Графические модели: факторные графы и вывод в графических моделях

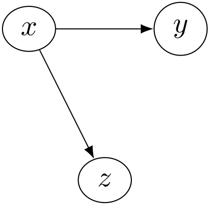
Александр Адуенко

13е апреля 2022

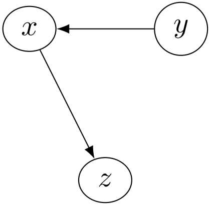
Содержание предыдущих лекций

- 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.

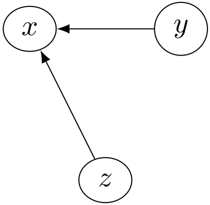
Идея: Представим совместное распределение переменных в виде графа.



Граф 1



Граф 2

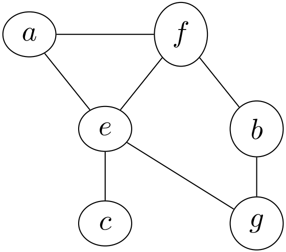


Граф 3

Вопрос: Одинаковые ли совместные распределения соответсвуют графическим представлениям выше?

Граф 1: $p(x,\ y,\ z) = p(x)p(y|x)p(z|x)$;
Граф 2: $p(x,\ y,\ z) = p(y)p(x|y)p(z|x)$;
Граф 3: $p(x,\ y,\ z) = p(z)p(y)p(x|z,\ y)$.

Замечание: Если x_i и x_j не соединены ребром, то они независимы при условии всех остальных переменных:
$p(x_i,\, x_j|\mathbf{x}_{\setminus\{i,\,j\}}) = p(x_i|\mathbf{x}_{\setminus\{i,\,j\}})p(x_j|\mathbf{x}_{\setminus\{i,\,j\}})$.



$$p(a,\, b,\, c,\, e,\, f,\, g) = \frac{1}{Z}\psi_{afe}(a, f, e)\psi_{ec}(e, c)\psi_{eg}(e, g)\psi_{bg}(b, g)\psi_{bf}(b, f).$$

$$Z = \int \prod_i \psi_{C_i}(\mathbf{x}_{C_i}).$$

Вопрос: Какими свойствами должны обладать $\psi_{C_i}(\mathbf{x}_{C_i})$, чтобы задать корректное распределение?

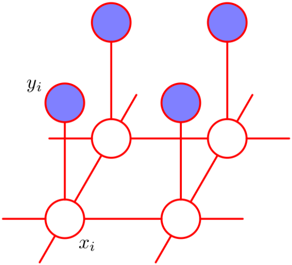
Теорема (Hammersley-Clifford). Предположим, что все потенциалы строго положительны $\psi_C(\mathbf{x}_C) > 0\,\forall\,\mathbf{x}_C$. Тогда факторизация по максимальным кликам графа задает то же множество распределений, что и набор условий условной независимости в терминах графовой разделимости.

Задание потенциала через энергию

Идея: $\psi_{C_i}(\mathbf{x}_{C_i}) = \exp(-E_{C_i}(\mathbf{x}_{C_i}))$.
Тогда $\log p(\mathbf{x}) \propto -E(\mathbf{x}) = -\sum_i E_{C_i}(\mathbf{x}_{C_i})$.

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]

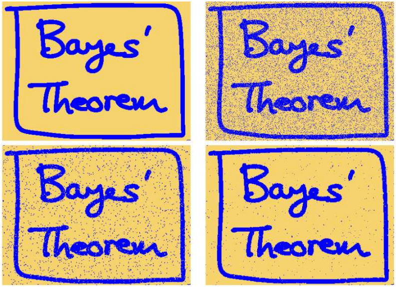


Иллюстрация шумоподавления [Bishop, 2006]

$$x_1 \longrightarrow x_2 \longrightarrow \cdots \longrightarrow x_{N-1} \longrightarrow x_N$$

$$p(\mathbf{x}) = p(x_1)p(x_2|x_1)\cdot\ldots\cdot p(x_{N-1}|x_{N-2})p(x_N|x_{N-1}).$$

$$x_1 \;—\; x_2 \;—\; \cdots \;—\; x_{N-1} \;—\; x_N$$

$$p(\mathbf{x}) = \frac{1}{Z}\psi_{12}(x_1,\, x_2)\psi_{23}(x_2,\, x_3)\cdot\ldots\cdot\psi_{N-1,N}(x_{N-1},\, x_N).$$

Пусть для простоты $x_i \in \{1,\, \ldots,\, K\}$ и требуется найти $p(x_n)$.

$$p(x_n) = \sum_{x_1}\sum_{x_2}\ldots\sum_{x_{n-1}}\sum_{x_{n+1}}\ldots\sum_{x_N} p(\mathbf{x}).$$

Вопрос: Сколько нужно операций, чтобы выполнить суммирование для подсчета $p(x_n)$, то есть $\mathsf{P}(x_n = k)$, $k = 1,\, \ldots,\, K$?

