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 моделей

Определение 2.1. *Смесь моделей — мультимодель, ответы которой представляют собой взвешенную сумму ответов всех задействованных моедлей независимо от объекта.*

$$\hat{\boldsymbol{f}} = \sum_{k=1}^{K} \pi_k \boldsymbol{f}_k, \qquad \pi_k = const, \tag{2.2}$$

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

*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

Логистическая регресия. Рассматривается следующая вероятностная модель:

1. Веса моделей в смеси $\boldsymbol{\pi}$ получены из априорного распределения
$$p\left(\boldsymbol{\pi}|\mu\right);$$

2. Векторы параметров моделей $\mathbf{w}_k$ получены из нормального распределения
$$p\left(\mathbf{w}_k|\mathbf{A}_k\right) = \mathcal{N}\left(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k^{-1}\right),\ k = 1, \cdots K;$$

3. Для каждого объекта $\mathbf{x}_i$ выбрана модель $\mathbf{f}_{k_i}$, которой он описывается, причем $p\left(k_i = k\right) = \pi_k$;

4. Для каждого объекта $\mathbf{x}_i$ класс y_i определен в соответсвии с моделью
$$\mathbf{f}_{k_i}:\ y_i \sim \text{ Be}\left(\sigma\left(\mathbf{w}_{k_i}^{\mathsf{T}}\mathbf{x}_i\right)\right),$$
где σ — сигмоидная функция.

Совместное правдоподобие модели:
$$\begin{aligned} p\left(\mathbf{y}, \mathbf{w}_1, \cdots, \mathbf{w}_K, \boldsymbol{\pi}|\mathbf{X}, \mathbf{A}_1, \cdots, \mathbf{A}_K, \boldsymbol{\mu}\right) = \\ = p\left(\boldsymbol{\pi}|\boldsymbol{\mu}\right)\prod_{k=1}^{K}\mathcal{N}\left(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k^{-1}\right)\prod_{i=1}^{N}\left(\sum_{l=1}^{K}\pi_l\sigma\left(y_i\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i\right)\right)\end{aligned} \tag{2.3}$$

Введем скрытые переменные $\mathbf{Z} = ||z_{ik}||$, тогда совместное правдоподобие будет иметь вид:

$$p\left(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}_1, \cdots, \mathbf{A}_K, \boldsymbol{\mu}\right) = \text{Dir}\left(\boldsymbol{\pi}|\boldsymbol{\mu}\right)\prod_{k=1}^{K}\mathcal{N}\left(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k^{-1}\right)\prod_{i=1}^{N}\prod_{l=1}^{K}\left(\pi_l\sigma\left(y_i\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i\right)\right)^{z_{il}} \tag{2.4}$$

Используем вариационное приближение (E-шаг):
$$q\left(\mathbf{Z}, \mathbf{w}_1, \cdots, \mathbf{w}_K, \boldsymbol{\pi}\right) = q\left(\mathbf{Z}\right)q\left(\mathbf{w}_1, \cdots, \mathbf{w}_K\right)q\left(\boldsymbol{\pi}\right) \tag{2.5}$$

Найдем $q\left(\mathbf{Z}\right)$:
$$\begin{aligned}\log q\left(\mathbf{Z}\right) \approx \mathsf{E}_{q/\mathbf{Z}}\log p\left(\mathbf{y}, \mathbf{w}_1, \cdots, \mathbf{w}_K, \mathbf{Z}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{A}_1, \cdots, \mathbf{A}_K, \boldsymbol{\mu}\right) \approx \\ \approx \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\left(\mathsf{E}\log\pi_k + \mathsf{E}\log\sigma\left(y_i\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i\right)\right)\end{aligned} \tag{2.6}$$

Получили распределение $q\left(\mathbf{Z}\right)$:

$$p\left(z_{ik}=1\right)=\frac{\exp \mathsf{E}\log\pi_k+\mathsf{E}\log\sigma\left(y_i\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i\right)}{\sum_p\left(z_{ik}=1\right)} \tag{2.7}$$

Найдем $q\left(\boldsymbol{\pi}\right)$:

$$\begin{aligned}\log q\left(\boldsymbol{\pi}\right)\approx \mathsf{E}_{q/\boldsymbol{\pi}}\log p\left(\mathbf{y},\mathbf{w}_1,\cdots,\mathbf{w}_K,\mathbf{Z},\boldsymbol{\pi}|\mathbf{X},\mathbf{A}_1,\cdots,\mathbf{A}_K,\boldsymbol{\mu}\right)\approx\\ \approx\sum_{k=1}^{K}\log\pi_k\left(\mu_k-1+\sum_{i=1}^{N}\mathsf{E}z_{ik}\right)\end{aligned} \tag{2.8}$$

