Теоретическое задание 3

Королев Кирилл

4 декабря 2023 г.

Пусть $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^D$ – 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:

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

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

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

Задание 1

Выписать формулы пересчёта для компонент вариационного приближения $q_T(T)$ и $q_Z(Z)$.

Алгоритм mean-field говорит нам, что компоненты в приближении апостериорного распределения пересчитываются как

$$\log q_T(T) = \mathbb{E}_{q_Z} \left(\log p(X, T, Z \mid \boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu) \right) + const$$
$$\log q_Z(Z) = \mathbb{E}_{q_T} \left(\log p(X, T, Z \mid \boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu) \right) + const$$

Тогда найдем $q_T(T)$, воспользовавшимся следующим равенством.

$$\log \left(\det \left(\Sigma_k / z_n \right) \right) = \log \left(\frac{1}{z_n^D} \det \Sigma_k \right) = -D \log z_n + \log \left(\det \Sigma_k \right) \tag{3}$$

$$\begin{aligned}
&\log q_T(T) = \mathbb{E}_{q_Z} \left(\log p(X, T, Z \mid \boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu) \right) + const = \\
&\mathbb{E}_{q_Z} \Bigg[\sum_{n,k=1}^{N,K} t_{nk} \Bigg(\log w_k - \frac{D}{2} \log 2\pi - \frac{1}{2} \log \left(\det \left(\Sigma_k / z_n \right) \right) - \\
&\frac{1}{2} (x_n - \mu_k)^T (\Sigma_k / z_n)^{-1} (x_n - \mu_k) + \nu/2 \log (\nu/2) - \log \Gamma(\nu/2) + \\
&(\nu/2 - 1) \log z_n - \frac{\nu}{2} z_n \Bigg) \Bigg] + const = \sum_{n,k=1}^{N,K} t_{nk} \Bigg(\log w_k - \frac{D}{2} \log 2\pi - \\
&\frac{1}{2} \log \left(\det \Sigma_k \right) + \frac{D}{2} \mathbb{E}_{q_Z} (\log z_n) - \frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) \mathbb{E}_{q_Z}(z_n) + \\
&\nu/2 \log (\nu/2) - \log \Gamma(\nu/2) + (\nu/2 - 1) \mathbb{E}_{q_Z} (\log z_n) - \frac{\nu}{2} \mathbb{E}_{q_Z}(z_n) \Bigg) + const
\end{aligned} \tag{4}$$

Так как $\sum_{k=1}^{K} t_{nk} = 1$, то все что внутри скобки, независящее от индекса k, можно перенести в константу, ведь такие слагаемые можно просуммировать сначала по k, избавившись зависимости от T.

$$\log q_T(T) = \sum_{n,k=1}^{N,K} t_{nk} \left(\log w_k - \frac{1}{2} \log\left(\det \Sigma_k\right) - \frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) \mathbb{E}_{q_Z}(z_n) \right) + const_1 \tag{5}$$

Прибавим и отнимем $\frac{D}{2} \log \mathbb{E}_{q_Z}(z_n)$, чтобы получить необходимый вид для матрицы ковариации многомерного нормального распределения. То что добавили лишнее все равно уйдет в константу, так как от t_{nk} не зависит.

$$\log q_T(T) = \sum_{n,k=1}^{N,K} t_{nk} \left(\log w_k - \frac{1}{2} \log\left(\det\left(\Sigma_k / \mathbb{E}_{q_Z}(z_n)\right)\right) - \frac{1}{2}(x_n - \mu_k)^T \left(\Sigma_k / \mathbb{E}_{q_Z}(z_n)\right)^{-1} (x_n - \mu_k) \right) + const_2 \tag{6}$$

Таким образом, свернув все обратно, получим числитель $q_T(T)$.

$$q_T(T) \propto \prod_{n,k=1}^{N,K} \left[w_k \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \Sigma_k / \mathbb{E}_{q_Z}(z_n)) \right]^{t_{nk}}$$

Знаменатель это сумма по всем комбинациям t_{nk} таких произведений. Из-за структуры t_{nk}, ведь в каждой строке этой матрицы ровно 1 единица, слагаемые можно сгруппировать.

$$\sum_{T} \prod_{n,k=1}^{N,K} \left[w_k \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \Sigma_k / \mathbb{E}_{q_Z}(z_n)) \right]^{t_{nk}} = \prod_{n=1}^{N} \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \Sigma_k / \mathbb{E}_{q_Z}(z_n))$$

