Section Handout 7 Solutions

CS188 Spring 2019 Section 7: Probability and Bayes Nets (Representation and Inference)

Probability

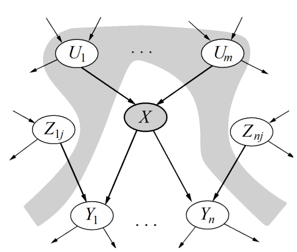
A **random variable** represents an event whose outcome is unknown. A **probability distribution** is an assignment of weights to outcomes A **joint distribution** over discrete random variables is a table of probabilities which captures the likelihood of each possible **outcome**, also known as an **assignment** of values to the random variables.

To write that random variables X and Y are **marginally independent**, we write $X \perp\!\!\!\perp Y$. To write that random variables X and Y are **conditionally independent** given another random variable Z, we write $X \perp\!\!\!\perp Y|Z$.

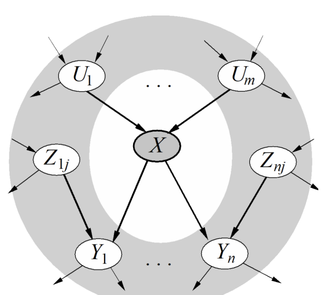
Bayesian Network Representation

In a Bayesian network, rather than storing information in a giant table, probabilities are instead distributed across a large number of smaller local probability tables along with a **directed acyclic graph** (DAG) which captures the relationships between variables. Thus, if we have a node representing variable X, we store $P(X|A_1, A_2, ..., A_N)$, where $A_1, ..., A_N$ are the parents of X.

- **Each node is conditionally independent of all its ancestor nodes (non-descendents) in the graph, given all of its parents**.



- **Each node is conditionally independent of all other variables given its Markov blanket.** A variable's Markov blanket consists of parents, children, children's other parents.



Bayesian Network Inference

Inference by enumeration

Given a joint PDF, we can trivially perform compute any desired probablity distribution $P(Q_1 \dots Q_k \mid e_1 \dots e_k)$ using a simple and intuitive procedure known as **inference by enumeration**, for which we define three types of variables we will be dealing with:

1. **Query variables** Q_i, which are unknown and appear on the left side of the probability distribution we are trying to compute.
2. **Evidence variables** e_i, which are observed variables whose values are known and appear on the right side of the probability distribution we are trying to compute.
3. **Hidden variables**, which are values present in the overall joint distribution but not in the distribution we are currently trying to compute.

In this procedure, we collect all the rows consistent with the observed evidence variables, sum out all the hidden variables, and finally normalize the table so that it is a probability distribution (i.e. values sum to 1).

Inference by elimination

Let us suppose we have a joint distribution $p(X_1, X_2, X_3, X_4) = p(X_1)p(X_2|X_1)p(X_3|X_1)p(X_4|X_2, X_3)$. We want to compute $p(X_2|X_4 = x_4)$. The query variable is X_2, the evidence variable is X_4, and the hidden variable are X_1 and X_3. The **elimination ordering** we will use is to first eliminate X_3, then X_1. When we eliminate a variable X_i, we will create a **factor** f_{X_i}, which is an unnormalized probability distribution. Here are the steps.

1. Write out the marginal distribution between the query and evidence variables $p(X_2, X_4 = x_4)$ by summing over the values of the hidden variables.

$$p(X_2, X_4 = x_4) = \sum_{x_1} \sum_{x_3} p(X_1 = x_1)p(X_2|X_1 = x_1)p(X_3 = x_3|X_1 = x_1)p(X_4 = x_4|X_2, X_3 = x_3)$$

2. Distributed the summations according to the elimination ordering.

$$p(X_2, X_4 = x_4) = \sum_{x_1} p(X_1 = x_1)p(X_2|X_1 = x_1) \sum_{x_3} p(X_3 = x_3|X_1 = x_1)p(X_4 = x_4|X_2, X_3 = x_3)$$

3. Eliminate variables by summing over the conditional probability distributions containing the variable.
   (a) Eliminate X_3 to create factor $f_{X_3}(X_4 = x_4|X_1 = x_1, X_2) = \sum_{x_3} p(X_3 = x_3|X_1 = x_1)p(X_4 = x_4|X_2, X_3 = x_3)$ to obtain

$$p(X_2, X_4 = x_4) = \sum_{x_1} p(X_1 = x_1)p(X_2|X_1 = x_1)f_{X_3}(X_4 = x_4|X_1 = x_1, X_2).$$