Получили распределение $q\left(\boldsymbol{\pi}\right)$:

$$\log q\left(\boldsymbol{\pi}\right)=\mathrm{Dir}\left(\boldsymbol{\pi}|\boldsymbol{\mu}+\boldsymbol{\gamma}\right),\qquad \gamma_k=\sum_{i=1}^{N}z_{ik} \tag{2.9}$$

Найдем $q\left(\mathbf{W}\right)$:

$$\begin{aligned}\log q\left(\mathbf{W}\right)\approx \mathsf{E}_{q/\mathbf{W}}\log p\left(\mathbf{y},\mathbf{w}_1,\cdots,\mathbf{w}_K,\mathbf{Z},\boldsymbol{\pi}|\mathbf{X},\mathbf{A}_1,\cdots,\mathbf{A}_K,\boldsymbol{\mu}\right)\approx\\ \approx\sum_{k=1}^{K}\left(-\frac{1}{2}\mathbf{w}_k\mathbf{A}_k\mathbf{w}_k+\sum_{i=1}^{N}\mathsf{E}z_{ik}\log\sigma\left(y_i\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i\right)\right)\end{aligned} \tag{2.10}$$

Сделаем М-шаг:

$$\begin{aligned}\mathsf{E}_{q(\boldsymbol{\pi},\mathbf{W},\mathbf{Z})}\log p(\mathbf{y},\mathbf{W},\boldsymbol{\pi},\mathbf{Z}|\mathbf{X},\mathbf{A},\boldsymbol{\mu})=\mathcal{F}(\mathbf{A})\propto\\ \propto\sum_{k=1}^{K}\left[\left(\mu_k+2\gamma_k-1\right)\mathsf{E}\log\pi_k+\frac{1}{2}\log\det\mathbf{A}_k^{-1}-\frac{1}{2}\mathsf{E}\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k\right]+\\ +\sum_{k=1}^{K}\left[\sum_{i=1}^{N}\mathsf{E}z_{ik}\left(\mathsf{E}\log\pi_k+\mathsf{E}\log\sigma\left(y_i\mathbf{w}_l^{\mathsf{T}}\mathbf{x}_i\right)\right)\right]\end{aligned} \tag{2.11}$$

Оптимизируем $\mathbf{A}_k$:

$$\frac{\partial\mathcal{F}}{\partial\mathbf{A}_k^{-1}}=\frac{1}{2}\mathbf{A}_k-\frac{1}{2}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}=\mathbf{0}\Rightarrow\mathbf{A}_k^{new}=\mathrm{diag}(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}) \tag{2.12}$$

Линейная регрессия. Рассматривается следующая вероятностная модель:

1. Веса моделей в смеси $\boldsymbol{\pi}$ получены из априорного распределения

$$p\left(\boldsymbol{\pi}|\mu\right);$$

2. Векторы параметров моделей $\mathbf{w}_k$ получены из нормального распределения

$$p\left(\mathbf{w}_k|\mathbf{A}_k\right)=\mathcal{N}\left(\mathbf{w}_k|\mathbf{0},\mathbf{A}_k\right),\ k=1,\cdots K;$$

3. Для каждого объекта $\mathbf{x}_i$ выбрана модель $\mathbf{f}_{k_i}$, которой он описывается, причем $p\left(k_i = k\right) = \pi_k$;

4. Для каждого объекта $\mathbf{x}_i$ класс y_i определен в соответсвии с моделью

$$\mathbf{f}_{k_i}:\ y_i \sim\ \mathcal{N}\left(\mathbf{w}_{k_i}^{\mathsf{T}}\mathbf{x}_i + b_k, \beta^{-1}\right)$$

Совместное правдоподобие модели:

$$p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{A}, \mathbf{b}, \beta, \mu) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\mu})\prod_{k=1}^{K} N(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k)\prod_{i=1}^{N}\left(\sum_{k=1}^{K}\pi_k N(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i + b_k, \beta^{-1})\right) \tag{2.13}$$

Введем скрытые переменные $\mathbf{Z} = ||z_{ik}||$, тогда совместное правдоподобие будет иметь вид:

$$p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}, \mathbf{b}, \beta, \mu) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\mu})\prod_{k=1}^{K} N(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k)\prod_{i=1}^{N}\prod_{j=1}^{K}\left(\pi_k N(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i + b_k, \beta^{-1})\right)^{z_{ik}} \tag{2.14}$$

Используем вариационное приближение (Е-шаг):

$$q(\boldsymbol{\pi}, \mathbf{W}, \mathbf{Z}) = q(\boldsymbol{\pi})q(\mathbf{W})q(\mathbf{Z}) \tag{2.15}$$

