

Байесовский выбор моделей II: введение

Александр Адуенко

17е февраля 2024

- Воспоминания из части 1:
 - EM-алгоритм и вариационный EM-алгоритм;
 - Методы сэмплирования (MCMC);
- Понятие графической модели. Байесовские и марковские сети.
- Ациклические графические модели и алгоритм Belief Propagation.
- Скрытые марковские модели (HMM) и их расширения.
- Линейные динамические системы. Фильтр Калмана и его расширения.
- Циклические графические модели и вывод в них.
- Сегментация изображений: Алгоритмы на основании разрезов графов и алгоритм α – расширение.
- Приближенный вывод в графических моделях, алгоритм Expectation Propagation.
- Понятие разладки и методы ее нахождения.

Система оценивания (пример)

- 12 лекций + 1-2 небольших теста на них (суммарно до 100 баллов);
- 3 задания:
 - 1 теоретическое (150 баллов);
 - 1 практическое (150 баллов);
 - 1 групповое соревнование (до 200 баллов);
- Экзамен:
 - Письменная часть (150 баллов),
 - Устная часть (200 баллов).

Замечания:

- Задания содержат задачи более, чем на 150 баллов, поэтому можно выбрать, что выполнять;
- В каждом задании баллы лучшей работы удваиваются, если она оценена более, чем в 120 баллов (не более 250 баллов);
- За каждую неделю опоздания балл за задание снижается в 2 раза. Задание не принимается после его разбора или объявления об этом.

EM-алгоритм

Пусть $\mathbf{D} = (\mathbf{X}, \mathbf{y})$ – наблюдаемые переменные, \mathbf{Z} – скрытые переменные.
 $p(\mathbf{D}, \mathbf{Z}|\Theta) = p(\mathbf{D}|\mathbf{Z}, \Theta)p(\mathbf{Z}|\Theta)$.

Вопрос 1: как решить задачу $p(\mathbf{D}|\Theta) = \int p(\mathbf{D}, \mathbf{Z}|\Theta)d\mathbf{Z} \rightarrow \max_{\Theta}$?

Пример 1. $\mathbf{y} = \mathbf{X}\mathbf{w} + \varepsilon$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1})$, $\varepsilon \sim \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I})$

$p(\mathbf{y}, \mathbf{w}|\mathbf{X}, \mathbf{A}, \beta) = p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \beta)p(\mathbf{w}|\mathbf{A})$.

$\log p(\mathbf{y}|\mathbf{X}, \underbrace{\mathbf{A}, \beta^{-1}}_{\Theta}) \propto -\frac{1}{2} \log \det(\beta^{-1}\mathbf{I} + \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^{\top}) - \frac{1}{2}\mathbf{y}^{\top}(\beta^{-1}\mathbf{I} + \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^{\top})^{-1}\mathbf{y}$.

EM-алгоритм

Введем $F(q, \Theta) = - \int q(\mathbf{Z}) \log q(\mathbf{Z})d\mathbf{Z} + \int q(\mathbf{Z}) \log p(\mathbf{D}, \mathbf{Z}|\Theta)d\mathbf{Z} =$
 $- \int q(\mathbf{Z}) \log q(\mathbf{Z})d\mathbf{Z} + \int q(\mathbf{Z}) \log p(\mathbf{Z}|\mathbf{D}, \Theta)d\mathbf{Z} + \int \log p(\mathbf{D}|\Theta)q(\mathbf{Z})d\mathbf{Z} =$
 $\log p(\mathbf{D}|\Theta) - \int q(\mathbf{Z}) \log \frac{q(\mathbf{Z})}{p(\mathbf{Z}|\mathbf{D}, \Theta)}d\mathbf{Z} = \log p(\mathbf{D}|\Theta) - D_{\text{KL}}(q||p(\mathbf{Z}|\mathbf{D}, \Theta))$.

Идея 1: $p(\mathbf{D}|\Theta) \rightarrow \max_{\Theta}$ заменим на $F(q, \Theta) \rightarrow \max_{q, \Theta}$.

