Completed

Principles of Machine Learning

Exercises

Victor Verreet *victor.verreet@cs.kuleuven.be*
Laurens Devos *laurens.devos@cs.kuleuven.be*

Fall, 2021

Exercise Session 6: Graphical Models

6.1 Independence Rules

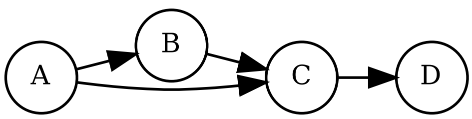
Let $A \perp B$ denote that A and B are independent stochastic variables. Prove that $A \perp (B, C)$ implies that $A \perp B$ and $A \perp C$. Also prove that $A \perp (B, C)$ implies that $A \perp B|C$.

6.2 Bayesian Car Trouble

Whether a car starts or not depends on whether there is fuel in the tank and whether the battery is dead or not. The battery is more likely to be dead in winter than at other times of the year. The car has a fuel meter indicating empty, half full or full. The fuel meter may be broken, and the tank may have a leak. If fuel is leaking from the tank, we may smell this. There have been recent reports about a fuel thief in the area. The car owner always fills up the tank on Saturday.

1. Provide the graph of a Bayesian network modeling this domain.
2. Which conditional probability tables (CPTs) need to be specified to turn your graph into a BN?
3. How many parameters are needed to specify the CPTs of your Bayesian network?
4. How many parameters would be needed to directly specify the full joint distribution over the random variables you used?

6.3 Inference in Bayesian Networks



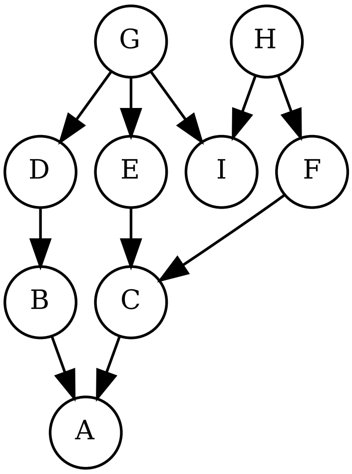
$P[B=0 \vert A]$	$A=0$	$A=1$
	0.6	0.3

$P[C=0 \vert A,B]$	$A=0$	$A=1$
$B=0$	0.8	0.1
$B=1$	0.2	0.8

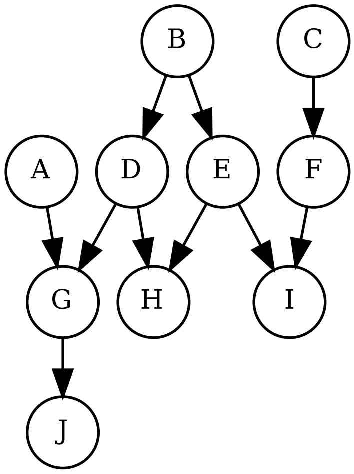
$P[D=0 \vert C]$	$C=0$	$C=1$
	0.4	0.5

Consider the Bayesian network of boolean variables and the conditional probability tables above. We know that $P[A=0] = 0.9$. Calculate the probability $P[D=1|B=0]$.

6.4 D-Separation in Bayesian Networks



(a)

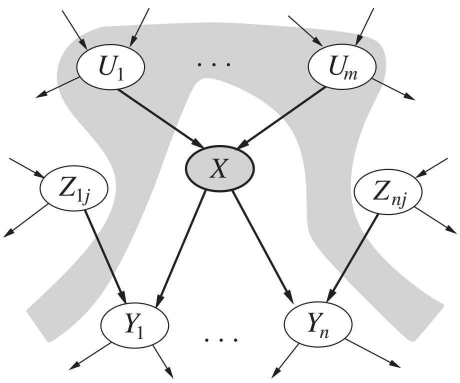


(b)

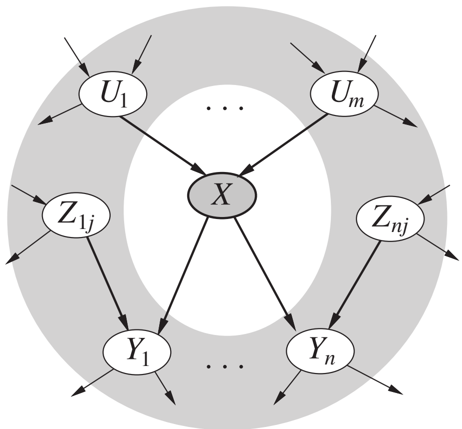
(a) Assume B is the only observed node. Which nodes are d-separated from A?

