Solutions

ECE 45234 - Spring 2018 Problem Set #3 Due: 4/5/18

The purpose of this problem set is to gain experience with probabilistic inference methods. In the engineering design section you will be building a generic tool for probabilistic inference using approximations.

You can complete the exercises by either directly marking up this pdf, or by printing, completing, and scanning as a pdf. You should complete the Engineering Design Problems by writing the python code as instructed. The resulting pdf and python files should be uploaded to Canvas via the assignment tab by the due date and time.

Exercises

1. Suppose that $f_X(x)$ is a probability density for a random variable X, with $x \in R$; and $P[Y]$ is a probability mass function for a random variable $Y \in \{-3, -2, -1, 0, 1, 2, 3\}$.

   (a) (2 points) What is the expression that will compute the probability that X is positive?

   $$P[X > 0] = \int_0^\infty f_X(x)\, dx$$

   (b) (2 points) What is the expression that will compute the probability that Y is positive?

   $$P[Y > 0] = \sum_{y=1}^{3} P[Y = y]$$

   (c) (2 points) Suppose that $P[Y]$ is uniform distributed. What is it's probability mass function?

   $$P[Y = y] = \frac{1}{7}$$

   (d) (2 points) What are the possible factorings of the joint density of X and Y? Use a lowercase f to denote density functions and P to denote mass functions.

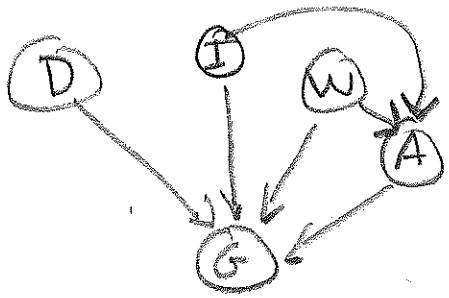
   $f(x|y)P[y]$ and $P[y|x]f_X(x)$

2. Suppose that there are four factors that affect a student's grade in a course, the average grade in prerequisite courses, the difficulty of the course, the intelligence of the student, and the work ethic of the student.

   (a) (2 points) Describe how you would model each of the variables to account for uncertainty.

   A = average grade in prereq in $[0, 4]$
   D = course difficulty in $[0, \infty)$
   I = intelligence in $[0, \infty)$
   W = work ethic in $\mathbb{R}$
   G = grade, categorical $\{F, D-, D, D+, \ldots, A-, A\}$

   (b) (2 points) Sketch a Bayesian Network showing the dependencies among the variables.



   (c) (2 points) Suppose the relationship of factors to grades was a noisy-or. What might the node probabilities look like for each variable? You can use unspecified constants, but specify their semantic interpretation.

A changes to low/high
D changes to easy/hard
I changes to dumb/smart
W changes to lazy/contientious
G changes to pass/fail

let f_a, f_d, f_i, f_w be reliability factors for each variable, and f the overall reliability.

Example:

A	D	I	W	P(G = fail | A, D, I, W)
0	0	0	0	f
0	0	0	1	$f(1-f_w)$

3. Suppose you are trying to decide whether to take a job or not. You are primarily concerned about two issues, the pay and your overall interest in the job's duties. Casting the above situation as an inference problem,

   (a) (2 points) Define the random variables involved and their domain, giving a brief description of their semantic interpretation.

Let pay be in units of \$/year denoted $A \in [0, \infty)$

Interest be rated on a scale 0-10

$I \in \{0, 1, 2, 3 \ldots, 10\}$

   (b) (2 points) Define a reasonable probability relationship among the variables, giving a brief rationale for your choices.

There are two models capturing different preferences.

Interest determines pay uncertainty

pay determines interest level

   (c) (2 points) What is the expression required to help you make your decision?

We need to define a utility over A, I, $U(A, I)$

Given alternate jobs with different $P(A, I)$

Take job that maximizes,

$$\sum_{i=0}^{10} \int_{0}^{\infty} U(A=a, I=i)\, P(A=a, I=i)\, da$$

4. (5 points) Suppose two agents have identical internal state variables, X, and evidence variables, E, but different models, M_1 and M_2. Which of the following probabilities is most important for deciding which agent should perform better (circle one)? For simplicity, assume all random variables are discrete.

   (a) $P[X]$

   (b) $P[E]$

   (c) $P[M_1]$ and $P[M_2]$

   (d) $P[X, E, M_1]$ and $P[X, E, M_2]$

   (e) $P[M_1|E]$ and $P[M_2|E]$

   (f) $P[X|E, M_1]$ and $P[X|E, M_2]$

   Provide a rationale for your choice:

In Bayesian Inference.

$$P(X|E,m) = \frac{p(E|X,m)\,p(X|m)}{P[E|m]}$$

Given the evidence and two models m_1 & m_2, we prefer model 1 (m_1) if

$$P[m_1|E] > P[m_2|E]$$

and vice-versa.

