دانشکده مهندسی کامپیوتر

هوش مصنوعی و سیستم‌های خبره

تمرین تشریحی ششم[1]

نام و نام خانوادگی - شماره دانشجویی . . . .

مدرس . . . . محمدطاهر پیله‌ور - سید صالح اعتمادی

طراحی و تدوین . . . . سارا کدیری - غزاله محمودی - پریسا یل سوار

تاریخ انتشار . . . . ۱۶ آذر ۱۳۹۹

تاریخ تحویل . . . . ۲۵ اذر ۱۳۹۹

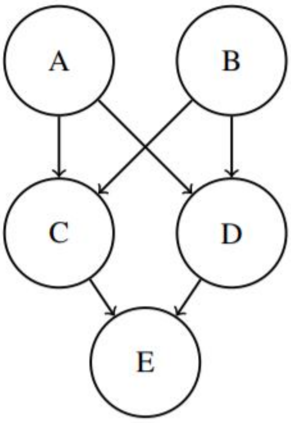
---

[1] در طراحی این تمرین از منابع کورس CS188 دانشگاه برکلی استفاده شده است.

۱ Bayes' Nets

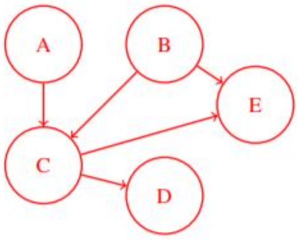
۱.۱ Bayes' Nets and Joint Distributions

Write down the joint probability distribution associated with the following Bayes Net. Express the answer as a product of terms representing individual conditional probabilities tables associated with this Bayes Net:



Draw the Bayes net associated with the following joint distribution:

P(A).P(B).P(C|A,B).P(D|C).P(E|B,C)



Do the following products of factors correspond to a valid joint distribution over the variables A, B, C, D? (Circle FALSE or TRUE.)

True ● False P(A).P(B).P(C|A).P(C|B).P(D|E)

●True False P(A).P(B|A).P(C).P(D|B, C)

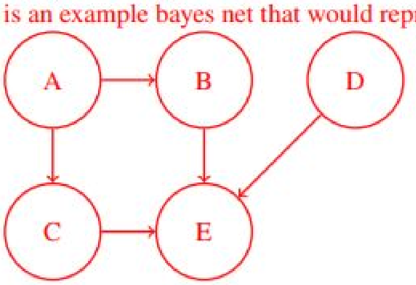
True ● False P(A).P(B|A).P(C).P(C|A).P(D)

True ● False P(A|B).P(B|C).P(C|D).P(D|A)

What factor can be multiplied with the following factors to form a valid joint distribution? (Write "none" if the given set of factors can't be turned into a joint by the inclusion of exactly one more factor.)

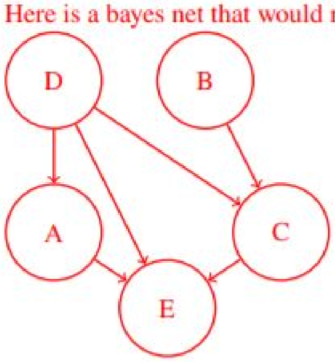
P(A).P(B|A).P(C|A).P(E|B, C, D) P(D).P(B).P(C|D, B).P(E|C, D, A)

P(D) is missing. D could also be conditioned on A,B, and/or C without creating a cycle (e.g. $P(D|A,B,C)$). Here is an example bayes net that would represent the distribution after adding in $P(D)$:



P(D).P(B).P(C|D, B).P(E|C, D, A)

P(A) is missing to form a valid joint distributions. A could also be conditioned on B, C, and/or D (e.g. $P(A|B,C,D)$. Here is a bayes net that would represent the distribution is $P(A|D)$ was added in.

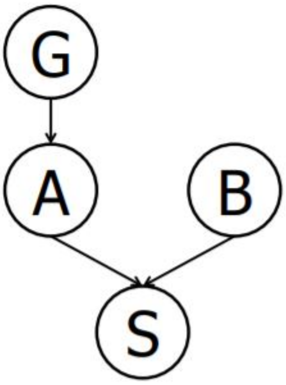


۲.۱ Bayes' Nets Representation and Probability

Suppose that a patient can have a symptom (S) that can be caused by two different diseases (A and B). It is known that the variation of gene G plays a big role in the manifestation of disease A. The Bayes' Net and corresponding conditional probability tables for this situation are shown below. For each part, you may leave your answer as an arithmetic expression.

