Variable Elimination

Dhruv Batra,
10-708 Recitation
10/16/2008
Overview

• Variable Elimination
  – Example on chain networks
  – Intuition
  – Tools for VE, factor product, marginalization
  – Implementation hints
Chain Networks

- BN:

\[ A \rightarrow B \rightarrow C \rightarrow D \]

- Goal: Need all marginals \( P(X) \)
Chain Networks

• Naïve solution

\[
\begin{align*}
P(a^1) & \quad P(b^1 | a^1) & \quad P(c^1 | b^1) & \quad P(d^1 | c^1) \\
+ P(a^2) & \quad P(b^1 | a^2) & \quad P(c^1 | b^1) & \quad P(d^1 | c^1) \\
+ P(a^1) & \quad P(b^2 | a^1) & \quad P(c^1 | b^2) & \quad P(d^1 | c^1) \\
+ P(a^2) & \quad P(b^2 | a^2) & \quad P(c^1 | b^2) & \quad P(d^1 | c^1) \\
+ P(a^1) & \quad P(b^1 | a^1) & \quad P(c^2 | b^1) & \quad P(d^1 | c^2) \\
+ P(a^2) & \quad P(b^1 | a^2) & \quad P(c^2 | b^1) & \quad P(d^1 | c^2) \\
+ P(a^1) & \quad P(b^2 | a^1) & \quad P(c^2 | b^2) & \quad P(d^1 | c^2) \\
+ P(a^2) & \quad P(b^2 | a^2) & \quad P(c^2 | b^2) & \quad P(d^1 | c^2) \\
+ P(a^1) & \quad P(b^1 | a^1) & \quad P(c^1 | b^1) & \quad P(d^2 | c^1) \\
+ P(a^2) & \quad P(b^1 | a^2) & \quad P(c^1 | b^1) & \quad P(d^2 | c^1) \\
+ P(a^1) & \quad P(b^2 | a^1) & \quad P(c^1 | b^2) & \quad P(d^2 | c^1) \\
+ P(a^2) & \quad P(b^2 | a^2) & \quad P(c^1 | b^2) & \quad P(d^2 | c^1) \\
+ P(a^1) & \quad P(b^1 | a^1) & \quad P(c^2 | b^1) & \quad P(d^2 | c^2) \\
+ P(a^2) & \quad P(b^1 | a^2) & \quad P(c^2 | b^1) & \quad P(d^2 | c^2) \\
+ P(a^1) & \quad P(b^2 | a^1) & \quad P(c^2 | b^2) & \quad P(d^2 | c^2) \\
+ P(a^2) & \quad P(b^2 | a^2) & \quad P(c^2 | b^2) & \quad P(d^2 | c^2)
\end{align*}
\]
Chain Networks

• A little smarter solution: Phased Computation

\[ P(B) = \sum_a P(a)P(B \mid a). \]
\[ P(C) = \sum_b P(b)P(C \mid b). \]

• General Chains:

\[ P(X_{i+1}) = \sum_{x_i} P(X_{i+1} \mid x_i)P(x_i) \]

• Why is this better?

\[ O(nk^2) \quad \text{vs} \quad O(k^n) \]

• What’s the intuition?
Chain Networks

- A lot of structure
Chain Networks

• Let’s cache
Let’s cache again

\[ \begin{align*}
\tau_2(c^1) & = \tau_1(b^1)P(c^1 \mid b^1) + \tau_1(b^2)P(c^1 \mid b^2) \\
\tau_2(c^2) & = \tau_1(b^1)P(c^2 \mid b^1) + \tau_1(b^2)P(c^2 \mid b^2)
\end{align*} \]
Chain Networks

- Intuitions from this process
  - Group common things/terms/factors based on scope
  - Dynamic programming ideas: cache computations

- VE extends/formalizes these intuitions to general graphs
  - but separates the elimination ordering from the process
Another Idea

• A commonly used idea

• Goal

\[ P(Y \mid E = e) = \frac{P(Y, e)}{P(e)} \]

\[ P(y, e) = \sum_w P(y, e, w). \]

\[ P(e) = \sum_y P(y, e) \]

• Can forget about denominator; just renormalize when done
Example

- Let’s do an example for general graphs

\[ P(S = T | C = T) = ? \]

\[ P(F = T | G = T) \]
Implementing VE

• What do you need to implement VE?
  – Reuse some code from HW2
Tools for VE

- Factors

\[ \pi_1[A, B] : Val(A, B) \rightarrow \mathbb{R}^+ \]

- Special kind of factors: CPTs
Operations on Factors

• Factor Product
  – Consider two factors
    \[ \phi_1(X, Y) \text{ and } \phi_2(Y, Z) \]
  – Define factor product
    \[ \psi : \text{Val}(X, Y, Z) \rightarrow \mathbb{R} \]
  – such that
    \[ \psi(X, Y, Z) = \phi_1(X, Y) \cdot \phi_2(Y, Z) \]
Operations on Factors

- Factor Product
Operations on Factors

• Factor Marginalization
  – Consider a factor
    \( \phi(X, Y) \)
  – Define factor marginal
    \( \psi(X) : \text{Val}(X) \leftrightarrow \mathbb{I}R. \)
  – such that
    \[ \psi(X) = \sum_Y \phi(X, Y). \]
Operations on Factors

- Factor Marginalization
Factors

• Are factors always distributions?
  – Obviously not

• Are factors produced in VE always distributions?
  – Yes, always conditional distributions
  – In SOME graph, not necessarily the original graph
  – HW3, prob 2. Hint: read 8.3.1.3
Implementing VE

• What do you need to implement VE?
  – Reuse some code from HW2

• Representation
  – BN as an array of factors
  – table_factor.m
  – assignment.m

• VE
  – multiply_factors.m
  – marginalize_factor.m
  – min_fill.m
Variable elimination algorithm

- Given a BN and a query $P(X|e) / P(X,e)$
- Instantiate evidence $e$
- Prune non-active vars for $\{X,e\}$
- Choose an ordering on variables, e.g., $X_1, ..., X_n$
- Initial factors $\{f_1, ..., f_n\}$: $f_i = P(X_i|\text{Pa}_{X_i})$ (CPT for $X_i$)
- For $i = 1$ to $n$, If $X_i \notin \{X,E\}$
  - Collect factors $f_1, ..., f_k$ that include $X_i$
  - Generate a new factor by eliminating $X_i$ from these factors
    \[
g = \sum_{X_i} \prod_{j=1}^{k} f_j
\]
  - Variable $X_i$ has been eliminated!
- Normalize $P(X,e)$ to obtain $P(X|e)$
Questions?