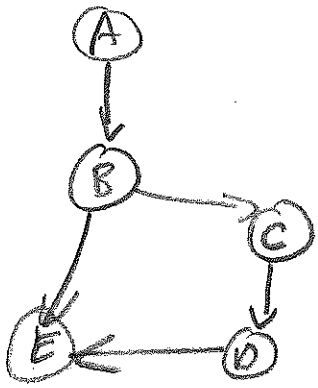
These are computed as,

$$P[m_1|E] = \frac{p[E|m_1]\,p[m_1]}{p[E|m_1]P[m_1] + p[E|m_2]\,p[m_2]}$$

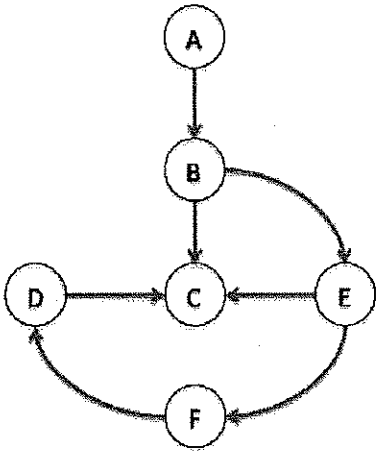
Similarly for m_2.

5. (5 points) Given the following factorization of discrete random variables A, B, C, D, E what is the associated Bayesian network structure?

$$P(A, B, C, D, E) = P(E|B, D)P(D|C)P(C|B)P(B|A)P(A)$$



6. (5 points) Given the following graph structure, is it a valid Bayesian network (justify your answer)? If so, what is the associated factorization of the random variables?



Yes, no cycles.

$P(C \mid B, D, E)P(D|F)P(F|E)P(E|B)P(B|A)$ is the joint $P(A, B, C, D, E, F)$

7. Suppose you are given the following Bayesian Network for binary random variables

$$P(A,B,C,D) = P(D|B,C)P(B|A)P(C|A)P(A)$$

where

$P(A=1) = 0.4$,

A	$P(C=1\|A)$
0	0.2
1	0.8

,

A	$P(B=1\|A)$
0	0.1
1	0.9

,

B	C	$P(D=1\|B,C)$
0	0	0.1
0	1	0.3
1	0	0.2
1	1	0.4

Using exact inference determine the following (show your work)

(a) (2 points) Supposing that $B=0$ and $D=1$, what is the $P[C=1]$?

$$P(C=1|B=0,D=1) = \alpha \sum_{a=0}^{1} P(D=1|B=0,C=1)P(B=0|A=a)P(C=1|A=a)P(A=a)$$

$$= \alpha \cdot (0.042)$$

$$P(C=0|B=0,D=1) = \alpha \sum_{a=0}^{1} P(D=1|B=0,C=0)P(B=0|A=a)P(C=0|A=a)P(A=a)$$

$$= \alpha \cdot (0.044)$$

$$P(C=1|B=0,D=1) = \frac{0.042}{0.042+0.044} \approx \frac{1}{2}$$

(b) (2 points) Supposing that $A=1$, $B=0$, and $C=0$, what is the $P[D=0]$?

No marginalization is required.

$$P(D=0|A=1,B=0,C=0) = \alpha\, P[D=0|B=0,C=0)P(B=0|A=1)P(C=0|A=1)P(A=1)$$

$$= \alpha \cdot (0.0072)$$

$$P(D=1|A=1,B=0,C=0) = \alpha\, P[D=1|B=0,C=0)P(B=0|A=1)P(C=0|A=1)P(A=1)$$

$$= \alpha\,(0.0008)$$

$$P(D=0|A=1,B=0,C=0) = \frac{0.0072}{0.0072+0.0008} \approx 0.9$$

8. Considering the likelihood-weighting and Gibbs sampling algorithms, answer the following questions:

(a) (2 points) In both algorithms, what is the purpose of the NORMALIZE function?

To convert the histograms to probability densities.

(b) (2 points) In the likelihood-weighting algorithm, when would you expect the weight to be high?

for values of the query variable that agree with the evidence variables.

(c) (2 points) In the Gibbs sampling algorithm, how is the Markov blanket defined?

parents, children, children's parents

(d) (2 points) Suppose I have a Bayesian Network that consists of only one variable. What (if any) difference is there between the algorithms?

None, sampling is same a perturbing the node.

9. Suppose that your old clunker of a car breaks down and you have it towed to a mechanic. The mechanic believes the problem could be the battery, the starter, or some unknown cause and asks you what you want to do: replace the battery, replace the starter, or sell him the car. Now that you have taken this course and know how to be a rational decision maker,

   (a) (5 points) What questions would you ask him/her to inform your decision?

Let B=1 mean battery failed
S=1 " starter "
K=1 " some other cause
C=1 " car is working

B, S, K	P(C=1 | B, S, K)
000	
001	
010	
...	

ask reliability of battery, starter, etc to fill in CPT.

   (b) (5 points) Formulate the expression you would use to make your decision.

We need the cost of the battery, starter, unknown cause.

$U(B, S, K)$

Maximize expected utility over decisions.

e.g.

$$EU[\text{replace battery}] = \sum_{S=0}^{1} \sum_{K=0}^{1} U(B=1, S=s, K=k) P(C=1 \mid B=1, S=s, K=k)$$