Смесь экспертов*

А. В. Грабовой[1], В. В. Стрижов[2]

Аннотация:

Ключевые слова: ; .

DOI: 00.00000/00000000000000

1 Введение

2 Постановка задачи

Задана выборка:

$$\mathbf{X} \in \mathbb{R}^{N \times n}, \tag{2.1}$$

где N — количество объектов в выборке, а n — размерность признакового пространства.

2.1 Смесь экспертов

Определение 2.1. *Смесь экспертов — мультимодель, определяющая правдоподобие каждой π_k каждой модели $\boldsymbol{f}_k$ на объекте $\boldsymbol{x}$ на основе его признакового опсиания.*

$$\hat{\boldsymbol{f}} = \sum_{k=1}^{K} \pi_k \boldsymbol{f}_k, \qquad \pi_k\left(\boldsymbol{x}, \boldsymbol{V}\right) : \mathbb{R}^{2\times n} \to [0, 1], \qquad \sum_{k=1}^{K} \pi_k\left(\boldsymbol{x}, \boldsymbol{V}\right) = 1 \tag{2.2}$$

где $\boldsymbol{f}$ — мультимодель, а $\boldsymbol{f}_k$ является некоторой моделью, π_k — параметрическая модель.

*Работа выполнена при поддержке РФФИ и правительства РФ.

[1]Московский физико-технический институт, grabovoy.av@phystech.edu

[2]Московский физико-технический институт, strijov@ccas.ru

Правдоподобие модели:

$$p(\mathbf{y}, \mathbf{Z}|\mathbf{X}, \mathbf{W}, \mathbf{V}, \beta) = \prod_{i=1}^{N}\prod_{k=1}^{K}\left[\pi_k(\mathbf{x}_i, \mathbf{V})\,\mathrm{N}\left(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}\right)\right]^{z_{ik}} \tag{2.3}$$

Логарифм правдоподобия модели:

$$\log p(\mathbf{y}, \mathbf{Z}|\mathbf{X}, \mathbf{W}, \mathbf{V}, \beta) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\left[\log \pi_k(\mathbf{x}_i, \mathbf{V}) - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)^2 + \frac{1}{2}(\log\beta - \log 2\pi)\right] \tag{2.4}$$

E-step

Найдем $q(\mathbf{Z})$:

$$\begin{aligned} q(\mathbf{Z}) &= \mathsf{E}_{q/\mathbf{Z}} \log p(\mathbf{y}, \mathbf{Z}|\mathbf{X}, \mathbf{W}, \mathbf{V}, \beta) \\ p(z_{ik} = 1) &= \frac{\exp\left(\log\pi_k(\mathbf{x}_i, \mathbf{V}) - \frac{\beta}{2}\left[y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]^2\right)}{\sum_{k'=1}^{K}\exp\left(\log\pi_{k'}(\mathbf{x}_i, \mathbf{V}) - \frac{\beta}{2}\left[y_i - \mathbf{w}_{k'}^{\mathsf{T}}\mathbf{x}_i\right]^2\right)} \end{aligned} \tag{2.5}$$

M-step

$\mathsf{E}_q p(\mathbf{y}, \mathbf{Z}|\mathbf{X}, \mathbf{W}, \mathbf{V}, \beta) = \mathcal{F}(\mathbf{W}, \mathbf{V}, \beta)$

$$\mathcal{F}(\mathbf{W}, \mathbf{V}, \beta) = \sum_{i=1}^{N}\sum_{k=1}^{K} \mathsf{E}z_{ik}\left[\log \pi_k(\mathbf{x}_i, \mathbf{V}) - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)^2 + \frac{1}{2}(\log\beta - \log 2\pi)\right] \tag{2.6}$$

Найдем β:

$$\begin{aligned} \frac{\partial\mathcal{F}(\mathbf{W}, \mathbf{V}, \beta)}{\partial\beta} &= \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}\mathsf{E}z_{ik}\left[-\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)^2 + \frac{1}{\beta}\right] = 0 \\ \frac{1}{\beta} &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left[y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]^2\mathsf{E}z_{ik} \end{aligned} \tag{2.7}$$

Найдем $\mathbf{W}$:

$$\begin{aligned} \frac{\partial\mathcal{F}(\mathbf{W}, \mathbf{V}, \beta)}{\partial\mathbf{w}_k} &= -\beta\sum_{i=1}^{N}\mathsf{E}z_{ik}\left[-y_i\mathbf{x}_i + \mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathbf{w}_k\right] = 0 \\ \mathbf{w}_k &= \left[\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}z_{ik}\right]^{-1}\left[\sum_{i=1}^{N}y_i\mathsf{E}z_{ik}\mathbf{x}_i\right] \end{aligned} \tag{2.8}$$

Найдем $\mathbf{V}$:

Аналитически решение не ищется, поэтому воспользуемся градиентным спуском для максимизации правдоподобия модели:

$$\mathbf{V}^{j+1} = \mathbf{V}^j + \alpha\frac{\partial\mathcal{F}\left(\mathbf{W}, \mathbf{V}^j, \beta\right)}{\partial\mathbf{V}} \tag{2.9}$$

3 Вычислительный эксперимент

4 Заключение

Список литературы