(b) Assume J is the only observed node. Which nodes are d-separated from A?

6.5 Independence in Bayesian Networks



(a)



(b)

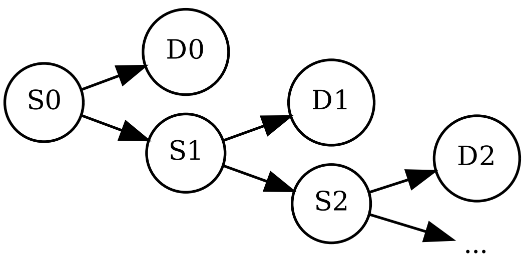
(figures: Russel & Norvig)

Show that each of the two characterizations below follows from the more general notion of d-separation.

(a) Each node is conditionally independent of its **non-descendants** given its **parents**.

(b) Each node is conditionally independent of **all other nodes** in the network given its **Markov blanket**.

6.6 Hidden Markov Models



A hidden Markov model is a special type of Bayesian network that takes on the structure as shown in the figure above. The variables S_n represent hidden states of some process and D_n are measurable quantities depending on the states S_n. First show for a set $I \subseteq \{0, 1, ..., n-1\}$ that $P[S_n|S_i \text{ for } i \in I] = P[S_n|S_{\max(I)}]$. Then show that $P[D_n|S_i \text{ for } i \in I] = P[D_n|S_{\max(I)}]$ using the previous result.

6.7 Markov Networks

Consider a Markov network over variables $X = \{X_i \,|\, i = 1, ..., n\}$ which has a set of maximal cliques C. For every maximal clique $c \in C$ there is a potential $\phi_c \in \mathbb{R}_0^+$ which takes as arguments the variables occurring in c. The network structure then defines a joint probability distribution over its n variables X_i as

$$P[X_1, ..., X_n] = \frac{1}{Z} \prod_{c \in C} \phi_c(X_i \in c)$$

with Z the normalization. If a set of nodes σ separates two variables A and B, in the sense that every path from A to B has to run through σ, then A and B are conditionally independent given σ. This is known as the Markov property. Show that the above factorization of the probability distribution implies the Markov property.

Hint: It is useful to split up the network into the region that can be reached from A and the region that can be reached from B. Then reason about the potentials in these separate regions.

6.1

$$P(A,B) = \sum_c P(A,B,c)$$

$$= \sum_c P(A)P(B,c)$$

$$= P(A)\sum_c P(B,c)$$

$$= P(A)P(B)$$

$$P(A,B|c) = \frac{P(A,B,c)}{P(c)} = \frac{P(A)\,P(B,c)}{P(c)}$$

$$= P(A|c)\cdot\frac{P(B,c)}{P(c)}$$

$$= P(A|c)P(B|c)$$

$A \perp B \mid c$

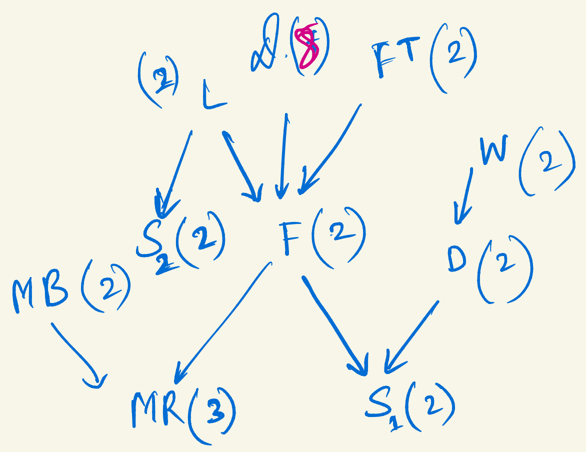
6.2

S → The car starts.

F → There's fuel in the tank.

D → The battery is dead.

W → Winter season.



