Теоретическое задание №3
«Вариационный вывод»

курс «Байесовские методы в машинном обучении»
кафедра ММП ВМК МГУ

Михеев Борис, 417 группа

29 ноября 2022 г.

Постановка задачи

Пусть $X = \{x_1, \dots, x_N\}$, $x_n \in \mathbb{R}^D$ – независимая выборка из смеси распределений Стьюдента

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

Рассмотрим следующую вероятностную модель со скрытыми переменными:

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

Здесь $t_{nk} \in \{0, 1\}$, $\sum_j t_{nj} = 1$ обозначает принадлежность n-го объекта k-ой компоненте смеси. Очевидно, что неполное правдоподобие $p(X|w, \mu, \Sigma, \nu)$ 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

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

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

Проведем на E-шаге оценку распределений величин T, Z итерационно с помощью вариационного байесовского вывода. Применим формулы mean-field аппроксимации, и т. к. присутствует условная сопряженность, то можно получить аналитические формулы для пересчета соответствующих распределений.

В дальнейших формулах пересчета используется логарифм совместной плотности $p(X, T, Z \mid w, \mu, \Sigma, \nu)$. Распишем его отдельно:

$$\log p(X, T, Z \mid w, \mu, \Sigma, \nu) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\log w_k + \log \mathcal{N}\left(x_n \mid \mu_k, \frac{\Sigma_k}{z_n} \right) + \log \mathcal{G}\left(z_n \mid \frac{\nu}{2}, \frac{\nu}{2} \right) \right).$$

В свою очередь:

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

$$\mathcal{G}\left(z_n \mid \frac{\nu}{2}, \frac{\nu}{2} \right) = \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} z_n^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2} z_n} \Rightarrow$$

$$\Rightarrow \log p(X, T, Z \mid w, \mu, \Sigma, \nu) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\log w_k - \frac{D}{2} \log 2\pi - \frac{1}{2} \log det \Sigma_k + \frac{D}{2} \log z_n - \right.$$
$$\left. - \frac{z_n}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \log z_n - \frac{\nu}{2} z_n \right).$$

Рассмотрим формулы пересчета для $q_T(T)$. Учтем, что распределения $q_T(T), q_Z(Z)$ можно факторизовать по объектам и компонентам смеси: $q_T(T) = \prod\limits_{n=1}^{N} \prod\limits_{k=1}^{K} q_T(t_{nk})$, $q_Z(Z) = \prod\limits_{n=1}^{N} q_Z(z_n)$.

$$\log q_T(T) = \mathbb{E}_{q_Z(Z)} \log p(X, T, Z \mid w, \mu, \Sigma, \nu) = \mathbb{E}_{q_Z(Z)} \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\log w_k - \frac{D}{2} \log 2\pi - \frac{1}{2} \log det \Sigma_k + \right.$$
$$\left. + \frac{D}{2} \log z_n - \frac{z_n}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \log z_n - \frac{\nu}{2} z_n \right) =$$
$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\log w_k - \frac{D}{2} \log 2\pi - \frac{1}{2} \log det \Sigma_k + \frac{D}{2} \mathbb{E}_{q_Z(z_n)} \log z_n - \right.$$
$$\left. - \frac{1}{2} \mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \mathbb{E}_{q_Z(z_n)} \log z_n - \frac{\nu}{2} \mathbb{E}_{q_Z(z_n)} z_n \right).$$

Распределение $q_T(T)$ является дискретным, величины t_{nk} являются по сути индикаторами. Распишем $\log q_T(t_{nk})$ для конкретного t_{nk}.

$$\log q_T(t_n k) = t_{nk} \left(\log w_k - \frac{1}{2} \log det \Sigma_k - \frac{1}{2} \mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) + \frac{D}{2} \mathbb{E}_{q_Z(z_n)} \log z_n - \frac{D}{2} \log 2\pi + \right.$$
$$\left. + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \mathbb{E}_{q_Z(z_n)} \log z_n - \frac{\nu}{2} \mathbb{E}_{q_Z(z_n)} z_n \right).$$

Заметим, что в данном выражении в скобке все слагаемые, кроме первых трех, по сути не зависят от n или k, поэтому все остальные слагаемые можем обозначить как $\log C$ для удобства. Учтем также, что $q_T(t_{nk}) = q_T(t_{nk} = 1)$, поэтому можем принять $t_{nk} = 1$. Проэкспоненцировав, получим:

$$q_T(t_{nk} = 1) = \frac{Cw_k}{\sqrt{det\Sigma_k}} e^{-\frac{1}{2}\mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k)}.$$

