Score matching for the exponential family

Набиев Мухаммадшариф Фуркатович

Московский физико-технический институт

9 ноября 2023 г.

Постановка задачи

[1]Пусть имеем статистическую модель, которая задана плотностью из экспоненциального семейства распределений

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} e^{\boldsymbol{\theta}^T \boldsymbol{F}(\boldsymbol{x})}, \tag{1}$$

где $Z(\boldsymbol{\theta})$ — нормализационная константа и $\boldsymbol{F}(\boldsymbol{x})$ — достаточная статистика.
Всюду далее будем рассматривать натуральный логарифм от плотности

$$\ln p(\boldsymbol{x};\boldsymbol{\theta}) = \sum_{k=1}^{K} \theta_k F_k(\boldsymbol{x}) - \log Z(\boldsymbol{\theta}) \tag{2}$$

[1]Michael U. Gutman - Exercises in Machine Learning

Постановка задачи

Нашей задачей является найти минимум следующей функции

$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left[\partial_j\psi_j(\boldsymbol{x}_i;\boldsymbol{\theta}) + \frac{1}{2}\psi_j(\boldsymbol{x}_i;\boldsymbol{\theta})^2\right], \tag{3}$$

где $\psi_j(\boldsymbol{x};\boldsymbol{\theta}) = \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial x_j}$ и $\partial_j\psi_j(\boldsymbol{x};\boldsymbol{\theta}) = \frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial x_j^2}$

Решение

Посчитаем производные

$$\psi_j(\boldsymbol{x};\boldsymbol{\theta}) = \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial x_j} = \sum_{k=1}^{K} \theta_k \frac{\partial F_k(\boldsymbol{x})}{\partial x_j} = \sum_{k=1}^{K} \theta_k K_{kj}(\boldsymbol{x}),$$

$$\partial_j \psi_j(\boldsymbol{x};\boldsymbol{\theta}) = \sum_{k=1}^{K} \theta_k \frac{\partial^2 F_k(\boldsymbol{x})}{\partial x_j^2} = \sum_{k=1}^{K} \theta_k H_{kj}(\boldsymbol{x}),$$

где $K_{kj}(\boldsymbol{x}) = \frac{\partial F_k(\boldsymbol{x})}{\partial x_j}$ и $H_{kj}(\boldsymbol{x}) = \frac{\partial^2 F_k(\boldsymbol{x})}{\partial x_j^2}$. По этим элементам построим матрицы $\boldsymbol{K}(\boldsymbol{x})$ и $\boldsymbol{H}(\boldsymbol{x})$.

Решение

Возведем в квадрат $\psi_j(\boldsymbol{x};\boldsymbol{\theta})$

$$\psi_j(\boldsymbol{x};\boldsymbol{\theta})^2 = \left[\sum_{k=1}^{K} \theta_k K_{kj}(\boldsymbol{x})\right]^2 = \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} K_{k_1 j}(\boldsymbol{x}) K_{k_2 j}(\boldsymbol{x}) \theta_{k_1}\theta_{k_2} \tag{4}$$

Тогда

$$\sum_{j=1}^{m}\psi_j(\boldsymbol{x};\boldsymbol{\theta})^2 = \sum_{j=1}^{m}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} K_{k_1 j}(\boldsymbol{x}) K_{k_2 j}(\boldsymbol{x}) \theta_{k_1}\theta_{k_2} =$$
$$= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{j=1}^{m} K_{k_1 j}(\boldsymbol{x}) K_{k_2 j}(\boldsymbol{x}) \theta_{k_1}\theta_{k_2} =$$
$$= \boldsymbol{\theta}^T \boldsymbol{K}(\boldsymbol{x}) \boldsymbol{K}(\boldsymbol{x})^T \boldsymbol{\theta} \tag{5}$$

Решение

Подставим

$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left[\sum_{k=1}^{K}\theta_k H_{kj}(\boldsymbol{x}) + \frac{1}{2}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} K_{k_1 j}(\boldsymbol{x})K_{k_2 j}(\boldsymbol{x})\theta_{k_1}\theta_{k_2}\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{\theta}^T\boldsymbol{h}_j(\boldsymbol{x}) + \frac{1}{2n}\sum_{i=1}^{n}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{j=1}^{m} K_{k_1 j}(\boldsymbol{x})K_{k_2 j}(\boldsymbol{x})\theta_{k_1}\theta_{k_2} =$$

$$= \boldsymbol{\theta}^T\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{h}_j(\boldsymbol{x}_i)\right] + \frac{1}{2}\boldsymbol{\theta}^T\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{K}(\boldsymbol{x}_i)\boldsymbol{K}(\boldsymbol{x}_i)^T\right]\boldsymbol{\theta} \quad (6)$$

Решение

Обозначив

$$\boldsymbol{r} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{h}_j(\boldsymbol{x}_i) \quad \boldsymbol{M} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{K}(\boldsymbol{x}_i)\boldsymbol{K}(\boldsymbol{x}_i)^T \tag{7}$$

Получим квадратичную форму следующего вида

$$J(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{M}\boldsymbol{\theta} + \boldsymbol{\theta}^T\boldsymbol{r} \quad \square \tag{8}$$

Выводы

- Представив $J(\boldsymbol{\theta})$ как квадратичную форму, мы можем аналитически найти минимум из уравнения $\nabla J(\boldsymbol{\theta}^*) = 0$.
- Для нормальных распределений со средним равным 0, оценка $\boldsymbol{\theta}^*$ полученная нашим методом, совпадает с оценкой по методу максимального правдоподобия.