

Probabilistic Graphical Models

Lectures 12

Introduction to Inference
Variable Elimination

Probabilistic Model Problems



K. N. Toosi
University of Technology

- Representation:
($Y|X$)

$P_{\theta}(X)$, $P_{\theta}(X, Y)$, P_{θ}

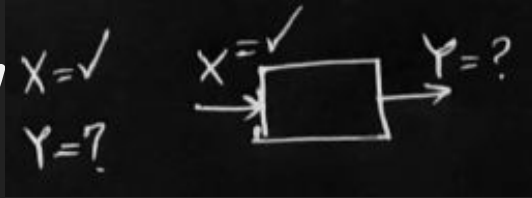
Probabilistic Model Problems



K. N. Toosi
University of Technology

- Representation: $P_{\theta}(X)$, $P_{\theta}(X, Y)$, $P_{\theta}(Y|X)$

- Inference (θ known)



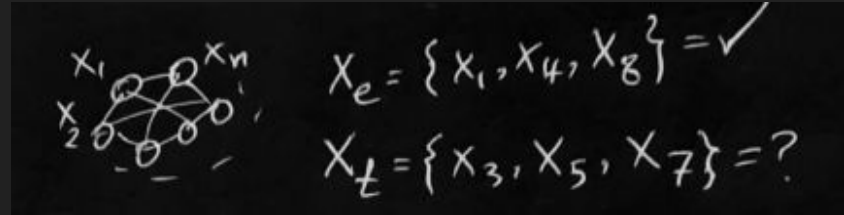
Probabilistic Model Problems



K. N. Toosi
University of Technology

- Representation: $P_{\theta}(X)$, $P_{\theta}(X, Y)$, $P_{\theta}(Y|X)$

- Inference (θ known)

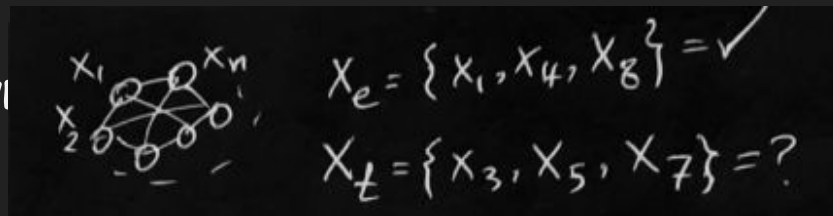




Probabilistic Model Problems

- Representation: $P_{\theta}(X)$, $P_{\theta}(X, Y)$, $P_{\theta}(Y|X)$

- Inference (θ known)



- Sampling (θ known) $P_{\theta}(X) \rightarrow$ generate data X^1, X^2, \dots, X^m

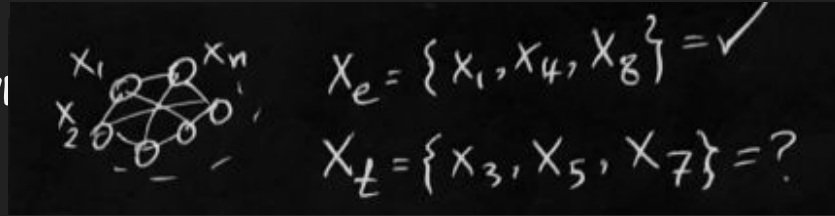
Probabilistic Model Problems



K. N. Toosi
University of Technology

- Representation: $P_{\theta}(X)$, $P_{\theta}(X, Y)$, $P_{\theta}(Y|X)$

- Inference (θ known)



- Sampling (θ known) $P_{\theta}(X) \rightarrow$ generate data X^1, X^2, \dots, X^m

- Learning (θ unknown) data X^1, X^2, \dots, X^m find θ , find structure

Inference



K. N. Toosi
University of Technology

- Given $P(X_1, X_2, \dots, X_N)$ find target variables given observed variables.