$P(S|F,D)$, $P(MR|MB,F)$,

$P(S|L)$, $P(F|L,D,FT)$

$P(D|W)$, $P(L)$, $P(D)$, $P(FT)$, $P(W)$, $P(MB)$

$$2^2 + 2\cdot 2^2 + 2 + 2\cdot 2\cdot 8 + 1 + 1 + 7 + 1 + 2 + 1$$

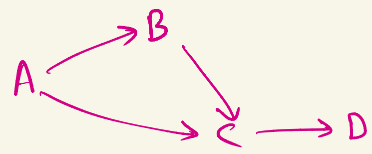
$$= 4+8+2+32+10+2+1 = 59$$

=

④

$$2^8\cdot 3\cdot 2^3 = 2^{11}\cdot 3$$

6.3



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

$$P(A=0) = 0.9$$

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

$$= \frac{\sum_{A,C} P(A)P(B|A)P(C|A,B)P(D|C)}{\sum_{A,C,D} P(A)P(B|A)P(C|A,B=0)P(D|C)}$$

$$= \frac{\sum_{A,C} P(A)P(B|A)P(C|A,B=0)P(D=1|C)}{\sum_A P(A)P(B=0|A)}$$

$$= \frac{\begin{array}{l} P(A=0)P(B=0|A=0)P(C=0|A=0,B=0)P(D=1|C=0) + \\ P(A=0)P(B=0|A=0)P(C=1|A=0,B=0)P(D=1|C=1) + \\ P(A=1)P(B=0|A=1)P(C=0|A=1,B=0)P(D=1|C=0) + \\ P(A=1)P(B=0|A=1)P(C=1|A=1,B=0)P(D=1|C=1) \end{array}}{P(A=0)P(B=0|A=0) + P(A=1)P(B=0|A=1)}$$

$$= \frac{0.9\times0.6\times0.8\times0.6 + 0.9\times0.6\times0.2\times0.5 + 0.1\times0.3\times0.1\times0.6 + 0.1\times0.3\times0.9\times0.5}{0.9\times0.6 + 0.1\times0.3}$$

$$= \frac{0.9\times0.6\times(0.48+0.1) + 0.1\times0.3\times(0.06+0.45)}{0.54+0.03}$$

$$= \frac{0.54\times0.58 + 0.03\times0.51}{0.57} = \frac{0.3285}{0.57}$$

$$= 0.576$$

6.4

X is d-separated from Y given Z if $\forall\, U(X,Y)$. if $\exists\, \omega$ is a collider then $\omega \notin Z$, $desc(\omega) \notin Z$

& if ω is a non-collider then $\omega \in Z$

a) $A_{d\text{-}sep} = \{\ \}$

b) $A_{d\text{-}sep} = \{\ \}$

6.5 $P(X, Z_{nj} \mid U_1, U_m)$

Consider all unblocked paths from x to Z_{nj}

(i) If the path goes through one of his parent, then the path's blocked.

(ii) If the path goes through one of the descendants, then $\exists$ a descendant of X, D_{n}. $+$ it's a collider & not observed.

So, D blocks the path.

b) let G_x be all the nodes.

$X \perp\!\!\!\perp g \mid M$

Every path from x to g will be

(i) Through one of its parents, then it's blocked

(ii) Through one of its children, then it's blocked.

(iii) Through the other parent of the children, then it will be blocked by the other parent.

$$P(S_n \mid S_i) = \frac{\sum_{\substack{0 \le K \le n-1 \\ K \notin I}} P(S_0, S_1, \ldots S_n)}{\sum_{\substack{0 \le K \le max(I) \\ K \notin I}} P(S_1, S_2 \ldots S_{max(I)})}$$

$$= \frac{\sum P(S_0) P(S_1|S_0) \ldots P(S_n|S_{n-1})}{\sum P(S_{i^*}|S_{i^*-1}) \ldots P(S_1|S_0) P(S_0)}$$

$$= \sum P(S_n|S_{n-1}) \ldots P(S_{i^*+1}|S_{i^*})$$

$$\frac{\sum P(S_n|S_{n-1}) \ldots P(S_{i^*+1}|S_{i^*}) \ldots P(S_1|S_0)}{P(S_i^*)}$$

$$= \frac{P(S_n, S_i^*)}{P(S_i^*)} = P(S_n / S_i^*)$$

6.7

$P(D_n \mid \cup S_i, i \in I)$

$$= \frac{\sum_{\substack{1 \le K \le n \\ K \notin I}} P(D_n, D_{n-1} \ldots D_0, S_n, S_{n-1} \ldots S_0)}{\sum_{1 \le K \le i^*-1} P(S_i)}$$

$$= \sum_{\substack{(K \mid 1 \le K \le n \\ K \notin I)}} P(D_n|S_n) \cdot P(D_{n-1}|S_{n-1}) \ldots P(D_{i^*}|S_{i^*})\, P(S_n|S_{n-1}) \ldots P(S_{i^*+1}|S_{i^*})$$

$$= P(D_n / S_{i^*})$$

6.7

$X = \{X_i \mid i = 1, \ldots n\}$

$A = \{X_i \mid i \text{ in } I_A, I_A \subseteq \{1, \ldots n\}\}$

$B = \{X_i \mid i \text{ in } I_B, I_B \subseteq \{1, \ldots n\} \setminus I_A\}$

