EECS 391: Introduction to AI (Spring 2012) Written Homework 6 (Max Points: 100)

Assigned Tuesday March 27, due 5pm Thursday April 5. Write your answers neatly and remember to show all relevant work. Before turning in your work, staple your answer sheets together and write your name(s) and Case ID(s) on the front page.

1. Write down specifications for learning systems that learn to (a) understand a language and (b) play tennis. Do these tasks fit well into the supervised learning framework we have been studying? Explain your answer. (10 points)

2. This question is best answered with the help of a mathematics package such as *R*, *Matlab* or *Mathematica*. The data used for Figure 20.1 in the book can be viewed as being generated by hypothesis h_5. For hypotheses h_3 and h_4, generate one dataset each of size 100 and plot graphs for: (i) $P(h_i|d_1,\ldots d_N)$, $i=1,..,5$,(ii) $P(D_{N+1}=lime|d_1,\ldots d_N)$, (iii) $P(D_{N+1}=lime|h_{MAP})$ and (iv) $P(D_{N+1}=lime|h_{ML})$ as N varies from 1 to 100. For (iii) and (iv), pick h_{MAP} / h_{ML} just from h_3 and h_4 and assume the prior probability of h_3 to be 0.9. Put the lines for (i) in one graph and (ii), (iii) and (iv) in a second graph, so you generate four graphs in total. (30 points)

3. Suppose that Alice's utilities for cherry and lime candies are c_A and l_A, while Bob's are c_B and l_B. Alice has a bag of x candies generated by one of the h_3, h_4 or h_5 hypotheses above. At each point, she can either unwrap a candy and eat it, or she can sell the remaining bag of candies to Bob. Bob will pay the "true" price (i.e. the value according to him once he unwraps all the remaining candies). Determine a good selling criterion for Alice. (10 points)

4. Construct two naïve Bayes models for the following set of examples over three binary features, one for each class label. Then use the models to classify each example. What does this tell you about naïve Bayes classifiers? Of all the functions which can be constructed over three binary features, how many are perfectly separable by naïve Bayes? (It may help if you plot the examples in 3D and label them positive/negative.) (20 points)

F1	F2	F3	$Class_1$	$Class_2$
0	0	0	0	0
0	0	1	1	1
0	1	0	1	1
0	1	1	1	0
1	0	0	1	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

5. A *Gaussian naïve Bayes* model represents the conditional density of a continuous random variable X given the label Y as a Gaussian PDF: $p(X|Y)=N(\mu_{X|Y}, \sigma_{X|Y})$ where $\mu_{X|Y}$ is the mean and $\sigma_{X|Y}$ the standard deviation of the Gaussian. Derive the MLE estimate for $\mu_{X|Y}$ for a binary classification problem. (10 points)

6. Suppose you are running a binary classification experiment with a new algorithm and a dataset with 100 positive and 100 negative examples. You use leave-one-out cross validation and a baseline of a majority class classifier. You expect the majority class classifier to be about 50% accurate, but it ends up being 0%. Explain why. (10 points)

7. Suppose you have a neural network with one hidden layer, in which every (non-input) unit has linear activation functions. That is, the output of a node is $\boldsymbol{w}\cdot\boldsymbol{x}+b$, where $\boldsymbol{x}$ is the input vector to that unit. Write down the output of the network as a function of $\boldsymbol{x}$, and show that the same output can be computed by a perceptron with a linear activation function at the output node. (10 points)