RMIT University
School of Science
COSC2110/COSC2111 Data Mining
Tutorial Problems Week 8

1. Suppose that a 4-2-3 network is being trained for the iris data. The following encoding of the class variable has been used:

Class	O_1	O_2	O_3
setosa	1	0	0
versicolor	0	1	0
virginica	0	0	1

   Below is a segment of the result file. A line commencing with # identifies the pattern, the next three lines give the inputs, desired outputs and actual outputs.

   ```
   #1.1
   4.4 3 1.3 0.2
   1 0 0
   0.90041 0.27903 0.00001
   #2.1
   6 3.4 4.5 1.6
   0 1 0
   0.82013 0.28581 0.00019
   #3.1
   6 2.9 4.5 1.5
   0 1 0
   0.02983 0.29019 0.70105
   ```

   For these 3 instances, give

   (a) The TSS (Total Sum Squared Error)

   (b) The MSE (Mean Squared Error)

   (c) The classification error

   (d) The confusion matrix

2. For the following result file segment:

   ```
   #1.1
   4.4 3 1.3 0.2
   1 0 0
   0.90041 0.27903 0.00001
   #2.1
   6 3.4 4.5 1.6
   0 1 0
   0.22013 0.28581 0.00019
   #3.1
   6 2.9 4.5 1.5
   ```

```
0 1 0
0.02983 0.29019 0.00105
#4.1
6.3 2.9 5.6 1.8
0 0 1
0 0.30939 0.62236
#5.1
6.7 3 5.2 2.3
0 0 1
0 0.31168 0.79593
#6.1
5.7 4.4 1.5 0.4
1 0 0
0.7005 0.27902 0.00001
#7.1
4.7 3.2 1.6 0.2
1 0 0
0.0042 0.27903 0.00001
#8.1
4.7 3.2 1.6 0.2
1 0 0
0.6042 0.27903 0.00001
#9.1
4.7 3.2 1.6 0.2
1 0 0
0.5142 0.27903 0.00001
#10.1
4.7 3.2 1.6 0.2
1 0 0
0.4042 0.27903 0.00001
```

For each pattern give the classification that would result from

(a) A 402040 strategy

(b) A 206020 strategy

(c) A 500050 strategy

(d) A winner-takes-all strategy. (The node with the highest output is chosen, no matter what its value.)

3. Which of these strategies do you think is the best?

4. Describe how you would develop a neural network classifier for the following data:

```
@relation mushroom-small
@attribute cap-shape {b,c,f,k,s,x}
@attribute cap-surface {f,g,s,y}
@attribute cap-color {b,c,e,g,n,p,r,u,w,y}
@attribute bruises? {f,t}
@attribute odor {a,c,f,l,m,n,p,s,y}
@attribute gill-attachment {a,d,f,n}
@attribute 'class' { 'e', 'p'}
@data
x,s,n,t,p,f,e
x,s,y,t,a,f,p
```

(a) Show how you would encode the inputs.

(b) How many inputs will there be altogether?

(c) How would you encode the output if one output node is used?

(d) How would you encode the output if two output nodes are used?

(e) How would you determine if a test case is classified correctly if one output node is used? Two output nodes?

(f) How would you find the best number of hidden units?

Logistic

(1/(1+exp(-x)))