HW2:
Name: Steven Gong
ID: 999780855

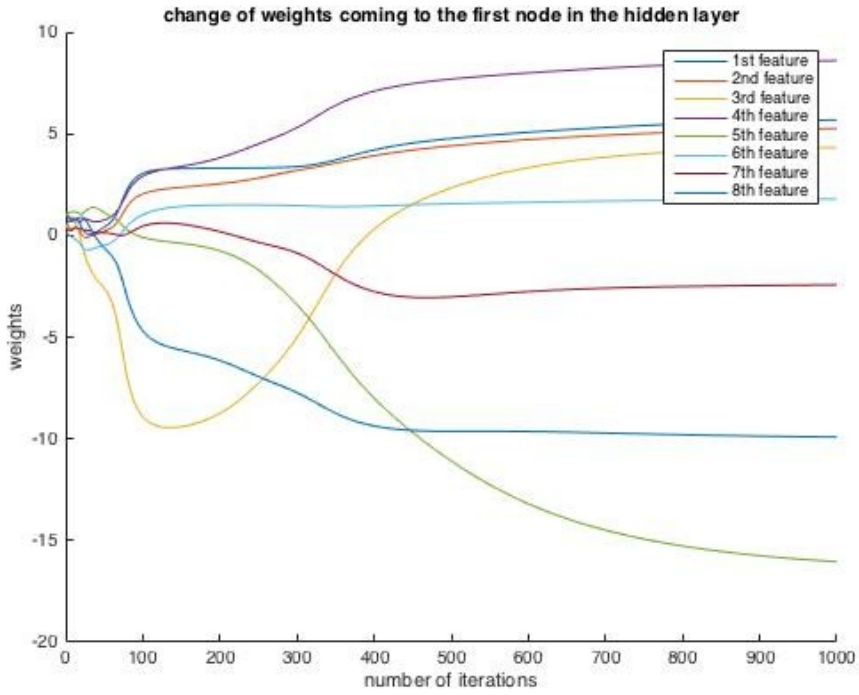
Instructions: There are 7 files:

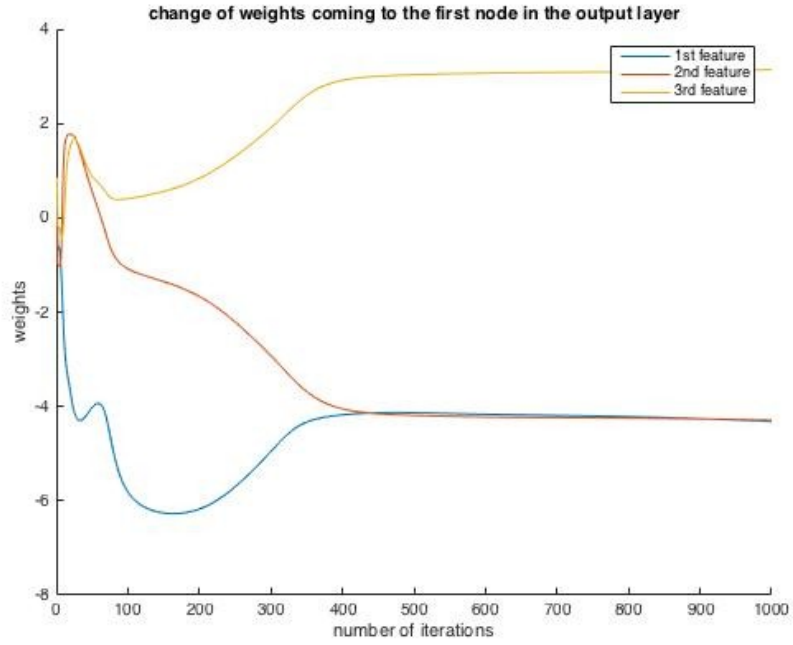
1. “problem1.m” is the code for problem #1,
2. “problem2.m” is the code for problem #2
3. “hidden1.m”, “hidden2.m” and “hidden3.m” are the codes for problem#4.
   “hidden1.m” applies to the situation has 1 hidden layer,
   “hidden2.m” applies to the situation has 2 hidden layers,
   “hidden3.m” applies to the situation has 3 hidden layers,
4. “sigmoid.m” is the function file for sigmoid function
5. “data.txt” is the data file. I delete the 1st column, and I label the last column as the following:

```
%CYT (cytosolic or cytoskeletal)                    1
%NUC (nuclear)                                      2
%MIT (mitochondrial)                                3
%ME3 (membrane protein, no N-terminal signal)       4
%ME2 (membrane protein, uncleaved signal)           5
%ME1 (membrane protein, cleaved signal)             6
%EXC (extracellular)                                7
%VAC (vacuolar)                                     8
%POX (peroxisomal)                                  9
%ERL (endoplasmic reticulum lumen)                  10
```