Inference



Inference $P(X) = P(X_1, X_2, \dots, X_n) = P(X_e, X_t, X_o)$

$X_e \subseteq X$ (evidence, observation, input) known

$X_t \subseteq X$ (target, output) unknown

$X_o = X \setminus (X_e \cup X_t)$ (other variables) we don't care about them

in many cases $X_o = \{\}$ $\Rightarrow X = X_e \cup X_t$

given X_e find X_t

Example



Example

$$P(X_1, X_2, X_3, X_4, X_5) = \checkmark$$

$$X_t = \{X_1, X_3\}$$

$$X_e = \{X_2, X_5\}$$

$$X_0 = X_4$$

given X_2, X_5 find X_1, X_3

15 (II)

Example



Example: X input
 Y output

$P(X, Y)$
or
 $P(Y|X)$

$X_e = X$
 $X_t = Y$

given X find Y

Example



Given X_e find X_t

What does it mean to find

the target variables X_t ?

- In probabilistic sense
(As a random variable) \Rightarrow Find $P(X_t)$
 $P(\text{Flu})=?$ $P(\text{Covid})=?$ $P(\text{Tb})=?$

- In deterministic sense \Rightarrow Find the most likely X_t ^(MAP inference)
or Find a reasonable X_t

Distance = 1.2 m , Temperature = 19.3 °C

Examples: Semantic Segmentation / Speech Recognition /

message encoding

message \rightarrow media \rightarrow noisy message

Inference



K. N. Toosi
University of Technology

$$X = (X_t, X_e, X_o)$$



$$X_e = \sqrt{\quad}$$

$$P(X_t) = ?$$

Simple Case



Simple Case: $X_e = \{\}$ $X = X_t, X_0$

$$P(X) = P(X_t, X_0)$$
$$P(X_t) = \sum_{X_0} P(X) = \sum_{X_0} P(X_t, X_0)$$

↪ number of summation terms grows exponentially with the size of X_0 .

no. of variables in X_0

Example: Y (v)

Simple Case



Example: $X_0 = \{X_{0,1}, X_{0,2}, \dots, X_{0,m}\}$ no. of variables in X_0 $X_{0,i} \in \{0, 1\}$

$$\sum_{X_0} P(X) = \sum_{\substack{X_{0,1}=0 \\ X_{0,1}=1}} \sum_{\substack{X_{0,2}=0 \\ X_{0,2}=1}} \dots \sum_{\substack{X_{0,m}=0 \\ X_{0,m}=1}} P(X, X_{0,1}, X_{0,2}, \dots, X_{0,m})$$

2^M terms
Intractable for large M

General Case



$$P(X) = P(X_t, X_e, X_0) \quad X_e = x_e \checkmark$$

$$P(X_t | X_e = x_e) = \frac{P(X_t, X_e = x_e)}{P(X_e = x_e)} = \frac{\sum_{X_0} P(X_t, X_0, X_e = x_e)}{\sum_{X_0} \sum_{X_t} P(X_t, X_0, X_e = x_e)}$$

might be intractable

Exploit Factorization



$$P(X) = P(X_t, X_0)$$

$$P(X_t) = \sum_{X_0} P(X) \Rightarrow \text{might be intractable}$$

\Rightarrow Use the fact that for the PGMs (BN/MRF/CRF) the joint distribution is in the form of product of factors

$$P(X) = \prod_c \phi_c(X_c) \quad X_c \text{ comprises a small no. of variables}$$

Exploit Factorization



$P(A, B, C, D, E) \Rightarrow$ find $P(E)$

$$X_e = \{ \} \quad X_o = \{A, B, C, D\} \quad X_t = \{E\}$$

$$P(E) = \sum_A \sum_B \sum_C \sum_D P(A, B, C, D, E)$$

$2^4 = 16$ summation terms if $A, B, C, D \in \{0, 1\}$

Exploit Factorization



K. N. Toosi
University of Technology

$$\sum_D \sum_C \sum_B \sum_A \phi_0(A) \phi_1(A,B) \phi_2(B,C) \phi_3(C,D) \phi_4(D,E)$$

$$\sum_D \sum_C \sum_B \phi_2(B,C) \phi_3(C,D) \phi_4(D,E) \sum_A \phi_0(A) \phi_1(A,B)$$

