UNIVERSITY AT BUFFALO

The State University of New York

Introduction to Machine Learning

(CSE 574)

Report of Programming Assignment 1

Group Members

Ting Zhou

Xuan Han

March 8, 2017

Explanation of Selected Hyper-Parameters Based on nnScript

1. **Analysis of the Regularization Coefficient, λ**

   To avoid overfitting problem, the regularization term could prevent the objective function from getting too accurate and being unable to predict the test data. However, if the regularization parameter λ is too large, the prediction model would be underfitting. In order to discover the effect of λ on the performance of the Neural Network, various λ from 0 to 1 in increments of 0.2 are tried. We use the hidden layer with 20 hidden units to test different λ. The results are summarized in Table 1.

   As shown in Figure 1, when λ is equal to 1 the objective function converges at 0.85, which means the regularization coefficient keep the model from too accurate. Meanwhile, the validation accuracy is kept at 92% when λ is equal to 1. The values of λ below 1 have no obvious difference for this Neural Network. After trying several times, we choose $\lambda = 1$ as an expected value to avoid overfitting and doesn't make the Neural Network underfitting.

Table1. Summary of λ Effect

lambda	Objective Value	Validation Accuracy(%)
0	0.458622087	92.28
0.2	0.443799146	92.72
0.4	0.420379365	92.99
0.6	0.526565408	91.68
0.8	0.42143523	93.71
1	0.849008211	92.28

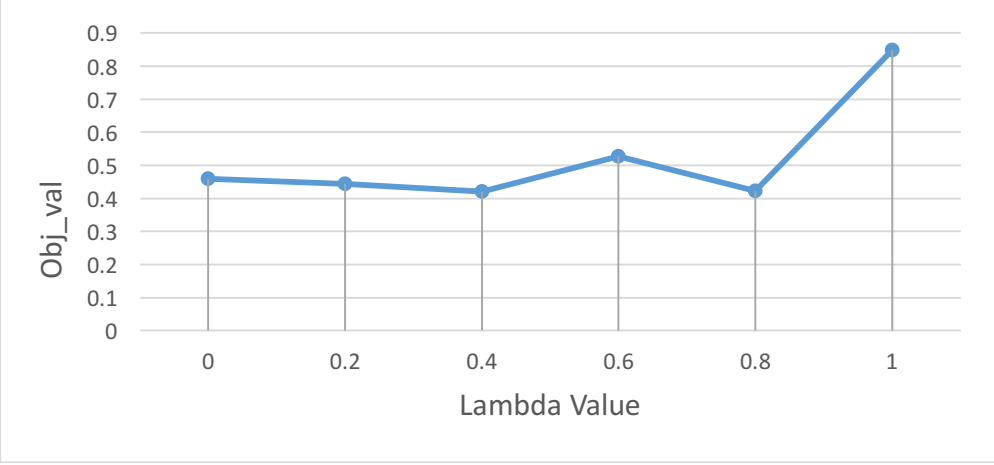

Figure1. Effect of λ on the Objective Value

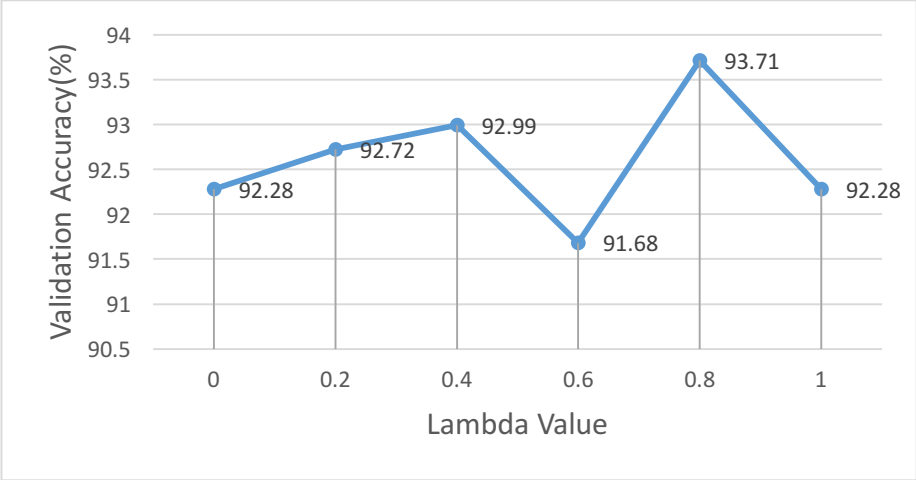

Figure2. Effect of λ on the Validation Accuracy

2. Analysis of the Number of Hidden Units

The performance of Neural Network highly depends on how many units in the hidden layers. For this single hidden layer Neural Network, different number of hidden units are tried. With the increasing number of hidden units, the prediction should be more accurate. And the Neural Network need more training time with more hidden units. Table 2 summarizes the results of the Neural Network with various number of hidden units. Using enough hidden layers and hidden units, the training accuracy should be 100%, in this assignment we only use single hidden layer and 20 hidden units at most.