В итоге получаем

$$\boxed{q_T(T) = \prod_{n,k=1}^{N,K} \left[\frac{w_k \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \Sigma_k / \mathbb{E}_{q_Z}(z_n))}{\sum_{j=1}^{K} w_j \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_j, \Sigma_j / \mathbb{E}_{q_Z}(z_n))} \right]^{t_{nk}}}$$

Теперь найдем $q_Z(Z)$, используя факт, что $\sum_{k=1}^{K} \mathbb{E}_{q_T}(t_{nk}) = 1$, так как вне зависимости от распределения в строке должна быть 1 единица.

$$\begin{aligned}
&\log q_Z(Z) = \mathbb{E}_{q_T}\left(\log p(X, T, Z \mid \boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu)\right) + const = \\
&\mathbb{E}_{q_T}\left(\sum_{n,k=1}^{N,K} t_{nk} \left(\frac{D}{2} \log z_n - \frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) z_n + (\nu/2 - 1) \log z_n - \frac{\nu}{2} z_n \right) \right) + const_1 = \\
&\sum_{n,k=1}^{N,K} \mathbb{E}_{q_T}(t_{nk}) \left(\frac{D}{2} \log z_n - \frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) z_n + (\nu/2 - 1) \log z_n - \frac{\nu}{2} z_n \right) + const_1 = \\
&\sum_{n=1}^{N} \frac{D}{2} \log z_n - \sum_{n,k=1}^{N,K} \frac{\mathbb{E}_{q_T}(t_{nk})}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) z_n + \sum_{n=1}^{N} (\nu/2 - 1) \log z_n - \sum_{n=1}^{N} \frac{\nu}{2} z_n + const_1
\end{aligned} \tag{7}$$

По функциональному виду заметно, что это гамма распределение, где $q_{z_n}(z_n) = \mathcal{G}(z_n \mid a_n, b_n)$.

$$a_n = \frac{D + \nu}{2}$$

$$b_n = \sum_{k=1}^{K} \frac{\mathbb{E}_{q_T}(t_{nk})}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) + \frac{\nu}{2}$$

Таким образом, $\boxed{q_Z(Z) = \prod_{n=1}^{N} \mathcal{G}(z_n \mid \frac{D + \nu}{2}, \sum_{k=1}^{K} \frac{\mathbb{E}_{q_T}(t_{nk})}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) + \frac{\nu}{2})}$.

Задание 2

Выписать формулы пересчёта параметров w_k, μ_k, Σ_k на М-шаге. Убедиться, что эти формулы переходят в соответствующие формулы с семинара по ЕМ-алгоритму для случая $K = 1$.

$$\begin{gathered}\mathbb{E}_{q(Z,T)}\Bigg[\log p(X, T, Z|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu)\Bigg] = \\ \mathbb{E}_{q(Z,T)}\Bigg[\sum_{n,k=1}^{N,K} t_{nk}\Bigg(\log w_k + \log \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \Sigma_k/z_n) + \log \mathcal{G}(z_n|\nu/2, \nu/2)\Bigg)\Bigg]\end{gathered} \tag{8}$$

Поиск w_k

$$\mathbb{E}_{q(Z,T)}\Bigg[\sum_{n,k=1}^{N,K} t_{nk} \log w_k\Bigg] + const \to \max_{w_k}$$

$$\mathbb{E}_{q(Z,T)}\Bigg[\sum_{n,k=1}^{N,K} t_{nk} \log w_k\Bigg] = \sum_{n,k=1}^{N,K} \mathbb{E}_{q(T)}[t_{nk}] \log w_k$$

Также не забудем условие $\sum_{k=1}^{K} w_k = 1$. Тогда решим задачу условной оптимизации методом Лагранжа.

$$L(w, \mu) = \sum_{n,k=1}^{N,K} \mathbb{E}_{q(T)}[t_{nk}] \log w_k - \mu\Bigg(\sum_{k=1}^{K} w_k - 1\Bigg)$$

$$\frac{\partial L}{\partial w_k} = \sum_{n=1}^{N} \mathbb{E}_{q(T)}[t_{nk}]\frac{1}{w_k} - \mu = 0$$

$$w_k = \frac{\sum_{n=1}^{N} \mathbb{E}_{q(T)}[t_{nk}]}{\mu}$$

Найдем μ из ограничения на равенство.

$$1 = \sum_{k=1}^{K} w_k = \sum_{k=1}^{K} \frac{\sum_{n=1}^{N} \mathbb{E}_{q(T)}[t_{nk}]}{\mu} \Rightarrow \mu = \sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{E}_{q(T)}[t_{nk}] = N$$

Таким образом, $\boxed{w_k = \frac{\sum_{n=1}^{N} \mathbb{E}_{q(T)}[t_{nk}]}{N}}$.

