Part A: Classification Problem

1. The 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:

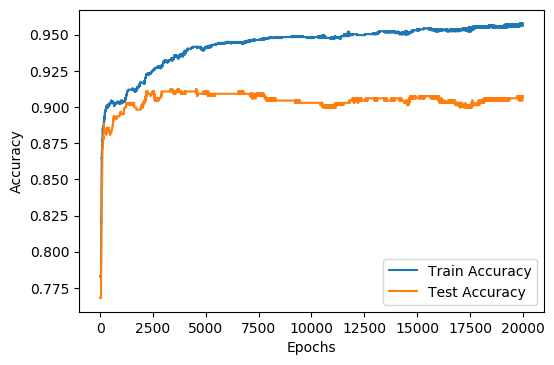


Figure 1

b) The approximate number of epochs where the test error converges is around 4000 epochs. After 4000 epochs, the test accuracy stops increasing and reaches a steady state.

2. This part is to find the optimal batch size from 4, 8, 16, 32, 64. The batch size 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:

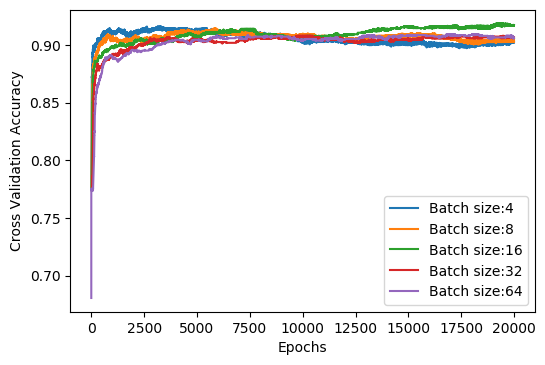


Figure 2

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

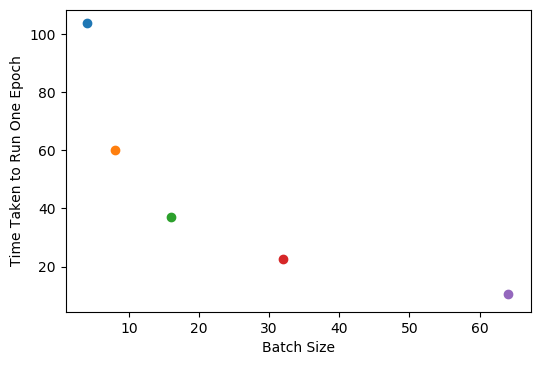


Figure 3

b) Based on results obtained in Figure 2 and Figure 3, the batch size of 64 is selected to be the optimal batch size. The reasons it is chosen include:

* Although a smaller batch size leads to a faster increase in the cross-validation accuracy for the first 2500 epochs, the increase becomes slower and slower after that. At around 8000 epochs, the system reaches a state such that all batch sizes have a similar cross-validation accuracy. In the end, the cross-validation accuracies for all the batch sizes do not differ significantly.
* When batch size is 64, the time taken to train the FFN for one epoch is significantly lower than other batch sizes.

c) When the batch size is changed to 64, the accuracies on both training and testing data against epochs are shown in the Figure 4 below:

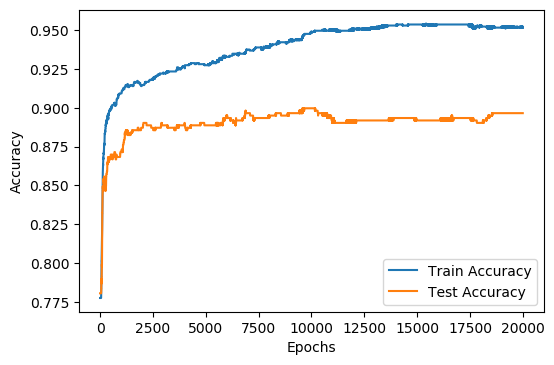


Figure 4

3.