STAT 535 Homework 3 and 4 extra credit problems

Out October 20, 2011

Due October 27,2011

©Marina Meilă
mmp@stat.washington.edu

These problems are for extra credit. The solutions of some (maybe all) of these problems can be found in the literature, sometimes in the references I hand out in class. If you solve it by finding the answer in the literature, please cite the source: author, article, publication venue, year, theorem number. A solution found in the literature a is a solution deserving credit, provided you were the one to find the solution by reading the respective article.

Problem 1 - Tarjan elimination and chordal graphs

a. Removing simplicial nodes.
Prove that if $\mathcal{G}$ is a chordal graph, and a simplicial node is removed, the resulting graph remains chordal.

b. A chordal graph as a Bayes net (with no V-structures).
Write in pseudocode an algorithm that directs the edges of a chordal graph to obtain a DAG with no V-structures. This algorithm will take a decomposable MRF and find a perfect Bayes Net map for it. Look at the footnote for a hint[1].

c. Prove that a chordal graph with n nodes cannot have more than $n-1$ cliques. What is the chordal graph that attains this value?

d.* Give a condition for the uniqueness of the junction tree.

e.* [Try to] give a characterization of the set of all junction trees that correspond to a given chordal graph. (This set represents an equivalence class of junction trees).

f. If an edge is added/removed from a chordal graph, is the resulting graph still chordal? If not, give a characterization of (i) what edge removals preserve chordality, and (ii) what edges can be added to preserve chordality. (These "moves" could be used to traverse the space of all chordal graphs over n nodes.)

[1] Modify the Tarjan elimination algorith m.

Problem 2 – The "minors" of the incidence matrix of a graph

The *incidence matrix* of a graph $\mathcal{G} = (V, \mathcal{E})$ with $|V| = n$, $|\mathcal{E}| = m$ is a $n \times m$ matrix M with elements in $\{\pm 1, 0\}$. Each edge $e = (i, j)$ of the graph is assigned a column, and in this column $M_{ie} = 1$, $M_{je} - 1$ and the rest of the elements are 0. Note that by this, each edge e is actually directed, but for the present problem it's not important how they are directed. That is, you can consider that the plus and minus of each edge is assigned arbitrarily.

We assume the graph $\mathcal{G}$ is connected, i.e it has a path between any two nodes of V. This implies $m \geq n - 1$.

Let $S \subset \mathcal{E}$ be a set of $n - 1$ edges and consider the submatrix M_S formed by the colums in S. This will be a $n \times n - 1$ matrix. Remove an arbitrary row of M_S to obtain B, a $(n-1) \times (n-1)$ matrix.

Prove that $det\, B = \pm 1$ if the edges in S form a spanning tree of $\mathcal{G}$ and $det\, B = 0$ otherwise.

1. This property, beautiful in itself, is the main ingredient for proving another beautiful and remarkable result in graph theory that will be mentioned later in the course.

2. This combinatorics problem, as well as the better known result I'm alluding to, have some interesting relations to statistics. Hidden in this problem is a process of counting. Do you see it? And counting is a special form of measure (in the measure theory sense). Hence, whenever you can count something, there likely is a way to turn this into a probabilistic model. Examples later in the course.

Problem 3 - Value of information

Suppose that we are interested in a query variable Q taking value q. So far, we have observed variables $E \subset V$ taking values e. The question is which other variable to observe from the ones that could be observed (this may not be all of $V \setminus E \setminus \{Q\}$). Therefore, we will be interested in computing a *score* for each candidate variable X, that reflects how valuable X would be in reducing uncertainty in $P[Q = q|E = e]$.

a. One possible way to arrive at a score for X is the *sensitivity*

$$S(Q = q, X = x) \;=\; \frac{P[Q = q|X = x, E = e]}{P[Q = q|E = e]} \;=\; \frac{P[X = x|Q = q, E = e]}{P[X = x|E = e]} \quad (1)$$

The first equality above is the definition. Show that the second (called *outward* formulation, because it goes from Q to X) inequality follows from the first (called *inward* formulation, because it goes from X to Q).

b. While the two defintions of sensitivity are equivalent mathematically, they may not be the same in terms of computation. Why may the outward formulation be more practical? Consider also the special case of a "disease and symptom" Bayesian network where Q is a disease and X is a symptom. Can you see an additional advantage in this case?

c. Sensitivity measures if observing $X = x$ increases or decreases the probability of $Q = q$.

Is this really a "good" criterion to choose X by? Why or why not?

Second, as you must have already realized, S depends on the particular value x that is observed. Is it possible to have $S(Q = q, X = x) > 1$ (or < 1) for *all* values $x \in \Omega_X$? (In other words, is it possible that no matter what X is, the probability of $Q = q$ increases?). Prove your answer.

d. *Mutual information* as score. The entropy of a (discrete) variable Z measures the "uncertainty" in the distribution P_Z. The mutual information[2] measures expected the reduction of *entropy* after observing X. Calculate the expression of the mutual information $I(Q, X) = H(Q) - H(Q|X)$ as a function of the conditional probabilities $P[Q = q|X = x, E = e]$, $P[Q = q|E = e], \ldots$. Note that this in an average over all possible values of Q.

Reformulate the expression of the mutual information in order to compute a score for the query of interest, i.e. for $Q = q$.

e. Note that $I(Q, X) = I(X, Q) = H(X) - H(X|Q)$. Similarly to question **b.**, are there among the various possible expressions for $I(Q = q, X)$ some that are more practical computationally?

(When) is the mutual information a good score? When not?

[2]For definitions of mutual information and entropy see `http://www.stat.washington.edu/courses/stat538/winter11/handouts.html` Lecture 7, section 1. These will be studied in more detail in STAT 538.