Tutorial 6 - Variable Elimination

COMP9418 – Advanced Topics in Statistical Machine Learning

Lecturer: Gustavo Batista

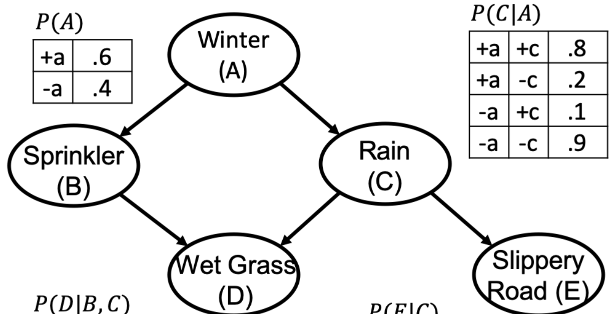
Lecture: Variable Elimination

Topic: Questions from lecture topics

Last revision: Sunday 4th October, 2020 at 13:00

Question 1

Consider the following Bayesian network.



$P(A)$

+a	.6
-a	.4

$P(C|A)$

+a	+c	.8
+a	-c	.2
-a	+c	.1
-a	-c	.9

$P(B|A)$

+a	+b	.2
+a	-b	.8
-a	+b	.75
-a	-b	.25

$P(D|B,C)$

+b	+c	+d	.95
+b	+c	-d	.05
+b	-c	+d	.9
+b	-c	-d	.1
-b	+c	+d	.8
-b	+c	-d	.2
-b	-c	+d	0
-b	-c	-d	1

$P(E|C)$

+c	+e	.7
+c	-e	.3
-c	+e	0
-c	-e	1

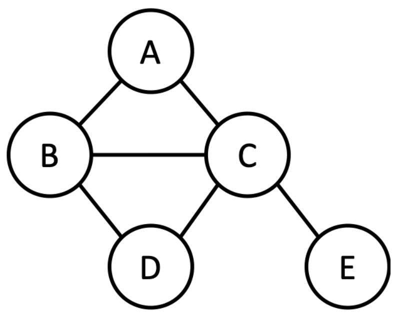
Use variable elimination to computer the marginals $P(Q, e)$ and $P(Q|e)$ where $\mathbf{Q} = \{E\}$ and $\mathbf{e} : D = false$. Use the min-degree heuristic for determining the elimination order breaking ties by choosing variables that come first in the alphabet.

Use the following algorithm for min-degree order:

Data: PGM: probabilistic grahical model
Data: $\mathbf{X}$: variables in the PGM
Result: an ordering π of variables $\mathbf{X}$
begin
    $G \leftarrow$ induced graph of the factors in PGM;
    for *i = 1 to number of variables in **X*** **do**
        $\pi(i) \leftarrow$ a variable in $\mathbf{X}$ with smallest number of neighbours in G;
        add an edge between every pair of non-adjacent neighbours of $\pi(i)$ in G;
        delete variable $\pi(i)$ from G and from $\mathbf{X}$;
    end
end
return π

Answer

The Bayesian network has the following induced graph:



As E is a query variable, we will not eliminate it. Considering the remaining variables, A and D have degree two. We chose A according to our tie-breaking policy. The removal of A from the induced graph makes the degree of B to decrease to two and the degree of C to reduce to three. Now, B and D have degree two, we chose B. The removal of B decreases the degree of D to one and C to two. We conclude choosing D and C.

The final elimination order is $\pi = A, B, D, C$.

To answer the query, we need first to set the evidence $D = false$.

B	C	D	$P(D\|B,C)$
+b	+c	-d	.05
+b	-c	-d	.1
-b	+c	-d	.2
-b	-c	-d	1

We have the following factors:

$\phi_A(A), \phi_B(B,A), \phi_C(C,A), \phi_D(D,B,C), \phi_E(E,C)$

We start eliminating A by calculating $\sigma_1(A,B,C) = \phi_A(A) \times \phi_B(B,A) \times \phi_C(C,A)$ and summing out A, resulting in a new factor $\tau_1(B,C) = \sum_A \sigma_1(A,B,C)$.