   (b) Eliminate X_1 to create factor $f_{X_1}(X_2, X_4 = x_4) = \sum_{x_1} p(X_1 = x_1)p(X_2|X_1 = x_1)f_{X_3}(X_4 = x_4|X_1 = x_1, X_2)$ to obtain

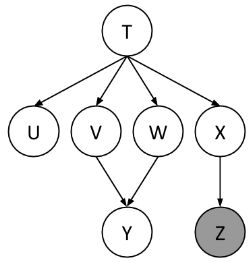
$$p(X_2, X_4 = x_4) = f_{X_1}(X_2, X_4 = x_4).$$

4. Obtain the conditional probability.

$$p(X_2 = x_2|X_4 = x_4) = \frac{f_{X_1}(X_2 = x_2, X_4 = x_4)}{\sum_{x_2'} f_{X_1}(X_2 = x_2', X_4 = x_4)}.$$

1 Variable Elimination

Using the same Bayes Net (shown below), we want to compute $P(Y \mid +z)$. All variables have binary domains. Assume we run variable elimination to compute the answer to this query, with the following variable elimination ordering: X, T, U, V, W.



Complete the following description of the factors generated in this process:

After inserting evidence, we have the following factors to start out with:

$$P(T), P(U|T), P(V|T), P(W|T), P(X|T), P(Y|V, W), P(+z|X)$$

(a) When eliminating X we generate a new factor f_1 as follows, which leaves us with the factors:

$$f_1(+z|T) = \sum_x P(x|T)P(+z|x) \qquad P(T), P(U|T), P(V|T), P(W|T), P(Y|V, W), f_1(+z|T)$$

(b) When eliminating T we generate a new factor f_2 as follows, which leaves us with the factors:

$$f_2(U, V, W, +z) = \sum_t P(t)P(U|t)P(V|t)P(W|t)f_1(+z|t) \qquad P(Y|V, W), f_2(U, V, W, +z)$$

(c) When eliminating U we generate a new factor f_3 as follows, which leaves us with the factors:

$$f_3(V, W, +z) = \sum_u f_2(u, V, W, +z) \qquad P(Y|V, W), f_3(V, W, +z)$$

Note that U could have just been deleted from the original graph, because $\sum_u P(U|t) = 1$. We can see this in the graph: we can remove any leaf node that is not a query variable or an evidence variable.

(d) When eliminating V we generate a new factor f_4 as follows, which leaves us with the factors:

$$f_4(W, Y, +z) = \sum_v f_3(v, W, +z)P(Y|v, W) \qquad f_4(W, Y, +z)$$

(e) When eliminating W we generate a new factor f_5 as follows, which leaves us with the factors:

$$f_5(Y, +z) = \sum_w f_4(w, Y, +z) \qquad f_5(Y, +z)$$

(f) How would you obtain $P(Y \mid +z)$ from the factors left above:
Simply renormalize $f_5(Y, +z)$ to obtain $P(Y \mid +z)$. Concretely,

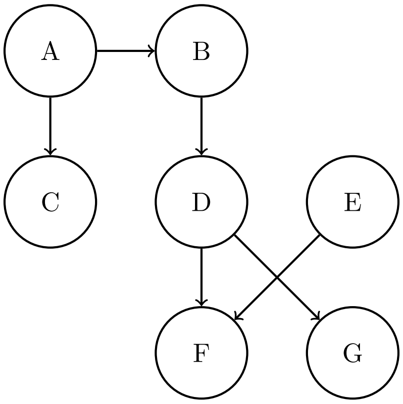
$$P(y \mid +z) = \frac{f_5(y, +z)}{\sum_{y'} f_5(y', +z)}$$

(g) What is the size of the largest factor that gets generated during the above process?
$f_2(U, V, W, +z)$. This contains 3 unconditioned variables, so it will have $2^3 = 8$ probability entries (U, V, W are binary variables, and we only need to store the probability for $+z$ for each possible setting of these variables).

(m) Does there exist a better elimination ordering (one which generates smaller largest factors)?
Yes. One such ordering is X, U, T, V, W. All factors generated with this ordering contain at most 2 unconditioned variables, so the tables will have at most $2^2 = 4$ probability entries (as all variables are binary).

