Exact Methods and Dynamic Programming

Note Title 4/2/2006

Suppose we have a joint distribution

$$\pi(\underline{x}) \propto e^{-\sum_{i=1}^{d} h_i(x_{i-1},x_i)}$$

Undirected Graph $\quad x_0 - x_1 - \circ - \circ - \circ - \circ - \circ - x_d$

Markov Random Field. (MRF)

The x_i's are discrete random variables (rv's) taking values in the finite set $S = \{s_1 \dots s_k\}$.

Dynamic Programming can be used to find the global maximum of $\pi(\underline{x})$, $\hat{x}$, and $\pi(\hat{x})$ in $O(dk^2)$ operations.

DP can also find the marginal distribution $\pi_i(x_i)$ and draw exact random samples from $\pi(\underline{x})$ efficiently.

Practicality depends on k.

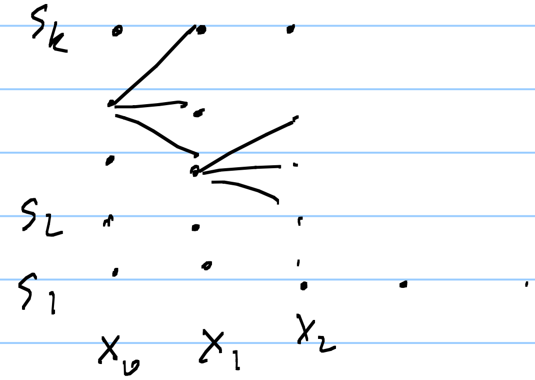
Page 2. Maximizing $\pi(\underline{x})$ is equivalent to minimizing $E(\underline{x}) = h_1(x_0, x_1) + \dots + h_d(x_{d-1}, x_d)$.

Forward pass

DP acts recursively:

- Define $m_1(x_1) = \min_{s_i \in S} h_1(s_i, x_1)$ for $x_1 = s_1, \dots s_k$
- Recursively compute $m_t(x_t) = \min_{s_i \in S} \{m_{t-1}(s_i) + h_t(s_i, x_t)\}$, for $x_t = s_1 \dots s_k$.

Claim: Optimal value $E(\hat{x})$ is obtained by $\min_{s \in \{s_1 \dots s_k\}} m_d(s)$



To compute $m_1(x_1)$ for $x_1 = s_1 \dots s_k$ requires $O(k^2)$ operations.

Computing all $m_t(x_t) \dots$ requires $O(k^2 d)$

Justify Claim: minimum of $m_1(x_1)$ is the minimum of $h_1(x_0, x_1)$

by induction $\min_{x_t \in S} m_t(x_t) = \min_{x_0, \dots x_t \in S} \{ h_1(x_0, x_1) + \dots + h_t(x_{t-1}, x_t) \}$.

Page 3.

To find the optimal path, we need ~~to~~ trace back.

Backward pass.

- Let $\hat{x}_d$ be the minimizer of $m_d(x_d)$

$$\hat{x}_d = \arg\min_{s_i \in S} m_d(s)$$

(Break ties arbitrarily)

- For $t = d-1, d-2, \ldots 1,$

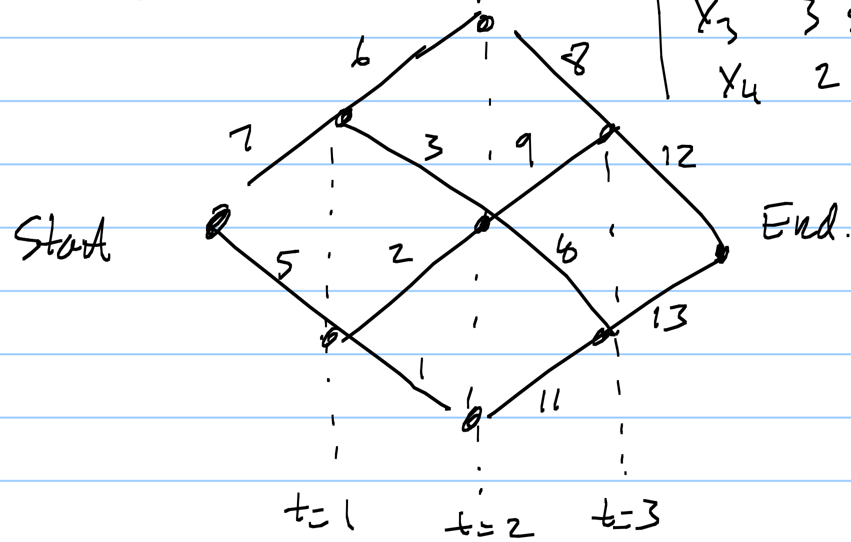
Let $$\hat{x}_t = \arg\min_{s_i \in S} \{m_t(s_i) + h_{t+1}(s_i, \hat{x}_{t+1})\}$$

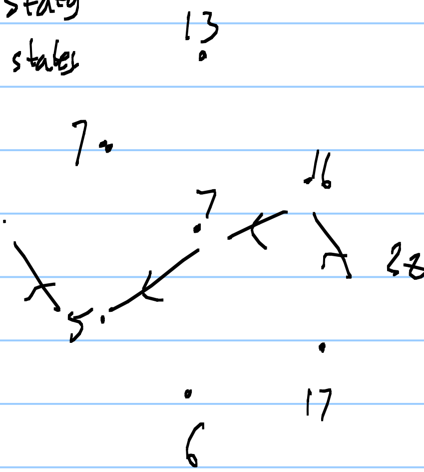
(Break ties arbitrarily)

Configuration $\underline{\hat{x}} = (\hat{x}_1, \ldots, \hat{x}_d)$ obtained in this way is the minimizer.