$P(G)$	
$+g$	0.1
$-g$	0.9

$P(A\|G)$		
$+g$	$+a$	1.0
$+g$	$-a$	0.0
$-g$	$+a$	0.1
$-g$	$-a$	0.9



$P(B)$	
$+b$	0.4
$-b$	0.6

$P(S\|A,B)$			
$+a$	$+b$	$+s$	1.0
$+a$	$+b$	$-s$	0.0
$+a$	$-b$	$+s$	0.9
$+a$	$-b$	$-s$	0.1
$-a$	$+b$	$+s$	0.8
$-a$	$+b$	$-s$	0.2
$-a$	$-b$	$+s$	0.1
$-a$	$-b$	$-s$	0.9

Compute the following entry from the joint distribution:

$$P(+g)P(+a|+g)P(+b)P(+s|+b,+a) = (0.1)(1.0)(0.4)(1.0) = 0.04$$

What is the probability that a patient has disease A?

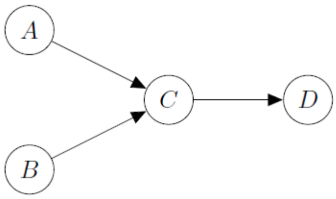
$$P(+a|+g)P(+g) + P(+a|-g)P(-g) = (1.0)(0.1) + (0.1)(0.9) = 0.19$$

What is the probability that a patient has disease A given that they have disease B?

$P(+a|+b) = P(+a) = 0.19$ The first equality holds true as we have $A \perp\!\!\!\perp B$, which can be inferred from the graph of the Bayes' net.

۲ Bayes' Net Sampling

Assume you are given the following Bayes' net and the corresponding distributions over the variables in the Bayes' net.



$P(A)$	
+a	0.1
-a	0.9

$P(B)$	
+b	.7
-b	.3

$P(C\|A,B)$			
+c	+a	+b	.25
-c	+a	+b	.75
+c	-a	+b	.6
-c	-a	+b	.4
+c	+a	-b	.5
-c	+a	-b	.5
+c	-a	-b	.2
-c	-a	-b	.8

$P(D\|C)$		
+d	+c	.5
-d	+c	.5
+d	-c	.8
-d	-c	.2

2.1 Assume we receive evidence that A = +a. If we were to draw samples using rejection sampling, on expectation what percentage of the samples will be rejected?

Since $P(+a) = \frac{1}{10}$, we would expect that only 10% of the samples could be saved. Therefore, expected 90% of the samples will be rejected.

2.2 Next, assume we observed both A = +a and D = +d. What are the weights for the following samples under likelihood weighting sampling?

Sample	Weight
$(+a, -b, +c, +d)$	$P(+a) \cdot P(+d\| + c) = 0.1 * 0.5 = 0.05$
$(+a, -b, -c, +d)$	$P(+a) \cdot P(+d\| - c) = 0.1 * 0.8 = 0.08$
$(+a, +b, -c, +d)$	$P(+a) \cdot P(+d\| - c) = 0.1 * 0.8 = 0.08$

2.3 Given the samples in the previous question, estimate P(-b | +a, +d).

$$P(-b|+a,+d) = \frac{P(+a)\cdot P(+d|+c) + P(+a)\cdot P(+d|-c)}{P(+a)\cdot P(+d|+c) + 2\cdot P(+a)\cdot P(+d|-c)} = \frac{0.05+0.08}{0.05+2\cdot 0.08} = \frac{13}{21}$$

2.4 Assume we need to (approximately) answer two different inference queries for this graph: P(C | + a) and P(C | + d). You are required to answer one query using likelihood weighting and one query using Gibbs sampling. In each case you can only collect a relatively small amount of samples, so for maximal accuracy you need to make sure you cleverly assign algorithm to query based on how well the algorithm fits the query. Which query would you answer with each algorithm? Justify your answer.