A	B	C	$\sigma_1(A,B,C)$
+a	+b	+c	.096
+a	+b	-c	.024
+a	-b	+c	.384

A	B	C	$\sigma_1(A,B,C)$
+a	-b	-c	.096
-a	+b	+c	.03
-a	+b	-c	.27
-a	-b	+c	.01
-a	-b	-c	.09

and,

B	C	$\tau_1(B,C)$
+b	+c	.126
+b	-c	.294
+b	+c	.394
+b	-c	.186

Now, we have the factors:

$\phi_D(D,B,C), \phi_E(E,C), \tau_1(B,C)$

and we proceed to eliminate B by calculating $\sigma_2(B,C,D) = \phi_D(D,B,C) \times \tau_1(B,C)$ and summing out B, $\tau_2(C,D) = \sum_B \sigma_2(B,C,D)$

B	C	D	$\sigma_2(B,C,D)$
+b	+c	-d	.0063
+b	-c	-d	.0294
-b	+c	-d	.0788
-b	-c	-d	.186

and,

C	D	$\tau_2(C,D)$
+c	-d	.0851
-c	-d	.2154

Now, we have the factors:

$\phi_E(E,C), \tau_2(C,D)$

and we proceed to eliminate D by calculating $\tau_3(C) = \sum_D \tau_2(C,D)$

C	$\tau_3(C)$
+c	.0851
-c	.2154

Now, we have the factors:

$\phi_E(E,C), \tau_3(C)$

and we proceed to eliminate C by calculating $\sigma_4(C,E) = \phi_E(E,C) \times \tau_3(C)$ and summing out C, $\tau_4(E) =$

$\sum_C \sigma_4(C, E)$

C		E
+c	+e	.05957
+c	-e	.02553
-c	+e	0
-c	-e	.2154

and we proceed to eliminate C by calculating $\tau_4(E) = \sum_C \sigma_4(C, E)$

E	$\tau_4(E)$
+e	.05957
-e	.24093

Therefore $P(+e, -d) = 0.05957$ and $P(-e, -d) = 0.24093$. We can also calculate $P(-d) = 0.05957 + 0.24093 = 0.3005$.

Normalizing the results, we get $P(+e|-d) = \frac{P(+e,-d)}{P(-d)} \approx 0.198236$ and $P(-e|-d) = \frac{P(-e,-d)}{P(-d)} \approx 0.801764$.

Question 2

Consider a chain network $C_0 \rightarrow C_1 \rightarrow \ldots \rightarrow C_n$. Suppose that variable C_t, for $t \geq 0$, denotes the health state of a component at time t. In particular, let each C_t take on states ok and faulty. Let C_0 denote component birth where $P(C_0 = ok) = 1$ and $P(C_0 = faulty) = 0$. For each $t > 0$, let the CPT of C_t be $P(C_t = ok|C_{t-1} = ok) = \lambda$ and $P(C_t = faulty|C_{t-1} = faulty) = 1$. That is, if a component is healty at time $t-1$, then it remains healthy at time t with probability λ. If a component is faulty at time $t-1$, then it remains faulty at time t with probability 1.

a. Using variable elimination with variable ordering C_0, C_1 compute $P(C_2)$.
b. Using variable elimination with variable ordering $C_0, C_1, \ldots, C_{n-1}$ compute $P(C_n)$.

Answer

a. $P(C_2) = \sum_{C_0, C_1} P(C_0, C_1, C_2)$
   $= \sum_{C_0, C_1} P(C_2|C_1)P(C_1|C_0)P(C_0)$
   $= \sum_{C_1} P(C_2|C_1) \sum_{C_0} P(C_1|C_0)P(C_0)$

   Operating over factors, we have

C_0	$P(C_0)$
ok	1
faulty	0

and

C_{t-1}	C_t	$P(C_t\|C_{t-1})$
ok	ok	λ
ok	faulty	$1-\lambda$
faulty	ok	0

C_{t-1}	C_t	$P(C_t\|C_{t-1})$
faulty	faulty	1

Therefore,

C_0	C_1	$P(C_1\|C_0)$
ok	ok	λ
ok	faulty	$1-\lambda$
faulty	ok	0
faulty	faulty	0

Now, we eliminate C_0

C_1	$P(C_1)$
ok	λ
faulty	$1-\lambda$

We do the same computation to obtain $P(C2|C1)$.

C_1	C_2	$P(C_2\|C_1)$
ok	ok	λ^2
ok	faulty	$(1-\lambda)\lambda$
faulty	ok	0
faulty	faulty	$1-\lambda$

Now, we eliminate C_1

C_2	$P(C_2)$
ok	λ^2
faulty	$1-\lambda^2$

b. Using variable elimination, we can observe that each time transition from $t-1$ to t and elimination of variable C_{t-1} results in a multiplication of $P(C_{t-1} = ok)$ by λ. Therefore, we can use the following factor to represent $P(C_n)$

C_n	$P(C_n)$
ok	λ^n
faulty	$1-\lambda^n$

Question 3

Consider a Naive Bayes structure with edges $X \to Y_1, \ldots, X \to Y_n$.

a. What is the size of the largest factor of variable elimination order $Y_1, \dots, Y_n, X$?
b. What is the size of the largest factor of variable elimination order $X, \dots, Y_n$?

