Смесь экспертов*

А. В. Грабовой[1], В. В. Стрижов[2]

Аннотация:

Ключевые слова: ; .

DOI: 00.00000/00000000000000

1 Введение

2 Постановка задачи смеси экспертов

Задана выборка:
$$\mathbf{X} \in \mathbb{R}^{N\times n}, \tag{2.1}$$
где N — количество объектов в выборке, а n — размерность признакового пространства.

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

2.1 Общий случай

Правдоподобие модели:

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

Логарифм правдоподобия модели:

$$\begin{aligned}\log p\left(\mathbf{y},\mathbf{W},\mathbf{Z}|\mathbf{X},\mathbf{V},\mathbf{A}\right)=\sum_{i=1}^{N}\sum_{k=1}^{K}z_{ik}\left[\log\pi_k\left(\mathbf{x}_i,\mathbf{V}\right)+\log p_k\left(y_i|\mathbf{w}_k,\mathbf{x}_i\right)\right]+\\+\sum_{k=1}^{K}\log p^k\left(\mathbf{w}_k|\mathbf{A}_k\right)\end{aligned} \tag{2.4}$$

E-step

Найдем $q\left(\mathbf{Z}\right)$:

$$\begin{aligned}\log q\left(\mathbf{Z}\right)&=\mathsf{E}_{q/\mathbf{Z}}\log p\left(\mathbf{y},\mathbf{W},\mathbf{Z}|\mathbf{X},\mathbf{V},\mathbf{A}\right)\\p\left(z_{ik}=1\right)&=\frac{\exp\left(\log\pi_k\left(\mathbf{x}_i,\mathbf{V}\right)+\mathsf{E}\log p_k\left(y_i|\mathbf{w}_k,\mathbf{A}_k\right)\right)}{\sum_{k'=1}^{K}\exp\left(\log\pi_{k'}\left(\mathbf{x}_i,\mathbf{V}\right)+\mathsf{E}\log p_{k'}\left(y_i|\mathbf{w}'_k,\mathbf{A}'_k\right)\right)}\end{aligned} \tag{2.5}$$

Найдем $q\left(\mathbf{W}\right)$:

$$\begin{aligned}\log q\left(\mathbf{W}\right)&=\mathsf{E}_{q/\mathbf{W}}\log p\left(\mathbf{y},\mathbf{W},\mathbf{Z}|\mathbf{X},\mathbf{V},\mathbf{A}\right)\\&=\sum_{i=1}^{N}\sum_{k=1}^{K}\mathsf{E}z_{ik}\left[\log\pi_k\left(\mathbf{x}_{i,\mathbf{V}}\right)+\log p_k\left(y_i\left(y_i|\mathbf{w}_k,\mathbf{x}_i\right)\right)\right]+\\&+\sum_{k=1}^{K}\log p^k\left(\mathbf{w}_k,\mathbf{A}_k\right)\end{aligned} \tag{2.6}$$

M-step

$$\begin{aligned}\mathsf{E}_q p\left(\mathbf{y},\mathbf{W},\mathbf{Z}|\mathbf{X},\mathbf{V},\mathbf{A}\right)=\mathcal{F}\left(\mathbf{V},\mathbf{A}\right)\\\mathcal{F}\left(\mathbf{V},\mathbf{A}\right)=\sum_{i=1}^{N}\sum_{k=1}^{K}\mathsf{E}z_{ik}\left[\log\pi_k\left(\mathbf{x}_i,\mathbf{V}\right)+\mathsf{E}\log p_k\left(y_i|\mathbf{w}_k,\mathbf{x}_i\right)\right]\\+\sum_{k=1}^{K}\mathsf{E}\log p_k\left(\mathbf{w}_k|\mathbf{A}_k\right)\end{aligned} \tag{2.7}$$

Найдем $\mathbf{A}$ из условия:

$$\frac{\partial\mathcal{F}\left(\mathbf{V},\mathbf{A}\right)}{\partial\mathbf{A}^{-1}}=0 \tag{2.8}$$

Найдем $\mathbf{V}$:

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

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

2.2 Случай $p_k\left(y_i|\mathbf{X}, \mathbf{w}_k\right) = \mathbf{N}\left(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}\right)$

E-step

Для нахождения $q\left(\mathbf{Z}\right)$ требуется:

$$\mathsf{E}\log p_k\left(y_i|\mathbf{w}_k, \mathbf{A}_k\right) = -\frac{\beta}{2}\left(y_i^2 - 2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k + \mathbf{x}_i^{\mathsf{T}}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\right)\mathbf{x}_i\right) + \frac{1}{2}\left(\log\beta - \log 2\pi\right) \tag{2.10}$$

Найдем $q\left(\mathbf{w}_k\right)$:

$$\begin{aligned}\log q\left(\mathbf{w}_k\right) &\propto -\frac{1}{2}\left(\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k + \beta\sum_{i=1}^{N}\mathsf{E}z_{ik}\left[y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]^2\right) \propto \\ &\propto -\frac{1}{2}\left(\mathbf{w}_k - \mathbf{m}_k\right)^{\mathsf{T}}\mathbf{B}_k^{-1}\left(\mathbf{w}_k - \mathbf{m}_k\right)\end{aligned} \tag{2.11}$$

где введены обозначения:

$$\mathbf{B}_k = \left(\mathbf{A}_k^{-1} + \beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}z_{ik}\right)^{-1}, \qquad \mathbf{m}_k = \beta\mathbf{B}_k\left(\sum_{i=1}^{N}\mathbf{x}_iy_i\mathsf{E}z_{ik}\right). \tag{2.12}$$

