Графические модели: Скрытые марковские модели

Александр Адуенко

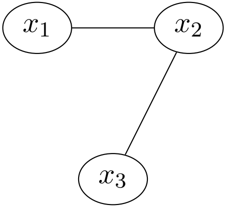
27е апреля 2022

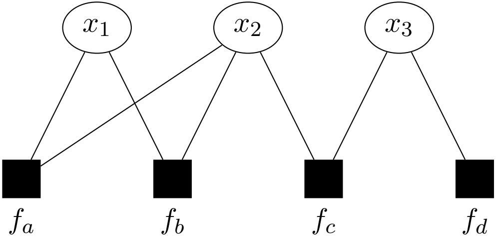
Содержание предыдущих лекций

- Выбор априорного распределения. Неинформативные распределения. Распределение Джефриса.
- EM-алгоритм. Использование EM-алгоритма для отбора признаков в байесовской линейной регрессии.
- Вариационный EM-алгоритм и его использование для вывода в смеси моделей линейной регрессии.
- Гамильтоновы методы Монте-Карло и сравнение с вариационным EM-алгоритмом.
- Ориентированные графические модели и их представление plate notation. Критерий условной независимости d-separation.
- Неориентированные графические модели и их связь с ориентированными.
- Факторные графы и алгоритм Sum-Product для вывода в ациклических графических моделях.

Идея: Построить общее представление для ориентированных и неориентированных моделей.

$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s).$$





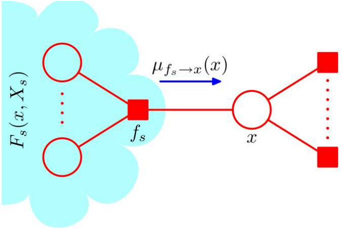
Вопрос: Задает ли граф справа другой набор условных независимостей, чем граф слева?

Утверждение: Если исходная графическая модель есть направленное или ненаправленное дерево, то для нее можно построить ациклический фактор-граф.

Найти: $p(x) = \sum_{\mathbf{x}\backslash x} p(\mathbf{x})$.

$p(\mathbf{x}) = \frac{1}{Z}\prod_s f_s(\mathbf{x}_s) =$

$\prod_{s\in N(x)} F_s(x,\ X_s) = \frac{1}{Z}\tilde{p}(x)$.



Фактор-граф в окрестности вершины x [Bishop, 2006]

$$\tilde{p}(x) = \sum_{\mathbf{x}\backslash x}\tilde{p}(\mathbf{x}) = \sum_{\mathbf{x}\backslash x}\prod_{s\in N(x)} F_s(x,\ X_s) = \prod_{s\in N(x)}\sum_{\mathbf{x}\backslash x} F_s(x,\ X_s) =$$

$$\prod_{s\in N(x)}\sum_{X_s} F_s(x,\ X_s) = \prod_{s\in N(x)} \mu_{f_s\to x}(x).$$

Алгоритм Sum-Product вывода в ациклических ГМ

Получаем следующие формулы пересчета сообщений:

- $\mu_{x_m \to f_s}(x_m) = \prod_{l \in N(x_m) \setminus f_s} \mu_{f_l \to x_m}(x_m);$
- $\mu_{f_s \to x}(x) = \sum_{x_{1:M}} f_s(x,\ x_{1:M}) \prod_{m \in N(f_s) \setminus x} \mu_{x_m \to f_s}(x_m).$

Алгоритм:

1. Объявляем вершину x корнем;
2. От листьев фактор-графа движемся к корню, пересылая сообщения по правилам выше;
3. По достижении корня имеем: $p(x) = \frac{1}{Z} \prod_{s \in N(x)} \mu_{f_s \to x}(x).$

База рекурсии (сообщения от листьев): $\mu_{x \to f} = 1,\ \mu_{f \to x} = f(x).$

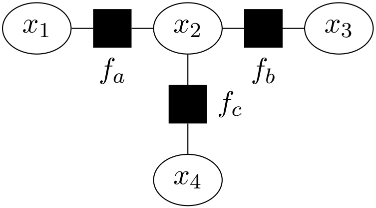
Вопрос 1: Как показать, что процедура работает, то есть все вершины получат достаточно сообщений, чтобы отправить своё?

