Московский Государственный
Университет

Вариационный вывод

Выполнил: Курцев Д.В.

Группа: 417

Факультет Вычислительной математики и кибернетики

Кафедра Математических методов прогнозирования

Ноябрь 2022

Пусть $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^D$ – независимая выборка из смеси распределений Стьюдента

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} w_k \mathcal{T}(\boldsymbol{x}|\boldsymbol{\mu}_k, \Sigma_k, \nu), \quad w_k \geq 0,\ \sum_j w_j = 1. \tag{1}$$

Рассмотрим следующую вероятностную модель со скрытыми переменными:

$$p(X, T, Z|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu) = \prod_{n,k=1}^{N,K} \Big[w_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \Sigma_k/z_n) \mathcal{G}(z_n|\nu/2, \nu/2) \Big]^{t_{nk}}. \tag{2}$$

Здесь $t_{nk} \in \{0, 1\}$, $\sum_j t_{nj} = 1$ обозначает принадлежность n-го объекта k-ой компоненте смеси. Очевидно, что неполное правдоподобие $p(X|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu)$ для модели (2) совпадает с правдоподобием выборки X для смеси (1). Поэтому оценки максимального правдоподобия $w_{ML,k}, \mu_{ML,k}, \Sigma_{ML,k}$ для смеси (1) можно искать с помощью вариационного EM-алгоритма для модели (2), в котором на E-шаге апостериорное распределение приближается в семействе

$$q_T(T) q_Z(Z) \approx p(T, Z|X, \boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu).$$

Так как нам везде понадобится $\log p(X, T, Z|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu$, то распишем его.

$$\mathcal{N}(x_n|\mu_k, \Sigma_k/z_n) = \left(\frac{z_n}{2\pi}\right)^{\frac{D}{2}} \frac{1}{\sqrt{\det \Sigma_k}} \exp(-\frac{z_n}{2}(x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k))$$

$$\mathcal{G}(z_n|\nu/2, \nu/2) = \frac{\nu/2}{\Gamma(\nu/2)} \left(\frac{\nu}{2} z_n\right)^{\frac{\nu}{2}-1} \exp(-\frac{\nu}{2} z_n)$$

Тогда

$$\log p(X, T, Z|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu) = \sum_{n,k} t_{n,k} (\log w_k - \frac{1}{2} \log \det \Sigma_k + \frac{D}{2} \log z_n - \frac{D}{2} \log(2\pi) -$$

$$- \frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k) z_n - \log \Gamma\left(\frac{\nu}{2}\right) + \frac{\nu}{2} \log \frac{\nu}{2} + \left(\frac{\nu}{2} - 1\right) \log z_n - \frac{\nu}{2} z_n)$$

E-шаг

$q_T(T)$

Для начала выведем формулу для $q_T(T)$:

$$\log q_T(t_{nk} = 1) = \mathbb{E}_{q_Z(Z)} \log p(X, T, Z|w, \mu, \Sigma, \nu) = \log w_k - \frac{1}{2} \log \det \Sigma_k + \frac{D}{2} \mathbb{E}_{q_Z(Z)} \log z_n -$$

$$-\frac{D}{2}\log(2\pi)-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\mathbb{E}_{q_Z(Z)}z_n-\log\Gamma(\frac{\nu}{2})+\frac{\nu}{2}\log\frac{\nu}{2}+(\frac{\nu}{2}-1)\mathbb{E}_{q_Z(Z)}\log z_n-$$

$$-\frac{\nu}{2}\mathbb{E}_{q_Z(Z)}z_n=$$

Оставим только те слагаемые, которые зависят от k и n:

$$=\log w_k-\frac{1}{2}\log\det\Sigma_k+\frac{D}{2}\mathbb{E}_{q_Z(Z)}\log z_n-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\mathbb{E}_{q_Z(Z)}z_n+$$

$$+(\frac{\nu}{2}-1)\mathbb{E}_{q_Z(Z)}\log z_n-\frac{\nu}{2}\mathbb{E}_{q_Z(Z)}z_n+const$$

Возведя в экспоненту, получим:

$$q_T(t_{nk}=1)=\frac{w_k}{\sqrt{\det\Sigma_k}}\exp((\frac{\nu+D}{2}-1)\mathbb{E}_{q_Z(Z)}\log z_n-(\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+$$

$$+\frac{\nu}{2})\mathbb{E}_{q_Z(Z)}z_n)\ /\ const$$

Из того, что $\sum_k q_T(t_{nk}=1)$ надём константу:

$$const=\sum_j\frac{w_j}{\sqrt{\det\Sigma_j}}\exp((\frac{\nu+D}{2}-1)\mathbb{E}_{q_Z(Z)}\log z_n-(\frac{1}{2}(x_n-\mu_j)^T\Sigma_j^{-1}(x_n-\mu_j)+$$

$$+\frac{\nu}{2})\mathbb{E}_{q_Z(Z)}z_n)$$

Таким образом, получим, что:

$$q_T(t_{nk}=1)=\frac{\frac{w_k}{\sqrt{\det\Sigma_k}}\exp(-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\mathbb{E}_{q_Z(Z)}z_n)}{\sum\limits_{j=1}^{K}\frac{w_j}{\sqrt{\det\Sigma_j}}\exp(-\frac{1}{2}(x_n-\mu_j)^T\Sigma_j^{-1}(x_n-\mu_j)\mathbb{E}_{q_Z(Z)}z_n)}$$