Найти константу C можно из условия $\sum\limits_{k=1}^{K} q_T(t_{nk}) = 1$, т. к. $q_T(t_{nk})$ — вероятность принадлежности объекта x_n компоненте смеси с номером k. Тогда:

$$\sum_{k=1}^{K} q_T(t_{nk}) = C \sum_{k=1}^{K} \frac{w_k}{\sqrt{det\Sigma_k}} e^{-\frac{1}{2}\mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k)} \Rightarrow$$
$$\Rightarrow C = \frac{1}{\sum\limits_{k=1}^{K} \frac{w_k}{\sqrt{det\Sigma_k}} e^{-\frac{1}{2}\mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k)}} \Rightarrow$$
$$\Rightarrow q_T(t_{nk}) = \frac{\frac{w_k}{\sqrt{det\Sigma_k}} e^{-\frac{1}{2}\mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k)}}{\sum\limits_{k=1}^{K} \frac{w_k}{\sqrt{det\Sigma_k}} e^{-\frac{1}{2}\mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k)}}, \qquad q_T(T) = \prod_{n=1}^{N} \prod_{k=1}^{k} q_T(t_{nk}).$$

Рассмотрим формулы пересчета для $q_Z(Z)$.

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

Т. к. t_{nk} — по сути индикатор, то $\mathbb{E}_{q_T(t_{nk})} t_{nk} = q_T(t_{nk} = 1) = q_T(t_{nk})$. Также $\sum\limits_{k=1}^{K} q_T(t_{nk}) = 1 \Rightarrow$

$$\Rightarrow \log q_Z(Z) = \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \Bigg(\log w_k - \frac{D}{2} \log 2\pi - \frac{1}{2} \log det\Sigma_k + \frac{D}{2} \log z_n - \frac{1}{2} z_n (x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k) +$$
$$+ \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \log z_n - \frac{\nu}{2} z_n \Bigg) = \sum_{n=1}^{N} \Bigg(\left(\frac{\nu + D}{2} - 1 \right) \log z_n -$$
$$- \left(\frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk}) (x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k) \right) z_n + \sum_{k=1}^{K} q_T(t_{nk}) \Bigg(\log w_k - \frac{1}{2} \log det\Sigma_k - \frac{D}{2} \log 2\pi +$$
$$+ \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) \Bigg) \Bigg).$$

Т. к. $q_Z(Z) = \prod\limits_{n=1}^{N} q_Z(z_n)$, можем записать $q_Z(z_n)$ для конкретного z_n:

$$\Rightarrow \log q_Z(z_n) = \left(\frac{\nu + D}{2} - 1\right) \log z_n - \left(\frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk})(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) z_n +$$
$$+ \sum_{k=1}^{K} q_T(t_{nk}) \left(\log w_k - \frac{1}{2} \log det \Sigma_k - \frac{D}{2} \log 2\pi + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right)\right).$$

Рассмотрим гамма-распределение $\mathcal{G}(x \mid a, b)$ в общем виде:

$$\mathcal{G}(x \mid a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \Rightarrow \log \mathcal{G}(x \mid a, b) = (a-1) \log x - bx + a \log b - \log \Gamma(a).$$

Сопоставив две формулы выше и рассмотрев функциональный вид $\log q_Z(z_n)$, можно сказать, что $q_Z(z_n) = \mathcal{G}\left(z_n \mid \frac{\nu+D}{2}, \frac{\nu}{2} + \frac{1}{2} \sum\limits_{k=1}^{K} q_T(t_{nk})(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right)$, $q_Z(Z) = \prod\limits_{n=1}^{N} q_Z(z_n)$.

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 $K = 1$.

На M-шаге производится максимизация нижней оценки логарифма неполного правдоподобия:

$$\mathcal{L}(q, w, \mu, \Sigma) = \mathbb{E}_{q_T(T), q_Z(Z)} \log p(X, T, Z \mid w, \mu, \Sigma, \nu) - \mathbb{E}_{q_T(T), q_Z(Z)} \log\left(q_T(T) q_Z(Z)\right) \to \max_{w, \mu, \Sigma}.$$

Распишем данное выражение с точностью до аддитивной константы, не зависящей от w, Σ, μ:

$$\mathbb{E}_{q_T(T), q_Z(Z)} \log p(X, T, Z \mid w, \mu, \Sigma, \nu) = \mathbb{E}_{q_T(T), q_Z(Z)} \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\log w_k - \frac{D}{2} \log 2\pi - \frac{1}{2} \log det \Sigma_k + \frac{D}{2} \log z_n - \right.$$
$$\left. - \frac{z_n}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \log z_n - \frac{\nu}{2} z_n\right) \propto$$
$$\propto \mathbb{E}_{q_T(T), q_Z(Z)} \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\log w_k - \frac{1}{2} \log det \Sigma_k - \frac{1}{2} z_n (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) =$$
$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \mathbb{E}_{q_T(t_{nk})} t_{nk} \left(\log w_k - \frac{1}{2} \log det \Sigma_k - \frac{1}{2} \mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) =$$
$$= \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \left(\log w_k - \frac{1}{2} \log det \Sigma_k - \frac{1}{2} \mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right).$$

