

Xie Zhou 912143385
Haozhe Gu 999200555

Assignment3

Problem 1)

Our Neural Network is fully connected with 1 hidden layer. The first layer of our Neural Network has 960×16 number of nodes, each representing an intensity of grayness of pixels in the image. The number of nodes for the hidden layer was 30. We initially chose 100 nodes and figured that it would cause overfitting. The generalization error for 100 hidden nodes was much lower than that for 30 hidden nodes. There are 2 output nodes: Female and Male. The activation functions for the output nodes generates the input data's probabilities of belonging to each of the nodes.

Problem 2)

Generalization Error:

```
----- Round: 0-----  
Training mean: 0.7204567   Training std. dev.: 0.21140319  
Testing mean: 0.7074074   Testing std. dev.: 0.26330182  
----- Round: 1-----  
Training mean: 0.8107306   Training std. dev.: 0.113719724  
Testing mean: 0.7711111   Testing std. dev.: 0.1502627  
----- Round: 2-----  
Training mean: 0.8125571   Training std. dev.: 0.09945537  
Testing mean: 0.78962964   Testing std. dev.: 0.15252785  
----- Round: 3-----  
Training mean: 0.8008219   Training std. dev.: 0.10400539  
Testing mean: 0.7474074   Testing std. dev.: 0.13167694  
----- Round: 4-----  
Training mean: 0.8308219   Training std. dev.: 0.03028881  
Testing mean: 0.77814814   Testing std. dev.: 0.095581405  
----- Round: 5-----  
Training mean: 0.8136986   Training std. dev.: 0.014554678  
Testing mean: 0.7888889   Testing std. dev.: 0.025119726  
----- Round: 6-----  
Training mean: 0.8456621   Training std. dev.: 0.028501349  
Testing mean: 0.7740741   Testing std. dev.: 0.035908747  
----- Round: 7-----  
Training mean: 0.7479452   Training std. dev.: 0.13326451  
Testing mean: 0.6851852   Testing std. dev.: 0.15843889  
----- Round: 8-----  
Training mean: 0.8191781   Training std. dev.: 0.031050231  
Testing mean: 0.8011111   Testing std. dev.: 0.051586635  
----- Round: 9-----
```

Training mean: 0.8091324 Training std. dev.: 0.021879725

Testing mean: 0.8111111 Testing std. dev.: 0.048855215

Average error = 0.78110046

Problem3)

See ashbo.prediction.txt

Problem4)

The plots below briefly show how 10 weights (from the input layer to 4 of the hidden layer nodes) change and reaches the best state during the course of training.