Q2. Bayes' Nets: Representation and Independence

Parts (a), (b), and (c) pertain to the following Bayes' Net.



(a) Express the joint probability distribution as a product of terms representing individual conditional probabilities tables associated with the Bayes Net.

$P(A)P(C|A)P(B|A)P(D|B)P(E)P(F|D,E)P(G|D)$

(b) Assume each node can take on 4 values. How many entries do the factors at A, D, and F have?

A: 4

D: 4^2

F: 4^3

(c) Mark the statements that are guaranteed to be true. Recall that every variable is conditionally independent of its non-descendants given its parents, and every variable is conditionally independent of all other variables given its Markov blanket.

- [ ] $B \perp\!\!\!\perp C$
- [ ] $A \perp\!\!\!\perp F$
- [ ] $D \perp\!\!\!\perp E|F$
- [x] $E \perp\!\!\!\perp A|D$
- [x] $F \perp\!\!\!\perp G|D$
- [x] $B \perp\!\!\!\perp F|D$
- [ ] $C \perp\!\!\!\perp G$
- [x] $D \perp\!\!\!\perp E$

Parts (d) and (e) pertain to the following probability distribution tables. The joint distribution $P(A, B, C, D)$ is equal to the product of these probability distribution tables.

A	$P(A)$
+a	0.8
–a	0.2

A	B	$P(B\|A)$
+a	+b	0.9
+a	–b	0.1
–a	+b	0.6
–a	–b	0.4

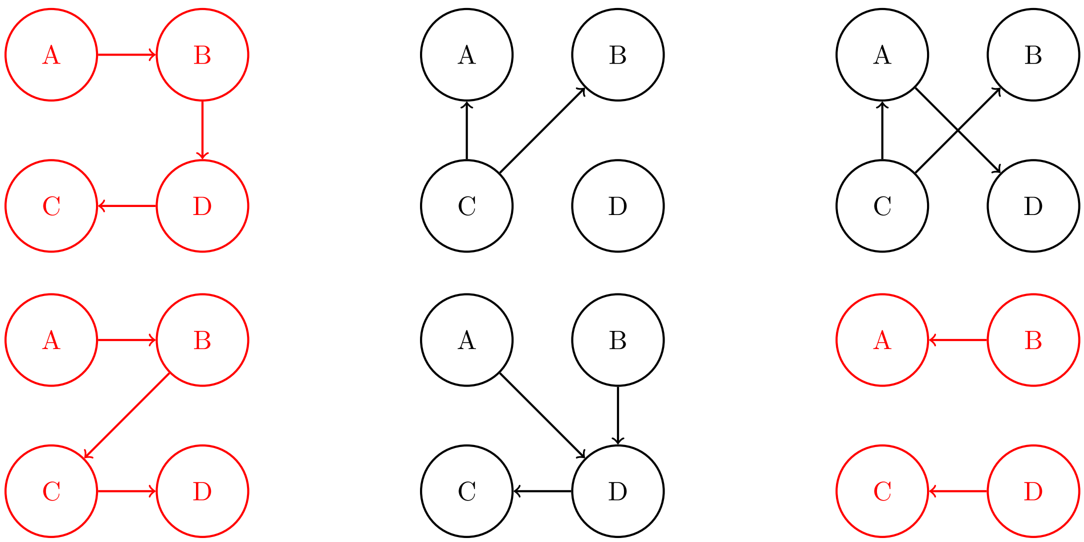
B	C	$P(C\|B)$
+b	+c	0.8
+b	–c	0.2
–b	+c	0.8
–b	–c	0.2

C	D	$P(D\|C)$
+c	+d	0.25
+c	–d	0.75
–c	+d	0.5
–c	–d	0.5

(d) State all non-conditional independence assumptions that are implied by the probability distribution tables.

From the tables, we have $A \not\perp\!\!\!\perp B$ and $C \not\perp\!\!\!\perp D$. Then, we have every remaining pair of variables: $A \perp\!\!\!\perp C, A \perp\!\!\!\perp D, B \perp\!\!\!\perp C, B \perp\!\!\!\perp D$

(e) Circle all the Bayes net(s) that can represent a distribution that is consistent with the tables given.