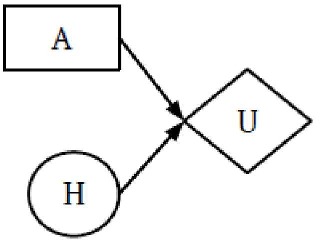
Algorithm	Query
Likelihood Weighting	$P(C|+a)$

Algorithm	Query
Gibbs Sampling	$P(C|+d)$

Justify your answer:
You should use Gibbs sampling to find the query answer $P(C|+d)$. This is because likelihood weighting only takes upstream evidence into account when sampling. Therefore, Gibbs, which utilizes both upstream and downstream evidence, is more suited to the query $P(C|+d)$ which has downstream evidence.

۳ Decision Networks

After years of battles between the ghosts and Pacman, the ghosts challenge Pacman to a winner-take-all showdown, and the game is a coin flip. Pacman has a decision to make: whether to accept the challenge *(accept)* or decline *(decline)*. If the coin comes out heads $(+h)$ Pacman wins. If the coin comes out tails $(-h)$, the ghosts win. No matter what decision Pacman makes, the outcome of the coin is revealed.



H	$P(H)$
+h	0.5
-h	0.5

H	A	U(H,A)
+h	*accept*	100
-h	*accept*	-100
+h	*decline*	-30
-h	*decline*	50

3.1 Maximum Expected Utility

Compute the following quantities:

$EU(accept) = P(+h)U(+h, accept) + P(-h)U(-h, accept) = 0.5 * 100 + 0.5 * -100 = 0$

$EU(decline) = P(+h)U(+h, decline) + P(-h)U(-h, decline) = 0.5 * -30 + 0.5 * 50 = 10$

$MEU(\{\}) = max(0, 10) = 10$

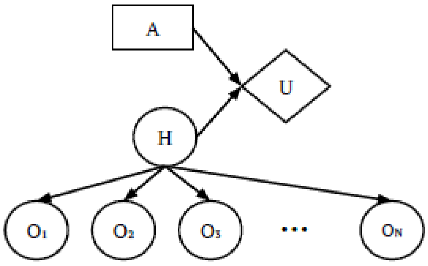
Action that achieves $MEU(\{\}) =$ *decline*

3.2 VPI relationships

When deciding whether to accept the winner-take-all coin flip, Pacman can consult a few fortune tellers that he knows. There are N fortune tellers, and each one provides a prediction O_n for H. For each of the questions below, select **all** of the VPI relations that are guaranteed to be true, or select *None of the above.*

3.2.1 **In this situation, the fortune tellers give perfect predictions. Specifically, $P(O_n = +h|H = +h) = 1$, $P(O_n = -h|H = -h) = 1$ for all n from 1 to N.**

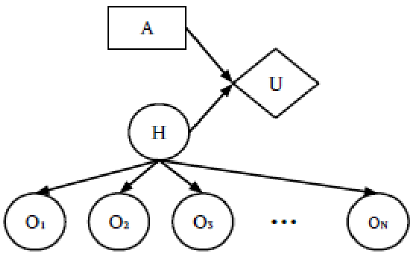
In this situation, the fortune tellers give perfect predictions.
Specifically, $P(O_n = +h \mid H = +h) = 1$, $P(O_n = -h \mid H = -h) = 1$, for all n from 1 to N.



- [ ] $\text{VPI}(O_1,\ O_2) \geq \text{VPI}(O_1) + \text{VPI}(O_2)$
- [x] $\text{VPI}(O_i) = \text{VPI}(O_j)$ where $i \neq j$
- [ ] $\text{VPI}(O_3 \mid O_2,\ O_1) > \text{VPI}(O_2 \mid O_1)$.
- [ ] $\text{VPI}(H) > \text{VPI}(O_1, O_2, \ldots O_N)$
- [ ] None of the above.

3.2.2 **In another situation, the fortune tellers are pretty good, but not perfect. Specifically, $P(O_n = +h|H = +h) = 0.8$, $P(O_n = -h|H = -h) = 0.5$ for all n from 1 to N.**

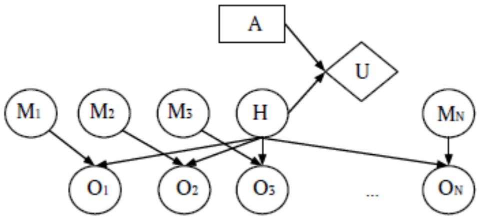
In another situation, the fortune tellers are pretty good, but not perfect.
Specifically, $P(O_n = +h \mid H = +h) = 0.8$, $P(O_n = -h \mid H = -h) = 0.5$, for all n from 1 to N.



