Problem 1. 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

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

Для выполнения задания требуется:

1. Выписать формулы пересчёта для компонент вариационного приближения $q_T(T)$ и $q_Z(Z)$;
2. Выписать формулы пересчёта параметров w_k, μ_k, Σ_k на M-шаге. Убедиться, что эти формулы переходят в соответствующие формулы с семинара по EM-алгоритму для случая $K = 1$;
3. Расписать функционал $\mathcal{L}(q, \boldsymbol{w}, \boldsymbol{\mu}, \Sigma)$ – нижнюю оценку на $\log p(X|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu)$;
4. Найти формулы для статистик распределений $q_T(T)$ и $q_Z(Z)$, требуемых в предыдущих трёх пунктах.

1. • $q_Z(Z)$

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

$$= \sum_{n,k} \mathbb{E}_{q_T(T)}[t_{nk}] \Bigg[\log w_k - \frac{D}{2}\log(2\pi) - \frac{1}{2}\log\det(\Sigma_k/z_n) - \frac{1}{2}(x_n - \mu_k)^\top (\Sigma_k/z_n)^{-1}(x_n - \mu_k) +$$

$$+ \frac{\nu}{2}\log\frac{\nu}{2} - \log\Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right)\log z_n - \frac{\nu}{2}z_n \Bigg] + C = \{\log\det(\Sigma_k/z_n) = -D\log z_n + \log\det(\Sigma_k)\} =$$

$$= \sum_{n,k} \mathbb{E}_{q_T(T)}[t_{nk}] \left[\frac{D}{2}\log z_n - \frac{1}{2}(x_n - \mu_k)^\top \Sigma_k^{-1}(x_n - \mu_k) z_n + \left(\frac{\nu}{2} - 1\right)\log z_n - \frac{\nu}{2}z_n \right] + C =$$

$$= \sum_n \left(\frac{D + \nu}{2} - 1\right)\log z_n - \sum_n \left(\frac{\nu}{2} + \sum_k \frac{1}{2}(x_n - \mu_k)^\top \Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_T(T)}[t_{nk}]\right) z_n + C$$

$$\implies q_Z(Z) = \prod_n \mathcal{G}\left(z_n \middle| \frac{D + \nu}{2}, \frac{\nu}{2} + \sum_k \frac{1}{2}(x_n - \mu_k)^\top \Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_T(T)}[t_{nk}] \right)$$

• $q_T(T)$

$$\log q_T(T) = \mathbb{E}_{q_Z(Z)} \left[\log p(X, Z, T \mid w, \mu, \Sigma_k)\right] + C =$$

$$= \sum_{n,k} t_{nk}\mathbb{E}_{q_Z(Z)} \Bigg[\log w_k - \frac{D}{2}\log(2\pi) - \frac{1}{2}\log\det(\Sigma_k/z_n) - \frac{1}{2}(x_n - \mu_k)^\top (\Sigma_k/z_n)^{-1}(x_n - \mu_k) +$$

$$+ \frac{\nu}{2}\log\frac{\nu}{2} - \log\Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right)\log z_n - \frac{\nu}{2}z_n \Bigg] + C = \{\text{то, что не зависит от } k\text{, уйдёт в константу}\} =$$

$$= \sum_{n,k} t_{nk} \left[\log w_k - \frac{1}{2}\log\det(\Sigma_k) - \frac{1}{2}(x_n - \mu_k)^\top \Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_Z(Z)} z_n \right] + C =$$

$$= \left\{ \text{из константы достанем } \frac{D}{2}\log\mathbb{E}_{q_Z(Z)} z_n \right\} =$$

$$= \sum_{n,k} t_{nk} \left[\log w_k - \frac{1}{2} \log\det(\Sigma_k / \mathbb{E}_{q_Z(Z)} z_n) - \frac{1}{2} (x_n - \mu_k)^\top \Sigma_k^{-1} (x_n - \mu_k) \mathbb{E}_{q_Z(Z)} z_n \right] + C$$

Теперь становится видно, что

$$q_T(t_{nk} = 1) = \frac{w_k \cdot \mathcal{N}(x_n | \mu_k, \Sigma_k / \mathbb{E}_{q_Z(Z)} z_n)}{\sum_i w_i \cdot \mathcal{N}(x_n | \mu_i, \Sigma_i / \mathbb{E}_{q_Z(Z)} z_n)}$$

