**Iowa State University**
**Department of Computer Science**
**Machine Learning (Com S 573)**
**Vasant Honavar**
Spring 2007
Problem Set 1
Due Jan 26, 2007

1. (**25 pts**) *Prove* the fundamental theorem of probability (See lecture notes, week 1 for a precise statement of the theorem).

2. (**25 pts.**) After winning a race, an Olympic runner is tested for the presence of steroids. The test comes up positive, and the athlete is accused of doping. Suppose it is known that 5 percent of all Olympics winners use performance-enhancing drugs. For this particular test, the probability of a positive result given that drugs are used is 95 percent. The probability of a positive result given that drugs are not used is 2 percent. What is the (posterior) probability that the athlete in fact did use steroids, given the positive outcome of the test?

3. (**25 pts**) It is quite often useful to consider the effect of some propositions in the context of some background evidence $\mathbf{E} = \mathbf{e}$ Let $\mathbf{e}$ be background evidence. Prove the conditionalized version of the general product rule: $P((\mathbf{X}, \mathbf{Y}|\mathbf{e}) = P(\mathbf{X}|\mathbf{Y}, \mathbf{e})P(\mathbf{Y}|\mathbf{e}$. In a similar vein, prove a conditionalized version of the Bayes theorem.

4. (**25 pts.**) Suppose you are given a bag containing $n$ coins. You are told that $n-1$ of these coins are normal, and one coin is fake, with heads on both sides.

   (a) Suppose you reach into the bag, and pick out a coin uniformly at random, flip it, and get a head. What is the (conditional) probability that the coin that you chose is a fake coin?

   (b) Suppose you continue flipping the coin you picked above for a total of $k$ times and see $k$ heads. Now what is the (conditional) probability that you picked a fake coin?

   (c) Suppose you wanted to decide whether the chosen coin was fake by flipping it $k$ times. Your decision procedure returns *FAKE* of a;; $k$ flips come up heads, otherwise it returns *NORMAL*. What is the probability that this procedure makes an error?

5. (**25 pts.**) This exercise investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations.

   (a) Suppose we wish to calculate $P(h|e_1, e_2)$ and we have no independence information. Which of the following sets of probabilities are sufficient for the calculation?

      i. $P(E_1, E_2)$,$P(H)$, $P(E_1|H)$, $P(E_2|H)$

      ii. $P(E_1, E_2)$, $P(H)$, $P(E_1, E_2|H)$

      iii. $P(H)$, $P(E_1|H)$, $P(E_2|H)$

(b) Suppose we know that $P(H|E_1, E_2) = P(E_1|H)$ for all values of $E_1$, $E_2$, and $H$. Now which of the above sets of probabilities is sufficient?

6. (**25 pts.**) Suppose you are a witness to a night-time hit-and-run accident involving a taxi in Athens. Suppose all taxis in Athens are blue or green. You swear, under oath, that the taxi was blue. Extensive testing shows that under the dim lighting conditions, discrimination between blue and green is 75% reliable. Is it possible to calculate the most likely color of the taxi? (Distinguish *carefully* between the propositions that the taxi is blue and that it *appears* blue). What about now, given that 9 out of 10 Athenian taxies are green?

7. (**25 pts.**) Let $X$, $Y$, and $Z$ be Boolean random variables. Label the entries in the joint distribution $P(X, Y, Z)$ with letters $a$ through $h$. Express the statement that $X$ and $Y$ are conditionally independent given $Z$ as a set of equations relating $a$ through $h$. How many *non-redundant* equations are there?

8. (**25 pts.**) Two statisticians go to the doctor and are both given the same prognosis: A 40% chance that the problem is a deadly disease $A$, and a 60% chance that the problem is a fatal disease $B$. Fortunately, there are anti-$A$ and anti-$B$ drugs that are inexpensive, 100% effective, and free of side effects. The statisticians have the choice of taking one drug, both, or neither. The first statistician is an avid Bayesian; the second statistician subscribes to the maximum likelihood principle.

   (a) What would the two statisticians do? Explain.

   (b) The doctor does some additional research and discovers that the disease $B$ actually comes in two types - dextro-$B$ and levo-$B$, which are equally likely, and equally treatable by the anti-$B$ drug. Now that there are three hypotheses - $A$, dextro-$B$ and levo-$B$, what will the two statisticians do? Explain.