Exploit Factorization



$$\sum_D \sum_C \sum_B \phi_2(B, C) \phi_3(C, D) \phi_4(D, E) \sum_A \phi_0(A) \phi_1(A, B)$$

$\phi_0(A)$	<table border="1"><tr><td>0.2</td><td>$\phi_0(A=0)$</td></tr><tr><td>0.8</td><td>$\phi_0(A=1)$</td></tr></table>	0.2	$\phi_0(A=0)$	0.8	$\phi_0(A=1)$	$\phi_1(A, B)$				
0.2	$\phi_0(A=0)$									
0.8	$\phi_0(A=1)$									
	<table border="1"><tr><td></td><td>$B=0$</td><td>$B=1$</td></tr><tr><td>$A=0$</td><td>0.7</td><td>0.3</td></tr><tr><td>$A=1$</td><td>0.5</td><td>0.5</td></tr></table>		$B=0$	$B=1$	$A=0$	0.7	0.3	$A=1$	0.5	0.5
	$B=0$	$B=1$								
$A=0$	0.7	0.3								
$A=1$	0.5	0.5								

Exploit Factorization



$$\sum_D \sum_C \sum_B \phi_2(B, C) \phi_3(C, D) \phi_4(D, E) \sum_A \phi_0(A) \phi_1(A, B)$$

$\phi_0(A)$

0.2	$\phi_0(A=0)$
0.8	$\phi_0(A=1)$

$\phi_1(A, B)$

	B=0	B=1
A=0	0.7	0.3
A=1	0.5	0.5

$\psi_1(A, B) = \phi_0(A) \phi_1(A, B)$

\rightarrow

	B=0	B=1
A=0	0.7x0.2	0.3x0.2
A=1	0.5x0.8	0.5x0.8

$= \psi_1(A, B) =$

0.14	0.06
0.4	0.4

Exploit Factorization



$$\sum_D \sum_C \sum_B \phi_2(B, C) \phi_3(C, D) \phi_4(D, E) \sum_A \phi_0(A) \phi_1(A, B)$$

$$\tau_1(B) = \sum_A \psi_1(A, B) = \psi_1(0, B) + \psi_1(1, B)$$

	B=0	B=1	
A=0	0.14	0.06	$\psi_1(A, B)$
A=1	0.4	0.4	

↓ +

$\tau_1(B)$	→	0.54	0.46
		B=0	B=1

Exploit Factorization



$$\sum_D \sum_C \sum_B \phi_2(B, C) \phi_3(C, D) \phi_4(D, E) \sum_A \psi_1(A, B)$$

$$\sum_D \sum_C \sum_B \phi_2(B, C) \phi_3(C, D) \phi_4(D, E) \tau_1(B)$$

$$\sum_D \sum_C \phi_3(C, D) \phi_4(D, E) \sum_B \underbrace{\phi_2(B, C) \tau_1(B)}_{\psi_2(B, C)}$$

$$\sum_D \sum_C \phi_3(C, D) \phi_4(D, E) \tau_2(C)$$

$$\sum_D \phi_4(D, E) \sum_C \phi_3(C, D) \tau_2(C)$$

Exploit Factorization



K. N. Toosi
University of Technology

$$\sum_D \underbrace{\phi_4(D, E) \tau_3(D)}_{\psi_3(D, E)} = \tau_4(D) = P(D)$$

Variable Elimination

Variable Elimination



Variable Elimination

$\phi_0(A) \quad \phi_1(A, B) \quad \phi_2(B, C) \quad \phi_3(C, D) \quad \phi_4(D, E)$