Поиск μ_k

Поставим задачу на минимум, убрав минус, который появится из экспоненты.

$$\mathbb{E}_{q(Z,T)}\Bigg[\sum_{n,k=1}^{N,K} t_{nk} \frac{z_n \mu_k^T \Sigma_k^{-1} \mu_k - 2 z_n \mu_k^T \Sigma_k^{-1} x_n}{2}\Bigg] + const \to \min_{\mu_k}$$

$$L(\mu) = \sum_{n,k=1}^{N,K} \mathbb{E}_{q(T)}(t_{nk}) \frac{\mathbb{E}_{q(Z)}(z_n)\mu_k^T \Sigma_k^{-1} \mu_k - 2\mathbb{E}_{q(Z)}(z_n)\mu_k^T \Sigma_k^{-1} x_n}{2}$$

$$\begin{gathered}\frac{\partial L}{\partial \mu_k} = \sum_{n=1}^{N} \mathbb{E}_{q(T)}(t_{nk})(\mathbb{E}_{q(Z)}(z_n)\Sigma_k^{-1}\mu_k - \mathbb{E}_{q(Z)}(z_n)\Sigma_k^{-1} x_n) = \\ \Bigg[\sum_{n=1}^{N} \mathbb{E}_{q(T)}(t_{nk})\mathbb{E}_{q(Z)}(z_n)\Bigg]\Sigma_k^{-1}\mu_k - \sum_{n=1}^{N} \mathbb{E}_{q(T)}(t_{nk})\mathbb{E}_{q(Z)}(z_n)\Sigma_k^{-1} x_n = 0\end{gathered}$$

Домножим уравнение на Σ_k и получим.

$$\boxed{\mu_k = \frac{\sum_{n=1}^{N} \mathbb{E}_{q(T)}(t_{nk})\mathbb{E}_{q(Z)}(z_n) x_n}{\sum_{n=1}^{N} \mathbb{E}_{q(T)}(t_{nk})\mathbb{E}_{q(Z)}(z_n)}}$$

Поиск Σ_k

$$\mathbb{E}_{q(Z,T)}\left[\sum_{n,k=1}^{N,K} t_{nk}\left(\frac{-z_n(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)}{2}-\frac{\log(\det\Sigma_k)}{2}\right)\right]+const\to\max_{\Sigma_k}$$

$$L(\Sigma)=\sum_{n,k=1}^{N,K}\mathbb{E}_{q(T)}(t_{nk})\left(\frac{-\mathbb{E}_{q(Z)}(z_n)(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)}{2}-\frac{\log(\det\Sigma_k)}{2}\right)$$

Воспользуемся следующими равенствами:

$$d(x^TAy)=d(tr(d(x^TAy)))=tr(d(x^TAy))=tr(x^TdAy)=tr(yx^TdA)=\langle xy^T,dA\rangle$$

$$\frac{\partial\log(\det\Sigma_k)}{\partial\Sigma_k^{-1}}=-\frac{\partial\log(\det\Sigma_k^{-1})}{\partial\Sigma_k^{-1}}=-(\Sigma_k^{-1})^{-T}=-\Sigma_k$$

Тогда производная по Σ_k^{-1} равна

$$\frac{\partial L}{\partial\Sigma_k^{-1}}=\sum_{n=1}^{N}\mathbb{E}_{q(T)}(t_{nk})\left(\frac{-\mathbb{E}_{q(Z)}(z_n)(x_n-\mu_k)(x_n-\mu_k)^T}{2}+\frac{\Sigma_k}{2}\right)=0$$

$$\boxed{\Sigma_k=\frac{\sum_{n=1}^N\mathbb{E}_{q(T)}(t_{nk})\mathbb{E}_{q(Z)}(z_n)(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^N\mathbb{E}_{q(T)}(t_{nk})}}$$

Задание 3

Расписать функционал $\mathcal{L}(q,\boldsymbol{w},\boldsymbol{\mu},\Sigma)$ – нижнюю оценку на $\log p(X|\boldsymbol{w},\boldsymbol{\mu},\Sigma,\nu)$.