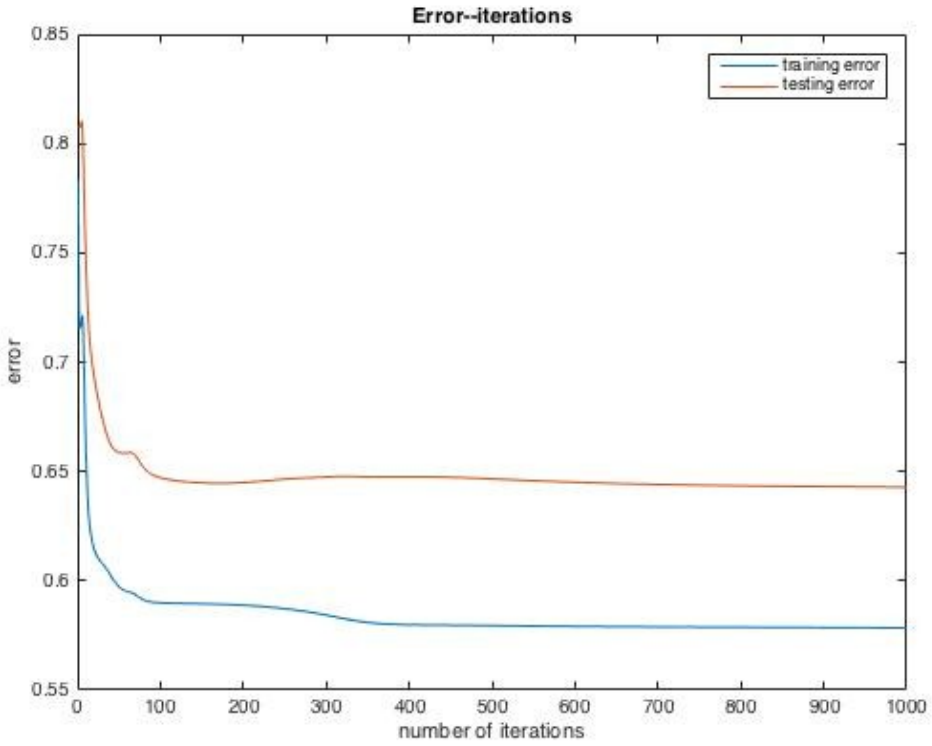
In order to run the code for problem#4, please change the number of nodes on the 1st line:
I make it a variable that you can easily change it from 3 to 6, 9 ,12 etc.

