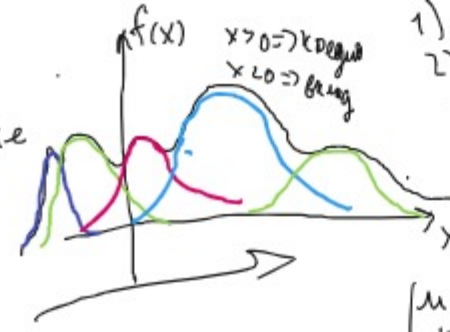
EM - алгоритм
20 - квиз
20 - EM на практике
40 - $\begin{cases} N(\mu_1, \sigma_1^2) \\ N(\mu_2, \sigma_2^2) \end{cases}$



1) многомерный
2) бинарные индикаторы
3) pos
4) 500 - линий
5) вещественный

$\mu_1, \mu_2, \ldots, \mu_5, \sigma_1^2, \ldots, \sigma_5^2, p_1, p_2, p_3, p_4, p_5 = 1 - \ldots$

x_i	$z_i \in \{1, \ldots 5\}$	p_1	p_2	p_3	p_4	p_5
		0,8	0,1	0,05	0,04	0,01
-18	1					
-5	2					
2	3					
4	3					
...	...					

z - то, что хотелось бы знать, но не знаем

$L(X|\theta) \rightarrow L(X, z|\theta)$

$L(X|\theta)$ - целевое пр-ие

$L(X, z|\theta)$ - полное пр-ие (z - латентные)

$$L(X|\theta) = \sum_i \ln p(x_i|\theta) = \sum_i \sum_j p(z_i=j) \cdot \ln \left[\frac{p(x_i, z_i=j|\theta)}{p(z_i=j|x_i,\theta)} \right] = \ln p(x|\theta)$$

$$= \sum_i \sum_j p(z_i=j) \cdot \ln \left[\frac{p(x_i, z_i=j|\theta)}{p(z_i=j|x_i,\theta)} \cdot \frac{p(z_i=j)}{p(z_i=j)} \right] =$$

$$= \sum_i \sum_j p(z_i=j) \cdot \ln \frac{p(x_i, z_i=j|\theta)}{p(z_i=j)} + \sum_i \sum_j p(z_i=j) \cdot \ln \left\{ \frac{p(z_i=j)}{p(z_i=j|x_i,\theta)} \right\}$$

$p(x, z) = p(z|x) p(x)$

M - Вова

L - Соколов

LB

$LB(p_z, p_{z,x|\theta})$

$-D_{KL}[z \| (z|x,\theta)] \geq 0$

$p(z) = p(z|x,\theta)$

EM-алгоритм

Init $\theta_0 = \begin{cases} \mu_1 = \min(x) \\ \mu_2 = x_{0,1} \\ \sigma_1^2 = \frac{\max x - \min x}{5} \\ p_1 = \frac{\# x}{5} \end{cases}$

1) $p(z|x,\theta) = p(z)$

E-шаг $p(z) = p(z|x, \theta_{old})$ 2) $E_z[L(x,z|\theta)|x, \theta_{old}] = Q(\theta, \theta_{old})$

M-шаг $Q(\theta, \theta_{old}) \rightarrow \max_\theta$

обратно к E

$\mu_1 = \frac{\sum p_1 x_i}{\sum p_1}$

$\sigma_1^2 = \frac{\sum p_1 (x_i - \bar{x})^2}{\sum p_1}$

$p(z|x,\theta) \rightarrow p(z)$

$$p(z_i=1|x_i,\theta) = \frac{p(z_i=1, x_i|\theta)}{p(x_i|\theta)} = \frac{p(x_i|z_i=1,\theta) \cdot p(z_i=1)}{p(x_i|\theta, z_i=1) \cdot p(z_i=1) + p(x_i|\theta, z_i=2) \cdot p(z_i=2)} = g(\theta_{old})$$