Variable Elimination (Koller)

José Eduardo Ochoa Luna
Dr. Ciencias - Universidade de São Paulo

Maestría C.C. Universidad Católica San Pablo
Sistemas Inteligentes

29 de noviembre 2018

Distributions

Joint Distribution

- Intelligence (I) ← 2
  - i^0 (low), i^1 (high),
- Difficulty (D) ← 2
  - d^0 (easy), d^1 (hard)
- Grade (G) ← 3
  - g^1 (A), g^2 (B), g^3 (C)

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

Joint Distribution

- Intelligence (I) ← 2
  - i^0 (low), i^1 (high),
- Difficulty (D) ← 2
  - d^0 (easy), d^1 (hard)
- Grade (G) ← 3
  - g^1 (A), g^2 (B), g^3 (C)

parameters
$2 \times 2 \times 3 = 12$

independent params
11

$P(I, D, G)$

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

1

Conditioning

condition on g^1

I	D	G	Prob.
i^0	d^0	g^1	0.126
~~i^0~~	~~d^0~~	~~g^2~~	~~0.168~~
~~i^0~~	~~d^0~~	~~g^3~~	~~0.126~~
i^0	d^1	g^1	0.009
~~i^0~~	~~d^1~~	~~g^2~~	~~0.045~~
~~i^0~~	~~d^1~~	~~g^3~~	~~0.126~~
i^1	d^0	g^1	0.252
~~i^1~~	~~d^0~~	~~g^2~~	~~0.0224~~
~~i^1~~	~~d^0~~	~~g^3~~	~~0.0056~~
i^1	d^1	g^1	0.06
~~i^1~~	~~d^1~~	~~g^2~~	~~0.036~~
i^1	d^1	g^3	0.024

Conditioning: Reduction

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^1	g^1	0.009
i^1	d^0	g^1	0.252
i^1	d^1	g^1	0.06

Conditioning: Renormalization

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^1	g^1	0.009
i^1	d^0	g^1	0.252
i^1	d^1	g^1	0.06

$P(I, D, g^1)$

unnormalized measure

Conditioning: Renormalization

I	D	G	Prob.
i^0	d^0	g^1	0.126 /0.447
i^0	d^1	g^1	0.009
i^1	d^0	g^1	0.252
i^1	d^1	g^1	0.06

$P(I, D, g^1)$ — 0.447

unnormalized meas

I	D	Prob.
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

$P(I, D \mid g^1)$

Marginalization

Marginalize I

P(I, D)

I	D	Prob.
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

D	Prob.
d^0	0.846
d^1	0.154

Factors

Factors

- A factor $\phi(\underline{X_1},\ldots,\underline{X_k})$

  $\phi : \underbrace{\mathrm{Val}(X_1,\ldots,X_k)} \to \mathrm{R}$

- Scope = $\{X_1,\ldots,X_k\}$

Joint Distribution

$P(I,D,G)$

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

Unnormalized measure $P(I,D,g^1)$

$P(I,D,g^1)$

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^1	g^1	0.009
i^1	d^0	g^1	0.252
i^1	d^1	g^1	0.06

Unnormalized measure $P(I,D,g^1)$

$P(I,D,g^1)$

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^1	g^1	0.009
i^1	d^0	g^1	0.252
i^1	d^1	g^1	0.06

Scope = I, D

Conditional Probability Distribution (CPD)

$P(G \mid I,D)$

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

Conditional Probability Distribution (CPD)

$P(G \mid I, D)$

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2
	A	B	C

Conditional Probability Distribution (CPD)

$P(G \mid I,D)$

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2
	A	B	C

context

General factors

A	B	ϕ
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

General factors

A	B	ϕ
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

Scope = $\{A, B\}$

Factor Product

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

Daph

Factor Product

$\phi_1(A, B)$

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

$\phi_2(B, C)$

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

A,B,C

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

Da

Factor Marginalization

Scope A,B,C

a^1	b^1	c^1	0.25
a^1	b^1	c^2	0.35
a^1	b^2	c^1	0.08
a^1	b^2	c^2	0.16
a^2	b^1	c^1	0.05
a^2	b^1	c^2	0.07
a^2	b^2	c^1	0
a^2	b^2	c^2	0
a^3	b^1	c^1	0.15
a^3	b^1	c^2	0.21
a^3	b^2	c^1	0.09
a^3	b^2	c^2	0.18