Q1:
Change of weights:

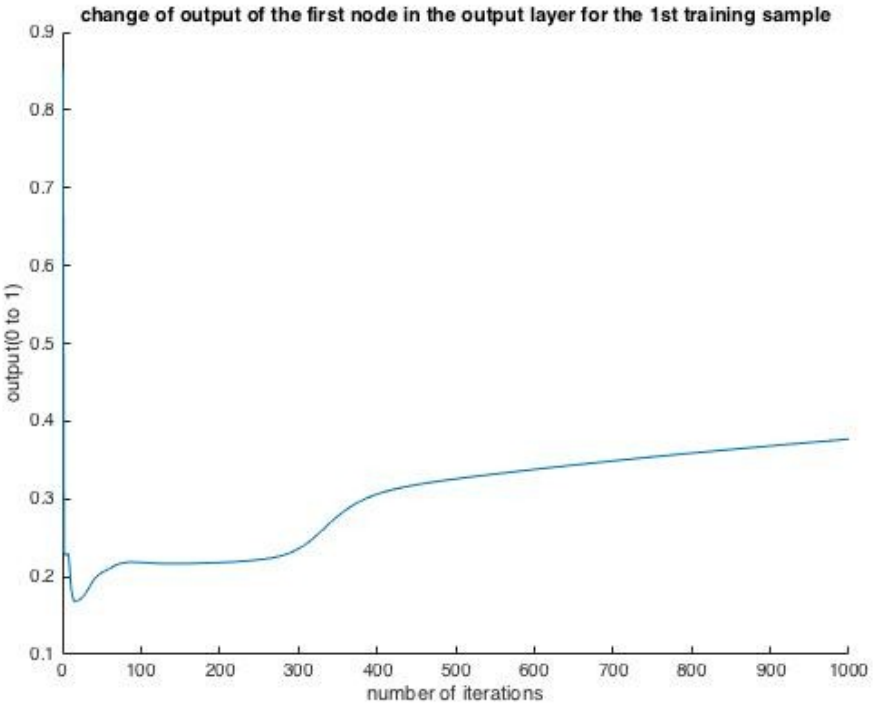




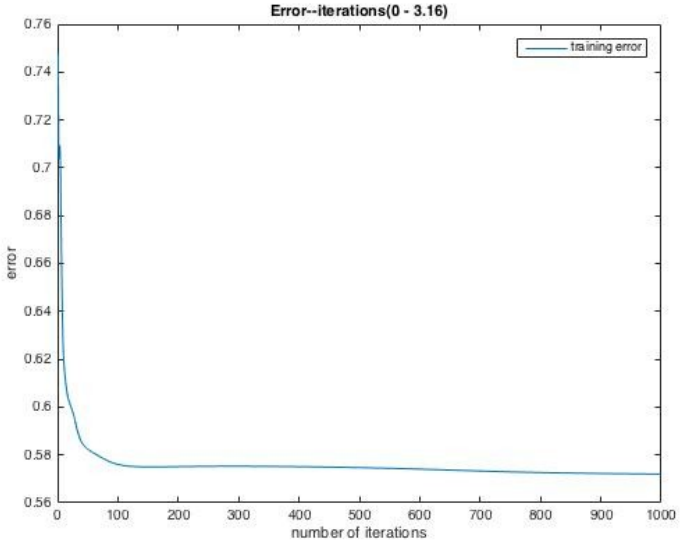
Change of mean square error:



Change of output:



Q2:
The final training mean square error is 0.5720, and the correct classification rate is 56.8%



My final activation function should be
$1 / (1 + \exp(-(w2*a^2 + bias2)))$

a^2 is calculated from x
$a2 = 1/(1 + \exp(-(w1 * x + bias1)))$ x is all the input values.

in which

w1 =

Columns 1 through 7

5.6856	5.2608	4.3197	8.5995	-16.0407	1.7870	-2.4243
-6.0991	-5.0208	-1.3049	-7.5186	7.4377	-2.9353	3.0368
-7.2187	-1.5289	35.1533	2.6656	-27.5665	-5.9460	-0.5002

