33%	33.31%
tanh accuracy	33.19%	33.37%	33.34%

From the test accuracy, we can see that the accuracy of tanh is better than ReLU in this three testings. From the loss figure, we can see that the initialization helps tanh a lot since its beginning cost is lower than ReLU. But the biggest problem of tanh is that it is very time consuming. Since there are not much difference between their test accuracy, I think it's better to choose ReLU in large scale training.