Answer

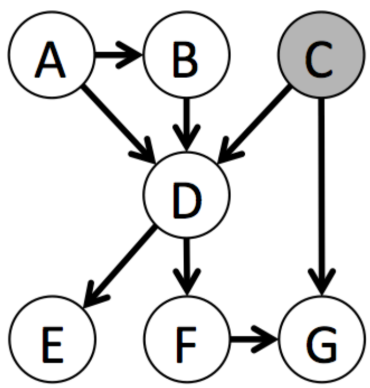
The Naive Bayes structure has the following factors: $\phi_{Y_1}(Y_1, X), \phi_{Y_2}(Y_2, X), \dots, \phi_{Y_n}(Y_n, X), \phi_X(X)$.

a. If we start eliminating Y_1, we will compute $\tau_1(X) = \sum_{Y_1} \phi_{Y_1}(Y_1, X)\phi_x(X)$, which is a factor of size $O(d)$, where d is the size of the outcome space of each variable. The other factors that involve variables Y_i will have the same size. After eliminating variables $Y_1, \dots, Y_n$, we end with factors $\tau_1(X), \dots \tau_n(X)$. The multiplication of these factors also results in a final factor of size $O(d)$.

b. If we start eliminating X, we will need to multiply all factors of the network since the variable X is present in all of them. This computation will result in a factor of size $O(d^n)$, which is exponential in the number of variables. Therefore, the first elimination order is more efficient than the second one.

Question 4

For the Bayesian network below, all variables are binary. Assume we run variable elimination to compute the answer to the query $P(A, E| + c)$, with the following elimination order: B, D, G, F.



a. What is the size of the largest computed factor?
b. Can the min-degree heuristic help to find an ordering that generates a smaller largest factor?

Answer

Initially, we have the following factors:

$\phi_A(A), \phi_B(B, A), \phi_C(C), \phi_D(D, A, B, C), \phi_E(E, D), \phi_F(F, D), \phi_G(G, F, C)$

We start by noting the evidence, so we will not consider it when computing the factor sizes:

$\phi_A(A), \phi_B(B, A), \phi_C(+c), \phi_D(D, A, B, +c), \phi_E(E, D), \phi_F(F, D), \phi_G(G, F, +c)$

We eliminate variable B. The elimination involves the factors: $\phi_B(B, A), \phi_D(D, A, B, +c)$, resulting in the new factor $\tau_1(D, A, +c)$. This new factor has two variables and therefore four entries.

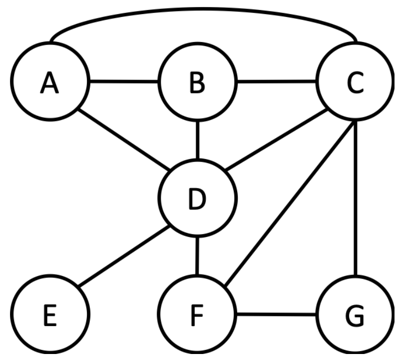
Now, we eliminate variable D. The elimination involves the factors: $\tau_1(D, A, +c), \phi_E(E, D), \phi_F(F, D)$, resulting in the new factor $\tau_2(A, +c, E, F)$ with three variables and eight entries.

Eliminating G involves only one factor $\phi_G(G, F, +c)$, creating the factor $\tau_3(F, +c)$.

Finally, the elimination of F includes the factors: $\tau_2(A, +c, E, F), \tau_3(F, +c)$ and creates the factor $\tau_4(A, +c, E)$ with two variables and four entries.

Therefore, the largest factor has three variables and eight entries.

Let us see whether min-degree can help to find a better ordering. In this case, the induced graph is the following:



We start with G since it has degree 2. The elimination of G involves a single factor $\phi_G(G, F, +c)$, resulting in a new factor $\tau_1(F, +c)$. This factor has a single variable and two entries.

Now, F has degree 2 and is the next to be eliminated. This involves the factors $\phi_F(F, D), \tau_1(F, +c)$ resulting in $\tau_2(+c, D)$ that also has a single variable and two entries.

The elimination of F reduces the degree of D to 4. However, B has degree 3 and is the following to be eliminated. The elimination has the factors $\phi_B(B, A), \phi_D(D, A, B, +c)$ and results on the factor $\tau_3(D, A, +c)$ that as two variables and four entries.

Finally, factor D is eliminated with the multiplication of $\tau_3(D, A, +c), \phi_E(E, D)$ creating a new factor $\tau_4(A, +c, E)$ that also has two variables and four entries.

In the end, the min-degree heuristic generated a smaller maximum factor of four entries compared to eight entries of the first elimination order.