Теорма: $N, G, \mathbb{E}$ росс. стар. G,

① Смесь распред. в общем виде

Генерация из смеси:

$t \sim \mathrm{Discrete}(\pi) \to x \sim p_t(x|\theta)$

БММО EM

$$p(x) = \sum_k \pi_k p_k(x|\theta) \qquad \sum_i \pi_i = 1 \quad \pi_i \geq 0$$

Общая вер. модель: $p(x,t) = \prod_k [\pi_k p_k(x|\theta)]^{t_k}$ $\quad t \in \{0,1\}^k \quad \sum_i t_i = 1$

Проверим: $\sum_t p(x,t) = p(x)$

$x_1, \ldots, x_N \sim p(x)$ $\qquad$ ММП: $p(x|\pi,\theta) \to \max_{\pi,\theta}$

$$p(X,T|\pi,\theta) = \prod_n \prod_k [\pi_k p_k(x_n|\theta)]^{t_{nk}}$$

E-шаг: $q(T) = p(T|X,\pi,\theta) \propto p(X,T|\pi,\theta) = \prod_n q_n(t_n)$

$$q_n(t_n) \propto \prod_k [\pi_k p_k(x_n|\theta)]^{t_{nk}} \to q(t_{nk}=1) = \frac{\pi_k p_k(x_n|\theta)}{\sum_j \pi_j p_j(x_n|\theta)} = \gamma_{nk}$$

M-шаг: $\mathbb{E}_q \log p(X,T|\pi,\theta) = \mathbb{E}_q \sum_n \sum_k t_{nk}(\log \pi_k + \log p_k(x_n|\theta)) =$

$= \{\mathbb{E} t_{nk} = \gamma_{nk}\} = \sum_{nk} \gamma_{nk}(\log \pi_k + \log p_k(x_n|\theta))$

$\pi_k = \frac{\sum_n \gamma_{nk}}{N}$ ← из Лагранжа $\sum_j \pi_j = 1$

Оптим. θ зависит от модели: $\sum_{nk} \gamma_{nk} \log p_k(x_n|\theta) \to \max_\theta$

② Конкретный пример смеси

$x_1, \ldots, x_N \sim \gamma p_1(x) + (1-\gamma) p_2(x)$ $\qquad$ $p_1(x)$: $\begin{array}{ccc} 1 & 2 & 3 \\ \alpha & (1-\alpha) & 0 \end{array}$ $\quad$ $p_2(x)$: $\begin{array}{ccc} 1 & 2 & 3 \\ 0 & (1-\beta) & \beta \end{array}$

$\alpha_0, \beta_0, \gamma_0 = \frac{1}{2}$ $\quad N_1 = 30 \quad N_2 = 20 \quad N_3 = 60$

Модель: $z_n = [x_n \in p_1]$ $\quad p(x,z|\gamma,\alpha,\beta) = \prod_n [\gamma p_1(x_n|\alpha)]^{z_n} [(1-\gamma) p_2(x_n|\beta)]^{(1-z_n)}$

E: $q(z) = \prod_n q_n(z_n)$ $\quad q_n(z_n=1) = \frac{\gamma p_1(x_n|\alpha)}{\gamma p_1(x_n|\alpha) + (1-\gamma) p_2(x_n|\beta)}$

$x_n = 1 \Rightarrow \lambda_1 = 1$, $x_n = 2 \Rightarrow \lambda_2 = \frac{1}{2} = \frac{(1-\alpha)\gamma}{(1-\alpha)\gamma + (1-\gamma)(1-\beta)}$ $\quad \lambda_3 = 0$. $\qquad \lambda_i = q(z_n=1)|_{x_n=i}$

M: $\mathbb{E}_z \log p(x,z|\alpha,\beta,\gamma) = 70 \log(1-\gamma) + 10\log(1-\beta) + 60 \log\beta + 40 \log\gamma + 30 \log\alpha + 10 \log(1-\alpha)$

$\to \alpha = \frac{3}{4} \quad \beta = \frac{6}{7} \quad \gamma = \frac{4}{11} \to \lambda_2^{new} = \frac{1}{2} \to$ сошлись.

③ Распред. Стьюдента.

