Смесь экспертов

Грабовой Андрей Валериевич

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Москва
2019 г

План

1. EM–алгоритм
2. Смесь моделей
3. Смесь экспертов

Максимизация обоснованости:

$$\boldsymbol{\Theta} = \arg\max_{\boldsymbol{\Theta}} p\left(\mathbf{y}|\mathbf{X}, \boldsymbol{\Theta}\right) \tag{1}$$

ELBO:

$$\begin{aligned}\mathcal{L}\left(q(\mathbf{Z}), \boldsymbol{\Theta}\right) &= \int q(\mathbf{Z}) \log \frac{p\left(\mathbf{y}, \mathbf{Z}|\boldsymbol{\Theta}, \mathbf{X}\right)}{q\left(\mathbf{Z}\right)} d\mathbf{Z} \\ &= p\left(\mathbf{y}|\mathbf{X}, \boldsymbol{\Theta}\right) - \mathsf{D}_{KL}\left(q\big(\mathbf{Z}\big)||p\big(\mathbf{Z}|\mathbf{y}, \mathbf{X}, \boldsymbol{\Theta}\big)\right)\end{aligned} \tag{2}$$

ЕМ–алгоритм:

1. Е–шаг:
$$q^s\left(\mathbf{Z}\right) = \arg\max_{q\left(\mathbf{z}\right) \in Q} \mathcal{L}\left(q\big(\mathbf{Z}\big), \boldsymbol{\Theta}^{s-1}\right) \tag{3}$$
2. М–шаг:
$$\boldsymbol{\Theta}^s = \arg\max_{\boldsymbol{\Theta}} \mathcal{L}\left(q^s\left(\mathbf{Z}\right), \boldsymbol{\Theta}\right) \tag{4}$$

Вариационный ЕМ–алгоритм (Mean Field Approximation[1]):

1. Е–шаг:
$$\log q\left(\mathbf{Z}_k^s\right) \propto \mathsf{E}_{q/k} \log p\left(\mathbf{y}, \mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}^{s-1}\right) \tag{5}$$
2. М–шаг:
$$\boldsymbol{\Theta}^s = \arg\max_{\boldsymbol{\Theta}} \mathbf{E}_{q^s} \log p\left(\mathbf{y}, \mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}\right) \tag{6}$$

[1] `https://github.com/andriygav/EMprior/blob/master/Lecture/Grabovoy2019MeanField.pdf`

Смесь моделей

Definition

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

$$\hat{\mathbf{f}} = \sum_{k=1}^{K} \pi_k \mathbf{f}_k, \qquad \pi_k = const, \quad \sum_{k=1}^{K} \pi_k = 1, \tag{7}$$

где $\mathbf{f}$ — мультимодель, а $\mathbf{f}_k$ — локальная модель.

Пример 1:

1. Веса моделей в смеси $\boldsymbol{\pi}$ получены из априорного распределения
$$p\left(\boldsymbol{\pi}|\mu\right); \tag{8}$$
2. Вектора параметров $\mathbf{w}_k$ получены из нормального распределения
$$p\left(\mathbf{w}_k|\mathbf{A}_k\right) = \mathcal{N}\left(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k\right), \;\; k = 1, \cdots K; \tag{9}$$
3. Для каждого объекта $\mathbf{x}_i$ существует 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
$$\mathbf{f}_{k_i}: \; y_i \sim \mathcal{N}\left(\mathbf{w}_{k_i}^{\mathsf{T}}\mathbf{x}_i + b_k, \beta^{-1}\right) \tag{10}$$

Правдоподобие модели:

$$p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{A}, \beta, \boldsymbol{\mu}) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\mu}) \prod_{k=1}^{K} N(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k) \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}) \right) \tag{11}$$

Введем скрытые переменные $\mathbf{Z} = ||z_{ik}||$, где $z_{ik} = 1 \Leftrightarrow k_i = k$:

$$p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}, \beta, \boldsymbol{\mu}) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\mu}) \prod_{k=1}^{K} N(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k) \prod_{i=1}^{N} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}) \right)^{z_{ik}} \tag{12}$$

Вариационный EM–алгоритм $q(\mathbf{Z}, \mathbf{W}, \boldsymbol{\pi}) = q(\mathbf{Z})\, q(\mathbf{W})\, q(\boldsymbol{\pi})$:

1. E–шаг:

$$\begin{aligned} \log q(\mathbf{Z}^s) &\propto \mathsf{E}_{q/\mathbf{Z}} \log p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}^{s-1}, \beta^{s-1}, \boldsymbol{\mu}) \\ \log q(\mathbf{W}^s) &\propto \mathsf{E}_{q/\mathbf{W}} \log p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}^{s-1}, \beta^{s-1}, \boldsymbol{\mu}) \\ \log q(\boldsymbol{\pi}^s) &\propto \mathsf{E}_{q/\boldsymbol{\pi}} \log p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}^{s-1}, \beta^{s-1}, \boldsymbol{\mu}) \end{aligned} \tag{13}$$

2. M–шаг:

$$\mathbf{A}^s, \beta^s = \arg\max_{\mathbf{A}, \beta} \mathbf{E}_{q^s} \log p(\mathbf{y}, \mathbf{W}, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}, \beta, \boldsymbol{\mu}) \tag{14}$$

Пример 1

Итерационные формулы EM–алгоритма[2]:

1. E–шаг:

$$p(z_{ik}=1)=\frac{\exp\left(\mathsf{E}\log\pi_k-\frac{\beta}{2}\left[y_i^2-2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^{\mathsf{T}}\left(\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\right)\mathbf{x}_i\right]\right)}{\sum_k p(z_{ik}=1)},$$

$$q(\boldsymbol{\pi})=\mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\mu}+\boldsymbol{\gamma}),\quad q(\mathbf{w}_k)=\mathcal{N}(\mathbf{w}_k|\mathbf{m}_k,\mathbf{B}_k),$$

$$\gamma_k=\sum_{i=1}^{N}\mathsf{E}z_{ik},\quad \mathbf{m}_k=\beta\mathbf{B}_k\left(\sum_{i=1}^{N}\mathbf{x}_iy_i\mathsf{E}z_{ik}\right),\quad \mathbf{B}_k=\left(\mathbf{A}_k^{-1}+\beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}z_{ik}\right)^{-1}. \tag{15}$$

2. M–шаг:

$$\mathbf{A}_k=\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}$$
$$\frac{1}{\beta}=\frac{\sum\sum\left[y_i^2-2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]\mathsf{E}z_{ik}}{\sum\sum\mathsf{E}z_{ik}} \tag{16}$$

Некоторые математические ожидания:

1. $\mathsf{E}z_{ik}=p(z_{ik}=1)$,
2. $\mathsf{E}\log\pi_k=\psi^0(\mu_k+\gamma_k)-\psi^0(K\mu_k+N)$,
3. $\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}=\mathbf{B}_k+\mathbf{m}_k\mathbf{m}_k^{\mathsf{T}}$.

[2] https://github.com/andriygav/EMprior/blob/master/paper/Grabovoy2019Draft.pdf

Пример 1

Картинки

Definition

Смесь экспертов — мультимодель, определяющая правдоподобие веса π_k каждой локальной модели $\mathbf{f}_k$ на признаковом описании объекта $\mathbf{x}$.

$$\hat{\mathbf{f}} = \sum_{k=1}^{K} \pi_k \mathbf{f}_k, \qquad \pi_k\left(\mathbf{x}, \mathbf{V}\right) : \mathbb{R}^{n \times |\mathbf{V}|} \to [0, 1], \qquad \sum_{k=1}^{K} \pi_k\left(\mathbf{x}, \mathbf{V}\right) = 1 \qquad (17)$$

где $\hat{\mathbf{f}}$ — мультимодель, а $\mathbf{f}_k$ является некоторой моделью,
π_k — параметрическая модель, $\mathbf{w}_k$ — параметры k-й локальной модели,
$\mathbf{V}$ — параметры шлюзовой функции.

Пример 2:

1. Правдоподобие k-й локальной модели $p_k\left(y_i|\mathbf{w}_k, \mathbf{x}_i\right) = \mathcal{N}\left(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}\right)$,
2. Априорное распределение параметров k-й локальной модели $p^k\left(\mathbf{w}_k\right) = \mathcal{N}\left(\mathbf{w}_k|\mathbf{w}_k^0, \mathbf{A}_k\right)$,
3. Шлюзовая функция $\boldsymbol{\pi}\left(\mathbf{x}, \mathbf{V}\right) = \mathrm{softmax}\left(\mathbf{V}_1^{\mathsf{T}}\boldsymbol{\sigma}\left(\mathbf{V}_2^{\mathsf{T}}\mathbf{x}\right)\right)$.