Найдем формулы для искомых точечных оценок для w_k, μ_k, Σ_k, максимизируя выражение выше по соответствующему параметру. Для удобства и наглядности записи для каждого параметра будем опускать слагаемые, не зависящие от него и не влияющие на оптимизацию:

w_k:

Имеется задача:

$$\begin{cases} \sum\limits_{n=1}^{N}\sum\limits_{k=1}^{K} q_T(t_{nk}) \log w_k \to \max_{w_k}, \\ \sum\limits_{k=1}^{K} w_k = 1. \end{cases}$$

Запишем функцию Лагранжа, продифференцируем ее по w_k и используем условия $\sum\limits_{k=1}^{K} w_k = 1$ (т. к. w_k — веса компонент смеси), $\sum\limits_{k=1}^{K} q_T(t_{nk}) = 1$:

$$L = \sum_{n=1}^{N}\sum_{k=1}^{K} q_T(t_{nk}) \log w_k + \lambda\left(\sum_{k=1}^{K} w_k - 1\right) \Rightarrow \frac{\partial L}{\partial w_k} = \sum_{n=1}^{N} \frac{q_T(t_{nk})}{w_k} + \lambda = 0 \Rightarrow w_k = -\frac{\sum\limits_{n=1}^{N} q_T(t_{nk})}{\lambda}.$$

$$\sum_{k=1}^{K} w_k = 1 \Rightarrow -\frac{\sum\limits_{n=1}^{N}\sum\limits_{k=1}^{K} q_T(t_{nk})}{\lambda} = 1 \Rightarrow \lambda = -\sum_{n=1}^{N}\sum_{k=1}^{K} q_T(t_{nk}) = -N \Rightarrow w_k = \frac{\sum\limits_{n=1}^{N} q_T(t_{nk})}{N}.$$

μ_k:

Имеем следующую задачу оптимизации:

$$-\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} q_T(t_{nk}) \mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) \to \max_{\mu_k}.$$

Продифференцируем по μ_k:

$$-\frac{1}{2}\sum_{n=1}^{N} q_T(t_{nk}) \mathbb{E}_{q_Z(z_n)} z_n \left(-2\Sigma_k^{-1}(x_n - \mu_k)\right) = \sum_{n=1}^{N} q_T(t_{nk}) \mathbb{E}_{q_Z(z_n)} z_n \Sigma_k^{-1}(x_n - \mu_k) = 0 \Rightarrow$$

$$\Rightarrow \sum_{n=1}^{N} q_T(t_{nk}) \mathbb{E}_{q_Z(z_n)} z_n \Sigma_k^{-1} x_n = \sum_{n=1}^{N} q_T(t_{nk}) \mathbb{E}_{q_Z(z_n)} z_n \Sigma_k^{-1} \mu_k \Rightarrow$$

$$\Rightarrow \mu_k = \frac{\sum\limits_{n=1}^{N} q_T(t_{nk}) \mathbb{E}_{q_Z(z_n)} z_n x_n}{\sum\limits_{n=1}^{N} q_T(t_{nk}) \mathbb{E}_{q_Z(z_n)} z_n}.$$

При $K = 1$ формула совпадает с соответствующей формулой с семинаров:

$$\mu = \frac{\sum\limits_{i=1}^{N} \mathbb{E}_{q_z(z_i)} z_i x_i}{\sum\limits_{i=1}^{N} \mathbb{E}_{q_z(z_i)} z_i}.$$

Σ_k :

Рассмотрим соответствующую задачу оптимизации:

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

Для удобства преобразуем часть данного выражения, использовав циклическое свойство следа:

$$(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k) = tr\left((x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\right) = tr\left(\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\right) \Rightarrow$$
$$\Rightarrow -\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} q_T(t_{nk})\left(\log det\Sigma_k + \mathbb{E}_{q_Z(z_n)} z_n (x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\right) =$$
$$= -\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} q_T(t_{nk})\left(\log det\Sigma_k + \mathbb{E}_{q_Z(z_n)} z_n tr\left(\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\right)\right) \to \max_{\Sigma_k}.$$