- [ ] $\text{VPI}(O_1,\ O_2) \geq \text{VPI}(O_1) + \text{VPI}(O_2)$
- [x] $\text{VPI}(O_i) = \text{VPI}(O_j)$ where $i \neq j$
- [ ] $\text{VPI}(O_3 \mid O_2,\ O_1) > \text{VPI}(O_2 \mid O_1)$.
- [x] $\text{VPI}(H) > \text{VPI}(O_1, O_2, \ldots O_N)$
- [ ] None of the above.

3.2.3 **In a third situation, each fortune teller's prediction is affected by their mood. If the fortune teller is in a good mood $(+m)$, then that fortune teller's prediction is guaranteed to be correct. If the fortune teller is in a bad mood $(-m)$ then that teller's prediction is guaranteed to be incorrect. Each fortune teller is happy with probability $P(M_n = +m) = 0.8$.**

In a third situation, each fortune teller's prediction is affected by their mood. If the fortune teller is in a good mood ($+m$), then that fortune teller's prediction is guaranteed to be correct. If the fortune teller is in a bad mood ($-m$), then that teller's prediction is guaranteed to be incorrect. Each fortune teller is happy with probability $P(M_n = +m) = 0.8$.

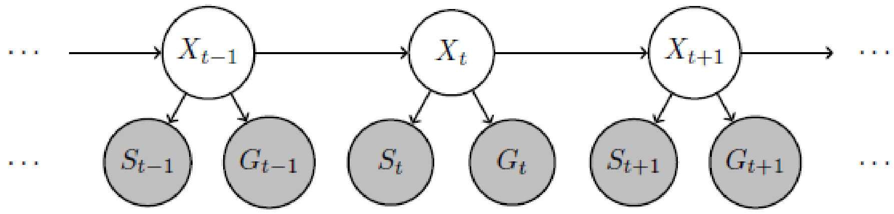


- ☐ $\text{VPI}(M_1) > 0$
- ■ $\forall i\ \text{VPI}(M_i|O_i) > 0$
- ☐ $\text{VPI}(M_1,\ M_2,\ \ldots,\ M_N) > \text{VPI}(M_1)$
- ■ $\forall i\ \text{VPI}(H) = \text{VPI}(M_i, O_i)$
- ☐ None of the above.

۴ HMM: Where is the Car?

Transportation researchers are trying to improve traffic in the city but, in order to do that, they first need to estimate the location of each of the cars in the city. They need our help to model this problem as an inference problem of an HMM. For this question, assume that only *one* car is being modeled.

The structure of this modified HMM is given below, which includes X, the location of the car; S, the noisy location of the car from the signal strength at a nearby cell phone tower; and G, the noisy location of the car from GPS.



We want to perform filtering with this HMM. That is, we want to compute the belief $P(x_t|s_{1:t}, g_{1:t})$, the probability of a state x_t given all past and current observations.

The **dynamics update** expression has the following form:

$$P(x_t|s_{1:t-1}, g_{1:t-1}) = \underline{\quad\text{(i)}\quad}\ \underline{\quad\text{(ii)}\quad}\ \underline{\quad\text{(iii)}\quad}\ P(x_{t-1}|s_{1:t-1}, g_{1:t-1}).$$

Complete the expression by choosing the option that fills in each blank.

Complete the expression by choosing the option that fills in each blank.

(i)	○ $P(s_{1:t}, g_{1:t})$	○ $P(s_{1:t-1}, g_{1:t-1})$	○ $P(s_{1:t-1})P(g_{1:t-1})$	○ $P(s_{1:t})P(g_{1:t})$	● 1
(ii)	○ $\sum_{x_t}$	● $\sum_{x_{t-1}}$	○ $\max_{x_{t-1}}$	○ $\max_{x_t}$	○ 1
(iii)	○ $P(x_{t-1} \mid x_{t-2})$	○ $P(x_{t-2}, x_{t-1})$	○ $P(x_{t-1}, x_t)$	● $P(x_t|x_{t-1})$	○ 1

The derivation of the dynamics update is similar to the one for the canonical HMM, but with two observation variables instead.