$x \in \mathbb{R}$ $\quad \tau(x|\mu,\sigma^2,\nu) \propto \left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$ $\qquad \mathbb{E}x = \mu, \ \nu > 1$; $\ \mathbb{D}x = \sigma^2 \cdot \frac{\nu}{\nu-2}, \ \nu > 2$

$\left(1+\frac{1}{n}\right)^{kn} \to e^k$; $n' = nk$, $\nu' = \nu/2$; $\left(1 + \frac{1}{2\nu'}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\nu'} \to e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$\nu \to \infty \Rightarrow \tau \to N(\mu, \sigma^2)$ $\qquad \nu \simeq 10 \to \tau$ и N очень близки

$x \in \mathbb{R}^D$ $\quad \tau(x|\mu,\Sigma,\nu) \propto \left(1 + \frac{1}{\nu}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)^{-\frac{\nu+D}{2}}$ $\qquad \mathbb{E}x = \mu$; $\ \mathrm{Cov}\,x = \frac{\nu}{\nu-2}\Sigma$

$$\tau(x|\mu,\Sigma,\nu) = \int_0^{+\infty} N\left(x|\mu, \frac{1}{z}\Sigma\right) G\left(z\Big|\frac{\nu}{2}, \frac{\nu}{2}\right) dz$$

$z \in \mathbb{R}^+$ — непрерывная смесь распред.

$\Rightarrow$ Генерация из Стьюдента: $z \sim G \to x \sim N(\ldots|z)$

④ $x_1, \ldots, x_N \sim \tau(x|\mu,\Sigma,\nu)$ $\qquad \mu_{ML}, \Sigma_{ML}, \nu_{ML}$ — ?

ММП напрямую сложно: $\log(1 + \text{Махаланобис})\ldots$

Применим EM. $p(x,z|\mu,\Sigma,\nu) = \prod_n N\left(x_n|\mu, \frac{1}{z_n}\Sigma\right) G\left(z_n\Big|\frac{\nu}{2},\frac{\nu}{2}\right)$

E-шаг: $q(z) = \prod_n q_n(z_n)$

$$q_n(z_n) \propto N\left(x_n|\mu, \frac{1}{z_n}\Sigma\right) G\left(z_n\Big|\frac{\nu}{2},\frac{\nu}{2}\right) = G(z_n|a_n, b_n)$$

$a_n = \frac{D+\nu}{2}$ $\qquad b_n = \frac{\nu}{2} + \frac{1}{2}(x_n-\mu)^T \Sigma^{-1}(x_n-\mu)$

M-шаг:

$$\mathbb{E}_q \log p(x,z|\mu,\Sigma,\nu) \propto \sum_n \Big[\frac{D}{2}\mathbb{E}\log z_n - \frac{1}{2}\log|\Sigma| - \frac{\mathbb{E}z_n}{2}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu) + \frac{\nu}{2}\log\frac{\nu}{2} - \log\Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2}-1\right)\mathbb{E}\log z_n - \frac{\nu}{2}\mathbb{E}z_n\Big]$$

Из дост. стат. G: $\mathbb{E} z_n = \frac{a_n}{b_n}$, $\mathbb{E}\log z_n = \psi(a_n) - \log b_n$

$\leftarrow$ дигамма ф-ция $= \frac{\partial}{\partial a_n}\log\Gamma(a_n)$

$$\frac{\partial}{\partial\mu} = \sum_n \mathbb{E} z_n \Sigma^{-1}(x_n-\mu) = 0 \Rightarrow \mu = \frac{\sum_n \mathbb{E} z_n x_n}{\sum_n \mathbb{E} z_n}$$

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

ν: нет аналит. формы $\Rightarrow$ одномерная оптим. выпуклой ф-ции.