Замечание: Формула суммирования не учитывает структуру модели.

Пример вывода в графической модели

$$x_1 - x_2 - \cdots - x_{n-1} - x_n - x_{n+1} - \cdots - x_{N-1} - x_N$$

$$p(\mathbf{x}) = \frac{1}{Z}\psi_1(x_1,\, x_2)\psi_2(x_2,\, x_3)\cdot\ldots\cdot\psi_{N-1}(x_{N-1},\, x_N).$$

$$p(x_n) = \sum_{x_1}\sum_{x_2}\ldots\sum_{x_{n-1}}\sum_{x_{n+1}}\ldots\sum_{x_N} p(\mathbf{x}).$$

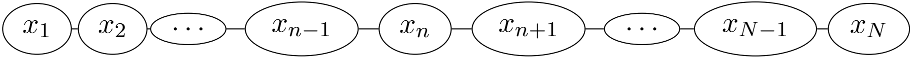
$$p(x_n) = \frac{1}{Z}\sum_{x_{n-1}}\psi_{n-1}(x_{n-1},\, x_n)\cdot\ldots\left[\sum_{x_2}\psi_2(x_2,\, x_3)\cdot\left[\sum_{x_1}\psi_1(x_1,\, x_2)\right]\right]\times$$

$$\sum_{x_{n+1}}\psi_n(x_n,\, x_{n+1})\cdot\ldots\left[\sum_{x_{N-1}}\psi_{N-2}(x_{N-2},\, x_{N-1})\left[\sum_{x_N}\psi_{N-1}(x_{N-1},\, x_N)\right]\right].$$

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

$\mu_\alpha(x_n)$ – сообщение вперед от x_{n-1} к x_n;

$\mu_\beta(x_n)$ – сообщение назад от x_{n+1} к x_n.



$$p(x_n) = \frac{1}{Z}\sum_{x_{n-1}} \psi_{n-1}(x_{n-1},\, x_n)\cdot\ldots\left[\sum_{x_2}\psi_2(x_2,\, x_3)\cdot\left[\sum_{x_1}\psi_1(x_1,\, x_2)\right]\right]\times$$

$$\sum_{x_{n+1}}\psi_n(x_n,\, x_{n+1})\cdot\ldots\left[\sum_{x_{N-1}}\psi_{N-2}(x_{N-2},\, x_{N-1})\left[\sum_{x_N}\psi_{N-1}(x_{N-1},\, x_N)\right]\right].$$

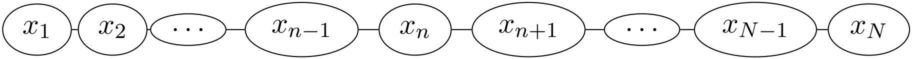
$$p(x_n) = \frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n).$$

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

$$\mu_\beta(x_n) = \sum_{x_{n+1}}\psi_n(x_n,\, x_{n+1})\left[\sum_{x_{n+2}}\ldots\right] = \sum_{x_{n+1}}\psi_n(x_n,\, x_{n+1})\mu_\beta(x_{n+1}).$$

Вопрос: Как определить базу рекурсии?

Пример вывода в графической модели



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

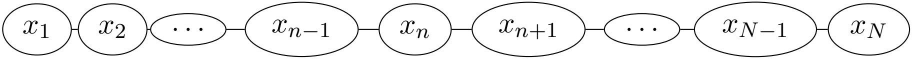
$$\mu_\alpha(x_n) = \sum_{x_{n-1}} \psi_{n-1}(x_{n-1},\, x_n)\mu_\alpha(x_{n-1}).$$

$$\mu_\beta(x_n) = \sum_{x_{n+1}} \psi_n(x_n,\, x_{n+1})\mu_\beta(x_{n+1}).$$

База рекурсии: $\mu_\alpha(x_1) = 1,\ \mu_\beta(x_N) = 1.$

Вопрос 1: Сколько операций требуется для подсчета $\mu_\alpha(x_n),\ \mu_\beta(x_n)$?
Вопрос 2: Как определить Z?
Вопрос 3: Как обобщить результат на случай непрерывных переменных?
Вопрос 4: Как изменится вывод, если часть переменных наблюдаемы, то есть ищем $p(x_n|x_{i_1},\, \ldots,\, x_{i_l})$?