Найдем $q\left(\mathbf{Z}\right)$:

$$\begin{aligned}\log q(\mathbf{Z}) = \mathsf{E}_{q/Z}\log p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}, \mathbf{b}, \beta, \boldsymbol{\mu}) \propto \\ \propto \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\left(\mathsf{E}_{\pi}\log\pi_k - \frac{\beta}{2}\left[y_i - b_k - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]^2 + \frac{1}{2}\left[\log\beta - \log 2\pi\right]\right) = \\ = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\left(\mathsf{E}\log\pi_k - \frac{\beta}{2}\left[\left(y_i - b_k\right)^2 - 2\left(y_i - b_k\right)\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k + \mathbf{x}_i^{\mathsf{T}}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\right)\mathbf{x}_i\right]\right) \Rightarrow \\ \Rightarrow p(z_{ik} = 1) = \frac{\exp\left(\mathsf{E}\log\pi_k - \frac{\beta}{2}\left[\left(y_i - b_k\right)^2 - 2\left(y_i - b_k\right)\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k + \mathbf{x}_i^{\mathsf{T}}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\right)\mathbf{x}_i\right]\right)}{\sum_k p(z_{ik} = 1)}\end{aligned} \tag{2.16}$$

Найдем $q\left(\boldsymbol{\pi}\right)$:

$$\begin{aligned}\log q(\boldsymbol{\pi}) = \mathsf{E}_{q/\boldsymbol{\pi}}\log p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}, \mathbf{b}, \beta, \boldsymbol{\mu}) \propto \sum_{k=1}^{K}\log\pi_k\left(\mu_k - 1 + \sum_{i=1}^{N}\mathsf{E}z_{ik}\right) \Rightarrow \\ \Rightarrow q(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\mu} + \boldsymbol{\gamma}), \qquad \gamma_k = \sum_{i=1}^{N}\mathsf{E}z_{ik}\end{aligned} \tag{2.17}$$

Найдем $q\left(\mathbf{W}\right)$:

$$\begin{gathered}\log q(\mathbf{W}) = \mathsf{E}_{q/\mathbf{W}} \log p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}, \mathbf{b}, \beta, \boldsymbol{\mu}) \propto \\ \propto \sum_{k=1}^{K} -\frac{1}{2}\left(\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k + \beta\sum_{i=1}^{N}\mathsf{E}z_{ik}\left[y_i - b_k - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]^2\right) \propto \\ \propto -\frac{1}{2}\sum_{k=1}^{K}\left(\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k + \mathbf{w}_k^{\mathsf{T}}\left[\beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}z_{ik}\right]\mathbf{w}_k - 2\beta\mathbf{w}_k^{\mathsf{T}}\left[\sum_{i=1}^{N}\mathbf{x}_i\left(y_i - b_k\right)\mathsf{E}z_{ik}\right]\right) \propto \\ \propto -\frac{1}{2}\sum_{k=1}^{K}\left(\mathbf{w}_k^{\mathsf{T}}\mathbf{B}_k^{-1}\mathbf{w}_k - 2\mathbf{w}_k^{\mathsf{T}}\mathbf{m}_k\right),\end{gathered} \tag{2.18}$$

где введены обозначения

$$\mathbf{B}_k = \left(\mathbf{A}_k^{-1} + \beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}z_{ik}\right)^{-1} \quad \mathbf{m}_k = \beta\mathbf{B}_k\left(\sum_{i=1}^{N}\mathbf{x}_i\left(y_i - b_k\right)\mathsf{E}z_{ik}\right) \tag{2.19}$$

Получаем $q\left(\mathbf{W}\right)$:

$$q(\mathbf{w}_k) = N(\mathbf{w}_k|\mathbf{m}_k, \mathbf{B}_k) \tag{2.20}$$

Сделаем М-шаг:

$$\begin{gathered}\mathsf{E}_{q(\boldsymbol{\pi},\mathbf{W},\mathbf{Z})} \log p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}, \mathbf{b}, \beta, \mu) = \mathcal{F}(\mathbf{A}, \mathbf{b}, \beta) \propto \\ \propto \sum_{k=1}^{K}\left[\left(\mu_k + 2\gamma_k - 1\right)\mathsf{E}\log\pi_k + \frac{1}{2}\log\det\mathbf{A}_k^{-1} - \frac{1}{2}\mathsf{E}\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k\right] + \\ + \sum_{k=1}^{K}\left[\sum_{i=1}^{N}\mathsf{E}z_{ik}\left(\mathsf{E}\log\pi_k + \log\beta - \log 2\pi - \frac{\beta}{2}\mathsf{E}\left(y_i - b_k - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)^2\right)\right]\end{gathered} \tag{2.21}$$

