Lecture 17: Sum-Product Algorithm

Scribes: Yaoshen, Ruiyang, Elizabeth

Homework 4: Released yesterday
Due Fri Apr 13
Today: Exact Marginals over Discrete Variables

Goal: Compute Marginals of the form

\[ p(x_i = x_i, X_j = x_j) = \sum_{x_k: k \neq i, j} p(x_i = x_i, \ldots, x_k = x_k) \]

Assumption: All variables \( X_w \) are discrete

Enables: Calculation of posterior

\[ p(x_i = x_i \mid X_j = x_j) = \frac{p(x_i = x_i, X_j = x_j)}{p(X_j = x_j)} \]
Example: Markov Chain

Idea: Rearrange Terms in Sum

\[ p(a, b, c, d) = p(a | b) p(b | c) p(c | d) p(d) \]

\[ p(a) = \sum_{b=1}^{B} \sum_{c=1}^{C} \sum_{d=1}^{D} p(a, b, c, d) \quad \text{Naive: ABCD} \]

\[ = \sum_{b=1}^{B} p(a | b) \left( \sum_{c=1}^{C} p(b | c) \left( \sum_{d=1}^{D} p(c | d) p(d) \right) \right) \]

\[ = \sum_{b=1}^{B} p(a | b) \left( \sum_{c=1}^{C} p(b | c) \gamma_a(c) \right) = \sum_{b=1}^{B} p(a | b) \gamma_a(b) \]

\[ \gamma_a(c) \]

\[ \gamma_a(b) \]
Example: Markov Chain

\[ p(a) = \sum_{a} p(a|b) \sum_{b} p(b|c) \sum_{c} p(c|d) \]

\[ Y_{c}(b) = \sum_{c} p(b|c) Y_{d}(c) \]

\[ Y_{d}(c) = \sum_{d} p(c|a) p(d) \]

\[ Y_{b}(a) = \sum_{b} p(a|b) Y_{c}(b) \]

Cost: CD

Question: What is the Computational Complexity of computing \( p(a) \)? AB + BC + CD
Sum-product on Factor Graphs

Bayesian Network (non-branching)

\[ p(a, b, c, d) = p(a | b) p(b | c) p(c | d) p(d) \]

Factor Graph (non-branching)

\[ p(a, b, c, d) = f_1(a, b) f_2(b, c) f_3(c, d) f_4(d) \]
Sum-product on Factor Graphs

Factor Graph (non-branching)

\[ p(a, b, c, d) = f_1(a, b) f_2(b, c) f_3(c, d) f_4(d) \]

Messages: Variable to Variable

\[ \mu_{a \rightarrow c}(c) = f_1(a, b) f_2(b, c) \sum_{d} f_3(c, d) f_4(d) \]

\[ p(a, b) = f_1(a, b) \sum_{c} f_2(b, c) \mu_{a \rightarrow c}(c) \]

\[ \mu_{c \rightarrow b}(b) = \]
Sum-product on Factor Graphs

Factor Graph (non-branching)

\[ p(a, b, c, d) = f_1(a, b) f_2(b, c) \]
\[ f_3(c, d) f_4(d) \]

Factor Graph (singly-connected) (a.k.a. Tree)

Factors with multiple edges

\[ p(a, b, c, d) = f_1(a, b) f_2(b, c, d) \]
\[ f_3(c) f_4(d) f_5(d) \]
Sum-product on Factor Graphs

Factor Graph (singly-connected)

$$p(a, b) = f_1(a, b) \sum_{c,d} f_2(b, c, d) f_3(c) f_5(a) \sum_{e} f_4(d, e)$$

Messages: Factor $\leftrightarrow$ Variable

- $$f_1(a, b)$$
- $$f_2(b, c, d)$$
- $$f_3(c)$$
- $$f_4(d, e)$$
- $$f_5(a)$$

$$p(a, b) = f_1(a, b) \sum_{c,d} f_2(b, c, d) f_3(c) f_5(a) \sum_{e} f_4(d, e)$$

$$M_{f_2 \rightarrow b}(b) = \sum_{c,d} f_2(b, c, d) M_{c \rightarrow f_3}(c) M_{d \rightarrow f_5}(d)$$

$$M_{c \rightarrow f_3}(c) = M_{f_2 \rightarrow c}(c)$$

$$M_{d \rightarrow f_5}(d) = M_{f_4 \rightarrow d}(d) M_{f_5 \rightarrow d}(d)$$

$$M_{f_5 \rightarrow d}(d) = f_5(d)$$

$$M_{f_4 \rightarrow d}(d) = \sum_{e} f_4(d, e) M_{e \rightarrow f_4}(e)$$

Factor Graph:

- Nodes: $$a, b, c, d, e$$
- Edges:
  - $$f_1(a, b)$$
  - $$f_2(b, c, d)$$
  - $$f_3(c)$$
  - $$f_4(d, e)$$
  - $$f_5(a)$$
Sum-product on Factor Graphs