A,C

a^1	c^1	0.33
a^1	c^2	0.51
a^2	c^1	0.05
a^2	c^2	0.07
a^3	c^1	0.24
a^3	c^2	0.39

Factor Marginalization

Scope A,B,C

a^1	b^1	c^1	0.25
a^1	b^1	c^2	0.35
a^1	b^2	c^1	0.08
a^1	b^2	c^2	0.16
a^2	b^1	c^1	0.05
a^2	b^1	c^2	0.07
a^2	b^2	c^1	0
a^2	b^2	c^2	0
a^3	b^1	c^1	0.15
a^3	b^1	c^2	0.21
a^3	b^2	c^1	0.09
a^3	b^2	c^2	0.18

A,C

a^1	c^1	0.33
a^1	c^2	0.51
a^2	c^1	0.05
a^2	c^2	0.07
a^3	c^1	0.24
a^3	c^2	0.39

Factor Reduction

a^1	b^1	c^1	0.25
a^1	b^1	c^2	0.35
a^1	b^2	c^1	0.08
a^1	b^2	c^2	0.16
a^2	b^1	c^1	0.05
a^2	b^1	c^2	0.07
a^2	b^2	c^1	0
a^2	b^2	c^2	0
a^3	b^1	c^1	0.15
a^3	b^1	c^2	0.21
a^3	b^2	c^1	0.09
a^3	b^2	c^2	0.18

Factor Reduction

a^1	b^1	c^1	0.25
a^1	b^1	c^2	0.35
a^1	b^2	c^1	0.08
a^1	b^2	c^2	0.16
a^2	b^1	c^1	0.05
a^2	b^1	c^2	0.07
a^2	b^2	c^1	0
a^2	b^2	c^2	0
a^3	b^1	c^1	0.15
a^3	b^1	c^2	0.21
a^3	b^2	c^1	0.09
a^3	b^2	c^2	0.18

$\Rightarrow$

a^1	b^1	c^1	0.25
a^1	b^2	c^1	0.08
a^2	b^1	c^1	0.05
a^2	b^2	c^1	0
a^3	b^1	c^1	0.15
a^3	b^2	c^1	0.09

A, B

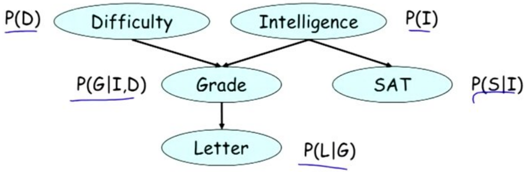
Why Factors

- Fundamental building block for defining distributions in high-dimensional spaces

Why Factors

- Fundamental building block for defining distributions in high-dimensional spaces
- Set of basic operations for manipulating these probability distributions

Chain Rule for Bayesian Networks



$$P(D,I,G,S,L) = P(D)\,P(I)\,P(G|I,D)\,P(S|I)\,P(L|G)$$

Distribution defined as a product of factors!

Inference in Bayesian Networks

- Given a joint probability distribution over $X = X_1, X_2, \ldots, X_n$

Inference in Bayesian Networks

- Given a joint probability distribution over $X = X_1, X_2, \ldots, X_n$
- Inference is $P(Y|Z)$, where $Y \subset X$ are query variables, $Z \subset X$ are evidence variables

Inference in Bayesian Networks

- Given a joint probability distribution over $X = X_1, X_2, \ldots, X_n$
- Inference is $P(Y|Z)$, where $Y \subset X$ are query variables, $Z \subset X$ are evidence variables
- The other variables H are called hidden variables

Inference in Bayesian Networks

- Given a joint probability distribution over $X = X_1, X_2, \ldots, X_n$
- Inference is $P(Y|Z)$, where $Y \subset X$ are query variables, $Z \subset X$ are evidence variables
- The other variables H are called hidden variables
- $X = Y \cup Z \cup H$

Inference in Bayesian Networks II

- Inference is performed by summing out hidden variables from joint probability distribution

Inference in Bayesian Networks II

- Inference is performed by summing out hidden variables from joint probability distribution
- $P(Y|Z) = \frac{P(Y,Z)}{P(Z)}$

Inference in Bayesian Networks II

- Inference is performed by summing out hidden variables from joint probability distribution
- $P(Y|Z) = \frac{P(Y,Z)}{P(Z)}$
- $P(Y,Z) = \sum_H P(Y,Z,H)$

Inference in Bayesian Networks II

- Inference is performed by summing out hidden variables from joint probability distribution
- $P(Y|Z) = \frac{P(Y,Z)}{P(Z)}$
- $P(Y, Z) = \sum_H P(Y, Z, H)$
- The full joint has 2^n entries

Inference in Bayesian Networks II

- Inference is performed by summing out hidden variables from joint probability distribution
- $P(Y|Z) = \frac{P(Y,Z)}{P(Z)}$
- $P(Y,Z) = \sum_H P(Y,Z,H)$
- The full joint has 2^n entries
- For large networks it is unreasonable to compute these many values

Normalizing

- Given A and B, and we want to compute $P(A|b)$, b denotes $B=$true.