Оптимизируем $\mathbf{A}_k$:

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

Оптимизируем b_k:

$$\begin{gathered}\frac{\partial\mathcal{F}}{\partial b_k} = \sum_{i=1}^{N}\mathsf{E}z_{ik}\left(y_i - b_k - \mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\right) = 0 \Rightarrow \\ \Rightarrow b_k^{new} = \frac{1}{S_k}\sum_{i=1}^{N}y_i\mathsf{E}z_{ik} - \frac{1}{S_k}\sum_{i=1}^{N}\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathsf{E}z_{ik}, \qquad S_k = \sum_{i=1}^{N}\mathsf{E}z_{ik}\end{gathered} \tag{2.23}$$

Оптимизируем β:

$$\frac{\partial \mathcal{F}}{\partial \beta} = \sum_{k=1}^{K}\sum_{i=1}^{N}\left(\frac{1}{\beta}\mathsf{E}z_{ik} - \frac{1}{2}\mathsf{E}z_{ik}\left[(y_i-b_k)^2 - 2(y_i-b_k)\mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k + \mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^\mathsf{T}\mathbf{x}_i\right]\right) = 0 \Rightarrow$$
$$\Rightarrow \frac{1}{\beta^{new}} = \frac{\sum\sum\left[(y_i-b_k)^2 - 2(y_i-b_k)\mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k + \mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^\mathsf{T}\mathbf{x}_i\right]\mathsf{E}z_{ik}}{\sum\sum\mathsf{E}z_{ik}} \tag{2.24}$$

Оптимизируем μ_k (пока нет):

$$\frac{\partial \mathcal{F}}{\partial \mu_k} = ??? \tag{2.25}$$

Важные матожидания:

$$\begin{aligned}\mathsf{E}z_{ik} &= p(z_{ik}=1)\\ \mathsf{E}\log\pi_k &= \psi^0(\mu_k+\gamma_k) - \psi^0(K\mu_k+N)\\ \mathsf{E}\mathbf{w}_k\mathbf{w}_k^\mathsf{T} &= \mathbf{B}_k + \mathbf{m}_k\mathbf{m}_k^\mathsf{T}\end{aligned} \tag{2.26}$$

2.2 Смесь экспертов

Определение 2.2. *Смесь экспертов — мультимодель, определяющая правдоподобие каждой π_k каждой модели $\boldsymbol{f}_k$ на объекте $\boldsymbol{x}$ на основе его признакового опсиания.*

$$\hat{\boldsymbol{f}} = \sum_{k=1}^{K}\pi_k\boldsymbol{f}_k, \qquad \pi_k(\boldsymbol{x},\boldsymbol{v}) : \mathbb{R}^{2\times n} \to [0,1], \tag{2.27}$$

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

Линейная регрессия. Рассматривается следующая вероятностная модель:

1. one

2. two

Совместное правдоподобие модели:

$$p(\mathbf{y},\mathbf{W},\mathbf{V},\mathbf{Z},\boldsymbol{\pi}|\mathbf{X},\mathbf{A},\mathbf{C},\beta) =$$
$$= \prod_{k=1}^{K}\mathcal{N}(\mathbf{v}_k|\mathbf{0},\mathbf{C}_k)\prod_{k=1}^{K}\mathcal{N}(\mathbf{w}_k|\mathbf{0},\mathbf{A}_k)\prod_{i=1}^{N}\left[\left(\sum_{k=1}^{K}\pi_{ik}\mathcal{N}\left(y_i|\mathbf{w}_k^\mathsf{T}\mathbf{x}_i,\beta^{-1}\right)\right)\mathrm{Dir}(\boldsymbol{\pi}_i|\boldsymbol{\mu}_i)\right], \tag{2.28}$$

где $\boldsymbol{\mu}_i = \exp(\mathbf{V}\mathbf{x}_i)$

Введем скрытые переменные $\mathbf{Z} = ||z_{ik}||$, тогда совместное правдоподобие будет иметь вид:

$$\begin{split} p\left(\mathbf{y}, \mathbf{W}, \mathbf{V}, \mathbf{Z}, \boldsymbol{\pi} | \mathbf{X}, \mathbf{A}, \mathbf{C}, \beta\right) = \\ = \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{v}_k | \mathbf{0}, \mathbf{C}_k\right) \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{w}_k | \mathbf{0}, \mathbf{A}_k\right) \prod_{i=1}^{N} \left[\left(\prod_{k=1}^{K} \left(\pi_{ik} \mathcal{N}\left(y_i | \mathbf{w}_k^{\mathsf{T}} \mathbf{x}_i, \beta^{-1}\right)\right)^{z_{ik}}\right) \mathrm{Dir}\left(\boldsymbol{\pi}_i | \boldsymbol{\mu}_i\right)\right], \end{split} \tag{2.29}$$

