CMPT310 Bonus Assignment

Sample output

```
Size 1: 0.6199999999999999
Size 2: 0.48
Size 3: 0.3600000000000004
Size 4: 0.2900000000000004
Size 5: 0.33
Size 6: 0.32
Size 7: 0.32
Size 8: 0.32
Size 9: 0.30999999999999994
Size 10: 0.33
Size 11: 0.31
Size 12: 0.3
Size 13: 0.29
Size 14: 0.29
Size 15: 0.31
Size 16: 0.3000000000000004
Size 17: 0.27999999999999997
Size 18: 0.31
Size 19: 0.32
Size 20: 0.3
```

Fig.1: Shows the hidden units and the resulting error

```
N          = 100   # number of examples to be used in experiments
k          =   5   # k parameter
epochs     = 100   # maximal number of epochs to be used in each training round
size_limit =  20   # maximal number of hidden units to be considered
```

Fig. 2: Parameters used in Fig 1

```
Size 1: 0.4800000000000001
Size 2: 0.13333333333333333
Size 3: 0.14666666666666667
Size 4: 0.15333333333333332
Size 5: 0.11333333333333331
Size 6: 0.10666666666666665
Size 7: 0.12666666666666665
Size 8: 0.11999999999999997
Size 9: 0.09999999999999998
Size 10: 0.09999999999999998
```

Fig 3: Shows the hidden units and the resulting error

```
N          = 150   # number of examples to be used in experiments
k          =  10   # k parameter
epochs     = 150   # maximal number of epochs to be used in each training round
size_limit =  20   # maximal number of hidden units to be considered
```

Fig. 4: Parameters used in Fig 3

We have the Size as the number of hidden units in the 1-layer feed forward network. We have overall error for each size average over all k-rounds in the output. We can see the general trend that as the hidden units increases, the error gets lower. However, during some experiments, the error fluctuates around the same values and computing with more hidden units starts to become computationally inefficient for the accuracy gained. With N=100, we can see that the optimal value is 17 but the computation took a long time. For a quicker computational time, 4 hidden units had low error values and were successful only in some cases. In most of my experiments, the errors stabilized by 7 hidden units and then the error value would start to fluctuate.

The optimal number of hidden units is 7 according to my results.