Графические модели: введение

Александр Адуенко

30е марта 2022

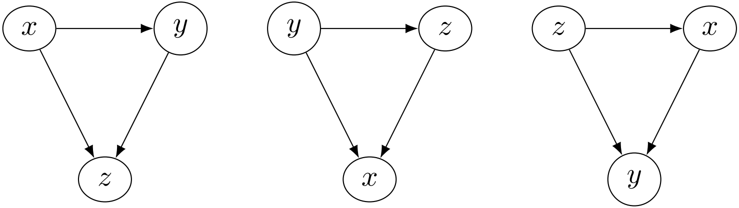
Содержание предыдущих лекций

- Выбор априорного распределения. Неинформативные распределения. Распределение Джефриса.
- ЕМ-алгоритм. Использование ЕМ-алгоритма для отбора признаков в байесовской линейной регрессии.
- Вариационный ЕМ-алгоритм и его использование для вывода в смеси моделей линейной регрессии.
- Гамильтоновы методы Монте-Карло и сравнение с вариационным ЕМ-алгоритмом.

Графические модели: примеры

Идея: Представим совместное распределение переменных в виде графа.
Пример: $p(x, y, z) = p(x)p(y|x)p(z|x, y)$.

Графическая вероятностная модель $p(x, y, z)$

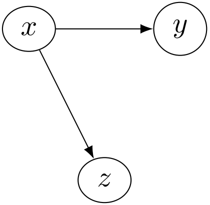


Вопрос 1: Чему соответствуют представления на средней и правой картинке?

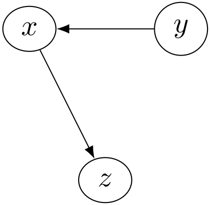
$p(x_1, \ldots, x_K) = p(x_1)p(x_2|x_1) \cdot \ldots \cdot p(x_K|x_1, \ldots, x_{K-1})$.
Вопрос 2: Для каких распределений выполнено разложение выше?
Вопрос 3: Какое представление получается для $p(x_1, \ldots, x_K)$ при таком разложении?

Графические модели: примеры

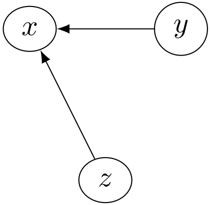
Идея: Представим совместное распределение переменных в виде графа.



Граф 1



Граф 2



Граф 3

Вопрос: Одинаковые ли совместные распределения соответсвуют графическим представлениям выше?

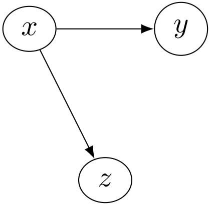
Граф 1: $p(x,\ y,\ z) = p(x)p(y|x)p(z|x)$;
Граф 2: $p(x,\ y,\ z) = p(y)p(x|y)p(z|x)$;
Граф 3: $p(x,\ y,\ z) = p(z)p(y)p(x|z,\ y)$.

Понятие условной независимости

Независимость: $p(y,\ z) = p(y)p(z)$.
Условная независимость: $p(y,\ z|x) = p(y|x)p(z|x)$.
Вопрос: Какое из определений более требовательное? Следует ли из независимости условная независимость и наоборот?

Свойство: $p(x|y,\ z) \propto p(x|y)p(x|z)$.

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: $p(x,\ y,\ z) = p(z)p(y)p(x|z,\ y)$.



Граф 1

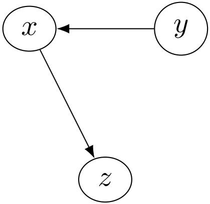
$p(y,\ z|x) = p(z|x)p(y|x,\ z) = p(z|x)p(y|x) \Longrightarrow$ $(y,\ z)$ условно независимы при x.

$p(y,\ z) = \int p(x)p(y|x)p(z|x)dx \neq p(y)p(z) \Longrightarrow$ $(y,\ z)$ зависимы.

Пример:

$x \to \mathbf{w},\ y \to y_1 = \mathbf{w}^{\mathsf{T}}\mathbf{x}_1 + \varepsilon_1,\ z \to y_2 = \mathbf{w}^{\mathsf{T}}\mathbf{x}_2 + \varepsilon_2.$

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: $p(x,\ y,\ z) = p(z)p(y)p(x|z,\ y)$.



Граф 2

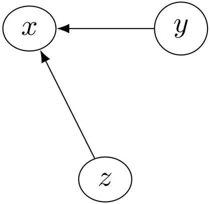
$p(y,\ z|x) = p(z|x)p(y|x,\ z) = p(z|x)p(y|x) \Longrightarrow$
$(y,\ z)$ условно независимы при x.

$$p(y,\ z) = \int p(y)p(x|y)p(z|x)dx \neq p(y)p(z) \Longrightarrow$$

$(y,\ z)$ зависимы.

