Homework Assignment 4

Due: March 13, 2018 at 12:01 am

Total points: 100

Deliverables: hw4.pdf containing typeset solutions to Problems 1-6.

Guidelines: All files must be submitted by Dropbox. You can brainstorm with others, but please solve the problems and write up the answers and code by yourself. You may use textbooks (Koller & Friedman, Russell & Norvig, etc.), and lecture notes. Please do NOT use any other resources or references (e.g., example code, online problem solutions, etc.) without asking.

1. Learning Bayesian Networks

(*20 points*) Consider learning the following Bayesian network: A → B ←C. And the following data table, with entries '?1' and '?2' missing at random:

A	B	C
?1	T	T
F	T	F
F	F	F
T	F	T
T	F	T
T	T	T
F	F	T
F	T	T
F	T	T
T	?2	F
T	T	F
T	F	F

1.1 Use the data to estimate initial parameters for this network, using maximum likelihood estimation for simplicity and ignoring missing data.

1.2 Perform two iterations of the EM algorithm (by hand) to estimate the values of the missing data, re-estimate the parameters, re-estimate the values of the missing data, and re-estimate the parameters once more. Show your calculations.

2. Learning Structure of Bayesian Networks

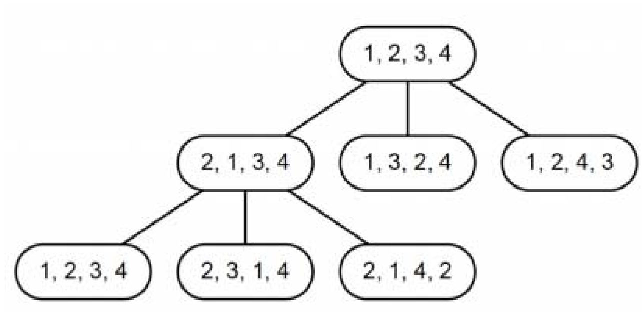

Figure 2: Partial search tree example for orderings over variables X_1, X_2, X_3, X_4. Successors to $\prec$= (1, 2, 3, 4) and $\prec$'= (2, 1, 3, 4) shown.

(*20 points*) Consider learning the structure of a Bayesian network for some given ordering, $\prec$, of the variables $X_1, \ldots, X_n$. (This can be done efficiently as described in section 18.5.2.1 of the textbook.) Now assume that we want to perform a search over the space of orderings; that is, we are searching for a network (with bounded in-degree k) that has the highest score. We do this by defining the score of an ordering as the score of the (bounded in-degree) network with the maximum score consistent with that ordering, and then we search for the ordering with the highest score. We bound the in-degree so that we have a smaller and smoother search space. We will define our search operator, o, to be "Swap X_i and X_{i+1}" for some $i = 1, \ldots, n-1$. Starting from some given ordering, $\prec$, we evaluate a decomposable structure score of all successor orderings, $\prec$', where a successor ordering is found by applying o to $\prec$ (see Figure 2). We now choose a particular successor, $\prec$'. Provide an algorithm for computing as efficiently as possible the score for all successors of the new ordering, $\prec$', given that we have already computed the scores for all successors of $\prec$. Note: A structure $score(G: D)$ is decomposable if the score of a structure G can be written as

$$score(G: D) = \sum_i FamScore(X_i | Pa_{X_i}^G : D),$$

where the family score $FamScore(X | U: D)$ is a score measuring how well a set of variables U serves as parents of X in the data set D.

3. Gibbs Sampling

(*10 points*) Consider the Bayesian network $A \rightarrow B$, where A and B are Boolean variables, $P = (A = 1) = 1/2$, and $P(B = 1 | A = 1) = P(B = 0 | A = 0) = 1$. Will Gibbs sampling applied to this network converge to the correct stationary distribution? Why?

4. Dynamic Bayesian Networks

(*20 points*) A patient has a disease N. Physicians measure the value of a parameter P to see the disease development. The parameter can take one of the following values: {low, medium, high}. The value of P is a result of patient's unobservable condition/state S. S can be {good, poor}. The state changes between two consecutive days in 1/5 of cases. If the patient is in good condition, the value for P is rather low (having 10 sample measurements, 5 of them are low, 3 medium and 2 high), while if the patient is in poor condition, the value is rather high (having 10 measurements, 3 are low, 3 medium and 4 high). On arrival to the hospital on day 0, the patient's condition was unknown, i.e., $p(S_0 = \text{good}) = 0.5$.

4.1 Draw the transition and sensor model of the dynamic Bayesian network modeling the domain under consideration.

4.2 Calculate probability that the patient is in good condition on day 2 given low P values on days 1 and day 2.

4.3 Can you determine the most likely patient state sequence in days 0, 1 and 2 without any additional computations? Justify.

5. Particle Filter

(1*0 points*) At each resampling step, particle filtering usually creates multiple copies of the highest-weight samples. Why does this not result in the number of different samples dwindling over time?

6. Decision Theory

(*20 points*) You are going to have pizza for dinner and are trying to decide whether to have to pizza delivered, or whether to pick it up yourself. In the end, all that matters to you is how much the pizza costs and whether the pizza is hot or cold (e.g., the trip to the parlor is irrelevant). The pizza costs $10. If it is delivered, you pay a $2 delivery charge, unless the pizza is cold when it arrives, in which case the pizza and the delivery are free. The pizza parlor delivers cold pizza 1 out of 50 times. If you decide to pick the pizza up, there is no delivery charge. However, there is a 1 in 10 chance that you will be late, and the pizza will be cold. There is also a 1-in-100 chance (independent of whether you are late) that you will be the 200th customer to go into the pizzeria today, in which case the pizza is free.

6.1 Write your decision as a choice between two lotteries.

6.2 If I only know that you like hot pizza more than cold pizza (other things equal) and cheap pizza more than expensive pizza (other things equal), can I determine your ranking of the possible outcomes in the two lotteries? Explain.

6.3 Is there any ranking of the outcomes consistent with the above (hot better than cold, etc.) such that having the pizza delivered stochastically dominates picking up the pizza? Explain.

Note: The simplest case of stochastic dominance is statewise dominance, defined as follows:

> Gamble A is statewise dominant over gamble B if A gives at least as good a result in every state (every possible set of outcomes), and a strictly better result in at least one state.

Statewise dominance is a special case of the canonical first-order stochastic dominance (FSD). FSD can be expressed as follows: if and only if A first-order stochastically dominates B, there exists some gamble y such that $x_B \cong (x_A + y)$ where $y \leq 0$ in all possible states (and strictly negative in at least one state); here $\cong$ means "is equal in distribution to" (that is, "has the same distribution as"). Thus, we can go from the graphed density function of A to that of B by, roughly speaking, pushing some of the probability mass to the left.

6.4 Give a ranking of the outcomes consistent with the above such that picking the pizza up first-order stochastically dominates having the pizza delivered.

6.5 Give a ranking of the outcomes consistent with the above such that picking the pizza up does not first-order stochastically dominate having the pizza delivered.