Идея 2: Пошагово оптимизируем по Θ и q , то есть

1 E-шаг: $q^s = F(q, \Theta^{s-1}) \rightarrow \max_q$;

2 M-шаг: $\Theta^s = F(q^s, \Theta) \rightarrow \max_{\Theta}$.

EM-алгоритм для максимизации обоснованности

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon}, \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1}), \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I})$$

$$p(\mathbf{y}, \mathbf{w} | \mathbf{X}, \mathbf{A}, \beta) = p(\mathbf{y} | \mathbf{X}, \mathbf{w}, \beta) p(\mathbf{w} | \mathbf{A}) = .$$

$$\log p(\mathbf{y}, \mathbf{w} | \mathbf{X}, \mathbf{A}, \beta) \propto \frac{m}{2} \log \beta - \frac{\beta}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 + \frac{1}{2} \log \det \mathbf{A} - \frac{1}{2} \mathbf{w}^\top \mathbf{A} \mathbf{w}.$$

$$F(q, \mathbf{A}, \beta) = - \int q(\mathbf{w}) \log q(\mathbf{w}) d\mathbf{w} + \int q(\mathbf{w}) \log p(\mathbf{y}, \mathbf{w} | \mathbf{X}, \mathbf{A}, \beta) d\mathbf{w} = \\ \log p(\mathbf{y} | \mathbf{X}, \mathbf{A}, \beta) - D_{\text{KL}}(q(\mathbf{w}) \| p(\mathbf{w} | \mathbf{y}, \mathbf{X}, \mathbf{A}, \beta)) \rightarrow \max_{q, \mathbf{A}, \beta}.$$

E-шаг (считаем \mathbf{A} , β фиксированными)

$$F(q, \mathbf{A}, \beta) \rightarrow \max_q \iff q(\mathbf{w}) = p(\mathbf{w} | \mathbf{y}, \mathbf{X}, \mathbf{A}, \beta) = \mathcal{N}(\mathbf{w}_0, \boldsymbol{\Sigma}_0^{-1}), \text{ где}$$

$$\boldsymbol{\Sigma}_0 = \mathbf{A} + \beta \mathbf{X}^\top \mathbf{X}, \mathbf{w}_0 = \beta \boldsymbol{\Sigma}_0^{-1} \mathbf{X}^\top \mathbf{y}.$$

M-шаг (считаем $q(\mathbf{w})$ фиксированным)

$$E_{q(\mathbf{w})} \log p(\mathbf{y}, \mathbf{w} | \mathbf{X}, \mathbf{A}, \beta) = \int q(\mathbf{w}) \log p(\mathbf{y}, \mathbf{w} | \mathbf{X}, \mathbf{A}, \beta) d\mathbf{w} \rightarrow \max_{\mathbf{A}, \beta}.$$

$$\tilde{F}(\mathbf{A}, \beta) = \frac{m}{2} \log \beta - \frac{\beta}{2} \mathbb{E} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 + \frac{1}{2} \sum_{j=1}^n \log \alpha_j - \frac{1}{2} \sum_{j=1}^n \alpha_j \mathbb{E} w_j^2 \rightarrow \max_{\mathbf{A}, \beta}$$

$$\frac{\partial F}{\partial \alpha_j} = \frac{1}{2\alpha_j} - \frac{1}{2} \mathbb{E} w_j^2 = 0 \iff \alpha_j = \frac{1}{\mathbb{E} w_j^2}.$$

$$\text{Hint: } 1 = \alpha_j (\mathbb{E}^2 w_j + D w_j) \implies \alpha_j^{\text{new}} = \frac{1 - \alpha_j^{\text{old}} D w_j}{\mathbb{E}^2 w_j}.$$

$$\frac{\partial F}{\partial \beta} = \frac{m}{2\beta} - \frac{1}{2} \mathbb{E} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 = 0 \iff \beta = \frac{m}{\mathbb{E} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2}.$$

EM-алгоритм для максимизации обоснованности

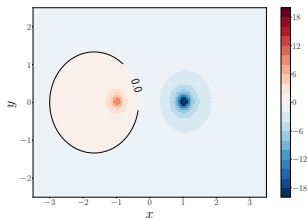
Потенциал поля точечного заряда: $\varphi = k \frac{q}{r}$.

Пусть имеется несколько зарядов q_1, \dots, q_l в точках $\mathbf{z}_1, \dots, \mathbf{z}_l$.