Пример:

$x \to \mathbf{w}$; $y \to \mathbf{A}$, $\alpha_j \sim \Gamma(\nu,\ \eta)$; $z \to y = \mathbf{w}^{\mathsf{T}}\mathbf{x} + \varepsilon$.



Граф 3

$p(y,\ z|x) \neq p(z|x)p(y|x)$, т.к.
$p(x|y,\ z) \neq p(x|y)p(x|z) \Longrightarrow$
$(y,\ z)$ условно зависимы при x.

$$p(y,\ z) = \int p(y)p(z)p(x|y,\ z)dx = p(y)p(z) \Longrightarrow$$

$(y,\ z)$ независимы.

Вопрос: Приведите пример модели с таким правдоподобием.

Смесь моделей линейной регрессии

Вероятностная модель генерации данных

- Веса моделей в смеси $\boldsymbol{\pi}$ получены из априорного распределения $p(\boldsymbol{\pi}|\mu)$;
- Векторы параметров моделей $\mathbf{w}_k$ получены из нормального распределения $p(\mathbf{w}_k|\mathbf{A}_k) = \mathcal{N}(\mathbf{w}_k|\mathbf{0},\ \mathbf{A}_k^{-1}),\ k = 1, \ldots, K$;
- Для каждого объекта $\mathbf{x}_i$ выбрана модель f_{k_i}, которой он описывается, причем $p(k_i = k) = \pi_k$;
- Для каждого объекта $\mathbf{x}_i$ целевая переменная y_i определена в соответствии с моделью f_{k_i}: $y_i \sim \mathcal{N}(y_i|\mathbf{w}_{k_i}^{\mathsf{T}}\mathbf{x}_i,\ \sigma_{k_i}^2)$.

Совместное правдоподобие модели

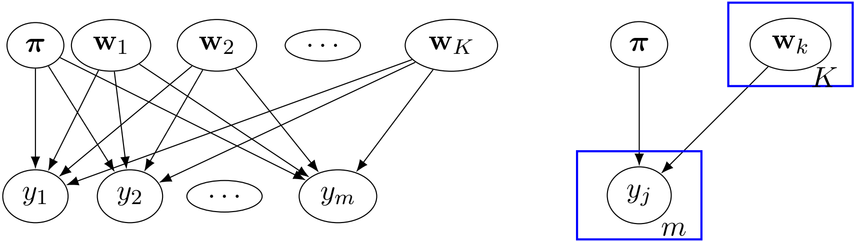
$$p(\mathbf{y},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K,\ \boldsymbol{\pi}|\mathbf{X},\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K,\ \sigma_1^2,\ \ldots,\ \sigma_K^2,\ \boldsymbol{\mu}) =$$
$$p(\boldsymbol{\pi}|\boldsymbol{\mu})\prod_{k=1}^{K}\mathcal{N}(\mathbf{w}_k|\mathbf{0},\ \mathbf{A}_k^{-1})\prod_{i=1}^{m}\left(\sum_{l=1}^{K}\pi_l\mathcal{N}(y_i|\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i,\ \sigma_l^2)\right).$$

Введем матрицу скрытых переменных $\mathbf{Z} = \|z_{ik}\|$, где $z_{ik} = 1 \iff k_i = k$.

$$p(\mathbf{y},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K,\ \boldsymbol{\pi},\ \mathbf{Z}|\mathbf{X},\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_K,\ \sigma_1^2,\ \ldots,\ \sigma_K^2,\ \boldsymbol{\mu}) =$$
$$p(\boldsymbol{\pi}|\boldsymbol{\mu})\prod_{k=1}^{K}\mathcal{N}(\mathbf{w}_k|\mathbf{0},\ \mathbf{A}_k^{-1})\prod_{i=1}^{m}\prod_{l=1}^{K}\left(\pi_l\mathcal{N}(y_i|\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i,\ \sigma_l^2)\right)^{z_{il}}.$$

Представление смеси моделей в виде графа

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



Вопрос 1: Как в представлении учитывается то, что смесь составлена из моделей линейной регрессии?

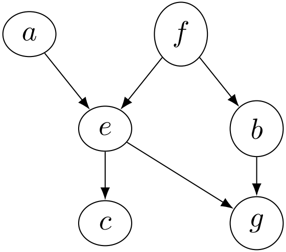
Вопрос 2: Как учесть, что $p(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\mu})$?