Another Example

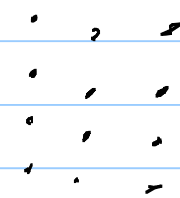
X_0 1 state
X_1 2 states
X_3 3 states
X_4 2 states





Compute shortest path to t=1
shortest to t=2
shortest to t=3 . . .

West Coast
→ East Coast.



Intuition — break the problem up into subcomponents

If you compute the shortest path from Los Angeles to Boston passing through Chicago.
Break it into two subproblems
- Find shortest path from LA to Chicago
- Shortest path from Chicago to Boston.

To simulate from $\pi(\underline{x})$ using DP.
Calculate the marginal $\pi(x_d)$ and the Conditionals $\pi(x_i \mid x_{i+1})$ $\quad i = 0, 1, \dots, d-1$.

Then sample $\hat{x}_d$ from $\pi(x_d)$

$\hat{x}_{d-1}$ from $\pi(x_{d-1} \mid \hat{x}_d)$

$\hat{x}_{d-2}$ from $\pi(x_{d-2} \mid \hat{x}_{d-1})$

$\dots$

How to calculate the marginals and conditionals?

- Define $V_1(x_1) = \sum_{x_0 \in S} e^{-h_1(x_0, x_1)}$
- Recursively compute for $t = 2, \dots, d$

$$V_t(x_t) = \sum_{y \in S} V_{t-1}(y)\, e^{-h_t(y, x_t)}$$

Then we can efficiently compute:

(A) The normalization constant

$$Z = \sum_{x_d \in S} V_d(x_d)$$

(B) The marginal

$$\pi_d(x_d) = V_d(x_d)/Z$$

(C) The conditionals

$$\pi(x_t \mid x_{t+1}) = \frac{V_t(x_t)\, e^{-h_{t+1}(x_t, x_{t+1})}}{\sum_y V_t(y)\, e^{-h_{t+1}(y, x_{t+1})}}$$

Recall $\pi_d(x_d) = \dfrac{\sum_{x_0 \dots x_{d-1}} e^{-E(x_0, \dots x_d)}}{\sum_{y_0 \dots y_d} e^{-E(y_0 \dots y_d)}}$

(Page 6) Special Case:

Ising Spin Model

$$\pi(\underline{x}) = \frac{1}{Z} e^{\beta(x_0 x_1 + \cdots + x_{d-1} x_d)}.$$

$x_i \in \{-1, +1\}$

$$V_1(x_1) = e^{\beta x_1} + e^{-\beta x_1} = e^{\beta} + e^{-\beta}$$

Very unusual that $V_1(x_1)$ is indep of x.

$$V_2(x_2) = \sum_{y \in S} V_1(y)\, e^{\beta y x_2}$$

$$= \{e^{\beta} + e^{-\beta}\} \sum_{y \in S} e^{\beta y x_2} = \{e^{\beta} + e^{-\beta}\}^2.$$

In general, $V_t(x_t) = \{e^{\beta} + e^{-\beta}\}^t$ // $V_d(x_d) = \{e^{\beta} + e^{-\beta}\}^d$

Hence $Z = \sum_{x_d \in S} V_d(x_d) = 2\{e^{\beta} + e^{-\beta}\}^d$

$S = \{-1, +1\}$

Marginal Density $\pi(x_d) = \frac{V_d(x_d)}{Z} = \frac{1}{2}$

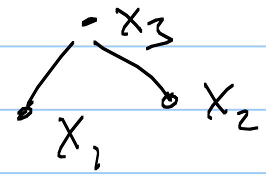
Conditional Distribution:

$$\pi(x_t \mid x_{t+1}) = \frac{\{e^{\beta} + e^{-\beta}\}^t e^{\beta x_t x_{t+1}}}{\sum_{y \in S} \{e^{\beta} + e^{-\beta}\}^t e^{\beta y x_{t+1}}}$$

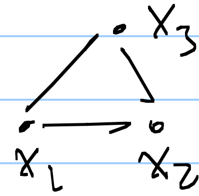
$$\pi(x_t \mid x_{t+1}) = \frac{e^{\beta x_t x_{t+1}}}{e^{\beta} + e^{-\beta}} \quad //$$

(Page 7) DP can be applied directly to probability distribution on a graph <u>without</u> closed loops.

<u>E.G.</u>



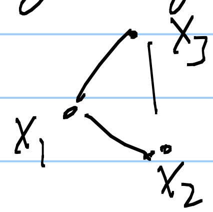
Yes



No.

If graphs have closed loops, then it is possible to transform the distribution into a new distribution <u>without</u> closed loops, by augmenting the variables

E.G



Define new variable (x_1, x_2, x_3)

Junction Tree Algorithm

<u>But</u>, augmenting variables risks making the set $S = \{S_1 .. S_k\}$ large. i.e. k large so $O(dk^2)$ may be enormous.

Gen<u>eral</u> limitations of DP, k is often too large.