Normalizing

- Given A and B, and we want to compute $P(A|b)$, b denotes $B =$true.
- $P(A|b) = P(A, b)/P(b)$, and $P(b) = P(a, b) + P(\neg a, b)$

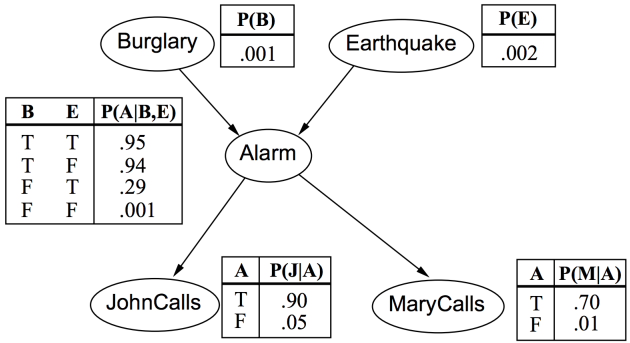
Normalizing

- Given A and B, and we want to compute $P(A|b)$, b denotes B =true.
- $P(A|b) = P(A, b)/P(b)$, and $P(b) = P(a, b) + P(\neg a, b)$
- Normalizing: $P(A|b) = \alpha P(A, b)$,

Normalizing

- Given A and B, and we want to compute $P(A|b)$, b denotes $B =$true.
- $P(A|b) = P(A, b)/P(b)$, and $P(b) = P(a, b) + P(\neg a, b)$
- Normalizing: $P(A|b) = \alpha P(A, b)$,
- α normalizing constant that makes the values in table $P(A|b)$ sum to 1.

Burglar Alarm Network



Enumeration

- Goal: $P(E|j,m) = \alpha P(E,j,m)$

Enumeration

- Goal: $P(E|j,m) = \alpha P(E,j,m)$
- Sum over hidden vars: $\alpha P(E,j,m) = \sum_a \sum_b P(E,j,m,b,a)$,

Enumeration

- Goal: $P(E|j,m) = \alpha P(E,j,m)$
- Sum over hidden vars: $\alpha P(E,j,m) = \sum_a \sum_b P(E,j,m,b,a)$,
- j,m means JonhCalls = true, MaryCalls = True

Enumeration

- Goal: $P(E|j,m) = \alpha P(E,j,m)$
- Sum over hidden vars: $\alpha P(E,j,m) = \sum_a \sum_b P(E,j,m,b,a)$,
- j, m means JonhCalls $=$ true, MaryCalls $=$ True
- $\sum_a \sum_b P(E,j,m,b,a) =$
  $\sum_b \sum_a P(b)P(E)P(a|b,E)P(j|a)P(m|a)$

Enumeration II

Save computations by pushing the $\sum$'s inward as much as possible:

$$\sum_b \sum_a P(b)P(E)P(a|b,E)P(j|a)P(m|a) = P(E)\sum_b P(b)\sum_a P(a|b,E)P(j|a)P(m|a)$$

We need to do this sum both for $E = true$ and $E = false$. For the first case:

$$P(e)\left(\begin{array}{c} P(b)\left(P(a|b,e)P(j|a)P(m|a) + P(\neg a|b,e)P(j|\neg a)P(m|\neg a)\right) + \\ P(\neg b)\left(P(a|\neg b,e)P(j|a)P(m|a) + P(\neg a|\neg b,e)P(j\neg a)P(m|\neg a)\right) \end{array}\right)$$

Variable Elimination

we will have one factor for each CPT term in our expression:

$$\alpha \underbrace{P(E)}_{E} \sum_b \underbrace{P(b)}_{B} \sum_a \underbrace{P(a|b,E)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

E	$f_E(E)$
T	.002
F	.998

B	$f_B(B)$
T	.001
F	.999

A	B	E	$f_A(A,B,E)$
T	T	T	.95
T	T	F	.94
T	F	T	.29
T	F	F	.001
F	T	T	.05
F	T	F	.06
F	F	T	.71
F	F	F	.999

A	$f_J(A)$
T	.9
F	.05

A	$f_M(A)$
T	.7
F	.01

Variable Elimination

Consider the product $f_{JM}(A) = f_J(A) f_M(A)$:

A	$f_{JM}(A)$
T	.9 * .7
F	.05 * .01

=

A	$f_J(A)$
T	.9
F	.05

A	$f_M(A)$
T	.7
F	.01

Variable Elimination

Consider the product $f_{AJM}(A, B, E) = f_A(A, B, E) f_{JM}(A)$:

A	B	E	$f_{AJM}(A, B, E)$
T	T	T	.95 * .63
T	T	F	.94 * .63
T	F	T	.29 * .63
T	F	F	.001 * .63
F	T	T	.05 * .0005
F	T	F	.06 * .0005
F	F	T	.71 * .0005
F	F	F	.999 * .0005

=

A	B	E	$f_{AJM}(A, B, E)$
T	T	T	.95
T	T	F	.94
T	F	T	.29
T	F	F	.001
F	T	T	.05
F	T	F	.06
F	F	T	.71
F	F	F	.999

A	$f_{JM}(A)$
T	.63
F	.0005

The entry for $A = T, B = T, E = T$ in $f_{AJM}(A, B, E)$ is the product of corresponding entries from $f_A(A, B, E)$ where $A = T, B = T, E = T$ and from $f_J M(A)$ where $A = T$ (since B, E do not appear in $f_J M(A)$. In general, we have

$$f(X_1, \ldots, X_j, Y_1, \ldots, Y_k, Z_1 \ldots, Z_l) = f_1(X_1, \ldots, X_j, Y_1, \ldots, Y_k) f_2(Y_1, \ldots, Y_k, Z_1, \ldots, Z_l)$$

Variable Elimination

Summing out A from $f_{AJM}(A, B, E)$, the resulting factor is:

B	E	$f_{\bar{A}JM}(B, E)$
T	T	$.95 * .63 + .05 * .0005 = .5985$
T	F	$.94 * .63 + .06 * .0005 = .5922$
F	T	$.29 * .63 + .71 * .0005 = .1830$
F	F	$.001 * .63 + .999 * .0005 = .001129$

Variable Elimination

To compute the answer to our query:

$$
\begin{aligned}
&\alpha P(E) \textstyle\sum_b P(b) \sum_a P(a|b,e) P(j|a) P(m|a) \\
&\quad = \alpha f_E(E) \textstyle\sum_B f_B(B) \sum_a f_A(A,B,E) f_J(A) f_m(A) \\
&\quad = \alpha f_E(E) \textstyle\sum_B f_B(B) \sum_a f_A(A,B,E) f_{JM}(A) \\
&\quad = \alpha f_E(E) \textstyle\sum_B f_B(B) \sum_a f_{AJM}(A,B,E) \\
&\quad = \alpha f_E(E) \textstyle\sum_B f_B(B) f_{\bar{A}JM}(B,E) \\
&\quad = \alpha f_E(E) \textstyle\sum_B f_{B\bar{A}JM}(B,E) \\
&\quad = \alpha f_E(E) f_{\bar{B}\bar{A}JM}(E) \\
&\quad = \alpha f_{E\bar{B}\bar{A}JM}(E)
\end{aligned}
$$

Variable Elimination

Now we compute $f_{B\hat{A}JM}(B, E)$:

B	E	$f_{B\bar{A}JM}(B,E)$
T	T	.5985 * .001
T	F	.5922 * .001
F	T	.1830 * .999
F	F	.001129 * .999

=

B	$f_B(B)$
T	.001
F	.999

B	E	$f_{\bar{A}JM}(B,E)$
T	T	.5985
T	F	.5922
F	T	.1830
F	F	.001129

Variable Elimination

Next, we sum out B to produce $f_{\bar{B}\bar{A}JM}(E)$.

E	$f_{\bar{B}\bar{A}JM}(E)$
T	.0005985 + .1828= .1834
F	.0005922 * .001128= .001720

Now we compute $f_{E\bar{B}\bar{A}JM}(E)$:

E	$f_{\bar{B}\bar{A}JM}(E)$
T	.1834 * .002 = .0003699
F	.001720 * .998 = .001717

=

E	$f_E(E)$
T	.002
F	.998

E	$f_{\bar{B}\bar{A}JM}(E)$
T	.1834
F	.001720

Variable Elimination

So we have:

$$\alpha f_{E\hat{B}\hat{A}JM}(E) = \alpha\langle .0003699, .001717\rangle = \langle .1772, .8228\rangle$$