Column 8

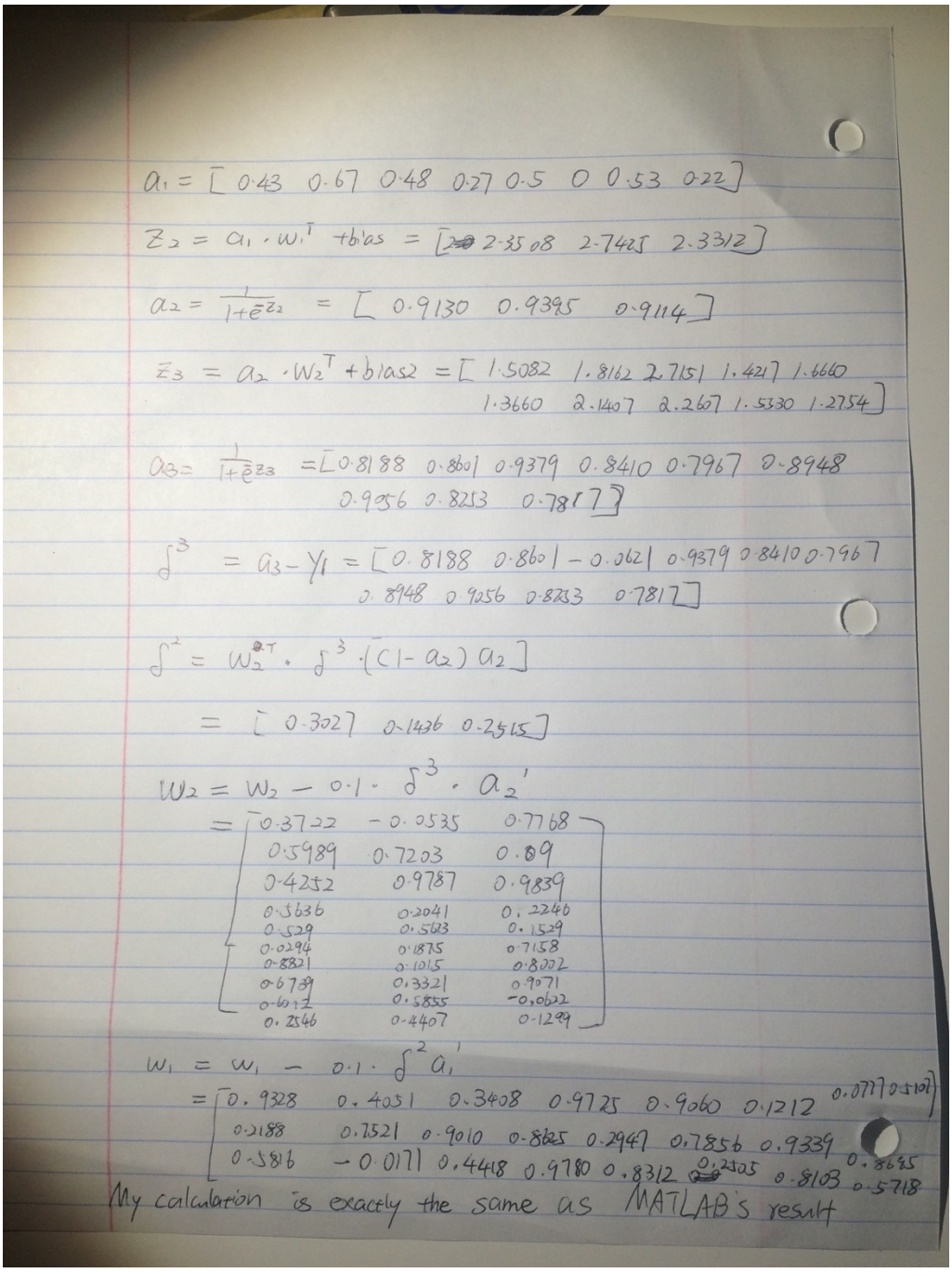
-9.9118
9.4053
3.3310

w2 =

-4.3185	-4.2927	3.1407
-6.5979	-2.5930	2.1676
-1.9949	-6.5002	1.2779
-5.9272	1.9818	-5.9722
-1.5669	-4.0526	-3.2896
0.2491	-8.1430	-10.8700
-2.3201	-10.1403	-0.6814
-5.1302	-4.6447	-0.0411
-3.6064	-6.2757	-1.4473
-11.8323	-6.0064	-11.2650

each row of w1 and w2 corresponds to one node in the next layer
bias is 0.6456 all the neurons in the the hidden layer
bias2 is 0.5107 for all the neurons in the output layer.