In this part, the regularization coefficient λ is fixed at 1. From Table 2, Figure 3 and Figure 4, just like the previous statements, the more hidden units the more accurate predictions and training time. When the Neural Network has 20 hidden units, we got the most accurate results also costing the most training time. The training time increases linearly respect to the number of hidden units. It is noteworthy that the objective function converges faster with more hidden units. Just like the data of the objective values in Table 2, when using 4 hidden units, the objective function converges at 1.632. When using 20 hidden units, the objective function converges at 0.445. Thus from this aspect, the Neural Network with 20 hidden units is actually trained more efficiently than other ones. Thus, the optimal number of hidden units of this Neural Network is 20.

Table 2. Summary of Hidden Units Effects

Number of Hidden Units	Training Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Training Time (sec)	Objective Value
4	60.396	59.07	59.96	730.7	1.63207
8	86.976	85.87	86.5	767.8	0.83865
12	92.17	91.13	91.81	786.6	0.53186
16	92.342	91.57	91.88	891	0.51019
20	94.426	93.68	93.93	958.4	0.44509

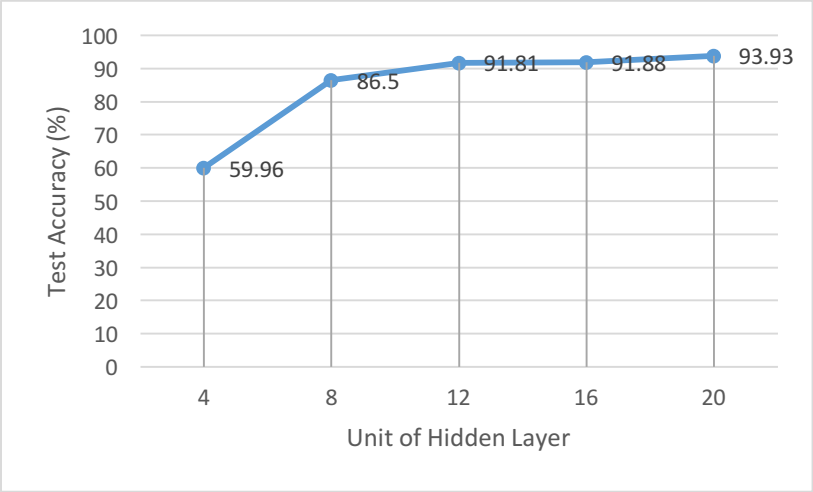

Figure 3. Trend of Test Accuracy

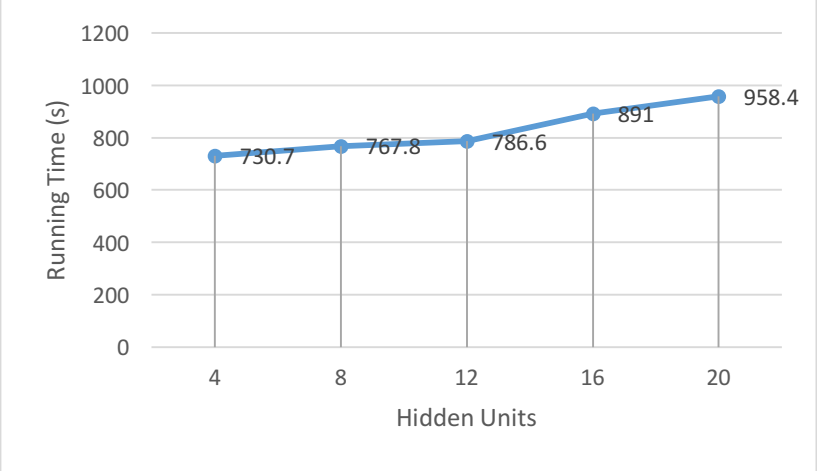

Figure 4. Trend of Training Time

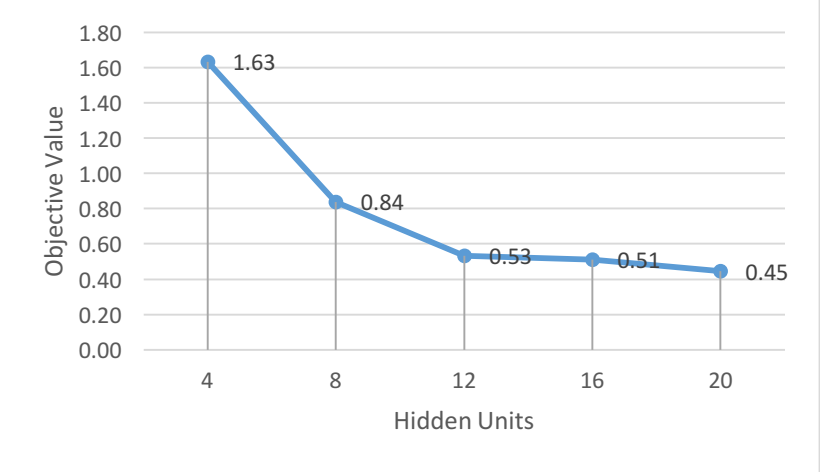

Figure 5. Convergence of Objective Function

Comparison of Our Neural Network with a Deep Neural Network

This part is based on the facennScript. The accuracy of classification method on the CelebA data set is summarized in Table 3. This data set need about three hours to complete the training, which is much longer than the handwritten digit data set. There are two reasons for the long running time. The first reason is the larger number of hidden units for the Neural Network of the CelebA data set. The second reason is that the CelebA data have 2376 features that is about three times than the handwritten digit data set.