Вопрос 3: Как указать наличие наблюдаемого признакового описания $\mathbf{x}_1, \ldots, \mathbf{x}_m$ и гиперпараметров модели $\mathbf{A}$, $\boldsymbol{\sigma}^2$?

Вопрос 4: Зачем нам графическое представление вероятностных моделей?

Рассмотрим две группы переменных A, B и проверим их условную независимость при условии группы переменных C.

D-separation. Группы переменных A и B условно независимы, если все неориентированные пути из A в B блокированы C.



Путь из вершины a в вершину b называется блокированным набором вершин C, если выполнено хотя бы одно из условий

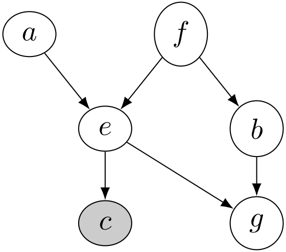
- Стрелки на пути встречаются перед-хвост или хвост-хвост в вершине $c \in C$;
- Стрелки на пути встречаются перед-перед в вершине x, такой, что $x \notin C$ и все ее ориентированные потомки $y \notin C$.

Пути из a **в** b**:** $a \to e \to g \leftarrow b$; $a \to e \leftarrow f \to b$.

Вопрос: Зависимы ли переменные a и b?

Рассмотрим две группы переменных A, B и проверим их условную независимость при условии группы переменных C.

D-separation. Группы переменных A и B условно независимы, если все неориентированные пути из A в B блокированы C.



Путь из вершины a в вершину b называется блокированным набором вершин C, если выполнено хотя бы одно из условий

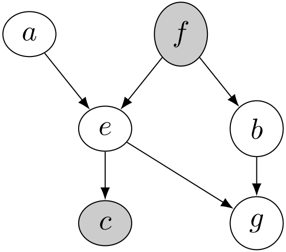
- Стрелки на пути встречаются перед-хвост или хвост-хвост в вершине $c \in C$;
- Стрелки на пути встречаются перед-перед в вершине x, такой, что $x \notin C$ и все ее ориентированные потомки $y \notin C$.

Пути из a **в** b**:** $a \to e \to g \leftarrow b$; $a \to e \leftarrow f \to b$.

Вопрос: Независимы ли переменные a и b при условии c?

Рассмотрим две группы переменных A, B и проверим их условную независимость при условии группы переменных C.

D-separation. Группы переменных A и B условно независимы, если все неориентированные пути из A в B блокированы C.



Путь из вершины a в вершину b называется блокированным набором вершин C, если выполнено хотя бы одно из условий

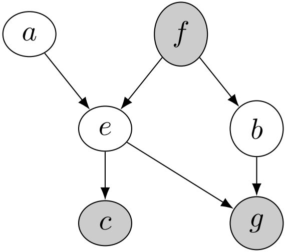
- Стрелки на пути встречаются перед-хвост или хвост-хвост в вершине $c \in C$;
- Стрелки на пути встречаются перед-перед в вершине x, такой, что $x \notin C$ и все ее ориентированные потомки $y \notin C$.

Пути из a **в** b**:** $a \to e \to g \leftarrow b$; $a \to e \leftarrow f \to b$.

Вопрос: Независимы ли переменные a и b при условии c, f?

Рассмотрим две группы переменных A, B и проверим их условную независимость при условии группы переменных C.

D-separation. Группы переменных A и B условно независимы, если все неориентированные пути из A в B блокированы C.



Путь из вершины a в вершину b называется блокированным набором вершин C, если выполнено хотя бы одно из условий

- Стрелки на пути встречаются перед-хвост или хвост-хвост в вершине $c \in C$;
- Стрелки на пути встречаются перед-перед в вершине x, такой, что $x \notin C$ и все ее ориентированные потомки $y \notin C$.

Пути из a **в** b: $a \to e \to g \leftarrow b$; $a \to e \leftarrow f \to b$.

Вопрос: Независимы ли переменные a и b при условии c, f, g?

Литература

1. Bishop, Christopher M. "Pattern recognition and machine learning". Springer, New York (2006). Pp. 359-383.
2. Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.
3. Bacchus, Fahiem, and Adam J. Grove. "Graphical models for preference and utility."arXiv preprint arXiv:1302.4928 (2013).
4. 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.
5. Pearl, Judea. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan kaufmann, 1988.
6. Pearl, Judea. "Graphical models for probabilistic and causal reasoning."Quantified representation of uncertainty and imprecision (1998): 367-389.