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

$$p(x)=\sum_k^K\omega_k\mathcal{T}(x|\mu_k,\Sigma_k,\nu),\qquad \omega_k\geq 0\quad \sum_k\omega_k=1$$

Вероятностная модель : $p(X,T,Z|\omega,\mu,\Sigma,\nu)=\prod_{n,k}^{N,K}\left[\omega_k\mathcal{N}(x_n|\mu_k,\frac{\Sigma_k}{z_n})\,\mathcal{G}(z_n|\nu/2,\nu/2)\right]^{t_{nk}}\Rightarrow$

$$\boxed{\begin{aligned}\log p(X,T,Z|\omega,\mu,\Sigma,\nu)=\sum_{k,n}^{K,N}t_{nk}\Big[&\log\omega_k-\tfrac12\log|\Sigma_k|+\tfrac12\log z_n-\tfrac D2\log 2\pi-\tfrac12(x_n-\mu_k)^T\Sigma_k^{-1}z_n(x_n-\mu_k)\\&+\tfrac\nu2\log\tfrac\nu2+(\tfrac\nu2-1)\log z_n-\tfrac\nu2 z_n-\log\Gamma(\tfrac\nu2)\Big]\end{aligned}}$$

1 E шаг :

Воспользуемся формулой для Mean-field approx. : $\log q_i(z_i)=\mathbb{E}_{\substack{q_j(z_j)\\ j\neq i}}\log p(X,Z)+\text{const}$

- $\log q_T(T)=\mathbb{E}_{q_Z(Z)}\log p(X,T,Z|\omega,\mu,\Sigma,\nu)+\text{const}=\mathbb{E}_Z\sum_{n=1}^N\sum_{k=1}^K t_{nk}\Big[(\tfrac\nu2-1+\tfrac12)\log z_n-\big(\tfrac12(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\tfrac\nu2\big)z_n+$ const – не зависящие от $z_n\Big]+\text{Const}=\sum_n\sum_k t_{nk}\Big[\underbrace{(\tfrac\nu2-1+\tfrac12)\mathbb{E}_{z_n}\log z_n-\big(\tfrac12(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\tfrac\nu2\big)\mathbb{E}_{z_n}z_n}_{\log p_{nk}}\Big]+$ $\text{Const}=\sum_n\sum_k t_{nk}\log p_{nk}+\text{Const}$

Значит, $q_T(T)=\prod_n^N q_T(T_n)$, где $q_T(T_n)=\frac{\prod_k p_{nk}^{t_{nk}}}{A}$, $A=\sum_k p_{nk}$ $\Rightarrow$ $q_T(T)=\prod_n\frac{\prod_k p_{nk}^{t_{nk}}}{\sum_k p_{nk}}$ а $q_T(t_{nk}=1)=\frac{p_{nk}}{\sum_k p_{nk}}=r_{nk}$

- $\log q_Z(Z)=\mathbb{E}_{q_T(T)}\log p(X,T,Z|\omega,\mu,\Sigma,\nu)+\text{const}=\mathbb{E}_T\sum_{n=1}^N\sum_{k=1}^K t_{nk}[\ldots\text{ не зависящее от } T]=\sum_n^N\sum_k^K[\ldots]\overset{r_{nk}}{\overset{\shortparallel}{\mathbb{E}_{T_n}t_{nk}}}=$

$\sum_{n=1}\sum_{k=1}r_{nk}\Big[(\tfrac\nu2-1+\tfrac12)\log z_n-\big(\tfrac12(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\tfrac\nu2\big)z_n+\text{const}$ – не завис. от $z_n\Big]$

$\sum_n\Big[(\tfrac\nu2-1+\tfrac12)\big(\overset{1}{\overset{\shortparallel}{\sum_k r_{nk}}}\big)\log z_n-\sum_k r_{nk}\big(\tfrac12(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\tfrac\nu2\big)z_n+\text{Const}\Big]+\text{Const}=$

$\sum_n\log\mathcal{G}\Big(z_n\,\Big|\,\underbrace{\tfrac{\nu+1}{2}}_{a},\ \underbrace{\sum_k r_{nk}\tfrac{(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\nu}{2}}_{b}\Big)$

Значит, $q_Z(Z)=\prod_n q_Z(z_n)$, $q_Z(z_n)=\mathcal{G}(z_n|a,b)$ $\quad a=\frac{\nu+1}{2}$, $b=\sum_k r_{nk}\frac{(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)+\nu}{2}$

2 M шаг :

$\mathbb{E}_{q(T,Z)}\log p(X,T,Z|\omega,\mu,\Sigma,\nu)=\mathbb{E}_{q(T)q(Z)}\sum_k^K\sum_n^N t_{nk}\Big[\log\omega_k+\log\mathcal{N}(x_n|\mu_k,\frac{\Sigma_k}{z_n})+\cancel{\log\mathcal{G}(z_n|\nu/2,\nu/2)}\Big]$ не зависит от ω,μ,Σ

$\geq\mathbb{E}_{q(T)q(Z)}\sum_k^K\sum_n^N t_{nk}\Big[\log\omega_k-\tfrac12\log|\Sigma_k|+\tfrac12\log z_n-\tfrac D2\log2\pi-\tfrac12(x_n-\mu_k)^T\Sigma_k^{-1}z_n(x_n-\mu_k)\Big]=$