Отсюда

$$q_T(T) = \prod_{n,k} \left[\frac{w_k \cdot \mathcal{N}(x_n | \mu_k, \Sigma_k / \mathbb{E}_{q_Z(Z)} z_n)}{\sum_i w_i \cdot \mathcal{N}(x_n | \mu_i, \Sigma_i / \mathbb{E}_{q_Z(Z)} z_n)} \right]^{t_{nk}}$$

2.

$$\mathbb{E}_{q(Z)q(T)}[\log p(X, Z, T | \theta)] \to \max_\theta$$

- w_k: Помним, что веса суммируются в 1, поэтому записываем Лагранжан

$$\mathcal{L} = \mathbb{E}_{q(Z)q(T)} \left[\sum_{n,k} t_{nk} \log w_k \right] + \lambda \left(\sum_k w_k - 1 \right) \to \max_w$$

$$\frac{\partial \mathcal{L}}{\partial w_k} = \sum_n \mathbb{E}_{q(T)} \left[t_{nk} \frac{1}{w_k} \right] + \lambda = 0$$

$$\implies w_k = \frac{1}{N} \sum_n \mathbb{E}_{q(T)}[t_{nk}]$$

- μ_k:

$$\mathcal{L} = \mathbb{E}_{q(Z)q(T)} \left[\sum_{n,k} t_{nk} \left(-\frac{1}{2} (x_n - \mu_k)^\top \Sigma_k^{-1} (x_n - \mu_k) z_n \right) \right] \to \max_\mu$$

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = \sum_n \mathbb{E}_{q(T)}[t_{nk}] \frac{\partial}{\partial \mu_k} \left(-\frac{1}{2} (x_n - \mu_k)^\top \Sigma_k^{-1} (x_n - \mu_k) \mathbb{E}_{q(Z)}[z_n] \right) =$$

$$= \sum_n \mathbb{E}_{q(T)}[t_{nk}] \mathbb{E}_{q(Z)}[z_n] \left(\Sigma_k^{-1} x_n - \Sigma_k^{-1} \mu_k \right) = 0$$

$$\implies \mu_k = \frac{\sum_n \mathbb{E}_{q(T)}[t_{nk}] \mathbb{E}_{q(Z)}[z_n] x_n}{\sum_n \mathbb{E}_{q(T)}[t_{nk}] \mathbb{E}_{q(Z)}[z_n]}$$

- Σ_k:

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

$$\frac{\partial \mathcal{L}}{\partial \Sigma_k^{-1}} = \left\{ \frac{\partial}{\partial \Sigma_k^{-1}} (\log\det \Sigma_k) = \frac{\partial}{\partial \Sigma_k^{-1}} \left(-\log\det \Sigma_k^{-1} \right) = -\Sigma_k \right\} =$$

$$= \sum_n \mathbb{E}_{q(T)}[t_{nk}] \left[\frac{1}{2} \Sigma_k - \frac{1}{2} (x_n - \mu_k)(x_n - \mu_k)^\top \mathbb{E}_{q(Z)}[z_n] \right] 0$$

$$\implies \Sigma_k = \frac{\sum_n \mathbb{E}_{q(T)}[t_{nk}] \mathbb{E}_{q(Z)}[z_n] (x_n - \mu_k)(x_n - \mu_k)^\top}{\sum_n \mathbb{E}_{q(T)}[t_{nk}]}$$

3.

$$\mathcal{L}(q, \theta) = \mathbb{E}_{q(T)q(X)}[\log p(X, Z, T | \theta) - \log q(T) - \log q(Z)] =$$

$$= \sum_{n,k} \mathbb{E}_{q(T)}[t_{nk}] \Bigg(\log w_k - \frac{D}{2} \log(2\pi) - \frac{1}{2} \log\det \Sigma_k + \frac{D}{2} \mathbb{E}_{q(Z)}[\log z_n] - \frac{1}{2} (x_n - \mu_k)^\top \Sigma_k^{-1} (x_n - \mu_k) \mathbb{E}_{q(Z)}[z_n] +$$

$$+ \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \mathbb{E}_{q(Z)}[\log z_n] - \frac{\nu}{2} \mathbb{E}_{q(Z)}[z_n] \Bigg) - \mathbb{E}_{q(Z)}[\log q(Z)] - \mathbb{E}_{q(T)}[\log q(T)]$$