Тогда $\varphi(\mathbf{x}) = k \sum_{i=1}^l \frac{q_i}{\|\mathbf{x} - \mathbf{z}_i\|}$. По набору точек $\mathbf{x}_1, \dots, \mathbf{x}_m$ и измеренным

$$y_i = \varphi(\mathbf{x}_i) - \underbrace{\varphi(\infty)}_{=0} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \beta^{-1})$$

требуется оценить $\varphi(\mathbf{x})$ для \mathbf{x} из тестовой выборки.



$$\mathbf{y} = \Phi \mathbf{w} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \beta^{-1} \mathbf{I}), \text{ где}$$

$$\Phi = \left\| \frac{1}{\delta + \|\mathbf{x}_i - \mathbf{x}_j\|} \right\|, \quad i, j = \overline{1, m};$$

$$\mathbf{w} \sim p(\mathbf{w} | \mathbf{A}) = \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1}).$$

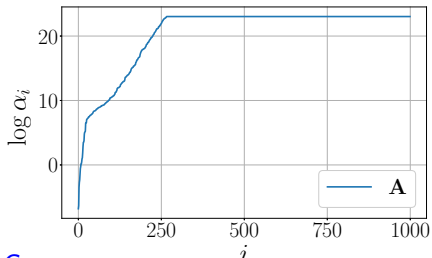
Шаг 1: $p(\mathbf{y}_{\text{train}} | \Phi_{\text{train}}, \mathbf{A}, \beta) \rightarrow \max_{\mathbf{A}, \beta}$ позволит отобрать признаки.

Шаг 2: Прогноз для тестовой выборки:

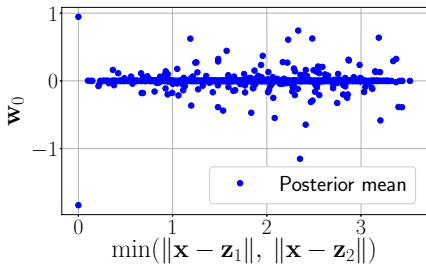
$$p(\mathbf{y}_{\text{test}} | \Phi_{\text{test}}, \Phi_{\text{train}}, \mathbf{y}_{\text{train}}) = \int p(\mathbf{y}_{\text{test}} | \mathbf{w}, \Phi_{\text{test}}) p(\mathbf{w} | \Phi_{\text{train}}, \mathbf{y}_{\text{train}}) d\mathbf{w}$$

Результаты для задачи восстановления потенциала

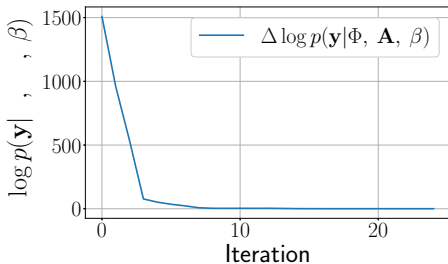
Оптимальный α



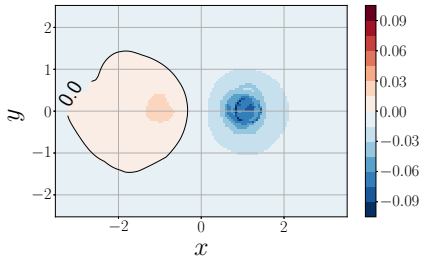
Среднее апостериорного распределения w_0



Обоснованность по итерациям



Восстановленный потенциал



- 1 Bishop, Christopher M. "Pattern recognition and machine learning". Springer, New York (2006). Pp. 113-120, 161-171.
- 2 MacKay, David JC. Bayesian methods for adaptive models. Diss. California Institute of Technology, 1992.
- 3 Gelman, Andrew, et al. Bayesian data analysis, 3rd edition. Chapman and Hall/CRC, 2013.
- 4 Chen, Ming-Hui, and Joseph G. Ibrahim. "Conjugate priors for generalized linear models." *Statistica Sinica* (2003): 461-476.
- 5 Koller, D., & Friedman, N. (2009). Probabilistic graphical models: principles and techniques. MIT press.
- 6 Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. Now Publishers Inc.