Exact Methods and Dynamic Programming

Suppose we have a joint distribution

\[ \mathcal{T}(\mathbf{X}) \propto e^{-\sum_{i=1}^{k} h_i(x_{i-1}, x_i)} \]

Undirected graph

\[ x_0 \quad x_1 \quad \cdots \quad x_k \]

Markov Random Field (MRF)

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

Dynamic Programming can be used to find the global maximum of \( \mathcal{T}(\mathbf{X}), \mathbf{X} \), and \( \mathcal{T}(\mathbf{X}) \) in \( O(d k^4) \) operations.

DP can also find the marginal distribution \( \mathcal{T}_i(x_i) \) and draw exact random samples from \( \mathcal{T}(\mathbf{X}) \) efficiently.

Practicality depends on \( k \).
Maximizing $\Pi(x)$ is equivalent to minimizing $E(x) = h_1(x_0, x_1) + \ldots + h_d(x_{d-1}, x_d)$.

Forward DP acts recursively:
- Define $m_1(x_i) = \min_{s_i \in S} h_i(s_i, x_i)$ for $x_i = s_1 \ldots s_k$
- Recursively compute $m_t(x_t) = \min_{s_t \in S} \{ m_{t-1}(s_t) + h_t(s_t, x_t) \}$ for $x_t = s_1 \ldots s_k$

Claim: Optimal value $E(x)$ is obtained by $\min_{s_0(s_1 \ldots s_k)} m_k(x_k)$

To compute $m_1(x_i)$ for $x_i = s_1 \ldots s_k$ requires $O(k^2)$ operations.
Computing all $m_t(x_t)$... requires $O(k^2d)$

Justify Claim: minimum of $m_i(x_i)$ is the minimum of $h_i(x_0, x)$ by induction $\min_{x_t \in S} m_t(x_t) = \min_{x_0, \ldots, x_t \in S} h_i(x_0, x_t) + \ldots + h_t(x_{t-1}, x_t)$.
To find the optimal path, we need to use the following steps:

1. Let $x_d$ be the minimizer of $\min_x \, m_d(x)$
   
   \[
   x_d = \arg \min_{s \in S} \, m_d(s) \tag{Break ties arbitrarily}
   \]

2. For $t = d-1, d-2, \ldots 1$
   
   Let $\hat{x}_t = \arg \min_{s \in S} \left( m_t(s_t) + h_{t+1}(s_t, \hat{x}_{t+1}) \right) \tag{Break ties arbitrarily}$

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

\[
\begin{align*}
X_0 & \quad \text{1 state} \\
X_1 & \quad \text{2 state} \\
X_2 & \quad \text{3 state} \\
X_3 & \quad \text{2 state} \\
\end{align*}
\]

|
|---|
| 8 |
| 6 |
| 11 |
| 13 |
| 7 |
| 12 |
| 9 |
| 3 |
| 7 |
| 16 |
| 17 |
| 6 |
| 1 |
| 2 |
| 3 |

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.

- Full shortest path from LA to Chicago
- Shortest path from Chicago to Boston.
To simulate from $\Pi_i(x)$ using DP:

Calculate the marginal $\Pi_i(x_i)$ and the conditionals $\Pi_i(x_i \mid x_{i+1})$ for $i = 0, 1, \ldots, d-1$.

Then sample $x_d$ from $\Pi_i(x_d)$

- $x_{d-1}$ from $\Pi_i(x_{d-1} \mid x_d)$
- $x_{d-2}$ from $\Pi_i(x_{d-2} \mid x_{d-1})$

How to calculate the marginals and conditionals:

- Define $V_t(x) = \sum_{x_o \in S} e^{-h_t(x_o, x_t)}$
- Recursively compute for $t = 2, \ldots, d$
  
  \[ V_t(x_t) = \sum_{y \in S} V_{t-1}(y) e^{-h_t(y, x_t)} \]

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:
  
  \[ \Pi_i(x_i \mid x_{i+1}) = V_t(x_t) e^{-h_{i+1}(x_t, x_{i+1})} \]
  
  \[ \sum_{y} V_t(y) e^{-h_{i+1}(y, x_{i+1})} \]
Special Case:

Ising Spin Model : 

\[ \Pi_i(x_i) = \frac{1}{Z} e^{\beta (x_0 x_1 + \ldots + x_{d-1} x_d)} \]

\[ V_1(x_i) = e^{\beta x_i} + e^{-\beta x_i} = e^{\beta} + e^{-\beta} \]

Very unusual:

That \( V_1(x_i) \) is widely of \( x_i \)

\[ V_2(x_i) = \sum_{y \in S} V_1(y) \cdot e^{\beta y x_i} \]

\[ = (e^{\beta} + e^{-\beta}) \sum_{y \in S} e^{\beta y x_i} = (e^{\beta} + e^{-\beta})^2. \]

In general:

\[ V_t(x_i) = (e^{\beta} + e^{-\beta})^t \prod_{y \in S \setminus \{x_i \}} \left( e^{\beta y} + e^{-\beta y} \right) \]

Hence:

\[ Z = \sum_{X_d \in S} V_d(x_d) = \sum_{X_d \in S} (e^{\beta} + e^{-\beta})^d \]

Marginal Density:

\[ \Pi_i(x_i) = V_d(x_d) = \frac{1}{Z} \]

Conditional Distribution:

\[ \Pi_i(x_t \mid x_{t+1}) = \frac{\left( e^{\beta} + e^{-\beta} \right)^t e^{\beta x_t x_{t+1}}}{\sum_{y \in S} \left( e^{\beta} + e^{-\beta} \right)^t e^{\beta y x_{t+1}}} \]

\[ \Pi_i(x_t \mid x_{t+1}) = \frac{e^{\beta x_t x_{t+1}}}{e^{\beta} + e^{-\beta}} \]
(Page 7) DP can be applied directly to
probability distributions on a graph
without closed loops.

\[ x_1 \xrightarrow{} x_3 \quad x_2 \quad \text{Yes} \]
\[ x_1 \xrightarrow{} x_3 \quad \text{No.} \]

If graphs have closed loops, then
it is possible to transform the distribu-
tion into a new distribution without
loops, by augmenting the variables

\[ E.6 - x_1 \xrightarrow{} x_3 \quad \text{Define new variable} \quad (x_1, x_3) \]

Junction Tree Algorithm

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

General limitations of DP, \( k \) is often too large.