Вопрос 2: Как получить $p(x_l)\ \forall\ x_l \neq x$?

Вопрос 3: Как определить нормировочную постоянную Z?

Вопрос 4: Как получить $p(\mathbf{x}_s)$?

Пример работы алгоритма Sum-Product



$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_2, x_3) f_c(x_2, x_4).$

Прямой проход (x_3 – корень):

$\mu_{x_1 \to f_a}(x_1) = 1,\ \mu_{x_4 \to f_c}(x_4) = 1,$

$\mu_{f_a \to x_2}(x_2) = \sum_{x_1} f_a(x_1,\ x_2) \mu_{x_1 \to f_a}(x_1) = \sum_{x_1} f_a(x_1,\ x_2),$

$\mu_{f_c \to x_2}(x_2) = \sum_{x_4} f_c(x_2,\ x_4),$

$\mu_{x_2 \to f_b}(x_2) = \mu_{f_a \to x_2}(x_2) \mu_{f_c \to x_2}(x_2),$

$\mu_{f_b \to x_3}(x_3) = \sum_{x_2} f_b(x_2,\ x_3) \mu_{x_2 \to f_b}(x_2).$

Обратный проход:

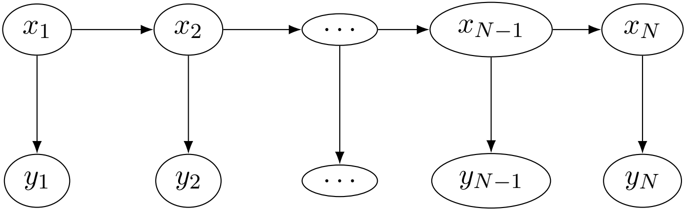
$\mu_{x_3 \to f_b}(x_3) = 1,\ \mu_{f_b \to x_2}(x_2) = \sum_{x_3} f_b(x_2,\ x_3),$

$\mu_{x_2 \to f_a}(x_2) = \mu_{f_b \to x_2}(x_2) \mu_{f_c \to x_2}(x_2),\ \ \mu_{x_2 \to f_c}(x_2) = \mu_{f_a \to x_2}(x_2) \mu_{f_b \to x_2}(x_2),$

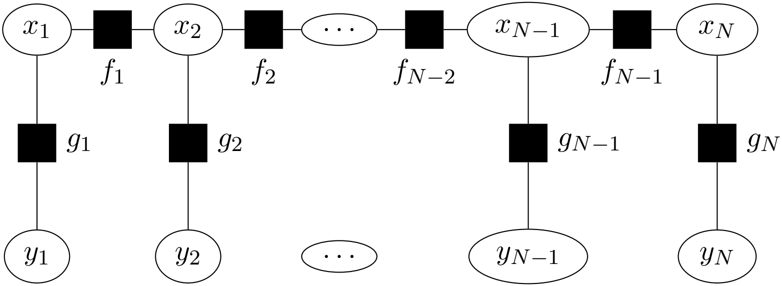
$\mu_{f_a \to x_1}(x_1) = \sum_{x_2} f_a(x_1,\ x_2) \mu_{x_2 \to f_a}(x_2),$

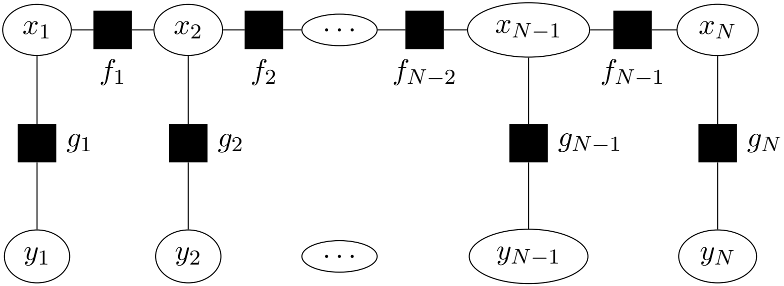
$\mu_{f_c \to x_4}(x_4) = \sum_{x_2} f_c(x_2,\ x_4) \mu_{x_2 \to f_c}(x_2).$