$=\sum_n^N\sum_k^K\mathbb{E}_{z_n}\underbrace{\mathbb{E}_{T_n}t_{nk}}_{r_{nk}}\Big[\log\omega_k-\tfrac12\log|\Sigma_k|+\tfrac12\log z_n-\tfrac D2\log2\pi-\tfrac12(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)z_n\Big]=\quad(*)$

- Обновление ω_k: добавляем к (*) условие $\sum_k\omega_k=1$ с помощью метода множителей Лагранжа: $+\lambda(\sum_k\omega_k-1)$

$\frac{\partial\ldots}{\partial\omega_k}=\sum_n^N r_{nk}\omega_k^{-1}-\lambda=0\Rightarrow\omega_k=\lambda^{-1}\sum_n r_{nk}$

Так как $\sum\omega_k=1$, значит $\lambda=\sum_k\sum_n r_{nk}=\sum_n\overset{1}{\overset{\shortparallel}{\sum_k r_{nk}}}=N\Rightarrow\omega_k=\overline{r_{nk}}$ – среднее по выборке

• Обновление μ_k:

$$\frac{\partial \dots}{\partial \mu_k} = \sum_n^N r_{nk} \mathbb{E}_{z_n} z_n \left(\Sigma_k^{-1} x_n - \Sigma_k^{-1}\mu_k\right) = 0 \Rightarrow \mu_k = \frac{\sum_n^N r_{nk} \overbrace{\mathbb{E}_{z_n} z_n}^{s_n} x_n}{\sum_n^N r_{nk} \mathbb{E}_{z_n} z_n} = \frac{\sum^N r_{nk} s_n x_n}{\sum_n r_{nk} s_n}$$

• Обновление Σ_k

$$\frac{\partial \dots}{\partial \Sigma_k^{-1}} = \sum^N \sum^K r_{nk} \left[\frac{1}{2}\Sigma_k - \frac{1}{2} s_n (x_n-\mu_k)(x_n-\mu_k)^T\right] = 0 \Rightarrow \Sigma_k = \frac{\sum_n^N r_{nk} s_n (x_n-\mu_k)(x_n-\mu_k)^T}{\sum_n^N r_{nk}}$$

3

$$\mathcal{L}(q, \omega_k, \mu_k, \Sigma_k) = \mathbb{E}_{q(T,Z)} \log p(X, T, Z \mid \omega_k, \mu_k, \Sigma_k, \nu) - \mathbb{E}_{q(T,Z)} \log q(T,Z) =$$

$$\mathbb{E}_{q(T,Z)} \sum_{k,n} t_{nk} \Big[\log \omega_k - \frac{1}{2}\log|\Sigma_k| + \frac{1}{2}\log z_n - \frac{D}{2}\log 2\pi - \frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1} z_n (x_n - \mu_k)$$

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

$$\boxed{\begin{aligned}\mathcal{L}(q, \omega_k, \mu_k, \Sigma_k) = & \sum_{k,n} r_{nk} \Big[\log\omega_k - \frac{1}{2}\log|\Sigma_k| + \left(\frac{1}{2}+\frac{\nu}{2}-1\right)\mathbb{E}_{z_n}\log z_n - \frac{D}{2}\log 2\pi - \frac{1}{2}\left((x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k) + \nu\right)\mathbb{E}_{z_n} z_n \\ & \frac{\nu}{2}\log\frac{\nu}{2} - \log\Gamma\left(\frac{\nu}{2}\right)\Big] - \mathbb{E}_T \log q(T) - \mathbb{E}_Z \log q(Z)\end{aligned}}$$

4

Статистики:

• $\mathbb{E}_{q(T_n)} t_{nk} = r_{nk} = \frac{p_{nk}}{\sum_k p_{nk}}$

• $\mathbb{E}_{q(z_n)} z_n = s_n = \{\text{для гамма-распределения}\} = \frac{1+\nu}{\sum_k r_{nk}\left((x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)+\nu\right)}$

• $\mathbb{E}_{q(z_n)} \log z_n = \Psi\left(\frac{1+\nu}{2}\right) - \log\left(\frac{\sum_k r_{nk}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)}{2}\right)$

• $\mathbb{E}_{q(T_n)} \log q(T_n) = \mathbb{E}_{T_n} \sum_k t_{nk} \log p_{nk} - \overbrace{\log\sum_k p_{nk}}^{C} = \sum_k r_{nk}\log p_{nk} - C$

• $\mathbb{E}_{q(z_n)} \log q(z_n) = \mathbb{E}_{z_n} \overbrace{\frac{\nu}{2}\log\frac{\nu}{2}}^{C} + \left(\frac{\nu}{2}-1\right)\log z_n - \frac{\nu}{2} z_n - \overbrace{\log\Gamma\left(\frac{\nu}{2}\right)}^{C} = \left(\frac{\nu}{2}-1\right)\mathbb{E}_{z_n}\log z_n - \frac{\nu}{2}\overbrace{\mathbb{E}_{z_n} z_n}^{s_n} - C$