CSE 574 Introduction to Machine Learning
Programming Assignment 1

Team 66:
Weiyi Jiang 50207995
Shih-Chia Chen 50207079

I. Explanation with supporting figures of how to choose the hyper-parameter for Neural Network: nnScript.py

- Approach to find the best performance

We have two variables, number of the hidden layer units (n_hidden) and regularization hyper-parameter (λ). First, we change the λ value from 0 to 60 in increments of 5 and number of hidden units(n_hiddent = 30) is not changed at this round. Then we change the number of hidden units from 4 to 20 in increments of 4 in order to observe its influence on training time.

- Relationship between λ and performance of Neural Network

From Figure 1, we can tell that the accuracy of three sets have slightly difference when we input different value of λ. However, the change of the accuracy is irregular, so we can not predict if we input a higher λ, will we also have a higher accuracy. Then, in our experiment, when λ is 5 we have the highest test accuracy. Although when λ = 0, we can obtain even higher accuracy, we cannot let λ = 0, since we need to consider the overfitting problem. Therefore, we assign λ = 5 in the next step.

λ	0	5	10	15	20	25	30	35	40	45	50	55	60
Training accuracy	94.56	94.328	93.758	94.128	94.138	93.982	94.034	93.864	93.722	93.872	93.608	93.63	93.294
Validation Accuracy	93.74	93.76	92.85	93.47	93.46	93.31	93.13	93.22	93.16	92.9	93.19	92.57	92.55
Test accuracy	94.04	93.97	93.45	93.75	93.88	93.64	93.79	93.7	93.55	93.39	93.51	93.35	93.38
Training time	11:58	10:42	17:36	15:26	14:57	16:04	15:12	16:55	17:07	11:04	11:27	09:50	11:32

Table 1. Relationship between λ and accuracy (number of hidden units is 30)

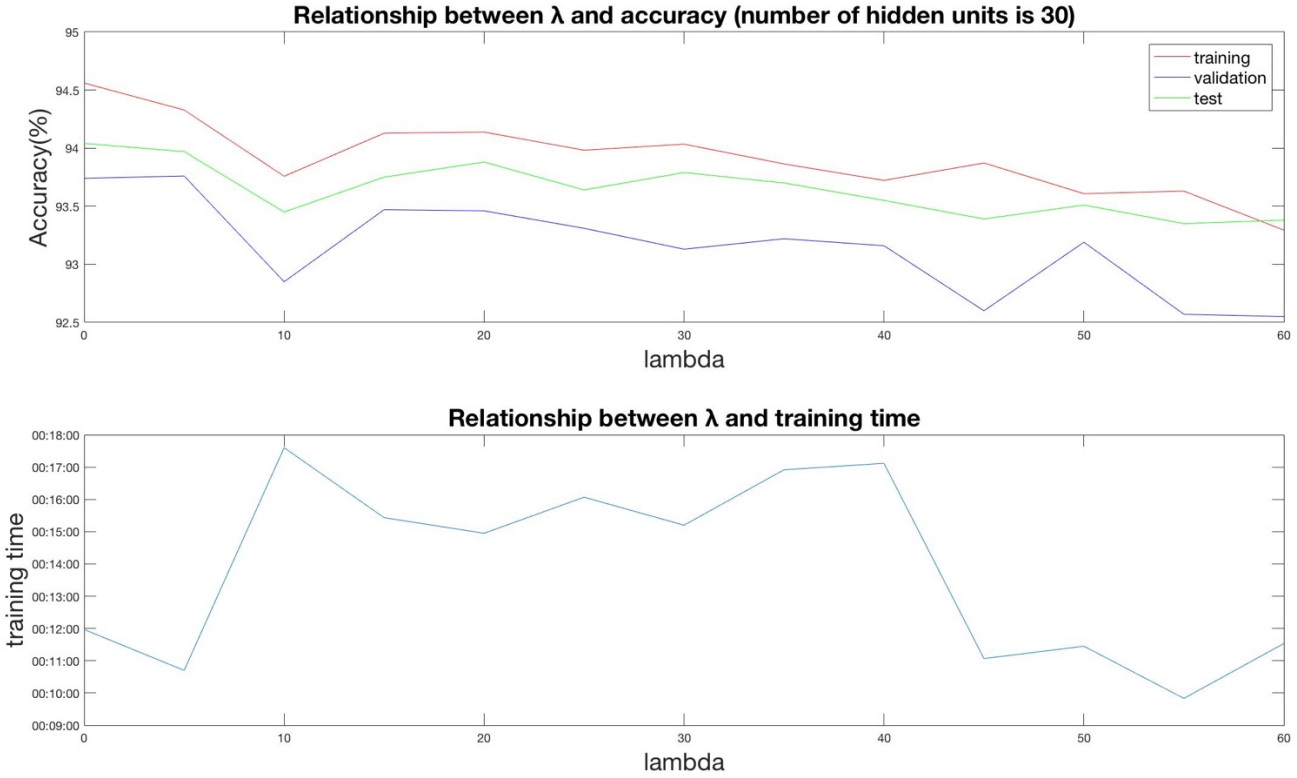

Figure 1.

