**Part A: Classification Problem**

1. The 3-layer feedforward neural network (FFN) is constructed with a learning rate of 𝛼 = 0.01, batch size of 32, hidden layer neuron number of 10, and weight decay parameter of 𝛽 = 10−6:

a) The accuracies on both training and testing data against epochs are shown in the Figure 1 below:

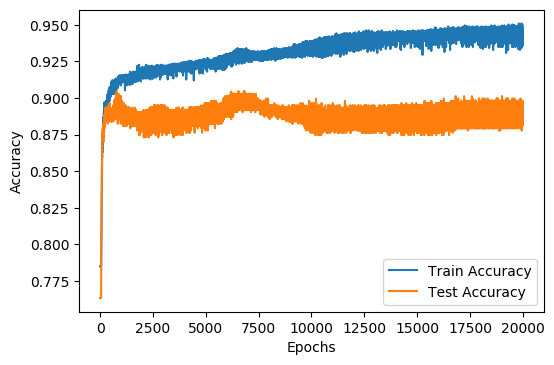


Figure 1

b) The approximate number of epochs where the test error converges is 7000 epochs. After 7000 epochs, the test accuracy starts dropping and finally reaches a steady state.

2. This part is to find the optimal batch size from 4, 8, 16, 32, 64. The batch sizes are evaluated based on their cross-validation accuracies and the time taken for training.

a) The cross-validation accuracies against epochs for five different batch sizes are shown in Figure 2 below:

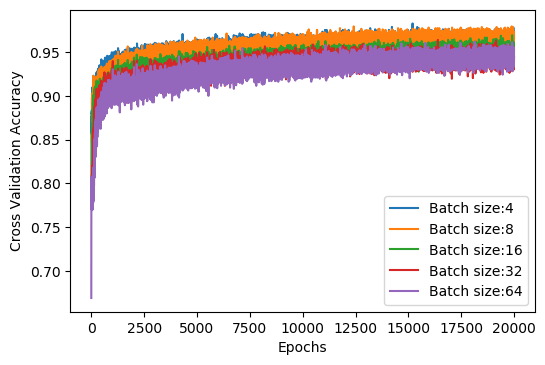


Figure 2

Besides, the time taken to train the FFN for one epoch against batch sizes are shown in Figure 3 below:

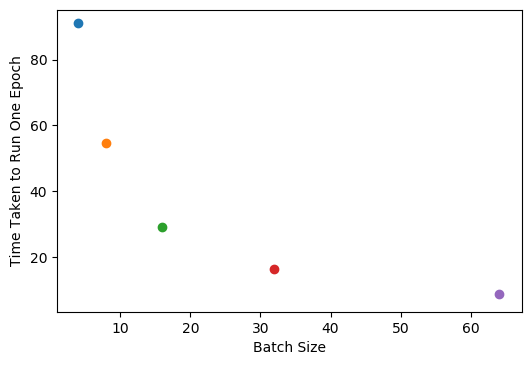


Figure 3

b) Based on results obtained in Figure 2 and Figure 3, the batch size of 8 is selected to be the optimal batch size. The reasons it is chosen include:

* A smaller batch size leads to a faster increase in the cross-validation accuracy at the beginning of the training.
* In the end, the cross-validation accuracy for batch size of 8 is similar to that for batch size of 4, and higher than the accuracies for other batch sizes.
* When batch size is 8, the time taken to train the FFN for one epoch is significantly lower than the time taken for batch size of 4.

c) When the batch size is changed to 8, the accuracies on both training and testing data against epochs are shown in the Figure 4 below:

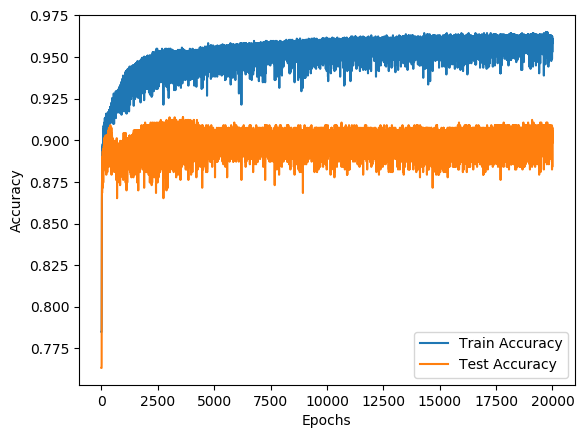


Figure 4

3. This part is to find the optimal number of neurons in the hidden-layer, the number is chosen from 5, 10, 15, 20, 25. The neuron numbers are evaluated based on their cross-validation accuracies.

a) The cross-validation accuracies against epochs for five different hidden-layer neuron numbers are shown in Figure 5 below:

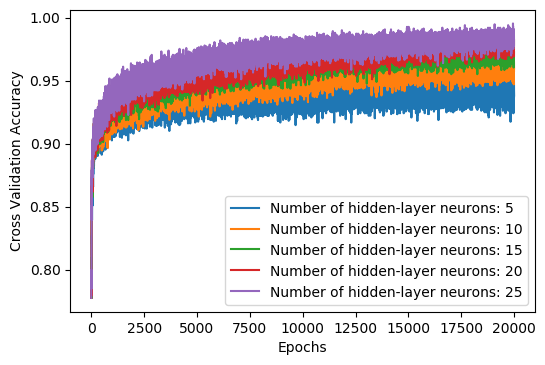


Figure 5

b) Based on Figure 5, it can be observed that the cross-validation accuracy for 25 hidden-layer neurons is above the cross-validation accuracies for other numbers of hidden neurons. Therefore, 25 is selected to be the optimal number of hidden neurons.

