CMP 242 Homework 2

20 pts Due Thursday October 30, work in groups of 1 to 3

1. (4 pts) Find a small dataset where the information gain split criteria fails to produce a minimum-node decision tree correctly classifying the data. List your data, the tree created by recursively selecting the test maximizing the information gain, the amount of the information gain due to each node, a smaller (fewer nodes) decision tree, and the information gain of each node in the smaller tree. Include An informal description of your dataset and why information gain fails to produce a minimum size tree.

2. (7 pts) Implement the backpropagation algorithm to run on a small 3-node, 3-input neural network (Each input plus a bias is connected to each of the two hidden nodes, both hidden nodes plus a bias is connected to the output node).

   Initialize the weights to small random values as suggested in the text and use your implementation to learn from the following data (majority):

x_1	x_2	x_3	y
0	0	0	0
0	1	0	0
0	0	1	0
0	1	1	1
1	0	0	0
1	1	0	1
1	0	1	1
1	1	1	1

   Loop through the data until the absolute error, $|y - \hat{y}|$ is small (less than 0.01 on all four data points). Repeat a total of 5 times with different initialization and report the average number of epochs (loops over the data) needed.

   Repeat the above for the following data ($x_1 \oplus x_2$):

x_1	x_2	x_3	y
x_1	x_2	x_3	y
0	0	0	0
0	1	0	1
0	0	1	0
0	1	1	1
1	0	0	1
1	1	0	0
1	0	1	1
1	1	1	0

   Finally, try to train on the following data ($x_1 \oplus x_2 \oplus x_3$):

x_1	x_2	x_3	y
x_1	x_2	x_3	y
0	0	0	0
0	1	0	1
0	0	1	1
0	1	1	0
1	0	0	1
1	1	0	0
1	0	1	0
1	1	1	1

What happens and why?

You may use an existing back-propagation package if you want.

3. (6 pts) Consider the following majority dataset:

x_1	x_2	x_3	y
-1	1	-1	-1
-1	-1	1	-1
-1	1	1	1
1	-1	-1	-1
1	1	-1	1
1	-1	1	1

Create a second dataset of 30 examples by adding 50 random features (iid with Pr(1)=Pr(-1)=1/2) to five copies of each of the above examples.

Implement Adaboost and the perceptron algorithm and run them on both the original and the expanded datasets. How do they compare? For Adaboost, use each feature as your weak learners, i.e. loop through the features and find the one that predicts the best with respect to the current distribution.

4. (3 pts) Recall the street hustler coin flip example from lecture. Use the following prior distribution on the probability q of heads: $P(q = 1/4) = 1/3$, $P(q = 1/2) = 1/3$, $P(q = 3/4) = 1/3$. Assume you see the data sequence: heads, tails, heads.

   Compute the posterior probabilites for each of the possible values of q.

   What is the MAP estimate for q?

   What is the mean a posteriori estimate for q?

For the experimental problems, be sure to describe your methodology. Feel free to play around with experimental problems and report anything interesting that happens. For example, one could add an incorrectly labeled example to problem 3.