Семинар 5

EM:

1. Init $\Theta_{old} = [\ldots]$
2. E1: $p(Z \mid X, \Theta_{old})$

   E2: $Q(\Theta, \Theta_{old}) = \mathbb{E}_{Z \mid X, \Theta_{old}} (\ell(x, z \mid \Theta) \mid x, \Theta_{old})$
3. M: $Q(\Theta, \Theta_{old}) \to \max\limits_{\Theta}$

Пример

$X_1 \ldots X_n \overset{iid}{\sim} \underbrace{f(x \mid \Theta)}$ вектор парам-в

Два кластера:

$Z \in \{1, 2\}$

$\mathbb{P}\{Z = 1\} = p_1$

$\mathbb{P}\{Z = 2\} = p_2 = 1 - p_1$

Z	1	2
$\mathbb{P}\{Z = z_i\}$	p_1	$1-p_1$

$1 \times k$

кол-во значений латентной перем

$$\ell(x \mid \Theta) = \sum_i \ln f(x_i \mid \Theta) =$$

$$= \sum_j \sum_i \underbrace{\mathbb{P}\{Z = j\}} \ln \underbrace{f(x_i \mid \Theta)} =$$

$$= \sum_i \sum_j \mathbb{P}\{Z=j\} \ln \frac{f(x_i, Z=j \mid \Theta)}{\underbrace{\mathbb{P}\{Z=j \mid x_i, \Theta\}}} =$$

$$p$$

$$\mathbb{P}\{Z=j \mid x_i, \Theta\} = \frac{\mathbb{P}\{Z=j, x_i, \Theta\}}{f(x_i \mid \Theta)}$$

$$= \sum_i \sum_j \mathbb{P}\{Z=j\} \ln \left[\frac{f(x_i, Z=j \mid \Theta) \cdot \mathbb{P}\{Z=j\}}{\mathbb{P}\{Z=j \mid x_i, \Theta\} \cdot \mathbb{P}\{Z=j\}} \right] =$$

$$\mathcal{L}(\mathbb{P}\{Z=j\}, \Theta)$$

$$= \sum_i \sum_j \mathbb{P}\{Z=j\} \ln \left[\frac{f(x_i, Z=j \mid \Theta)}{\mathbb{P}\{Z=j\}} \right] +$$

$$+ \underbrace{\sum_i \sum_j \mathbb{P}\{Z=j\} \ln \left[\frac{\mathbb{P}\{Z=j\}}{\mathbb{P}\{Z=j \mid x_i, \Theta)} \right]}$$

$$\underbrace{p \ln p - p \ln q}_{H \quad - \quad CE}$$

$$D_{KL}(\mathbb{P}\{Z=j\} \parallel \mathbb{P}\{Z=j \mid x_i, \Theta\})$$

$$\log_{\frac{1}{2}} : \ KL = CE - H \qquad \ln : \ KL = H - CE$$

$$\underbrace{\ell(x|\Theta)} = \underbrace{\mathcal{L}(\overset{=q_z}{\mathbb{P}\{z=j\}}, \Theta)}_{\text{нижняя оценка}} + \underbrace{D_{KL}(\mathbb{P}\{z=j\} \,\|\, \mathbb{P}\{z=j \mid x, \Theta\})}_{\geq 0}$$

EM: E : $\max \mathbb{P}\{z=j\}$

M : $\max \Theta$

$$\boxed{\Theta = (\ldots, p_1)}$$

<u>E-шаг</u>: $\max\limits_{\mathbb{P}\{z=j\}} \mathcal{L}(\ldots)$

$$\mathcal{L} = \underbrace{\ell(\ldots)}_{\text{не зав. } \mathbb{P}\{z=j\}} - D_{KL}(\ldots) \Rightarrow \mathcal{L} \to \max \Leftrightarrow D_{KL} \to \min$$

$$D_{KL}(A \| B) \overset{=0}{\min} \Leftrightarrow A = B$$

E-шаг $\longrightarrow$ $\boxed{q = \mathbb{P}\{z=j\} = \mathbb{P}\{z=j \mid x, \Theta\}}$