где $\boldsymbol{\mu}_i = \exp\left(\mathbf{V}\mathbf{x}_i\right)$

Используем вариационное приближение (Е-шаг):

$$q\left(\mathbf{V}, \mathbf{W}, \mathbf{Z}, \boldsymbol{\pi}\right) = q\left(\mathbf{V}\right) q\left(\mathbf{W}\right) q\left(\mathbf{Z}\right) q\left(\boldsymbol{\pi}\right) \tag{2.30}$$

Найдем $q\left(\mathbf{Z}\right)$:

$$\begin{split} \log q\left(\mathbf{Z}\right) = \mathsf{E}_{q/\mathbf{Z}} p\left(\mathbf{y}, \mathbf{W}, \mathbf{V}, \mathbf{Z} | \mathbf{X}, \mathbf{A}, \mathbf{C}, \beta\right) \propto \\ \propto \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left(\mathsf{E} \log \pi_{ik} - \frac{\beta}{2} \left[y_i - \mathbf{w}_k^{\mathsf{T}} \mathbf{x}_i\right]^2\right) = \\ = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left(\mathsf{E} \log \pi_{ik} - \frac{\beta}{2} \left[y_i^2 - 2 y_i \mathbf{x}_i^{\mathsf{T}} \mathsf{E} \mathbf{w}_k + \mathbf{x}_i^{\mathsf{T}} \left(\mathsf{E} \mathbf{w}_k \mathbf{w}_k^{\mathsf{T}}\right) \mathbf{x}_i\right]\right) \Rightarrow \\ \Rightarrow p\left(z_{ik} = 1\right) = \frac{\mathsf{E} \log \pi_{ik} - \frac{\beta}{2} \left[y_i^2 - 2 y_i \mathbf{x}_i^{\mathsf{T}} \mathsf{E} \mathbf{w}_k + \mathbf{x}_i^{\mathsf{T}} \left(\mathsf{E} \mathbf{w}_k \mathbf{w}_k^{\mathsf{T}}\right) \mathbf{x}_i\right]}{\sum_k p\left(z_{ik} = 1\right)} \end{split} \tag{2.31}$$

Найдем $q\left(\mathbf{W}\right)$:

$$\begin{split} \log q\left(\mathbf{W}\right) = \mathsf{E}_{q/\mathbf{W}} p\left(\mathbf{y}, \mathbf{W}, \mathbf{V}, \mathbf{Z} | \mathbf{X}, \mathbf{A}, \mathbf{C}, \beta\right) \propto \\ \propto \sum_{k=1}^{K} -\frac{1}{2} \left(\mathbf{w}_k^{\mathsf{T}} \mathbf{A}_k^{-1} \mathbf{w}_k + \beta \sum_{i=1}^{N} \mathsf{E} z_{ik} \left[y_i - \mathbf{w}_k^{\mathsf{T}} \mathbf{x}_i\right]^2\right) \propto \\ \propto -\frac{1}{2} \sum_{k=1}^{K} \left(\mathbf{w}_k^{\mathsf{T}} \mathbf{A}_k^{-1} \mathbf{w}_k + \mathbf{w}_k^{\mathsf{T}} \left[\beta \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \mathsf{E} z_{ik}\right] \mathbf{w}_k - 2\beta \mathbf{w}_k^{\mathsf{T}} \left[\sum_{i=1}^{N} \mathbf{x}_i y_i \mathsf{E} z_{ik}\right]\right) \propto \\ \propto -\frac{1}{2} \sum_{k=1}^{K} \left(\mathbf{w}_k^{\mathsf{T}} \mathbf{B}_k^{-1} \mathbf{w}_k - 2 \mathbf{w}_k^{\mathsf{T}} \mathbf{m}_k\right), \end{split} \tag{2.32}$$

где введены обозначения

$$\mathbf{B}_k = \left(\mathbf{A}_k^{-1} + \beta \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \mathsf{E} z_{ik}\right)^{-1} \quad \mathbf{m}_k = \beta \mathbf{B}_k \left(\sum_{i=1}^{N} \mathbf{x}_i y_i \mathsf{E} z_{ik}\right) \tag{2.33}$$