4. Небходимо найти $\mathbb{E}_{q(Z)}[z_n]$, $\mathbb{E}_{q(Z)}[\log z_n]$, $\mathbb{E}_{q(Z)}[\log q(Z)]$, $\mathbb{E}_{q(T)}[t_{nk}]$, $\mathbb{E}_{q(T)}[\log q(T)]$

   - $$\mathbb{E}_{q(Z)}[z_n] = \left\{ z_n \sim \mathcal{G}\left(z_n \middle| \frac{D+\nu}{2}, \frac{\nu}{2} + \sum_k \frac{1}{2}(x_n - \mu_k)^\top \Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_T(T)}[t_{nk}] \right) \right\} =$$
     $$= \frac{D+\nu}{\nu + \sum_k (x_n - \mu_k)^\top \Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_T(T)}[t_{nk}]}$$
   - Так как Гамма распределение лежит в классе экспоненцильных, то можно найти её достаточные статистики. Свёдём тогда его под известную формулу
     $$p(x|\theta) = \frac{f(x)}{g(\theta)} \exp(u(x)^\top \theta)$$
     $$p(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} = \frac{\frac{1}{x}}{\frac{\Gamma(a)}{b^a}} \exp(a \log x - bx)$$
     $$\implies g(\theta) = \frac{\Gamma(\theta_1)}{\theta_2^{\theta_1}}, \quad u(x) = \begin{bmatrix} \log x \\ -x \end{bmatrix}$$
     Известно, что
     $$\mathbb{E}[u_j(x)] = \frac{\partial}{\partial \theta_j} \log g(\theta).$$
     Тогда получаем
     $$\mathbb{E}[\log x] = \frac{\partial}{\partial \theta_1} (\log \Gamma(\theta_1) - \theta_1 \log \theta_2) = \psi(\theta_1) - \log \theta_2$$
     $$\implies \mathbb{E}_{q(Z)}[\log z_n] = \psi\left(\frac{D+\nu}{2}\right) - \log\left(\frac{\nu}{2} + \sum_k \frac{1}{2}(x_n - \mu_k)^\top \Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_T(T)} t_{nk}\right)$$
   - $$\mathbb{E}_{q(Z)}[\log q(Z)] = \sum_n \mathbb{E}_{q(Z)}\left[\log \mathcal{G}\left(z_n|a,b\right)\right] =$$
     $$= \sum_n \left[a \log b - \log \Gamma(b) + (a-1)\mathbb{E}_{q(Z)}[\log z_n] - b\mathbb{E}_{q(Z)}[z_n]\right],$$
     где
     $$a = \frac{D+\nu}{2}, \; b = \frac{\nu}{2} + \sum_k \frac{1}{2}(x_n - \mu_k)^\top \Sigma_k^{-1}(x_n - \mu_k)\mathbb{E}_{q_T(T)} t_{nk}$$
   - $$\mathbb{E}_{q(T)}[t_{nk}] = 1 \cdot q(t_{nk} = 1) + 0 \cdot q(t_{nk} = 0) = \frac{w_k \cdot \mathcal{N}(x_n|\mu_k, \Sigma_k / \mathbb{E}_{q_Z(Z)} z_n)}{\sum_i w_i \cdot \mathcal{N}(x_n|\mu_i, \Sigma_i / \mathbb{E}_{q_Z(Z)} z_n)}$$
   - $$\mathbb{E}_{q(T)}[\log q(T)] = \sum_{n,k} \mathbb{E}_{q_T(T)}[t_{nk}] \log\left(\frac{w_k \cdot \mathcal{N}(x_n|\mu_k, \Sigma_k / \mathbb{E}_{q_Z(Z)} z_n)}{\sum_i w_i \cdot \mathcal{N}(x_n|\mu_i, \Sigma_i / \mathbb{E}_{q_Z(Z)} z_n)}\right)$$