CMPS 142 Third Homework, Winter 2010

3 Problems, 12 pts, due start of class Thursday, Feb. 4

Each student should submit a homework and carefully acknowledge all sources of inspiration, techniques, and/or helpful ideas (web, people, books, etc.) other than the instructor, TA, and class text. The goal of this homework is to gain practice with Bayesian methods by working out some small examples by hand.

1. (3 pts)

   Consider the following experiment. A magician has three coins in his pocket, a two-headed coin, a two-tailed coin, and a fair coin. The magician picks a coin from his pocket with each coin equally likely. The magician then flips the coin twice, and sees what the comes up (either hh, ht, th, or tt). To make this more formal, consider three random variables $coin \in \{0, \frac{1}{2}, 1\}$, $flipA \in \{0, 1\}$, and $flipB \in \{0, 1\}$ where the value of $coin$gives the probability that a "head" results when the coin is flipped (and $1 - coin$ is the probability that a "tail" results) and $flipA$ and $flipB$ are indicator functions for the events "the first flip is a head" and "the second flip is a head" respectively. Let each triple of values $(coin, flipA, flipB)$ correspond to a point (atomic event) in Ω, and assume that $P(coin = 0) = P(coin = 1/2) = P(coin = 1) = 1/3$.

   First, (1 pt) how many points are in sample space $|\Omega|$? and how many points in Ω have zero probability? Second (2 pts), what is $P(flipB = 1|flipA = 1)$?

2. (4 pts) Consider building a decision tree from the following training set using the entropy impurity measure from the book (information gain, equation 6.2). What is the gain of each attribute if it is used at the root? Which boolean attribute will selected for the test at the root?

x_1	x_2	x_3	label
1	0	0	1
1	1	0	1
1	1	1	1
0	1	1	1
0	1	0	1
0	0	0	1
0	1	1	0
0	1	0	0
0	0	0	0

3. (5 pts) Consider a domain (instance space) with 3 kinds of points: a, b, and c, and a hypothesis class with three hypotheses: h_1, h_2, h_3 that assign the following probabilities of the label + to each kind of instance (the probability of the label − is 1 minus the

probability of +).

	h_1	h_2	h_3
a	0.8	0.25	0.5
b	0.8	0.75	0.5
c	0.2	0.75	0.5

Assume we have the prior probabilities $P(h_1) = 1/2$, $P(h_2) = 4/10$, and $P(h_3) = 1/10$ and then see the sample, $(a, +), (b, -), (c, +)$.

(a) What is the prior probability of + for each kind of point before seeing the data (based just on the hypotheses and their priors)?

(b) Which hypothesis is the maximum likelihood hypothesis after seeing the data?

(c) Which hypothesis is the maximum a'posteriori hypothesis after seeing the data?

(d) What is the mean posterior (full Bayesian) probabilities of + for each point after seeing the data?

(e) Consider the situation if we saw two of each kind of point, i.e. the sample was $(a, +), (b, -), (c, +), (a, +), (b, -), (c, +)$. Which of the preceding answers would change? (It might help to think of the the two a's, b's, and c's as different points.)