The question asks for Bayes Nets that **can** represent the distribution in the tables. So, in the nets we circle, the only requirement must be that A and B must not be independent, and C and D must not be independent.

The top left, bottom left, and bottom right nets have arrows between the A-B nodes and the C-D nodes, so we can circle those.

The top middle net has C and D as independent (D is not connected to anything), so we cannot circle it. The bottom middle net has A and B as independent (common cause), so we cannot circle it.

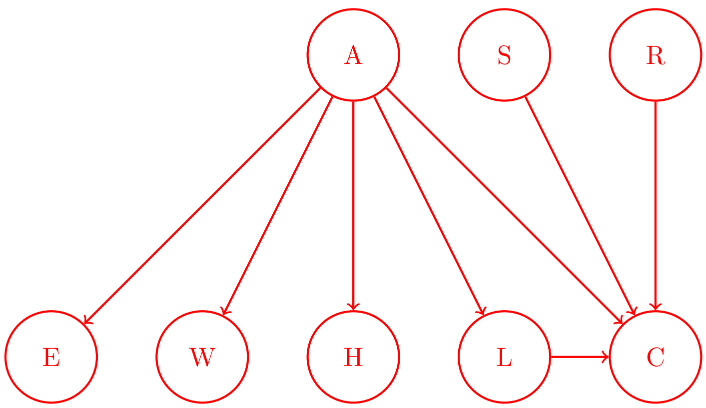
The top right net seems like it could represent the distribution, because D-separation finds that: A and B are not guaranteed to be independent (common effect), and C and D are not guaranteed to be independent (causal chain). However, according to Part D, $A \perp\!\!\!\perp C$, $A \perp\!\!\!\perp D$, and $B \perp\!\!\!\perp C$, so all of the arrows in the net are vacuous. That means, in this net, A and B are independent, and C and D are independent, so we cannot circle this net.

You are building advanced safety features for cars that can warn a driver if they are falling asleep (A) and also calculate the probability of a crash (C) in real time. You have at your disposal 6 sensors (random variables):

- E: whether the driver's eyes are open or closed
- W: whether the steering wheel is being touched or not
- L: whether the car is in the lane or not
- S: whether the car is speeding or not
- H: whether the driver's heart rate is somewhat elevated or resting
- R: whether the car radar detects a close object or not

A influences $\{E, W, H, L, C\}$. C is influenced by $\{A, S, L, R\}$.

(f) Draw the Bayes Net associated with the description above by adding edges between the provided nodes where appropriate.



(g) Mark all the independence assumptions that must be true.

- ■ $E \perp\!\!\!\perp S$
- ■ $W \perp\!\!\!\perp H|A$
- ■ $S \perp\!\!\!\perp R$
- ☐ $E \perp\!\!\!\perp L$
- ☐ $L \perp\!\!\!\perp R|C$
- ■ $W \perp\!\!\!\perp R$
- ☐ $A \perp\!\!\!\perp C$
- ☐ $E \perp\!\!\!\perp C|L$

(h) The car's sensors tell you that the car is in the lane ($L = +l$) and that the car is not speeding ($S = -s$). Now you would like to calculate the probability of crashing, $P(C| + l, -s)$. We will use the variable elimination ordering R, A, E, W, H. Write down the largest factor generated during variable elimination. Box your answer.

Our factors if we don't observe evidence are $P(A), P(S), P(R), P(E|A), P(W|A), P(H|A), P(L|A), P(C|L, A, S, R)$. We observe evidence, and we have: $P(A), P(R), P(E|A), P(W|A), P(H|A), P(C| + l, A, -s, R)$. We first eliminate R, so we select $P(R)$ and $P(C| + l, A, -s, R)$ to get $f_1(C| + l, A, -s)$. Now we eliminate A, so we select $P(A), P(E|A), P(W|A), P(H|A), f_1(C| + l, A, -s)$ and get $\boxed{f_2(C, E, W, H| + l, -s)}$. We see that this must be the largest factor because this is the only factor we have left at this point, and variable elimination is not yet finished.

(i) Write down a more efficient variable elimination ordering, i.e. one whose largest factor is smaller than the one generated in the previous question.

Any ordering of the five variables where at least one of $\{E, W, H\}$ is before A would be more efficient than the previous ordering. As an example, R, E, W, H, A would work.