Скрытые марковские модели



$$p(\mathbf{x},\ \mathbf{y}) = \prod_{i=1}^{N} p(x_i|x_{i-1}) \prod_{i=1}^{N} p(y_i|x_i), x_0 = \emptyset.$$





$$p(\mathbf{x},\ \mathbf{y}) = \prod_{i=1}^{N} p(x_i|x_{i-1}) \prod_{i=1}^{N} p(y_i|x_i),\, x_0 = \emptyset.$$

Замечание: С помощью алгоритма Sum-Product можем найти $p(x_i|\mathbf{y})\,\forall\, i$.

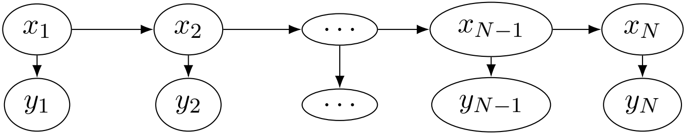
Вопрос 1: Как найти $\mathbf{x}^* = \arg\max\limits_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})$?

Идея: $\tilde{x}_i^* = \arg\max\limits_{x_i} p(x_i|\mathbf{y})$.

Вопрос 2: Верно ли, что $\mathbf{x}^* = \tilde{\mathbf{x}}^*$?

Пример скрытой марковской модели

Пусть $x_t \in [\text{Подъем},\ \text{Зависание},\ \text{Спуск}]$ есть состояние воздушного шара, а $\sqrt{y_t}$ полная скорость.



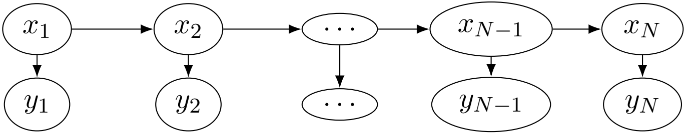
$$p(\mathbf{x},\ \mathbf{y}) = \prod_{i=1}^{N} p(x_i|x_{i-1}) \prod_{i=1}^{N} p(y_i|x_i), x_0 = \emptyset.$$

$$p(x_1) = \boldsymbol{\pi} = [1,\ 0,\ 0]^{\mathsf{T}},\ p(x_t|x_{t-1}) = \mathbf{A} = \begin{array}{c} \\ \text{П} \\ \text{З} \\ \text{С} \end{array} \!\!\begin{array}{c} \begin{array}{ccc} \text{П} & \text{З} & \text{С} \end{array} \\ \begin{pmatrix} 0.98 & 0.02 & 0 \\ 0 & 0.8 & 0.2 \\ 0 & 0 & 1 \end{pmatrix} \end{array}.$$

$$y_t|x_t = v^2_{\text{ветер}} + v^2_{\text{вертик.}} = \underbrace{\varepsilon_t}_{\sim\mathcal{N}(5,\, 3^2)^2} + \underbrace{v^2_{\text{вертик.}}|x_t}_{\text{П: }1,\ \text{З: }0,\ \text{С: }4}.$$

Вопрос: Что можно сказать про $x_t^* = \arg\max_{x_t} p(x_t|\mathbf{y})$?

Алгоритм Витерби



$$p(\mathbf{x},\ \mathbf{y}) = \prod_{i=1}^{N} p(x_i|x_{i-1}) \prod_{i=1}^{N} p(y_i|x_i), x_0 = \emptyset.$$

Задача: $p(\mathbf{x}|\mathbf{y}) \to \max\limits_{\mathbf{x}} \equiv p(\mathbf{x},\ \mathbf{y}) \to \max\limits_{\mathbf{x}}.$

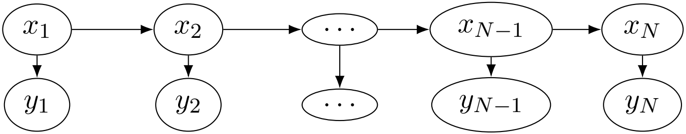
$$V_{1,\,k} = \pi_k p(y_1|x_1 = k),$$
$$V_{t,\,k} = \max_{j \in S} V_{t-1,\,j}\, a_{jk}\, p(y_t|x_t = k).$$

Вопрос 1: Что показывает $V_{t,\,k}$?