Правдоподобие модели:

$$p\bigl(\mathbf{y}, \mathbf{W}|\mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \beta\bigr) = \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{w}_k|\mathbf{w}_k^0, \mathbf{A}_k\right) \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k \mathcal{N}\left(y_i|\mathbf{w}_k^\mathsf{T}\mathbf{x}_i, \beta^{-1}\right)\right). \tag{18}$$

Введем скрытые переменные $\mathbf{Z} = ||z_{ik}||$, где $z_{ik} = 1 \Leftrightarrow k_i = k$:

$$p\bigl(\mathbf{y}, \mathbf{Z}, \mathbf{W}|\mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \beta\bigr) = \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{w}_k|\mathbf{w}_k^0, \mathbf{A}_k\right) \prod_{i=1}^{N} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}\left(y_i|\mathbf{w}_k^\mathsf{T}\mathbf{x}_i, \beta^{-1}\right)\right)^{z_{ik}}. \tag{19}$$

Вариационный ЕМ–алгоритм $q\left(\mathbf{Z}, \mathbf{W}\right) = q\left(\mathbf{Z}\right) q\left(\mathbf{W}\right)$:

1. Е–шаг:

$$\begin{aligned} \log q\left(\mathbf{Z}^s\right) &\propto \mathsf{E}_{q/\mathbf{Z}} \log p\bigl(\mathbf{y}, \mathbf{Z}, \mathbf{W}|\mathbf{X}, \mathbf{V}^{s-1}, \mathbf{A}^{s-1}, \mathbf{W}^{0,s-1}, \beta^{s-1}\bigr) \\ \log q\left(\mathbf{W}^s\right) &\propto \mathsf{E}_{q/\mathbf{W}} \log p\bigl(\mathbf{y}, \mathbf{Z}, \mathbf{W}|\mathbf{X}, \mathbf{V}^{s-1}, \mathbf{A}^{s-1}, \mathbf{W}^{0,s-1}, \beta^{s-1}\bigr) \end{aligned} \tag{20}$$

2. М–шаг:

$$\mathbf{W}^{0,s}, \mathbf{A}^s, \beta^s = \arg\max_{\mathbf{W}^0, \mathbf{A}, \beta} \mathbf{E}_{q^s} \log p\bigl(\mathbf{y}, \mathbf{Z}, \mathbf{W}|\mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \beta\bigr) \tag{21}$$

Итерационные формулы EM–алгоритма[3]:

1. E–шаг:

$$p\left(z_{ik}=1\right)=\frac{\exp\left(\log\pi_k\left(\mathbf{x}_i,\mathbf{V}\right)-\frac{\beta}{2}\left(\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i-\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\right)\right)}{\sum_{k'=1}^{K}\exp\left(\log\pi_{k'}\left(\mathbf{x}_i,\mathbf{V}\right)-\frac{\beta}{2}\left(\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_{k'}\mathbf{w}_{k'}^{\mathsf{T}}\mathbf{x}_i-\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_{k'}\right)\right)},$$

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

$$\mathbf{m}_k=\mathbf{B}_k\left(\mathbf{A}_k^{-1}\mathbf{w}_k^0+\beta\sum_{i=1}^{N}\mathbf{x}_iy_i\mathsf{E}z_{ik}\right),\quad \mathbf{B}_k=\left(\mathbf{A}_k^{-1}+\beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}z_{ik}\right)^{-1}. \tag{22}$$

2. M–шаг:

$$\mathbf{A}_k=\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}-\mathbf{w}_k^0\mathsf{E}\mathbf{w}_k^{\mathsf{T}}-\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{0\mathsf{T}}+\mathbf{w}_k^0\mathbf{w}_k^{0\mathsf{T}},$$

$$\frac{1}{\beta}=\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left[y_i^2-2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]\mathsf{E}z_{ik}, \tag{23}$$

$$\mathbf{w}_k^0=\mathsf{E}\mathbf{w}_k,$$

$$\mathbf{V}=\arg\max_{\mathbf{V}}\mathbf{E}_{q^s}\log p\big(\mathbf{y},\mathbf{Z},\mathbf{W}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\beta\big).$$

[3] `https://github.com/andriygav/EMprior/blob/master/paper/Grabovoy2019MixtureOfExpert.pdf`

Пример 2

Картинки