$$\ln f(\ldots) - \cancel{\ln \mathbb{P}\{z=j\}}$$

$$\mathcal{L} = \sum_i \sum_j \mathbb{P}\{z=j\} \ln \boxed{\frac{f(x_i, z=j \mid \Theta)}{\mathbb{P}\{z=j\}}} \to \max_\Theta$$

$$\sim \sum_i \sum_j \underbrace{\mathbb{P}\{z=j \mid x, \Theta_{old}\}} \ln \underbrace{f(x_i, z=j \mid \Theta)} =$$

$$= \sum_i \underbrace{\mathbb{E}_{z|x_i,\Theta_{old}}}\ \ln \underbrace{f(x_i, z=j|\Theta)} =$$

side-story

$$\boxed{\mathbb{E}(X) = \langle x, p \rangle}$$

$$= Q(\Theta, \Theta^{old}) \to \max_{\Theta}$$

$$Q(\Theta, \Theta_{old}) = \mathbb{E}_{z|X,\Theta_{old}} \ln p(x, z|\Theta)$$

$p(z)$: массив $1 \times k$ k - число знач. z

$p(z|X, \Theta_{old})$ массив $N \times k$

$$\begin{array}{c|ccc} & 1 & z_i \ldots & k \\ \hline 1 & \cdot & \cdot\ \cdot\ \cdot & \cdot \\ x_i\ \vdots & & & \\ N & & & \end{array}$$

$z=1$ $z=2$ $z=3$

$X=1$ $p(X=1|z=1)$ $p(X=1|z=2)$

$X=2$

Пример

Разделение смеси распределений Бернулли

$X_1 \dots X_n$

1кл. 2кл.

$X_i \sim Bern(p_1)$ $X_i \sim Bern(p_2)$

$Z \in \{1, 2\}$ – номер кластера

$\Theta = (p_1, p_2, \gamma)$ $\gamma := \mathbb{P}\{Z = 1\}$

$\hat{\Theta}$ при помощи EM-алгоритма

E1:

$$p(z_i = 1 \mid x_i, \Theta_{old}) =$$

$$= \boxed{\frac{p(x_i \mid z_i = 1, \Theta_{old}) \cdot \gamma}{\gamma\, p(x_i \mid z_i = 1, \Theta_{old}) + \underbrace{(1-\gamma) \cdot p(x_i \mid z_i = 2, \Theta_{old})}}}$$

$x_i = \begin{cases} 0 \\ 1 \end{cases}$ (пришли из Bern

x_i $1-x_i$

$$= \frac{p_1 \quad (1-p_1) \quad \cdot \gamma}{\gamma p_1^{x_i}(1-p_1)^{1-x_i} + (1-\gamma) p_2^{x_i}(1-p_2)^{1-x_i}} =$$

$$= g_i$$

side-story

$$X = \begin{cases} 1, & p \\ 0, & 1-p \end{cases} \quad \mathbb{P}\{X=k\} = p^k \cdot (1-p)^{1-k}$$

$$X=0 \Rightarrow \quad 1-p$$
$$X=1 \Rightarrow \quad p$$

$$Q(\Theta, \Theta^{old}) = \sum_i g_i \left[\ln \underbrace{p(x_i|\Theta)}_{} + \ln\gamma \right] +$$

1кл.

$$+ (1-g_i)\left[\ln \underbrace{p(x_i|\Theta)}_{} + \ln(1-\gamma)\right]$$

2кл.

$$\to \max_{\Theta}$$

$$(p_1, p_2, \gamma)$$

$$x_i \to 1, p$$
$$x_i \to 0, 1-p$$

$$p(x_i|\Theta) = p_1^{x_i} \quad (1-p_1)^{1-x_i}$$

$$P(x_i) = p_1^{x_i} \quad (1-p_1)^{1-x_i}$$

$x_i = 0 \Rightarrow 1-p_1$

$x_i = 1 \Rightarrow p_1$

$$Q = \sum_i \underline{\underline{g_i}} \left[x_i \ln p_1 + (1-x_i) \ln(1-p_1) + \ln \gamma \right] +$$

$$+ \underline{\underline{(1-g_i)}} \left[x_i \ln p_2 + (1-x_i) \ln(1-p_2) + \ln(1-\gamma) \right]$$