M-step

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

$$\begin{aligned}\mathsf{E}\log p_k\left(y_i|\mathbf{w}_k, \mathbf{x}_i\right) &= -\frac{\beta}{2}\left(y_i^2 - 2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k + \mathbf{x}_i^{\mathsf{T}}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\right)\mathbf{x}_i\right) + \frac{1}{2}\left(\log\beta - \log 2\pi\right) \\ \mathsf{E}\log p_k\left(\mathbf{w}_k|\mathbf{A}_k\right) &= -\frac{1}{2}\mathsf{E}\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k + \frac{1}{2}\log\det\mathbf{A}_k^{-1} - \frac{n}{2}\log 2\pi\end{aligned} \tag{2.13}$$

Найдем β:

$$\begin{aligned}\frac{\partial\mathcal{F}\left(\mathbf{V}, \mathbf{A}, \beta\right)}{\partial\beta} &= \sum_{k=1}^{K}\sum_{i=1}^{N}\left(\frac{1}{\beta}\mathsf{E}z_{ik} - \frac{1}{2}\mathsf{E}z_{ik}\left[y_i^2 - 2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k + \mathbf{x}_i^{\mathsf{T}}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]\right) = 0 \\ \frac{1}{\beta} &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left[y_i^2 - 2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k + \mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]\mathsf{E}z_{ik}\end{aligned} \tag{2.14}$$

Найдем $\mathbf{A}$:

$$\frac{\partial\mathcal{F}\left(\mathbf{V}, \mathbf{A}, \beta\right)}{\partial\mathbf{A}^{-1}} = \frac{1}{2}\mathbf{A}_k - \frac{1}{2}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}} = 0 \Rightarrow \mathbf{A}_k^{new} = \operatorname{diag}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\right) \tag{2.15}$$

Найдем $\mathbf{V}$:

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

$$\mathbf{V}^{j+1} = \mathbf{V}^j + \alpha \frac{\partial \mathcal{F}\left(\mathbf{W}, \mathbf{V}^j, \beta\right)}{\partial \mathbf{V}} \tag{2.16}$$

2.3 Случай $p_k(y_i|\mathbf{X}, \mathbf{w}_k) = \mathbf{Exp}\left(y_i - \mathbf{w}_k^{\mathsf{E}}\mathbf{x}_i | \frac{\beta}{2}\right)$

E-step

Для нахождения $q(\mathbf{Z})$ требуется:

$$\mathsf{E} \log p_k(y_i|\mathbf{w}_k, \mathbf{A}_k) = \log \frac{\beta}{2} - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right) \tag{2.17}$$

Найдем $q(\mathbf{w}_k)$:

$$\begin{aligned} \log q(\mathbf{w}_k) &\propto -\frac{1}{2}\left(\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k + \frac{\beta}{2}\sum_{i=1}^{N} \mathsf{E}z_{ik}\left[y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]\right) \propto \\ &\propto -\frac{1}{2}(\mathbf{w}_k - \mathbf{m}_k)^{\mathsf{T}} \mathbf{A}_k^{-1}(\mathbf{w}_k - \mathbf{m}_k) \end{aligned} \tag{2.18}$$

где введено обозначения:

$$\mathbf{m}_k = \frac{\beta}{2}\mathbf{A}_k\left(\sum_{i=1}^{N}\mathbf{x}_i \mathsf{E}z_{ik}\right). \tag{2.19}$$

M-step

$$\begin{aligned} &\mathsf{E}_q p(\mathbf{y}, \mathbf{W}, \mathbf{Z}|\mathbf{X}, \mathbf{V}, \mathbf{A}, \beta) = \mathcal{F}(\mathbf{V}, \mathbf{A}, \beta). \\ &\mathsf{E}\log p_k(y_i|\mathbf{w}_k, \mathbf{x}_i) = \log\frac{\beta}{2} - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right) \\ &\mathsf{E}\log p_k(\mathbf{w}_k|\mathbf{A}_k) = -\frac{1}{2}\mathsf{E}\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k + \frac{1}{2}\log\det\mathbf{A}_k^{-1} - \frac{n}{2}\log 2\pi \end{aligned} \tag{2.20}$$

Найдем β:

$$\begin{aligned} \frac{\partial \mathcal{F}(\mathbf{V}, \mathbf{A}, \beta)}{\partial \beta} &= \sum_{k=1}^{K}\sum_{i=1}^{N}\mathsf{E}z_{ik}\left[\frac{1}{\beta} - \frac{y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i}{2}\right] = 0 \\ \frac{1}{\beta} &= \frac{1}{2N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)\mathsf{E}z_{ik} \end{aligned} \tag{2.21}$$

Найдем $\mathbf{A}$:

$$\frac{\partial \mathcal{F}(\mathbf{V}, \mathbf{A}, \beta)}{\partial \mathbf{A}^{-1}} = \frac{1}{2}\mathbf{A}_k - \frac{1}{2}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}} = 0 \Rightarrow \mathbf{A}_k^{new} = \operatorname{diag}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\right) \tag{2.22}$$

Найдем $\mathbf{V}$:

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

$$\mathbf{V}^{j+1} = \mathbf{V}^j + \alpha \frac{\partial \mathcal{F}\left(\mathbf{W}, \mathbf{V}^j, \beta\right)}{\partial \mathbf{V}} \tag{2.23}$$

3 Вычислительный эксперимент

4 Заключение

Список литературы