$$\begin{aligned} P(x_t|s_{1:t-1}, g_{1:t-1}) &= \sum_{x_{t-1}} P(x_{t-1}, x_t|s_{1:t-1}, g_{1:t-1}) \\ &= \sum_{x_{t-1}} P(x_t|x_{t-1}, s_{1:t-1}, g_{1:t-1})P(x_{t-1}|s_{1:t-1}, g_{1:t-1}) \\ &= \sum_{x_{t-1}} P(x_t|x_{t-1})P(x_{t-1}, x_t|s_{1:t-1}, g_{1:t-1}) \end{aligned}$$

In the last step, we use the independence assumption given in the HMM, $X_t \perp\!\!\!\perp S_{1:t-1}, G_{1:t-1}|X_{t-1}$.

The **observation update** expression has the following form:

$$P(x_t|s_{1:t}, g_{1:t}) = \underline{\quad \textbf{(iv)} \quad} \; \underline{\quad \textbf{(v)} \quad} \; \underline{\quad \textbf{(vi)} \quad} \; P(x_t|s_{1:t-1}, g_{1:t-1}).$$

Complete the expression by choosing the option that fills in each blank.

(iv)
- ○ $P(s_{1:t-1}|s_t)P(g_{1:t-1}|g_t)$
- ● $\frac{1}{P(s_t, g_t|s_{1:t-1}, g_{1:t-1})}$
- ○ $\frac{1}{P(s_{1:t-1}, g_{1:t-1}|s_t, g_t)}$
- ○ $P(s_t, g_t|s_{1:t-1}, g_{1:t-1})$
- ○ $P(s_{1:t-1}, g_{1:t-1}|s_t, g_t)$
- ○ $P(s_t|s_{1:t-1})P(g_t|g_{1:t-1})$
- ○ $\frac{1}{P(s_t|s_{1:t-1})P(g_t|g_{1:t-1})}$
- ○ $\frac{1}{P(s_{1:t-1}|s_t)P(g_{1:t-1}|g_t)}$
- ○ 1

(v)
- ○ $\sum_{x_t}$
- ○ $\sum_{x_{t-1}}$
- ○ $\max_{x_t}$
- ○ $\max_{x_{t-1}}$
- ● 1

(vi)
- ○ $P(x_{t-1}, s_{t-1})P(x_{t-1}, g_{t-1})$
- ○ $P(x_{t-1}, s_{t-1}, g_{t-1})$
- ○ $P(x_t|s_t)P(x_t|g_t)$
- ○ $P(s_{t-1}|x_{t-1})P(g_{t-1}|x_{t-1})$
- ○ $P(x_t, s_t)P(x_t, g_t)$
- ○ $P(x_t, s_t, g_t)$
- ○ $P(x_{t-1}|s_{t-1})P(x_{t-1}|g_{t-1})$
- ● $P(s_t|x_t)P(g_t|x_t)$
- ○ 1

Again, the derivation of the observation update is similar to the one for the canonical HMM, but with two observation variables instead.

$$\begin{aligned}
P(x_t|s_{1:t}, g_{1:t}) &= P(x_t|s_t, g_t, s_{1:t-1}, g_{1:t-1}) \\
&= \frac{1}{P(s_t, g_t|s_{1:t-1}, g_{1:t-1})} P(x_t, s_t, g_t|s_{1:t-1}, g_{1:t-1}) \\
&= \frac{1}{P(s_t, g_t|s_{1:t-1}, g_{1:t-1})} P(s_t, g_t|x_t, s_{1:t-1}, g_{1:t-1}) P(x_t|s_{1:t-1}, g_{1:t-1}) \\
&= \frac{1}{P(s_t, g_t|s_{1:t-1}, g_{1:t-1})} P(s_t, g_t|x_t) P(x_t|s_{1:t-1}, g_{1:t-1}) \\
&= \frac{1}{P(s_t, g_t|s_{1:t-1}, g_{1:t-1})} P(s_t|x_t) P(g_t|x_t) P(x_t|s_{1:t-1}, g_{1:t-1})
\end{aligned}$$

In the second to last step, we use the independence assumption $S_t, G_t \perp\!\!\!\perp S_{1:t-1}, G_{1:t-1}|X_t$; and in the last step, we use the independence assumption $S_t \perp\!\!\!\perp G_t|X_t$.