Для дальнейшего дифференцирования по Σ_k воспользуемся следующими фактами:

$$\frac{\partial}{\partial X} detX = detX \cdot X^{-1}, \quad \frac{\partial}{\partial X}\left(X^{-1}Y\right) = -X^{-T}Y^TX^{-T}.$$

Продифференцируем по Σ_k:

$$-\frac{1}{2}\sum_{n=1}^{N} q_T(t_{nk})\left(\frac{\partial}{\partial\Sigma_k}\log det\Sigma_k + \mathbb{E}_{q_Z(z_n)} z_n \frac{\partial}{\partial\Sigma_k} tr\left(\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\right)\right) =$$
$$= -\frac{1}{2}\sum_{n=1}^{N} q_T(t_{nk})\left(\Sigma_k^{-1} - \mathbb{E}_{q_Z(z_n)} z_n \Sigma_k^{-T}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-T}\right) = 0.$$

С учетом того, что ковариационная матрица симметричная и ненулевая, можем полученное выражение домножить слева и справа на Σ_k:

$$-\frac{1}{2}\sum_{n=1}^{N} q_T(t_{nk})\left(\Sigma_k - \mathbb{E}_{q_Z(z_n)} z_n (x_n-\mu_k)(x_n-\mu_k)^T\right) = 0 \Rightarrow$$
$$\Rightarrow \sum_{n=1}^{N} q_T(t_{nk})\Sigma_k = \sum_{n=1}^{N} q_T(t_{nk})\mathbb{E}_{q_Z(z_n)} z_n (x_n-\mu_k)(x_n-\mu_k)^T \Rightarrow$$
$$\Rightarrow \Sigma_k = \frac{\sum\limits_{n=1}^{N} q_T(t_{nk})\mathbb{E}_{q_Z(z_n)} z_n (x_n-\mu_k)(x_n-\mu_k)^T}{\sum\limits_{n=1}^{N} q_T(t_{nk})}.$$

При $K = 1$ формула совпадает с соответствующей формулой с семинаров:

$$\Sigma = \frac{\sum\limits_{i=1}^{N} \mathbb{E}_{q_z(z_i)} z_i (x_i-\mu)(x_i-\mu)^T}{N}.$$

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

Распишем выражение для нижней оценки логарифма неполного правдоподобия $\mathcal{L}(q, w, \mu, \Sigma)$ с учетом найденных на Е-шаге распределений $q_T(T)$ и $q_Z(Z)$. Для преобразований воспользуемся выражением для логарифма гамма-распределения в общем виде и фактом того, что $\sum\limits_{k=1}^{K} q_T(t_{nk}) = 1$:

$$
\begin{aligned}
&\mathcal{L}(q, w, \mu, \Sigma) = \mathbb{E}_{q_T(T), q_Z(Z)} \log p(X, T, Z \mid w, \mu, \Sigma, \nu) - \mathbb{E}_{q_T(T), q_Z(Z)} \log \left(q_T(T) q_Z(Z)\right) = \\
&= \mathbb{E}_{q_T(T), q_Z(Z)} \log p(X, T, Z \mid 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)} \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\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. + \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 \right) - \mathbb{E}_{q_T(T)} \log q_T(T) - \\
&- \mathbb{E}_{q_Z(Z)} \sum_{n=1}^{N} \log \mathcal{G}\left(z_n \mid \frac{\nu + D}{2}, \frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk}) (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) = \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \left(\log w_k - \frac{D}{2} \log 2\pi - \frac{1}{2} \log det\Sigma_k - \mathbb{E}_{q_Z(z_n)} z_n \left(\frac{\nu}{2} + \frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) + \right. \\
&\left. + \left(\frac{\nu + D}{2} - 1\right) \mathbb{E}_{q_Z(z_n)} \log z_n + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) \right) - \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \log q_T(t_{nk}) - \\
&- \sum_{n=1}^{N} \mathbb{E}_{q_Z(z_n)} \log \mathcal{G}\left(z_n \mid \frac{\nu + D}{2}, \frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk}) (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) = \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \left(\log w_k - \frac{D}{2} \log 2\pi - \frac{1}{2} \log det\Sigma_k - \mathbb{E}_{q_Z(z_n)} z_n \left(\frac{\nu}{2} + \frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) + \right. \\
&\left. + \left(\frac{\nu + D}{2} - 1\right) \mathbb{E}_{q_Z(z_n)} \log z_n + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) \right) - \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \log q_T(t_{nk}) - \\
&- \sum_{n=1}^{N} \left(\frac{\nu + D}{2} \log \left(\frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk}) (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) - \log \Gamma\left(\frac{\nu + D}{2}\right) + \right. \\
&\left. + \left(\frac{\nu + D}{2} - 1\right) \mathbb{E}_{q_Z(z_n)} \log z_n - \mathbb{E}_{q_Z(z_n)} z_n \left(\frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk}) (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) \right) = \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \left(\log w_k - \frac{1}{2} \log det\Sigma_k - \log q_T(t_{nk}) \right) - \\
&- \sum_{n=1}^{N} \frac{\nu + D}{2} \log \left(\frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk}) (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) - \frac{ND}{2} \log 2\pi + \frac{N\nu}{2} \log \frac{\nu}{2} + N \log \frac{\Gamma(\frac{\nu + D}{2})}{\Gamma(\frac{\nu}{2})} \Rightarrow \\
&\Rightarrow \mathcal{L}(q, w, \mu, \Sigma) = \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \left(\log w_k - \frac{1}{2} \log det\Sigma_k - \log q_T(t_{nk}) \right) - \\
&- \sum_{n=1}^{N} \frac{\nu + D}{2} \log \left(\frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk}) (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right) - \frac{ND}{2} \log 2\pi + \frac{N\nu}{2} \log \frac{\nu}{2} + N \log \frac{\Gamma(\frac{\nu + D}{2})}{\Gamma(\frac{\nu}{2})}.
\end{aligned}
$$