$q_Z(Z)$

Теперь получим формулу для $q_Z(Z)$:

$$\log q_Z(Z)=\mathbb{E}_{q_T(T)}\log p(X,T,Z|w,\mu,\Sigma,\nu)=\mathbb{E}_{q_T(T)}\sum_{n,k}t_{n,k}(\log w_k-\frac{1}{2}\log\det\Sigma_k+\frac{D}{2}\log z_n-$$

$$-\frac{D}{2}\log(2\pi)-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)z_n-\log\Gamma(\frac{\nu}{2})+\frac{\nu}{2}\log\frac{\nu}{2}+(\frac{\nu}{2}-1)\log z_n-\frac{\nu}{2}z_n)=$$

Внесём внутрь мат ожидание и оставим только те слагаемые, которые зависят от z_n

$$= \sum_{n,k} q_T(t_{n,k}=1)(\frac{D}{2}\log z_n - \frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)z_n + (\frac{\nu}{2}-1)\log z_n - \frac{\nu}{2}z_n)$$

Тогда получаем, что:

$$q_Z(Z) = \prod_{n=1}^{N} q_Z(z_n)$$

Где

$$q_Z(z_n) = \prod_k z_n^{(\frac{\nu+D}{2}-1)q_T(t_{n,k}=1)} \exp(-q_T(t_{n,k}=1)(\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\frac{\nu}{2})z_n)const$$

Заметим, что $q_Z(z_n) \sim \mathcal{G}(a,b)$, где $a = \frac{\nu+D}{2}, b = \sum_k q_T(t_{n,k}=1)(\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\frac{\nu}{2}$

М-шаг

На М-шаге решается задача

$$\mathbb{E}_{q_T(T),q_Z(Z)} \log p(X,T,Z|w,\mu,\Sigma,\nu) \to \max_{w,\mu,\Sigma}$$

Оставим тогда в итоговой формуле только слагаемые, зависящие от w, μ, Σ, так как от остальных наша оптимизационная задача не зависит.

$$\sum_{n,k} q(t_{n,k}=1)(\log w_k - \frac{1}{2}\log\det\Sigma_k - \frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\mathbb{E}_{q_Z(Z)}z_n) \to \max_{w,\mu,\Sigma} \quad (*)$$

w_k

Так как у нас имеется ограничение на $w_k(\sum_k w_k = 1)$, то будем решать задачу условной оптимизации. Для этого запишем лагранжиан системы.

$$L = \sum_{n,k} q(t_{n,k}=1)(\log w_k - \frac{1}{2}\log\det\Sigma_k - \frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\mathbb{E}_{q_Z(Z)}z_n) - \lambda(\sum_k w_k - 1)$$

$$\frac{\partial L}{\partial w_k} = \sum_n q(t_{nk}=1)\frac{1}{w_k} - \lambda = 0 \quad \Rightarrow \quad w_k = \frac{\sum_n q(t_{nk}=1)}{\lambda}$$

$$1 = \sum_k w_k = \frac{\sum_{n,k} q(t_{nk}=1)}{\lambda} = \frac{\sum_n 1}{\lambda} = \frac{N}{\lambda} \quad \Rightarrow \quad \lambda = \frac{1}{N}$$

Таким образом, получим, что

$$w_k = \frac{1}{N} \sum_n q(t_{nk} = 1)$$

Для случая $K = 1, \quad q(t_n) = 1$. То есть очевидным образом получим $w = \frac{1}{N} \sum_n 1 = 1$

μ_k

$$\frac{\partial(*)}{\partial \mu_k} = -\sum_n q(t_{nk} = 1)\Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_Z(Z)} z_n = 0 \quad \Rightarrow$$

$$\Rightarrow \quad \Sigma_k^{-1}\mu_k \sum_n q(t_{nk} = 1)\mathbb{E}_{q_Z(Z)} z_n = \Sigma_k^{-1} \sum_n q(t_{nk} = 1)\mathbb{E}_{q_Z(Z)} z_n x_n$$

Таким образом, получим, что

$$\mu_k = \frac{\sum_n q(t_{nk} = 1) x_n \mathbb{E}_{q_Z(Z)} z_n}{\sum_n q(t_{nk} = 1)\mathbb{E}_{q_Z(Z)} z_n}$$