- Relationship between number of hidden units and training time

In order to find best performance from the λ and number of hidden units, we fix the λ value to 5 and change the value of the number of hidden units. Based on Table 2 and Figure 2, we can find out that when we add more hidden units, the accuracy is growing fast, but at the same time, training takes more time. Although it does take more time to complete the training, the increase of the time is not too much. Therefore, we still choose 20 as the number of hidden units.

Number of Hidden units	4	8	12	16	20
Training accuracy	42.682	81.388	92.006	90.568	93.098
Validation accuracy	40.92	80.47	91.38	90.34	92.57
Test accuracy	42.15	81.47	91.77	90.33	92.9
Training time	06:06	06:35	07:36	08:32	09:27

Table 2. Relationship between number of hidden units and training time (λ = 5)

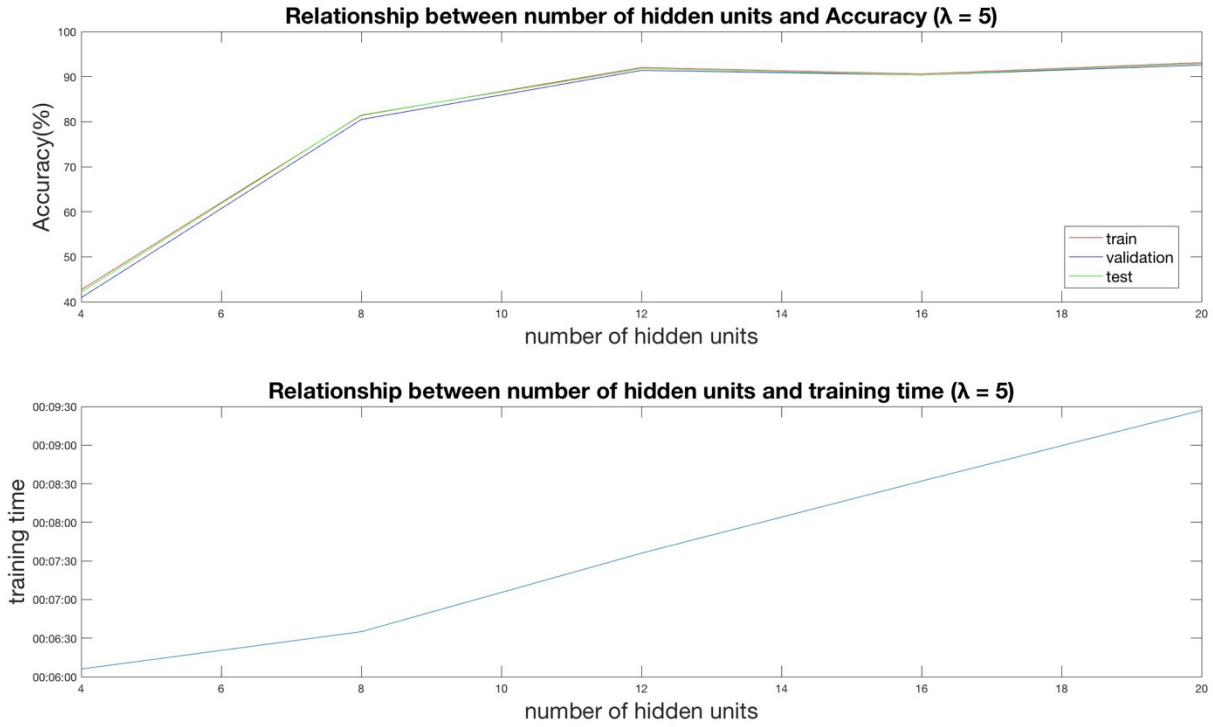

Figure 2.

- Chosen the Optimal variables

According to the statistic above, the optimal regularization hyper-parameter (λ) and number of the hidden layer units should be 5 and 20 respectively. These values can lead to the best performance of the neural network which also takes the time issue in consideration.

II. Accuracy of classification method on the handwritten digits test data: nnScript.py
Training set Accuracy:93.58%
Validation set Accuracy:92.97%
Test set Accuracy:92.98%

III. Accuracy of classification method on the CelebA data set: facennScript.py
Training set Accuracy: 85.649%
Validation set Accuracy: 85.103%
Test set Accuracy: 86.714%

IV. Compare the results of deep neural network and neural network with one hidden layre on the CelebA stat set: deepnnScripy.py

The single layer performance is around 94%.
Following is the Accuracy we get from the deep learn with several layers:
2 layers Accuracy: 0.806207
3 layers Accuracy: 0.779334
5 layers Accuracy: 0.766843
7 layers Accuracy: 0.718017

From the data shown above, we know that more hidden layers we have does not indicate higher performance of the neural network. The accuracy keeps going down when we add more hidden layers.