Вопрос 2: Как изменятся формулы для $V_{t,\,k}$, если y_t ненаблюдаемо?

Вопрос 3: Что мы получим в $V_{N,\,k}$?

Алгоритм Витерби 2



$$p(\mathbf{x},\ \mathbf{y}) = \prod_{i=1}^{N} p(x_i|x_{i-1}) \prod_{i=1}^{N} p(y_i|x_i), x_0 = \emptyset.$$

Задача: $p(\mathbf{x}|\mathbf{y}) \to \max\limits_{\mathbf{x}} \equiv p(\mathbf{x},\ \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \to \max\limits_{\mathbf{x}}.$

$$V_{1,\,k} = \pi_k p(y_1|x_1 = k),$$
$$V_{t,\,k} = \max_{j \in S} V_{t-1,\,j}\; a_{jk}\; p(y_t|x_t = k).$$

$V_{N,\,k}$ – вероятность наиболее вероятной последовательности состояний, оканчивающейся в $x_N = k$, то есть $V_{N,\,k} = \max\limits_{\mathbf{x} \setminus x_N} p(\mathbf{x},\ \mathbf{y}|x_N = k)$.

Замечание: $x_N^* = \arg\max\limits_{k} V_{N,\,k}$.

Вопрос: Как получить x_{N-1}^* из $\mathbf{x}^* = \arg\max\limits_{\mathbf{x}} p(\mathbf{x},\ \mathbf{y})$?

Идея: Запомнить j^* из $V_{N,\,k} = \max\limits_{j \in S} V_{N-1,\,j}\; a_{jk}\; p(y_N|x_N = k)$.

Задача Sum-Product

$$p(\mathbf{x}) = \frac{1}{Z}\prod_s f_s(\mathbf{x}_s)$$

Найти: $p(x) = \sum_{\mathbf{x}\backslash x} p(\mathbf{x})$.

Свойство:

$ab + ac = a(b + c)$.

Задача Max-Sum

$$g(\mathbf{x}) = \log p(\mathbf{x}) = C + \sum_s \log f_s(\mathbf{x}_s)$$

Найти: $g(x) = \max_{\mathbf{x}\backslash x} g(\mathbf{x})$.

Свойство:

$\max(a + b,\ a + c) = a + \max(b,\ c)$.

Формулы пересчета сообщений для Sum-Product:

- $\mu_{x_m\to f_s}(x_m) = \prod_{l\in N(x_m)\backslash f_s} \mu_{f_l\to x_m}(x_m)$;
- $\mu_{f_s\to x}(x) = \sum_{x_{1:M}} f_s(x,\ x_{1:M}) \prod_{m\in N(f_s)\backslash x} \mu_{x_m\to f_s}(x_m)$.

Формулы пересчета сообщений для Max-Sum:

- $\mu_{x_m\to f_s}(x_m) = \sum_{l\in N(x_m)\backslash f_s} \mu_{f_l\to x_m}(x_m)$;
- $\mu_{f_s\to x}(x) = \max_{x_{1:M}} \log f_s(x,\ x_{1:M}) + \sum_{m\in N(f_s)\backslash x} \mu_{x_m\to f_s}(x_m)$.

$$g(\mathbf{x}) = \log p(\mathbf{x}) = C + \sum_s \log f_s(\mathbf{x}_s)$$

Найти: $g(x) = \max\limits_{\mathbf{x}\backslash x} g(\mathbf{x})$.

Формулы пересчета сообщений для Max-Sum:

- $\mu_{x_m \to f_s}(x_m) = \sum\limits_{l \in N(x_m)\backslash f_s} \mu_{f_l \to x_m}(x_m)$;
- $\mu_{f_s \to x}(x) = \max\limits_{x_{1:M}} \log f_s(x,\ x_{1:M}) + \sum\limits_{m \in N(f_s)\backslash x} \mu_{x_m \to f_s}(x_m)$.