$C = \{X_i \mid i \text{ in } I_C, I_C \subseteq \{1, \ldots n\} \setminus (I_A \cup I_B)\}$

$P(A,B|\sigma)$

$$= \frac{\sum_{i \notin I_A \cup I_B \cup I_c} P(\wedge x_i)}{\sum_{i \notin I_c} P(\wedge x_i)}$$

Since $I_A \cap I_B \cap I_c = \varphi$

$$= \frac{\sum_{i \notin I_A \cup I_B \cup I_c} \prod_{c \in C} \phi_c(x_i \in c)}{\sum_{i \notin I_c} \prod_{c \in C} \phi_c(x_i \in c)}$$

$$= \frac{\sum_{\substack{x_i \in c \\ i \in I_A}} \prod_{c \in C} \phi_c(x_i \in c) \cdot \sum_{\substack{x_i \in c \\ i \in I_B}} \prod_{c \in C} \phi_c(x_i \in c)}{}$$

A —————— σ —————— B

let α be the nodes to reach from A without crossing σ.

let β be the nodes that can be reached from B without crossing σ.

let γ be the nodes that isn't present in α, β, σ

$$P(x_1, \ldots x_n) = \frac{1}{Z} \phi_\alpha(\alpha,\sigma)\, \phi_\beta(\beta,\sigma)\, \phi_\gamma(\gamma,\sigma)\, \phi_\sigma(\sigma)$$

$$P(A,B|\sigma) = \frac{P(A,B,\sigma)}{P(\sigma)}$$

$$= \frac{\sum_{\alpha,\beta,\gamma \setminus \{A,B\}} \phi_\alpha(\alpha,\sigma)\, \phi_\beta(\beta,\sigma)\, \phi_\gamma(\gamma,\sigma)\, \phi_\sigma(\sigma)}{\sum_{\alpha,\beta,\gamma} \phi_\alpha(\alpha,\sigma)\, \phi_\beta(\beta,\sigma)\, \phi_\gamma(\gamma,\sigma)\, \phi_\sigma(\sigma)}$$

$$= \frac{\sum_{\alpha \setminus \{A\}} \phi_\alpha(\alpha,\sigma)}{\sum_{\alpha} \phi_\alpha(\alpha,\sigma)} \times \frac{\sum_{\beta \setminus \{B\}} \phi_\beta(\beta,\sigma)}{\sum_{\beta} \phi_\beta(\beta,\sigma)}$$

$$= F_\alpha(A,\sigma)\, F_\beta(B,\sigma)$$

$P(A|\sigma) \cdot P(B|\sigma)$

$$= \sum \frac{P(\alpha,\beta,\gamma,\sigma)}{P(\sigma)} \sum \frac{P(\alpha,\beta,\gamma,\sigma)}{P(\sigma)}$$

$$= \frac{\sum_{\alpha \setminus \{A\},\beta,\gamma} \phi(\alpha,\sigma)\, \phi(\beta,\sigma) \cdot \phi(\gamma,\sigma) \cdot \phi(\sigma)}{\phi(\alpha,\sigma)\, \phi(\beta,\sigma)\, \phi(\gamma,\sigma) \cdot \phi(\sigma)} \times$$

$$\frac{\sum_{\alpha,\beta \setminus \{B\},\gamma} \phi(\alpha,\sigma)\, \phi(\beta,\sigma)\, \phi(\gamma,\sigma)\, \phi(\sigma)}{\sum_{\alpha,\beta,\gamma} \phi(\alpha,\sigma)\, \phi(\beta,\sigma)\, \phi(\gamma,\sigma)\, \phi(\sigma)}$$

$$= \frac{\sum_{\alpha \setminus \{A\}} \phi(\alpha,\sigma)}{\phi(\alpha,\sigma)} \times \frac{\sum_{\beta \setminus \{B\}} \phi(\beta,\sigma)}{\phi(\beta,\sigma)}$$

$$= F_\alpha(A,\sigma)\, F_\beta(B,\sigma)$$

T.1