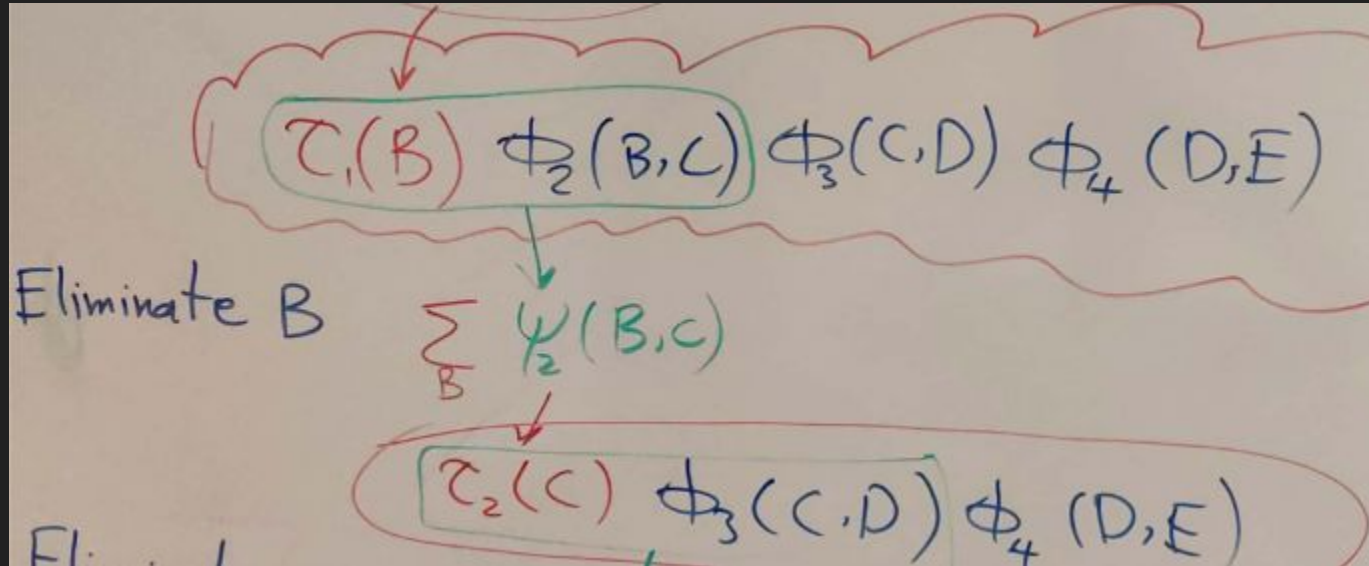
Eliminate A

$\sum_A \psi_1(A, B)$

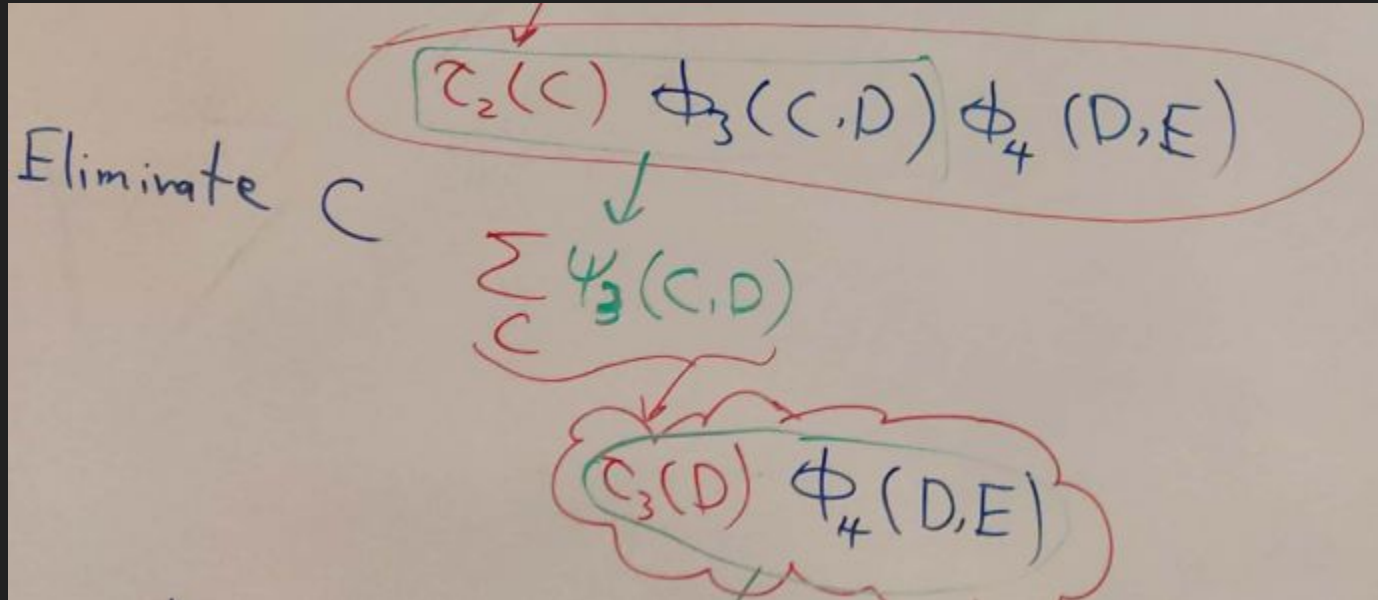
- take out $\phi_0(A) \phi_1(A, B)$
- replace $\tau(B)$