c) When the number of hidden neurons is changed to 25, the accuracies on both training and testing data against epochs are shown in the Figure 6 below:

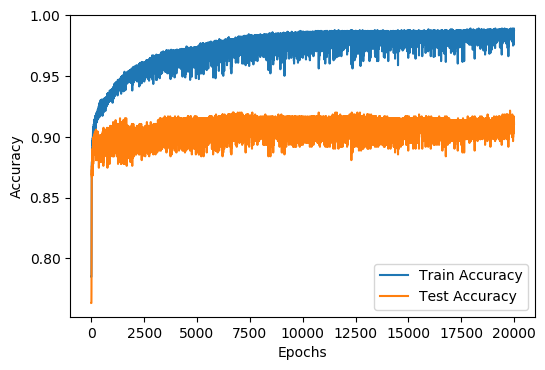


Figure 6

4. This part is to find the optimal weight decay parameter 𝛽 from 0, 10−3, 10−6, 10−9, 10−12. The weight decay parameters are evaluated based on their cross-validation accuracies.

a) The cross-validation accuracies against epochs for five different weight decay parameters are shown in Figure 7 below:

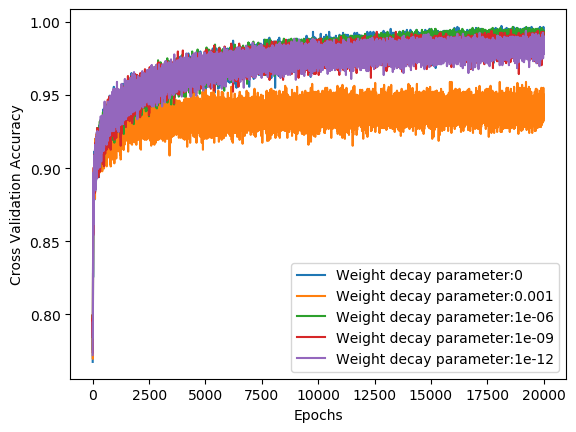


Figure 7

b) Based on Figure 7, it can be concluded that the optimal weight decay parameter is 0. The reason is that there is no significant difference between the cross-validation accuracies for weight decay parameters of 0, 1e-6, 1e-9 and 1e-12. Thus, there is no need for the additional computation using the weight decay parameter. Training the model with a weight decay parameter of 0 can reduce the usage of computational resources.

c) When the weight decay parameter is changed to 0, the accuracies on both training and testing data against epochs are shown in the Figure 8 below:

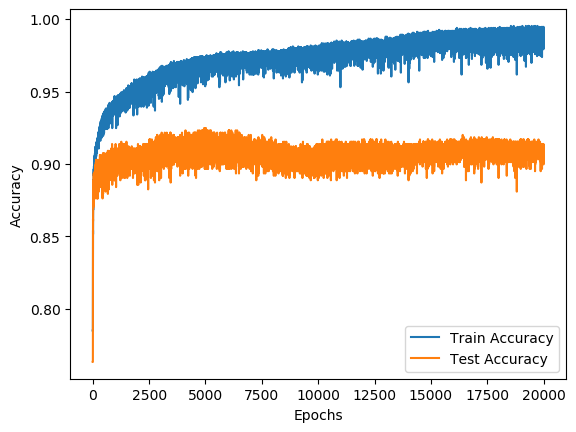


Figure 8

5. The 4-layer FFN is constructed with a learning rate of 𝛼 = 0.01, batch size of 32, hidden layer neuron number of 10 for the two hidden layers, and weight decay parameter of 𝛽 = 10−6:

a) The accuracies on both training and testing data against epochs for the 4-layer FFN are shown in the Figure 9 below:

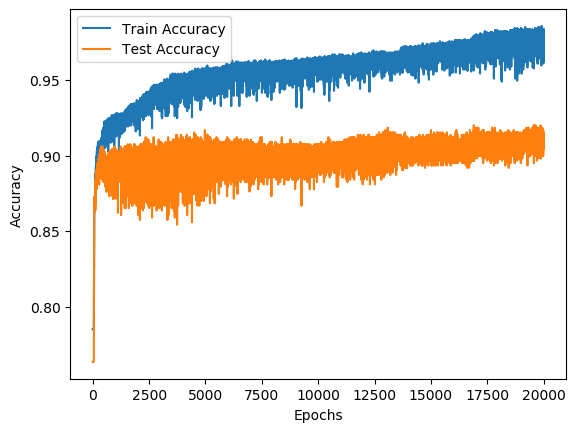


Figure 9

b) The optimal 3-layer FFN is constructed with a learning rate of 𝛼 = 0.01, batch size of 64, hidden layer neuron number of 25, and weight decay parameter of 𝛽 = 0, its accuracies on both training and testing data against epochs are shown in Figure 8 in part 4.

Compare Figure 9 with Figure 8, it can be observed that:

* In terms of the prediction accuracy on unseen data, 3-layer and 4-layer FFN have similar performance. For both of them, the test accuracies on unseen data converge to around 90%.
* During the training process, the train and test accuracies for the 4-layer FFN has a higher tendency to fluctuate up and down. At the beginning of the training, the test accuracy for the 4-layer FFN even fluctuates up and down in a range of around 5%, while the 3-layer FFN generally fluctuates within a range of around 2%.