Получаем $q\left(\mathbf{W}\right)$:

$$q(\mathbf{w}_k) = N(\mathbf{w}_k | \mathbf{m}_k, \mathbf{B}_k) \tag{2.34}$$

Найдем $q(\boldsymbol{\pi})$:

$$\log q(\boldsymbol{\pi}) = \mathsf{E}_{q/\boldsymbol{\pi}} p(\mathbf{y}, \mathbf{W}, \mathbf{V}, \mathbf{Z}, \boldsymbol{\pi} | \mathbf{X}, \mathbf{A}, \mathbf{C}, \beta) \propto$$
$$\propto \sum_{i=1}^{N}\sum_{k=1}^{K} \log \pi_{ik} \left(\exp\left(\mathbf{v}_k^\mathsf{T}\mathbf{x}_i\right) - 1 + \mathsf{E} z_{ik}\right) \Rightarrow \boldsymbol{\pi}_i \sim \mathrm{Dir}\left(\boldsymbol{\pi}_i | \boldsymbol{\mu}_i + \boldsymbol{\gamma}_i\right), \qquad \gamma_{ik} = \mathsf{E} z_{ik} \tag{2.35}$$

Найдем $q(\mathbf{V})$:

$$\log q(\mathbf{V}) = \mathsf{E}_{q/\mathbf{V}} p(\mathbf{y}, \mathbf{W}, \mathbf{V}, \mathbf{Z}, \boldsymbol{\pi} | \mathbf{X}, \mathbf{A}, \mathbf{C}, \beta) \propto$$
$$\propto \sum_{k=1}^{K} -\frac{1}{2}\left(\mathbf{v}_k^\mathsf{T}\mathbf{C}_k^{-1}\mathbf{v}_k + \sum_{i=1}^{N} \mathsf{E} \log \pi_{ik} \left[\exp\left(\mathbf{v}_k^\mathsf{T}\mathbf{x}_i\right) - 1\right]\right) \tag{2.36}$$

Сделаем М-шаг

Линейная регрессия без prior. Рассматривается следующая вероятностная модель:

1. one
2. two

3 Постановка задачи смеси экспертов

Задана выборка:

$$\mathbf{X} \in \mathbb{R}^{N \times n}, \tag{3.1}$$

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

Определение 3.1. *Смесь экспертов — мультимодель, определяющая правдоподобие каждой π_k каждой модели $\boldsymbol{f}_k$ на объекте $\boldsymbol{x}$ на основе его признакового опсиания.*

$$\hat{\boldsymbol{f}} = \sum_{k=1}^{K} \pi_k \boldsymbol{f}_k, \qquad \pi_k(\boldsymbol{x}, \boldsymbol{V}) : \mathbb{R}^{2\times n} \to [0, 1], \qquad \sum_{k=1}^{K} \pi_k(\boldsymbol{x}, \boldsymbol{V}) = 1 \tag{3.2}$$

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

3.1 Общий случай

Правдоподобие модели:

$$p(\mathbf{y}, \mathbf{W}, \mathbf{Z} | \mathbf{X}, \mathbf{V}, \mathbf{A}) = \prod_{i=1}^{N}\prod_{k=1}^{K}\left[\pi_k(\mathbf{x}_i, \mathbf{V}) p_k(y_i | \mathbf{w}_k, \mathbf{x}_i)\right]^{z_{ik}} \prod_{k=1}^{K} p^k(\mathbf{w}_k | \mathbf{A}_k) \tag{3.3}$$

Логарифм правдоподобия модели:

$$\log p\left(\mathbf{y}, \mathbf{W}, \mathbf{Z} | \mathbf{X}, \mathbf{V}, \mathbf{A}\right) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left[\log \pi_k\left(\mathbf{x}_i, \mathbf{V}\right) + \log p_k\left(y_i | \mathbf{w}_k, \mathbf{x}_i\right)\right] + \\ + \sum_{k=1}^{K} \log p^k\left(\mathbf{w}_k | \mathbf{A}_k\right) \tag{3.4}$$

E-step

Найдем $q\left(\mathbf{Z}\right)$:

$$\begin{aligned} \log q\left(\mathbf{Z}\right) &= \mathsf{E}_{q/\mathbf{Z}} \log p\left(\mathbf{y}, \mathbf{W}, \mathbf{Z} | \mathbf{X}, \mathbf{V}, \mathbf{A}\right) \\ p\left(z_{ik} = 1\right) &= \frac{\exp\left(\log \pi_k\left(\mathbf{x}_i, \mathbf{V}\right) + \mathsf{E} \log p_k\left(y_i | \mathbf{w}_k, \mathbf{A}_k\right)\right)}{\sum_{k'=1}^{K} \exp\left(\log \pi_{k'}\left(\mathbf{x}_i, \mathbf{V}\right) + \mathsf{E} \log p_{k'}\left(y_i | \mathbf{w}'_k, \mathbf{A}'_k\right)\right)} \end{aligned} \tag{3.5}$$