$$\begin{aligned}
&\mathcal{L}(q,\boldsymbol{w},\boldsymbol{\mu},\Sigma)=\mathbb{E}_{q(Z,T)}\Big[\log p(X,T,Z|\boldsymbol{w},\boldsymbol{\mu},\Sigma,\nu)\Big]-\mathbb{E}_{q(Z,T)}\Big[\log q(Z,T)\Big]=\\
&\mathbb{E}_{q(Z,T)}\left[\sum_{n,k=1}^{N,K}t_{nk}\Big(\log w_k+\log\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\Sigma_k/z_n)+\log\mathcal{G}(z_n|\nu/2,\nu/2)\Big)\right]-\mathbb{E}_{q(Z,T)}\Big[\log q(Z,T)\Big]=\\
&\mathbb{E}_{q(Z,T)}\left[\sum_{n,k=1}^{N,K}t_{nk}\Big(\log w_k-\frac{D}{2}\log 2\pi-\frac{1}{2}\log(\det\Sigma_k)+\frac{D}{2}\log z_n-\frac{z_n}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\right.\\
&\left.\nu/2\log(\nu/2)-\log\Gamma(\nu/2)+(\nu/2-1)\log z_n-\frac{\nu}{2}z_n\Big)\right]-\mathbb{E}_{q(Z)}\Big[\log q(Z)\Big]-\mathbb{E}_{q(T)}\Big[\log q(T)\Big]=\\
&\sum_{n,k=1}^{N,K}\mathbb{E}_{q_T}(t_{nk})\Big(\log w_k-\frac{D}{2}\log 2\pi-\frac{1}{2}\log(\det\Sigma_k)+\frac{D}{2}\mathbb{E}_{q_Z}(\log z_n)-\\
&\frac{\mathbb{E}_{q_Z}(z_n)}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\nu/2\log(\nu/2)-\log\Gamma(\nu/2)+\\
&(\nu/2-1)\mathbb{E}_{q_Z}(\log z_n)-\frac{\nu}{2}\mathbb{E}_{q_Z}(z_n)\Big)-\mathbb{E}_{q(Z)}\Big[\log q(Z)\Big]-\mathbb{E}_{q(T)}\Big[\log q(T)\Big]
\end{aligned}\tag{9}$$

Задание 4

Найти формулы для статистик распределений $q_T(T)$ и $q_Z(Z)$, требуемых в предыдущих трёх пунктах.

Найдем статистики представив гамма-распределение в виде экспоненциального семейства.

$$\mathcal{G}(z_n\mid a,b)=\frac{b^az_n^{a-1}e^{-bz_n}}{\Gamma(a)}=\frac{1}{z_n}\frac{1}{\frac{\Gamma(a)}{b^a}}e^{(a\log z_n-bz_n)}=\frac{1}{z_n}\frac{1}{\frac{\Gamma(a)}{b^a}}e^{(a,b)^T(\log z_n,-z_n)}$$

Следовательно, $\log z_n$ и $-z_n$ - это достаточные статистики и мы можем посчитать их ожидание через дифференцирование логарифма нормировочной константы.

$$g(a,b)=\frac{\Gamma(a)}{b^a}$$

$$\log g(a,b)=\log\Gamma(a)-a\log b$$

$$\mathbb{E}(\log z_n) = \frac{\partial \log g(a,b)}{\partial a} = \frac{\Gamma^{'}(a)}{\Gamma(a)} - \log b = \psi(a) - \log b, \text{ где } \psi(a) \text{ - дигамма функция.}$$

$$\mathbb{E}(z_n) = -\frac{\partial \log g(a,b)}{\partial b} = \frac{a}{b}$$

Таким образом, получаем

$$\boxed{\mathbb{E}_{q_Z}(z_n) = \frac{D+\nu}{\sum_{k=1}^K \mathbb{E}_{q_T}(t_{nk})(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\nu}}$$

$$\boxed{\mathbb{E}_{q_Z}(\log z_n) = \psi\left(\frac{D+\nu}{2}\right) - \log\left(\sum_{k=1}^K \frac{1}{2}\mathbb{E}_{q_T}(t_{nk})(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\frac{\nu}{2}\right)}$$

Теперь найдем $\boxed{\mathbb{E}_{q_T}(t_{nk}) = 1 \times q_T(t_{nk}=1) = \frac{w_k\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \Sigma_k/\mathbb{E}_{q_Z}(z_n))}{\sum_{j=1}^K w_j\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \Sigma_j/\mathbb{E}_{q_Z}(z_n))}}$.