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

В предыдущих формулах использовались следующие статистики распределений $q_T(T)$ и $q_Z(Z)$: $\mathbb{E}_{q_Z(z_n)} z_n, \mathbb{E}_{q_Z(z_n)} \log z_n, \mathbb{E}_{q_T(T)} \log q_T(t_{nk})$. Найдем формулы для них.

Рассмотрим $q_T(T)$.

$$q_T(t_{nk}) = \frac{\frac{w_k}{\sqrt{det \Sigma_k}} e^{-\frac{1}{2}\mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)}}{\sum\limits_{k=1}^{K} \frac{w_k}{\sqrt{det \Sigma_k}} e^{-\frac{1}{2}\mathbb{E}_{q_Z(z_n)} z_n (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)}}.$$

$$\mathbb{E}_{q_T(t_{nk})} \log q_T(T) = \sum_{n=1}^{N} \sum_{k=1}^{K} q_T(t_{nk}) \log q_T(t_{nk}).$$

$$\mathbb{E}_{q_T(t_{nk})} t_{nk} = q_T(t_{nk} = 1) = q_T(t_{nk}).$$

Рассмотрим $q_Z(Z)$.

$$q_Z(z_n) = \mathcal{G}\left(z_n \mid \frac{\nu + D}{2}, \frac{\nu}{2} + \frac{1}{2} \sum_{k=1}^{K} q_T(t_{nk}) (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right).$$

Рассмотрим матожидание для гамма-распределения в общем виде $\mathcal{G}(x \mid a, b)$. Вычислим его явно, используя замену $t = bx$ в интеграле, определение гамма-функции и ее свойство $\Gamma(a+1) = a\Gamma(a)$.

$$\mathbb{E}\mathcal{G}(x \mid a, b) = \int_0^{+\infty} x \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} dx = \int_0^{+\infty} \frac{b^a}{\Gamma(a)} x^a e^{-bx} dx = \frac{1}{b\Gamma(a)} \int_0^{+\infty} t^a e^{-t} dt = \frac{a\Gamma(a)}{b\Gamma(a)} = \frac{a}{b} \Rightarrow$$

$$\Rightarrow \mathbb{E}_{q_Z(z_n)} z_n = \frac{\nu + D}{\nu + \sum\limits_{k=1}^{K} q_T(t_n k)(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)}.$$

Аналогичным способом найдем $\mathbb{E}_{q_Z(z_n)} \log z_n$, используя свойство $\Gamma'(a) = \int_0^{+\infty} t^{a-1} \log t e^{-t} dt$.

$$\mathbb{E} \log \mathcal{G}(x \mid a, b) = \int_0^{+\infty} \frac{\log x b^a x^{a-1} e^{-bx}}{\Gamma(a)} dx = \int_0^{+\infty} \frac{b^a t^{a-1} \log \frac{t}{b} e^{-t}}{b^a \Gamma(a)} =$$

$$= \frac{1}{\Gamma(a)} \left(\int_0^{+\infty} t^{a-1} \log t e^{-t} dt - \int_0^{+\infty} t^{a-1} \log b e^{-t} dt \right) = \frac{\Gamma'(a)}{\Gamma(a)} - \log b = \Psi(a) - \log b,$$

где $\Psi(a)$ — дигамма-функция. Таким образом:

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