CSCI 590: Machine Learning

Lecture 18: Inference on a Chain and factor graphs

Instructor: Murat Dundar

Acknowledgement:
1. PRML by C. Bishop
Inference in Graphical Models

Once $y$ is observed the goal is to infer the corresponding posterior distribution over $x$. The joint distribution is expressed in terms of $p(y)$ and $p(x|y)$, which is represented by the last graph.

$$p(y) = \sum_{x'} p(y|x')p(x')$$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$
Inference on a Chain

N nodes represent discrete variables each having K states. Each potential function $\psi_{n-1,n}(x_{n-1}, x_n)$ comprises an K x K table. The joint distribution has $(N - 1)K^2$ parameters.

There are $K^N$ values for $x$. Evaluation and storage of the $p(x)$ and marginalization to obtain $p(x_n)$ scale exponentially with N.
Inference on a Chain

\[ p(x_n) = \frac{1}{Z} \left[ \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \cdots \sum_{x_1} \psi_{1,2}(x_1, x_2) \right] \cdots \]

\[ \mu_\alpha(x_n) \]

\[ \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \cdots \]

propagation of local messages

each summation removes a variable

\[ \mu_\beta(x_n) \]
Inference on a Chain

\[ \mu_\alpha(x_n) = \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \left[ \sum_{x_{n-2}} \cdots \right] \]

\[ = \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \mu_\alpha(x_{n-1}). \]

\[ \mu_\beta(x_n) = \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \left[ \sum_{x_{n+2}} \cdots \right] \]

\[ = \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \mu_\beta(x_{n+1}). \]
Inference on a Chain

\[ \mu_\alpha(x_n) = \mu_\alpha(x_{n-1}) \quad \mu_\beta(x_n) = \mu_\beta(x_{n+1}) \]

\[ \mu_\alpha(x_2) = \sum_{x_1} \psi_{1,2}(x_1, x_2) \quad \mu_\beta(x_{N-1}) = \sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \]

\[ p(x_n) = \frac{1}{Z} \mu_\alpha(x_n) \mu_\beta(x_n) \]

\[ Z = \sum_{x_n} \mu_\alpha(x_n) \mu_\beta(x_n) \]
Inference on a Chain

To compute local marginals:

• Compute and store all forward messages, $\mu_\alpha(x_n)$.
• Compute and store all backward messages, $\mu_\beta(x_n)$.
• Compute $Z$ at any node $x_n$
• Compute

$$p(x_n) = \frac{1}{Z} \mu_\alpha(x_n) \mu_\beta(x_n)$$

for all variables required.
Computational cost

Obtaining $p(x_n)$

**Naïve approach:** $O(K^N)$

**Message passing**

N-1 summations each of which is over K states and each of which involves a function of two variables.

For instance the summation over $x_1$ involves only the function $\psi_{1,2}(x_1, x_2)$, which is a table of $K \times K$ numbers. We have to sum this table over $x_1$ for each value of $x_2$. This has $O(K^2)$ cost. The resulting vector of K numbers is multiplied by the matrix of numbers $\psi_{2,3}(x_2, x_3)$, which is again $O(K^2)$ cost. Because there are N-1 summations and multiplications of this kind, the total cost of evaluating the marginal $p(x_n)$ is $O(NK^2)$
Trees

Undirected Tree

Directed Tree

Polytree
Factor Graphs

\[ p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3) \]

\[ p(\mathbf{x}) = \prod_{s} f_s(\mathbf{x}_s) \]
Factor Graphs from Directed Graphs

\[ p(x) = p(x_1)p(x_2) \]
\[ p(x_3|x_1, x_2) \]
\[ f(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2) \]
\[ f_a(x_1) = p(x_1) \]
\[ f_b(x_2) = p(x_2) \]
\[ f_c(x_1, x_2, x_3) = p(x_3|x_1, x_2) \]
Factor Graphs from Undirected Graphs

\[
\psi(x_1, x_2, x_3) = f(x_1, x_2, x_3) = \psi(x_1, x_2, x_3) = f_a(x_1, x_2, x_3)f_b(x_2, x_3)
\]