Q3



The results get from code:

```
w1 =
  Columns 1 through 7
    0.9476    0.4283    0.3574    0.9818    0.9233    0.1212    0.0960
    0.2265    0.7641    0.9097    0.8674    0.3037    0.7856    0.9435
    0.5944    0.0029    0.4561    0.9860    0.8461    0.2505    0.8261
  Column 8
    0.5183
```

```
    0.8735
    0.5784
w2 =
    0.4501    0.0265    0.8545
    0.6802    0.8037    0.1710
    0.4194    0.9728    0.9782
    0.6404    0.2829    0.3013
    0.6087    0.6491    0.2324
    0.1054    0.2656    0.7917
    0.9660    0.1877    0.8840
    0.7587    0.4190    0.9873
    0.6796    0.6660    0.0161
    0.3293    0.5174    0.2045
after the first iteration, w1 and w2 becomes: w1 =
  Columns 1 through 7
    0.9328    0.4051    0.3408    0.9725    0.9060    0.1212    0.0777
    0.2188    0.7521    0.9010    0.8625    0.2947    0.7856    0.9339
    0.5816   -0.0171    0.4418    0.9780    0.8312    0.2505    0.8103
  Column 8
    0.5107
    0.8695
    0.5718
w2 =
    0.3722   -0.0535    0.7768
    0.5989    0.7203    0.0900
    0.4252    0.9787    0.9839
    0.5636    0.2041    0.2246
    0.5290    0.5673    0.1529
    0.0294    0.1875    0.7158
    0.8821    0.1015    0.8002
    0.6739    0.3321    0.9027
    0.6012    0.5855   -0.0622
    0.2546    0.4407    0.1299
```

Those two are in agreement.

Q4:

Testing set mean square error	3 nodes	6 nodes	9 nodes	12 nodes
1 hidden layer	0.6428 (46.8% correct prediction)	0.5949 (57.6% correct prediction)	0.5948 (58.4% correct prediction)	0.6222 (55.7% correct prediction)
2 hidden	0.6217	0.6230	0.6377	0.6898

layer	(46.8% correct prediction)	(54.1% correct prediction)	(53.0% correct prediction)	(50.7% correct prediction)
3 hidden layer	0.6373 (47.6% correct prediction)	0.6438 (52.0% correct prediction)	0.6868 (48.4% correct prediction)	0.6592 (55.5% correct prediction)

The optimal configuration is 9 nodes 1 hidden layer
The more layer I have, and the more nodes I have, the training set and the training Set will be more accurate; however, if I have too many nodes, the testing set does not necessary to be more accurate. This is because of over fitting.

Q5

Plug in the Ws obtained from the optimal configuration obtain from Q4, which is 1 hidden Layer 9 nodes.

The following code is from the last part of “hidden1.m”. It is used to demonstrate my process.

```
test = [0.50 0.49 0.52 0.20 0.55 0.03 0.50 0.39]';

z2 = w1 * test + bias;
a2 = sigmoid(z2);

z3 = w2 * a2 + bias2;
a3 = sigmoid(z3);

[M, indice] = max(a3);
fprintf('The prediction is the %dth class', indice);
```

After running the code, we find that the data belongs to the 2nd class, which is NUC.
Q6:

If we classify the sample into the wrong class, there are still 9 possibilities that it can belong to. So in this case, we need to try the 2nd highest probability.

```
a3 =
    0.3100
    0.4102
    0.0115
    0.0996
    0.0013
    0.0000
    0.0017
```

0.0132
0.0001
0.0000

We can use cross validation.

P(class2| D) = p(D | class2) * P(class2)/P(D)
= 0.41 * (429/1484)/ (P(TP) + P(FN))