Найдем $q\left(\mathbf{W}\right)$:

$$\begin{aligned} \log q\left(\mathbf{W}\right) &= \mathsf{E}_{q/\mathbf{W}} \log p\left(\mathbf{y}, \mathbf{W}, \mathbf{Z} | \mathbf{X}, \mathbf{V}, \mathbf{A}\right) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} \mathsf{E} z_{ik} \left[\log \pi_k\left(\mathbf{x}_{i, \mathbf{V}}\right) + \log p_k\left(y_i\left(y_i | \mathbf{w}_k, \mathbf{x}_i\right)\right)\right] + \\ &+ \sum_{k=1}^{K} \log p^k\left(\mathbf{w}_k, \mathbf{A}_k\right) \end{aligned} \tag{3.6}$$

M-step

$$\begin{aligned} \mathsf{E}_q p\left(\mathbf{y}, \mathbf{W}, \mathbf{Z} | \mathbf{X}, \mathbf{V}, \mathbf{A}\right) &= \mathcal{F}\left(\mathbf{V}, \mathbf{A}\right) \\ \mathcal{F}\left(\mathbf{V}, \mathbf{A}\right) &= \sum_{i=1}^{N} \sum_{k=1}^{K} \mathsf{E} z_{ik} \left[\log \pi_k\left(\mathbf{x}_i, \mathbf{V}\right) + \mathsf{E} \log p_k\left(y_i | \mathbf{w}_k, \mathbf{x}_i\right)\right] \\ &+ \sum_{k=1}^{K} \mathsf{E} \log p_k\left(\mathbf{w}_k | \mathbf{A}_k\right) \end{aligned} \tag{3.7}$$

Найдем $\mathbf{A}$ из условия:

$$\frac{\partial \mathcal{F}\left(\mathbf{V}, \mathbf{A}\right)}{\partial \mathbf{A}^{-1}} = 0 \tag{3.8}$$

Найдем $\mathbf{V}$:

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

$$\mathbf{V}^{j+1} = \mathbf{V}^j + \alpha \frac{\partial \mathcal{F}\left(\mathbf{W}, \mathbf{V}^j, \beta\right)}{\partial \mathbf{V}} \tag{3.9}$$

3.2 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:

$$\mathsf{E}\log p_k\left(y_i|\mathbf{w}_k,\mathbf{A}_k\right)=-\frac{\beta}{2}\left(y_i^2-2y_i\mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^\mathsf{T}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^\mathsf{T}\right)\mathbf{x}_i\right)+\frac{1}{2}\left(\log\beta-\log 2\pi\right)\quad(3.10)$$

Найдем $q\left(\mathbf{w}_k\right)$:

$$\begin{aligned}\log q\left(\mathbf{w}_k\right)&\propto-\frac{1}{2}\left(\mathbf{w}_k^\mathsf{T}\mathbf{A}_k^{-1}\mathbf{w}_k+\beta\sum_{i=1}^{N}\mathsf{E}z_{ik}\left[y_i-\mathbf{w}_k^\mathsf{T}\mathbf{x}_i\right]^2\right)\propto\\&\propto-\frac{1}{2}\left(\mathbf{w}_k-\mathbf{m}_k\right)^\mathsf{T}\mathbf{B}_k^{-1}\left(\mathbf{w}_k-\mathbf{m}_k\right)\end{aligned}\quad(3.11)$$

где введены обозначения:

$$\mathbf{B}_k=\left(\mathbf{A}_k^{-1}+\beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^\mathsf{T}\mathsf{E}z_{ik}\right)^{-1},\qquad\mathbf{m}_k=\beta\mathbf{B}_k\left(\sum_{i=1}^{N}\mathbf{x}_iy_i\mathsf{E}z_{ik}\right).\quad(3.12)$$

M-step

$$\begin{aligned}&\mathsf{E}_qp\left(\mathbf{y},\mathbf{W},\mathbf{Z}|\mathbf{X},\mathbf{V},\mathbf{A},\beta\right)=\mathcal{F}\left(\mathbf{V},\mathbf{A},\beta\right).\\&\quad\mathsf{E}\log p_k\left(y_i|\mathbf{w}_k,\mathbf{x}_i\right)=-\frac{\beta}{2}\left(y_i^2-2y_i\mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^\mathsf{T}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^\mathsf{T}\right)\mathbf{x}_i\right)+\frac{1}{2}\left(\log\beta-\log 2\pi\right)\\&\quad\mathsf{E}\log p_k\left(\mathbf{w}_k|\mathbf{A}_k\right)=-\frac{1}{2}\mathsf{E}\mathbf{w}_k^\mathsf{T}\mathbf{A}_k^{-1}\mathbf{w}_k+\frac{1}{2}\log\det\mathbf{A}_k^{-1}-\frac{n}{2}\log 2\pi\end{aligned}\quad(3.13)$$