$\tau_1(B) \quad \phi_2(B, C) \quad \phi_3(C, D) \quad \phi_4(D, E)$

Variable Elimination



Variable Elimination



Variable Elimination



K. N. Toosi
University of Technology

Eliminate D

$$\sum_D \psi_4(D, E) = \zeta_4(E) = P(E) = \checkmark$$

The image shows a handwritten derivation on a piece of paper. At the top, the expression $C_3(D) \psi_4(D, E)$ is circled in green and enclosed in a red cloud-like shape. A red arrow points to the $C_3(D)$ term. Below this, the text "Eliminate D" is written. The main equation shows the summation over D of $\psi_4(D, E)$ (with a green arrow pointing from the circled expression to this term) equals $\zeta_4(E) = P(E)$, which is marked with a red checkmark.

Variable Elimination - Unnormalized



$$P(A, B, C, D) = \frac{1}{Z} \phi_1(A, B) \phi_2(B, C) \phi_3(C, D)$$

$$= \frac{1}{Z} \tilde{P}(A, B, C, D)$$

$$Z = \sum_A \sum_B \sum_C \sum_D \tilde{P}(A, B, C, D)$$

Eliminate A

$$P(B, C, D) = \sum_A P(A, B, C, D)$$

$$= \frac{1}{Z} \sum_A \tilde{P}(A, B, C, D)$$

$$= \frac{1}{Z} \tilde{P}(B, C, D)$$

$$= \frac{1}{Z} \underbrace{\zeta(B)}_{\text{VE}} \phi_2(B, C) \phi_3(C, D)$$

Z remains
the same

Variable Elimination - Unnormalized



Z remains
the same

$$= \frac{1}{Z} \tilde{P}(B, C, D) \xrightarrow{VE}$$

$$= \frac{1}{Z} \underbrace{\tau_1(B)}_{\downarrow VE} \phi_2(B, C) \phi_3(C, D)$$

eliminate B

$$P(C, D) = \frac{1}{Z}$$

eliminate C

$$P(D) = \frac{1}{Z}$$

eliminate D

$$1 = \sum_D P(D) = \frac{1}{Z} \sum_D \tau_3(D) \Rightarrow \boxed{\sum_D \tau_3(D) = Z}$$

$$\tau_2(C) \phi_3(C, D) = \frac{1}{Z} \tilde{P}(C, D)$$

$$\tau_3(D) = \frac{1}{Z} \tilde{P}(D)$$

Variable Elimination - Unnormalized



$$\begin{aligned} & \sum_D \tau_3(D) \\ = & \sum_D \sum_C \tau_2(C) \phi_3(C, D) \\ = & \sum_D \sum_C \sum_B \tau_1(B) \phi_2(B, C) \phi_3(C, D) \\ = & \sum_D \sum_C \sum_B \sum_A \phi_1(A, B) \phi_2(B, C) \phi_3(C, D) \\ = & \sum_{A, B, C, D} \tilde{P}(A, B, C, D) = Z \end{aligned}$$

