CSE 250A Assignment 1

This assignment is due at the start of class on Thursday October 11, 2012. You should do the assignment in partnership with exactly one other student.

Detailed instructions. The purpose of these instructions is to maximize the educational value of the assignment, while minimizing the grading burden, which will enable the teaching assistants to provide more feedback. Each pair of students should submit a single joint solution set. The answer to each question should be written by hand, by one person. Each student should write the answer for two of the questions. You should indicate who wrote the answer to each question, but the scores for the assignment will be joint for both students. Start the answer to each question on a separate sheet of standard letter paper. Write on one side only. Staple together the answers to all questions.

You may ask questions on Piazza. You may also use books, Wikipedia, and other reference materials. If you do use any of these, please say so in your answers, so that we can know what is useful. In the questions below, "prove" or "show" means write down a mathematical argument that is precise, concise, clear, and correct. Mathematical arguments do not have to include definitions. Rather than use symbols of formal logic, write sentences using words such as "therefore."

Acknowledgment: These questions are adapted from questions written by Lawrence Saul.

1. Conditioning on background evidence

(a) Denoting background evidence by E, prove the conditional version of Bayes' rule:

$$P(X|Y,E) = \frac{P(Y|X,E)P(X|E)}{P(Y|E)}.$$

(b) Suppose that 1% of all people have cancer, 2% of people over 50 have cancer, and that a particular test for cancer has a 10% false positive rate and a 20% false negative rate. John is over 50 and his test result is positive. Work out the probability that John has cancer.

2. Using your imagination

Describe commonsense events corresponding to the binary random variables X, Y, and Z that are consistent with the following scenarios. You may use different events for the different parts of the problem.

(a) Cumulative evidence: $P(X = 1) < P(X = 1|Y = 1) < P(X = 1|Y = 1, Z = 1)$.

(b) Explaining away: $P(X = 1) < P(X = 1|Y = 1) > P(X = 1|Y = 1, Z = 1)$.

(c) Conditional independence without independence:

$$\begin{aligned} P(X=1,Y=1) &\neq P(X=1)P(Y=1) \\ \text{but } P(X=1,Y=1|Z=1) &= P(X=1|Z=1)P(Y=1|Z=1) \end{aligned}$$

(d) Independence without conditional independence:

$$\begin{aligned} P(X=1,Y=1) &= P(X=1)P(Y=1) \\ \text{but } P(X=1,Y=1|Z=1) &\neq P(X=1|Z=1)P(Y=1|Z=1). \end{aligned}$$

3. Entropy

(a) Let X be a discrete random variable with $p(X=x_i)=p_i$ for $i=1$ to $i=n$. The entropy $H[X]$ is a measure of the overall uncertainty of X. It is defined to be

$$H[X] = -\sum_{i=1}^{n} p_i \log p_i.$$

Show that the entropy $H[X]$ is maximized when $p_i = 1/n$ for each i. You should do this by computing partial derivatives with respect to p_i and using Lagrange multipliers to enforce the constraint that $1=\sum_i p_i$. Later in 250A, we will use similar calculations for learning probabilistic models.

4. Kullback-Leibler (KL) divergence

Consider two discrete probability distributions, p_i and q_i, with $\sum_i p_i = \sum_i q_i = 1$. The Kullback-Leibler (note the spelling) divergence between these distributions, which is also called the relative entropy, is defined to be

$$KL(p,q) = \sum_i p_i \log \frac{p_i}{q_i}.$$

(a) Give an example to show that the KL divergence is not a symmetric function: $KL(p,q) \neq KL(q,p)$. Despite this asymmetry, it is still common to view $KL(p,q)$ as a measure of distance. Many learning algorithms are based on minimizing KL divergences between distributions.

(b) Consider the natural logarithm, with base e. By sketching graphs of $\log(x)$ and $x-1$, illustrate that $\log(x) \leq x-1$, with equality if and only if $x=1$. Prove this inequality by differentiating $\log(x)-(x-1)$.

(c) Use the result from part (b) to show that $KL(p,q) \geq 0$, with equality if and only if the two distributions p_i and q_i are identical.

(d) Using the result from part (b), as well as the fact $\log x = 2\log\sqrt{x}$, prove the lower bound

$$KL(p,q) \geq \sum_i (\sqrt{p_i} - \sqrt{q_i})^2.$$

Note that the square root of half of the right side above is called the Hellinger distance.