Пример вывода в графической модели



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

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

$$\mu_\beta(x_n) = \sum_{x_{n+1}} \psi_n(x_n,\, x_{n+1})\mu_\beta(x_{n+1}).$$

База рекурсии: $\mu_\alpha(x_1) = 1,\ \mu_\beta(x_N) = 1.$

Вопрос: Как найти $p(x_l),\ \forall\, l \neq n$?

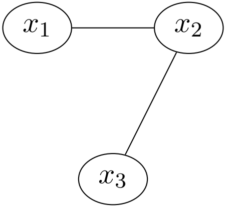
Идея: Сосчитать $\mu_\alpha(x_q),\ q = 1,\, \ldots,\, N$ и $\mu_\beta(x_q),\ q = N,\, \ldots,\, 1.$

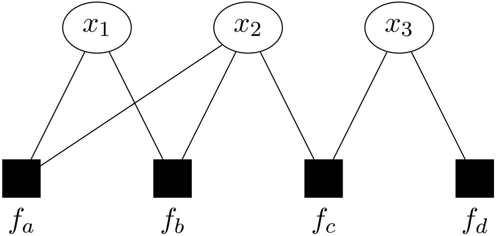
Задание: Показать $p(x_{n-1},\, x_n) = \dfrac{1}{Z}\mu_\alpha(x_{n-1})\psi_{n-1}(x_{n-1},\, x_n)\mu_\beta(x_n).$

Фактор-графы и их построение по графической модели

Идея: Построить общее представление для ориентированных и неориентированных моделей.

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





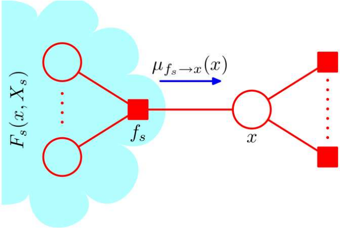
Вопрос: Задает ли граф справа другой набор условных независимостей, чем граф слева?

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]

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

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

$$F_s(x,\ X_s) = f_s(x,\ x_1,\ \dots,\ x_M) G_1(x_1,\ X_{s1}) \cdot \ldots \cdot G_M(x_M,\ X_{sM}).$$

$$\mu_{f_s \to x}(x) = \sum_{X_s} F_s(x,\ X_s) =$$

$$\sum_{x_{1:M}} f_s(x,\ x_{1:M}) \prod_{m \in N(f_s)\backslash x} \left[\sum_{X_{sm}} G_m(x_m,\ X_{sm}) \right] =$$

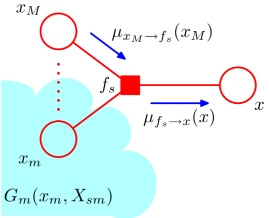
$$\sum_{x_{1:M}} f_s(x,\ x_{1:M}) \prod_{m \in N(f_s)\backslash x} \mu_{x_m \to f_s}(x_m),$$ где

$$\mu_{x_m \to f_s}(x_m) = \sum_{X_{sm}} G_m(x_m,\ X_{sm}).$$

$$G_m(x_m,\ X_{sm}) = \prod_{l \in N(x_m)\backslash f_s} F_l(x_m,\ X_{ml}).$$

$$\mu_{x_m \to f_s}(x_m) = \sum_{X_{sm}} \prod_{l \in N(x_m)\backslash f_s} F_l(x_m,\ X_{ml}) =$$

$$\prod_{l \in N(x_m)\backslash f_s} \left[\sum_{X_{ml}} F_l(x_m,\ X_{ml}) \right] = \prod_{l \in N(x_m)\backslash f_s} \mu_{f_l \to x_m}(x_m).$$



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

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: $p(x) = \dfrac{1}{Z} \prod_{s \in N(x)} \mu_{f_s \to x}(x)$.

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 $p(\mathbf{x}_s)$?

Литература

1. Bishop, Christopher M. "Pattern recognition and machine learning". Springer, New York (2006). Pp. 394-411.
2. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016. Pp. 549-576.
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.