(5) Смесь распред. Стьюдента

$$p(x) = \sum_k \pi_k \tau(x \mid \mu_k, \Sigma_k, \nu_k)$$

$$p(x, z, T \mid \mu, \nu, \Sigma, \pi) = \prod_n \prod_k \left[\pi_k N\left(x_n \mid \mu_k, \frac{1}{z_n}\Sigma_k\right) G\left(z_n \mid \frac{\nu_k}{2}, \frac{\nu_k}{2}\right)\right]^{t_{nk}}$$

E-шаг: $q(T, z) = \prod_n q_n(z_n, t_n) = \prod_n q(z_n \mid t_n) q(t_n)$

$$q(t_n, z_n) \propto \prod_k \left[\pi_k N\left(x_n \mid \mu_k, \frac{1}{z_n}\Sigma_k\right) G\left(z_n \mid \frac{\nu_k}{2}, \frac{\nu_k}{2}\right)\right]^{t_{nk}}$$

$$q(t_n) \propto \int_z q(t_n, z_n)\, dz \rightarrow q(t_{nk}=1) \propto \int_z \pi_k N \cdot G\, dz = \frac{\pi_k \tau(x_n \mid \mu_k, \Sigma_k, \nu_k)}{\sum_i \pi_i \tau(x_n \mid \mu_i, \Sigma_i, \nu_i)}, \gamma_{nk}$$

$$q(z_n \mid t_{nk}) \propto q(z_n, t_n)$$

$$q(z_n \mid t_{nk}=1) \propto \pi_k \cdot N \cdot G = G(z_n \mid a_{nk}, b_{nk})$$

$$a_{nk} = \frac{D+\nu_k}{2} \qquad b_{nk} = \frac{\nu_k}{2} + \frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)$$

$\rightarrow$ $\mathbb{E}_{q(t_n)} t_{nk}$, $\mathbb{E}_{q(z_n \mid t_{nk})} z_n$, $\mathbb{E}_{q(z_n \mid t_{nk})} \log z_n$

M-шаг:

$$\mathbb{E}_q \log p(x, z, T \mid \mu, \nu, \Sigma, \pi) = \mathbb{E}_q \sum_n \sum_k t_{nk}\left[\log\pi_k + \log N\left(x_n \mid \mu_k, \frac{1}{z_n}\Sigma_k\right) + \log G\left(z_n \mid \frac{\nu_k}{2}, \frac{\nu_k}{2}\right)\right] =$$

$$= \sum_n \sum_k \mathbb{E}_q\left[t_{nk}\left(\log\pi_k + \log N + \log G\right)\right] =$$

$$= \sum_k \sum_n \gamma_{nk}\Big(\log\pi_k + \frac{d}{2}\mathbb{E}\log z_n - \frac{1}{2}\log|\Sigma_k| - \frac{\mathbb{E} z_n}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k) + \frac{\nu_k}{2}\log\frac{\nu_k}{2} -$$

$$- \log\Gamma\left(\frac{\nu_k}{2}\right) + \left(\frac{\nu_k}{2}-1\right)\mathbb{E}\log z_n - \frac{\nu_k}{2}\mathbb{E} z_n\Big)$$

$\leftarrow$ $\mathbb{E}_{q(z_n \mid t_{nk}=1)} z_n$, $\mathbb{E}_{q(z_n \mid t_{nk}=1)} \log z_n$

$$\frac{\partial}{\partial\mu_k} = \sum_n \gamma_{nk}\mathbb{E} z_n \Sigma_k^{-1}(x_n-\mu_k) = 0 \qquad \mu_k = \frac{\sum_n \gamma_{nk}\mathbb{E} z_n x_n}{\sum_n \gamma_{nk}\mathbb{E} z_n}$$

— " — " —

$$\sum_n \mathbb{E}_q \sum_k t_{nk} f(z_n) = \sum_n \sum_{i=1}^{K} q(t_n=i)\int_0^{+\infty} q(z_n \mid t_n=i)\sum_k t_{nk} f(z_n)\, dz_n = \{t_{ni}=1, \text{остальные}=0\}$$

$$= \sum_n \sum_{i=1}^{K} \gamma_{ni}\int_0^{+\infty} q(z_n \mid t_n=i) f(z_n)\, dz_n = \sum_n \sum_k \gamma_{nk}\mathbb{E}_{q(z_n \mid t_n=k)} f(z_n)$$

Итого: за ту же сложность можем вместо разделения смеси гауссиан делать разделение смеси Стьюдентов.
Основная разница в необходимости одномерной оптимизации по ν, но ν можно и просто фиксированным взять.

$$\frac{\partial}{\partial x} x^T A x = (A + A^T)x$$

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

$$\frac{\partial}{\partial A} x^T A y = x y^T$$