$$\longrightarrow \max_{p_1, p_2, \gamma}$$

$$Q'_{p_1} = \sum_i g_i \left[\frac{x_i}{p_1} - \frac{(1-x_i)}{1-p_1} \right]$$

$$\frac{1}{\hat{p}_1} \sum_i g_i x_i = \frac{\sum_i g_i (1-x_i)}{1-\hat{p}_1}$$

$$\sum_i g_i x_i - \hat{p}_1 \sum_i g_i x_i = \hat{p}_1 \sum g_i (1-x_i)$$

$$\sum_i g_i x_i = \hat{p}_1 \left(\sum g_i (1-x_i) + \sum g_i x_i \right)$$

$$\hat{p}_1 = \frac{\sum g_i x_i}{\sum g_i (1-x_i) + \sum g_i x_i}$$

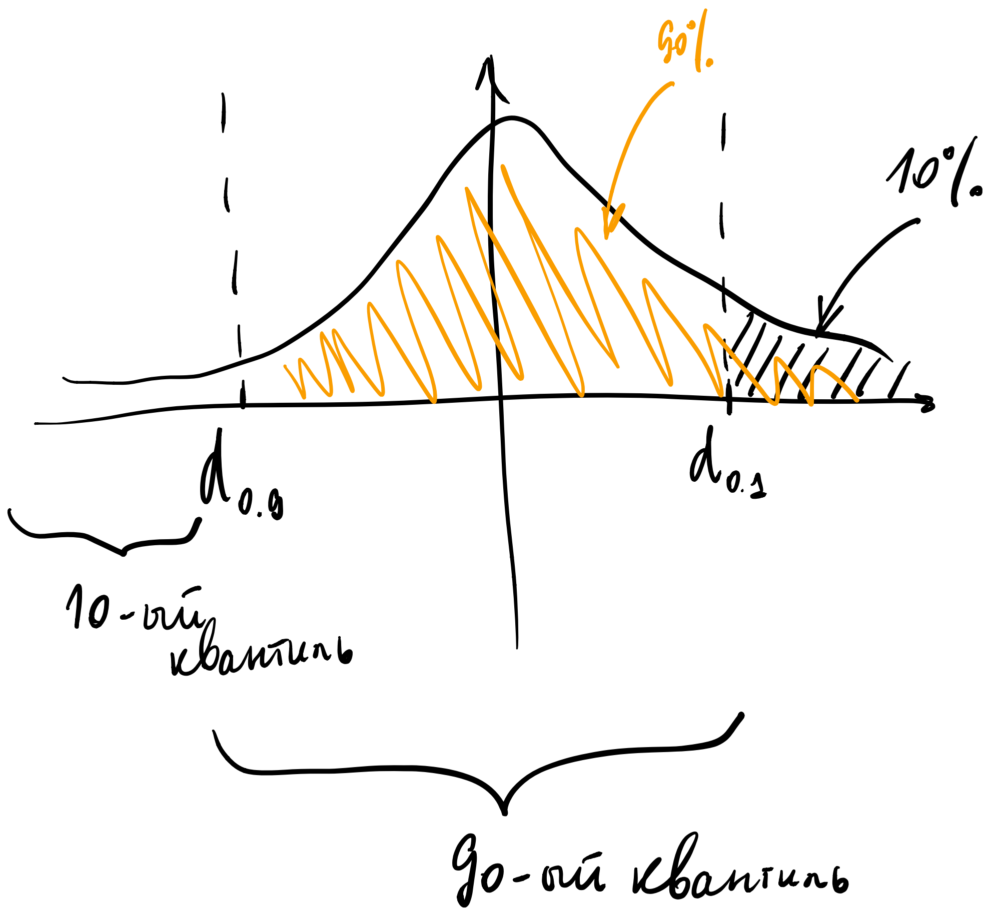
$$\hat{p}_2 = \frac{\sum_i (1-g_i) x_i}{\;}$$

$$\sum_i (1-g_i)(1-x_i) + \sum_i (1-g_i)x_i$$

вывести

$$\hat{\gamma} = \frac{\sum_i g_i}{n}$$

$$\Theta_{new} = (\hat{p}_1, \hat{p}_2, \hat{\gamma})$$


90%
10%
$d_{0.9}$
$d_{0.1}$
10-ый квантиль
90-ый квантиль