EECE 5698

Homework 4: due Wednesday, March 12, 2014

Reading:

- "Machine Learning" by Tom Mitchell, chapter 6.9-6.10 on Naïve Bayes
- 2.11 of Duda, Hart, and Stork

Problems:

Part I:

1. **Naïve Bayes:**
   Given the following table from Tom Mitchell's book as your training data:

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Goal: Predict the target attribute/class, which is whether to PlayTennis or not.

a. Determine all the class conditional probabilities $p(a_i|\omega_j)$'s you need to learn for a Naïve Bayes classifier (i.e., fill-in the following table):

Feature \ PlayTennis	Yes	No
Outlook = Sunny	P(Outlook = Sunny| Yes) = ________________	P(Outlook=Sunny|No) = _____________________
= Overcast		
= Rain		
Temperature = Hot		

b. When you encounter a test case:
<Outlook = Overcast, Temperature=Cool, Humidity =Normal, Wind=Weak>, what will your naïve Bayes classifier decide? (PlayTennis is yes or no?) Show all your steps.

c. If I do not make the naïve Bayes assumption (i.e, that the features are conditionally independent given the class), what would be the size of my class conditional probability table such as that in part a?

2. Bayesian Networks:
Book Problem Chapter 2, #50.

Section 2.11

50. Use the conditional probability matrices in Example 4 to answer the following separate problems.

(a) Suppose it is December 20—the end of autumn and the beginning of winter—and thus let $P(a_1) = P(a_4) = 0.5$. Furthermore, it is known that the fish was caught in the north Atlantic, that is, $P(b_1) = 1$. Suppose the lightness has not been measured but it is known that the fish is thin, that is, $P(d_2) = 1$. Classify the fish as salmon or sea bass. What is the expected error rate?

(b) Suppose all we know is that a fish is thin and medium lightness. What season is it now, most likely? What is your probability of being correct?

(c) Suppose we know a fish is thin and medium lightness and that it was caught in the north Atlantic. What season is it, most likely? What is the probability of being correct?