Table 3. Accuracy of the CelebA data set

Number of Hidden Units	Training Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Training Time (sec)	Objective Value
256	84.5355	82.4765	84.3301	11037	0.842867

Multi-layers Neural Network created by using tensorflow need less time to train the weights parameters, however with more hidden layers the accuracy of the Neural Network decreases. Compared the accuracy of the deep Neural Network with our single layer Neural Network, our Neural Network even more accurate than the multilayer one, but the training time is much longer than the deep neural network created by using tensorflow. The results are summarized in Table 4, Figure 6 and Figure 7.

Table 4. Accuracy of the Deep Neural Network

Number of Hidden Layers	Number of Hidden Units	Accuracy	Training Time (sec)
1	256 each layer	0.827403	218
2	256 each layer	0.808857	225
3	256 each layer	0.786904	263
4	256 each layer	0.760409	292
5	256 each layer	0.774035	326
6	256 each layer	0.732778	334
7	256 each layer	0.733535	350

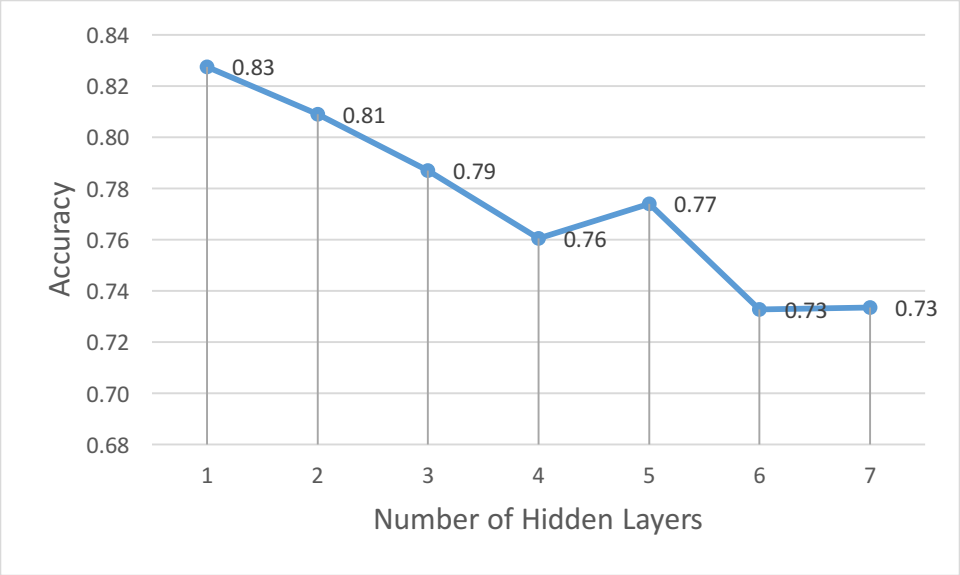

Figure 6. Trend of Accuracy Respect to the Number of Hidden Layers

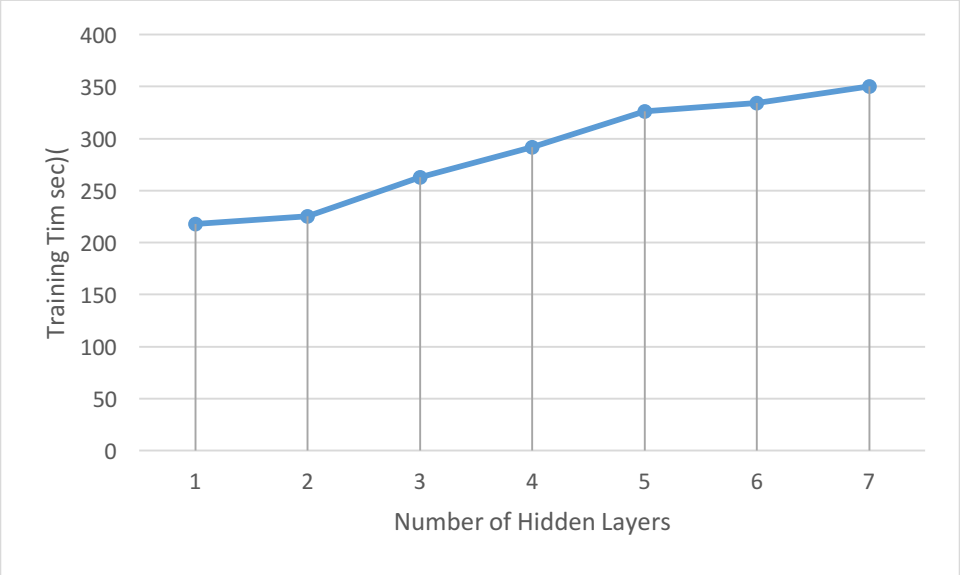

Figure 7. Trend of Training Time Respect to the Number of Hidden Layers