Сообщения из листьев: $\mu_{x \to f} = 0,\ \mu_{f \to x} = \log f(x)$.

Вопрос 1: Как получить $p(\mathbf{x}^*) = \max\limits_{\mathbf{x}} p(\mathbf{x})$?

Вопрос 2: Как получить $\mathbf{x}^* = \arg\max \log p(\mathbf{x})$?

Алгоритм Max-Sum 3

$$g(\mathbf{x}) = \log p(\mathbf{x}) = C + \sum_s \log f_s(\mathbf{x}_s)$$

Найти: $g(x) = \max\limits_{\mathbf{x}\backslash x} g(\mathbf{x})$.

Формулы пересчета сообщений для Max-Sum:

- $\mu_{x_m \to f_s}(x_m) = \sum\limits_{l \in N(x_m)\backslash f_s} \mu_{f_l \to x_m}(x_m)$;
- $\mu_{f_s \to x}(x) = \max\limits_{x_{1:M}} \left[\log f_s(x,\ x_{1:M}) + \sum\limits_{m \in N(f_s)\backslash x} \mu_{x_m \to f_s}(x_m)\right]$.

Сообщения из листьев: $\mu_{x \to f} = 0,\ \mu_{f \to x} = \log f(x)$.

$\max\limits_{\mathbf{x}} g(\mathbf{x}) = \max\limits_{x_R} g(x_R),\ x_R^* = \arg\max\limits_{x_R} g(x_R)$, где x_R – корень ф-дерева.

Вопрос 1: $x_i^* = \arg\max\limits_{x_i} g(x_i)$ для всех вершин для получения $\mathbf{x}^*$?

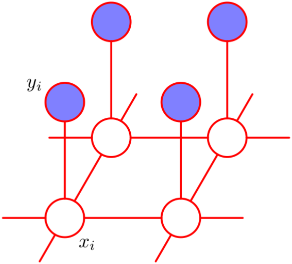
Идея: Хранить конфигурацию $x_{1:M}$, доставляющую максимум в $\mu_{f_s \to x}$.

Вопрос 2: Сколько потребуется памяти для хранения таких конфигураций?

Иллюстрация работы алгоритма Max-Sum

Пример: Пусть имеется бинарное изображение $\mathbf{y}$, $y_i \in \{-1, 1\}$, которое зашумлено. Требуется восстановить исходное изображение $\mathbf{x}$.

$$E(\mathbf{x}, \mathbf{y}) = h\sum_i x_i - \beta \sum_{(i, j)\in\varepsilon} x_i x_j - \eta \sum_i x_i y_i.$$



Графическая модель $p(\mathbf{x}, \mathbf{y})$ [Bishop, 2006]

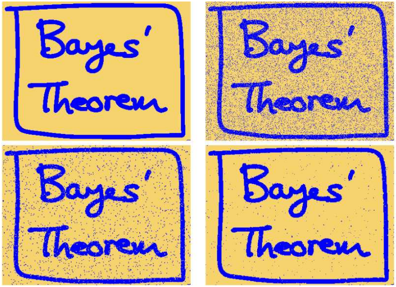


Иллюстрация шумоподавления [Bishop, 2006]

Литература

1. Bishop, Christopher M. "Pattern recognition and machine learning". Springer, New York (2006). Pp. 394-418.
2. Material on HMMs: https://web.stanford.edu/ jurafsky/slp3/A.pdf
3. Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.
4. Bacchus, Fahiem, and Adam J. Grove. "Graphical models for preference and utility."arXiv preprint arXiv:1302.4928 (2013).
5. Mor, Bhavya, Sunita Garhwal, and Ajay Kumar. "A systematic review of hidden markov models and their applications."Archives of computational methods in engineering 28.3 (2021): 1429-1448.
6. Pearl, Judea. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan kaufmann, 1988.
7. Pearl, Judea. "Graphical models for probabilistic and causal reasoning."Quantified representation of uncertainty and imprecision (1998): 367-389.