Аналогично для $K = 1$ получаем, что

$$\mu = \frac{\sum_n x_n \mathbb{E}_{q_Z(Z)} z_n}{\sum_n \mathbb{E}_{q_Z(Z)} z_n}$$

Σ_k

$$\frac{d(*)}{d\Sigma_k} = -\sum_n q(t_{nk} = 1)\left(\frac{\det \Sigma_k}{2 \det \Sigma_k} tr \Sigma_k^{-1} d\Sigma_k + \frac{1}{2}(x_n - \mu_k)^T d\Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_Z(Z)} z_n\right) =$$

$\{d\Sigma_k^{-1} = -\Sigma_k^{-1} d\Sigma_k \Sigma_k^{-1}\}$

$$= -\sum_n q(t_{nk} = 1)\frac{1}{2}(tr \Sigma_k^{-1} d\Sigma_k - \mathbb{E}_{q_Z(Z)} z_n tr[(x_n - \mu_k)^T \Sigma_k^{-1} d\Sigma_k \Sigma_k^{-1}(x_n - \mu_k)])$$

$$= \frac{1}{2} \sum_n q(t_{nk} = 1) tr[\mathbb{E}_{q_Z(Z)} z_n \Sigma_k^{-1}(x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1} - \Sigma_k^{-1}] d\Sigma_k$$

$$\Rightarrow \quad \nabla_{\Sigma_k}(*) = \frac{1}{2} \sum_n q(t_{nk} = 1)[\mathbb{E}_{q_Z(Z)} z_n \Sigma_k^{-1}(x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1} - \Sigma_k^{-1}] = 0$$

$$\Rightarrow \quad \sum_n q(t_{nk}=1)\mathbb{E}_{q_Z(Z)} z_n \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1} = \sum_n q(t_{nk}=1)\Sigma_k^{-1}$$

$$\Rightarrow \quad \sum_n q(t_{nk}=1)(x_n-\mu_k)(x_n-\mu_k)^T\mathbb{E}_{q_Z(Z)} z_n = \sum_n q(t_{nk}=1)\Sigma_k$$

$$\Rightarrow \quad \Sigma_k = \frac{\sum_n q(t_{nk}=1)(x_n-\mu_k)(x_n-\mu_k)^T\mathbb{E}_{q_Z(Z)} z_n}{\sum_n q(t_{nk}=1)}$$

И для $K=1$

$$\Sigma = \frac{\sum_n (x_n-\mu)(x_n-\mu)^T\mathbb{E}_{q_Z(Z)} z_n}{\sum_n 1} = \frac{1}{N}\sum_n (x_n-\mu)(x_n-\mu)^T\mathbb{E}_{q_Z(Z)} z_n$$

Правдободобие

$$\mathcal{L}(q,w,\mu,\Sigma) = \mathbb{E}_{q_T(T)q_Z(Z)} \log p(X,T,Z|w,\mu,\Sigma,\nu) - \mathbb{E}_{q_T(T)q_Z(Z)} \log q_T(T)q_Z(Z) =$$

$$= \mathbb{E}_{q_T(T)q_Z(Z)} \log p(X,T,Z|w,\mu,\Sigma,\nu) - \mathbb{E}_{q_T(T)} \log q_T(T) - \mathbb{E}_{q_Z(Z)} \log q_Z(Z)$$

$$\mathbb{E}_{q_T(T)q_Z(Z)} \log p(X,T,Z|w,\mu,\Sigma,\nu) = \sum_{n,k} q_T(t_{n,k}=1)(\log w_k - \frac{1}{2}\log\det\Sigma_k + \frac{D}{2}\mathbb{E}_{q_Z(Z)}\log z_n -$$

$$-\frac{D}{2}\log(2\pi) - \frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\mathbb{E}_{q_Z(Z)} z_n - \log\Gamma(\frac{\nu}{2}) + \frac{\nu}{2}\log\frac{\nu}{2} + (\frac{\nu}{2}-1)\mathbb{E}_{q_Z(Z)}\log z_n -$$

$$-\frac{\nu}{2}\mathbb{E}_{q_Z(Z)} z_n) =$$

$$\mathbb{E}_{q_T(T)} \log q_T(T) = \sum_{n,k} q_T(t_{nk}=1)\log q_T(t_{nk}=1)$$

$$\mathbb{E}_{q_Z(Z)} \log q_Z(Z) = \sum_n \mathbb{E}_{q_Z(Z)} \log q_Z(z_n)$$

Так как $q_Z(z_n) \sim \mathcal{G}(a,b)$ с выписанными ранне параметрами, то:

$$\mathbb{E}_{q_Z(Z)}\ q_Z(z_n) = \frac{a}{b}$$

$$\mathbb{E}_{q_Z(Z)} \log q_Z(z_n) = \psi(a) - \log b, \quad \psi(x) = \frac{d\ \log\Gamma(x)}{dx}$$