Advantage of Factor Graphs: Can compute any marginal

\[ p(e) = \sum_a f_4(d,e) \prod_{d' \in \text{source}(d)} f_{d'}(d) \]

\[ M_{d \rightarrow f_4}(d) = \prod_{f_z \rightarrow d}(d) \prod_{f_{1} \rightarrow d}(d) \]

\[ M_{f_z \rightarrow d}(d) = \sum_{b,c} f_z(b,c,d) M_{b \rightarrow f_z}(b) \prod_{c \in \text{dom}(f_z)} M_{c \rightarrow f_z}(c) \]
Sum-product on Factor Graphs

\[ ne(f_2) = \{b, c, d\} \]
\[ ne(d) = \{f_2, f_u, f_s\} \]
\[ x_f \equiv ne(f) \]

\[ \Sigma \]
\[ \Pi \]

Factors: General Form
\[ p(x) = \prod_{f} \phi_f(x_f) \]
\[ ne(x) \] (factors in which \( x \) occurs)
\[ ne(f) \] (variables that \( f \) depends on)

Factor → Variable:
\[ M_{f \rightarrow x}(x) = \sum \phi_f(x_f) \prod_{\{X_f \setminus x\} \in ne(f) \setminus \{x\}} M_{g \rightarrow x}(x) \]

Variable → Factor:
\[ M_{x \rightarrow f}(x) = \prod_{g \in ne(x) \setminus \{f\}} M_{g \rightarrow x}(x) \]
Belief Propagation: Compute Marginals for All Variables

General Form: Marginal

\[ p(x) \propto \prod_{f \in ne(x)} M_{f \rightarrow x}(x) \]

Algorithm: Compute All Messages

1. Pick any variable \( x \)
2. Compute incoming messages
3. Compute outgoing messages
Variable $\Rightarrow$ Variable $\Rightarrow$ vs Factor $\Rightarrow$ Variable Messages

Factor Graph (non-branching)

\[ \mu_{c \rightarrow b}(b) = \sum_c f_2(b,c) \mu_{c \rightarrow f_2(c)} \]
\[ = \mu_{f_2 \rightarrow b}(b) \]
Goal: Compute Marginals

\[ a_t(k) = \frac{P(h_t | V_t, \ldots, V_1) B_t(k)}{\sum_k P(h_t | V_t, \ldots, V_1) B_t(k)} \]

Example: Forward-Backward Algorithm (HMMs)
Example: Forward-Backward Algorithm (HMMs)

Factor Graph

Generative Model

\[ h_t \sim \text{Disc}(\pi_1, ..., \pi_K) \]
\[ h_t | h_{t-1} = h \sim \text{Disc}(A_{h_1}, ..., A_{h_K}) \]
\[ v_t | h_t = h \sim p(v_t | h_t = h) \]

Forward Pass (outgoing messages)

\[ \alpha_t(h_t) = \prod_{e_{t+1}} f_{k+1}(h_t | k) \]
\[ = \sum_l f_t(h_t, \ell) \sum_{h_{t-1}} p(h_t | h_{t-1} = \ell) \alpha_{t-1}(\ell) \]
\[ p(h_t = h | h_{t-1} = \ell) = A_{h_1} \alpha_{t-1}(\ell) \]
Example: Forward-Backward Algorithm (HMMs)

**Factor Graph**

\[ f, f_2, f_3, f_4 \]

\[ h_1 \rightarrow h_2 \rightarrow h_3 \rightarrow h_4 \]

\[ g_1, g_2, g_3, g_4 \]

\[ v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_4 \]

**Generative Model**

\[ h_t \sim \text{Disc}(n_1, \ldots, n_k) \]

\[ h_t | h_{t-1} = h \sim \text{Disc}(A_{n_1}, \ldots, A_{n_k}) \]

\[ v_t | h_t = h \sim p(v_t | h_t = h) \]

**Backward Pass (Incoming Messages)**

\[ B_t(h) = \mu_{f_t \rightarrow h_t}(h) = \sum_{\ell} f_t(l, h) \mu_{h_{t+1} \rightarrow f_t}(l) \]

\[ = \sum_{\ell} f_t(l, h) \mu_{g_{t+1} \rightarrow h_{t+1}}(l) \mu_{f_{t+1} \rightarrow h_{t+1}}(l) \]

\[ A_{h_t} \quad p(v_{t+1} | h_{t+1} = h) \quad B_{t+1}(l) \]
Example: Forward-Backward Algorithm (HMMs)

Forward Pass

\[
\alpha_t(k) = p(h_t = k) p(v_t, h_t = k) \\
\alpha_t(u) = p(v_t | h_t = u) \sum_{k=1}^{K} A_{ku} \alpha_{t-1}(k) \quad (t > 1) \quad K^2
\]

Backward Pass

\[
O(K + (t-1)K^2) \ll O(K^T)
\]

\[
\beta_T(u) = 1 \\
\beta_t(u) = \sum_{k=1}^{K} A_{ku} p(v_{t+1}, h_{t+1}=1) \beta_{t+1}(l) \quad (t < T) \quad K^2
\]

Marginals

\[
\gamma_t(u) \propto \alpha_t(u) \beta_t(u) = \mu_{f_{t-1} \rightarrow h_t}(u) \mu_{g_t = h_t}(l) \mu_{f_t \rightarrow h_t}(l)
\]