Variable Elimination - with evidence



$$P(X) = P(X_e, X_t, X_o) = \frac{1}{Z} \prod_c \phi_c(X_c)$$

$\underbrace{\hspace{10em}}_{\tilde{P}(X_t, X_e, X_o)}$

$X_e = x_e$ known

$$P(X_t | X_e = x_e) = \frac{\sum_{X_o} P(X_t, x_e, X_o)}{\sum_{X_t} \sum_{X_o} P(X_t, x_e, X_o)} = \frac{\sum_{X_o} \tilde{P}(X_t, x_e, X_o)}{\sum_{X_t} \sum_{X_o} \tilde{P}(X_t, x_e, X_o)}$$

Variable Elimination - with evidence



1- replace $X_e = x_e$

$$\tilde{P}(X_t, x_e, X_0)$$

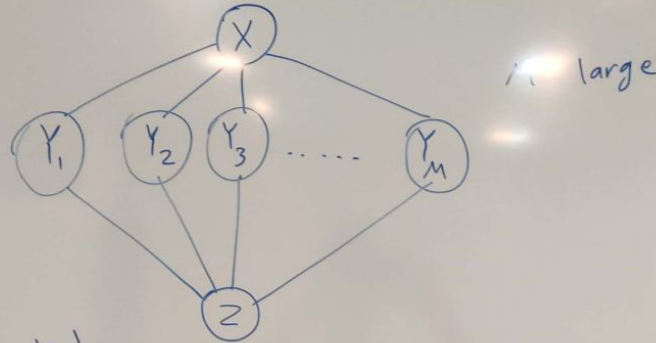
↓ Variable Elimination
Eliminate all $X_i \in X_0$

$$\tilde{P}(X_t, x_e)$$

↓ renormalize over X_t

$$P(X_t | X_e = x_e) = \frac{\tilde{P}(X_t, x_e)}{\sum_{X_t} \tilde{P}(X_t, x_e)}$$

Elimination order



$$P(X, Z, Y_1, Y_2, \dots, Y_m) = \frac{1}{Z} \phi_1(X, Y_1) \phi_2(X, Y_2) \dots \phi_m(X, Y_m) \psi_1(Z, Y_1) \psi_2(Z, Y_2) \dots \psi_m(Z, Y_m)$$

$P(Z) = ?$ eliminate X, Y_1, \dots, Y_m

first eliminate X

$$P(Z, Y_1, Y_2, \dots, Y_m) = \frac{1}{Z} \prod_{i=1}^m \psi_i(Z, Y_i) \sum_X \prod_{i=1}^m \phi_i(X, Y_i)$$

$$= \frac{1}{Z} \prod_{i=1}^m \psi_i(Z, Y_i) \sum_X \delta(X, Y_1, Y_2, \dots, Y_m) \rightarrow \text{too large to compute}$$

Elimination order



eliminate Y_i -s first

eliminate Y_1

$$P(X, Z, Y_2, Y_3, \dots, Y_m) = \frac{1}{2} \sum_{Y_1} \prod_{i=1}^m \psi_i(z, Y_i) \prod_{i=1}^m \phi_i(X, Y_i)$$

$$= \frac{1}{2} \prod_{i=2}^m \psi_i(z, Y_i) \prod_{i=2}^m \phi_i(X, Y_i) \sum_{Y_1} \psi(z, Y_1) \phi(X, Y_1)$$

⋮

$$\sum_{Y_1} \delta(X, z, Y_1)$$

$$\tau_1(X, z)$$

eliminate X

$$\sum_X \tau_1(X, z) \tau_2(X, z) \dots \tau_m(X, z)$$

$$\sum_X \delta_{n+1}(X, z)$$