Найдем β:

$$\begin{aligned}\frac{\partial\mathcal{F}\left(\mathbf{V},\mathbf{A},\beta\right)}{\partial\beta}&=\sum_{k=1}^{K}\sum_{i=1}^{N}\left(\frac{1}{\beta}\mathsf{E}z_{ik}-\frac{1}{2}\mathsf{E}z_{ik}\left[y_i^2-2y_i\mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^\mathsf{T}\mathbf{w}_k\mathbf{w}_k^\mathsf{T}\mathbf{x}_i\right]\right)=0\\\frac{1}{\beta}&=\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left[y_i^2-2y_i\mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^\mathsf{T}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^\mathsf{T}\mathbf{x}_i\right]\mathsf{E}z_{ik}\end{aligned}\quad(3.14)$$

Найдем $\mathbf{A}$:

$$\frac{\partial\mathcal{F}\left(\mathbf{V},\mathbf{A},\beta\right)}{\partial\mathbf{A}^{-1}}=\frac{1}{2}\mathbf{A}_k-\frac{1}{2}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^\mathsf{T}=0\Rightarrow\mathbf{A}_k^{new}=\mathrm{diag}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^\mathsf{T}\right)\quad(3.15)$$

Найдем $\mathbf{V}$:

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

$$\mathbf{V}^{j+1}=\mathbf{V}^j+\alpha\frac{\partial\mathcal{F}\left(\mathbf{W},\mathbf{V}^j,\beta\right)}{\partial\mathbf{V}}\quad(3.16)$$

3.3 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:

$$\mathsf{E}\log p_k(y_i|\mathbf{w}_k,\mathbf{A}_k) = \log\frac{\beta}{2} - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right) \tag{3.17}$$

Найдем $q(\mathbf{w}_k)$:

$$\begin{aligned}\log q(\mathbf{w}_k) &\propto -\frac{1}{2}\left(\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k + \frac{\beta}{2}\sum_{i=1}^{N}\mathsf{E}z_{ik}\left[y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]\right) \propto \\ &\propto -\frac{1}{2}(\mathbf{w}_k - \mathbf{m}_k)^{\mathsf{T}}\mathbf{A}_k^{-1}(\mathbf{w}_k - \mathbf{m}_k)\end{aligned} \tag{3.18}$$

где введено обозначения:

$$\mathbf{m}_k = \frac{\beta}{2}\mathbf{A}_k\left(\sum_{i=1}^{N}\mathbf{x}_i\mathsf{E}z_{ik}\right). \tag{3.19}$$

M-step

$$\begin{aligned}&\mathsf{E}_q p(\mathbf{y},\mathbf{W},\mathbf{Z}|\mathbf{X},\mathbf{V},\mathbf{A},\beta) = \mathcal{F}(\mathbf{V},\mathbf{A},\beta).\\ &\mathsf{E}\log p_k(y_i|\mathbf{w}_k,\mathbf{x}_i) = \log\frac{\beta}{2} - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)\\ &\mathsf{E}\log p_k(\mathbf{w}_k|\mathbf{A}_k) = -\frac{1}{2}\mathsf{E}\mathbf{w}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{w}_k + \frac{1}{2}\log\det\mathbf{A}_k^{-1} - \frac{n}{2}\log 2\pi\end{aligned} \tag{3.20}$$

Найдем β:

$$\begin{aligned}\frac{\partial\mathcal{F}(\mathbf{V},\mathbf{A},\beta)}{\partial\beta} &= \sum_{k=1}^{K}\sum_{i=1}^{N}\mathsf{E}z_{ik}\left[\frac{1}{\beta} - \frac{y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i}{2}\right] = 0\\ \frac{1}{\beta} &= \frac{1}{2N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)\mathsf{E}z_{ik}\end{aligned} \tag{3.21}$$

Найдем $\mathbf{A}$:

$$\frac{\partial\mathcal{F}(\mathbf{V},\mathbf{A},\beta)}{\partial\mathbf{A}^{-1}} = \frac{1}{2}\mathbf{A}_k - \frac{1}{2}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}} = 0 \Rightarrow \mathbf{A}_k^{new} = \operatorname{diag}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\right) \tag{3.22}$$

Найдем $\mathbf{V}$:

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

$$\mathbf{V}^{j+1} = \mathbf{V}^{j} + \alpha\frac{\partial\mathcal{F}\left(\mathbf{W},\mathbf{V}^{j},\beta\right)}{\partial\mathbf{V}} \tag{3.23}$$

4 Вычислительный эксперимент

5 Заключение

Список литературы