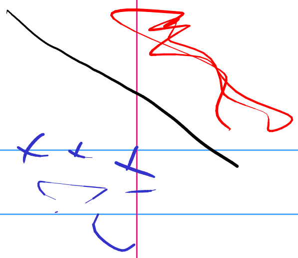
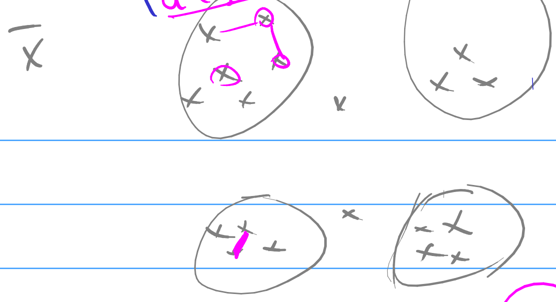
Unsupervised learning



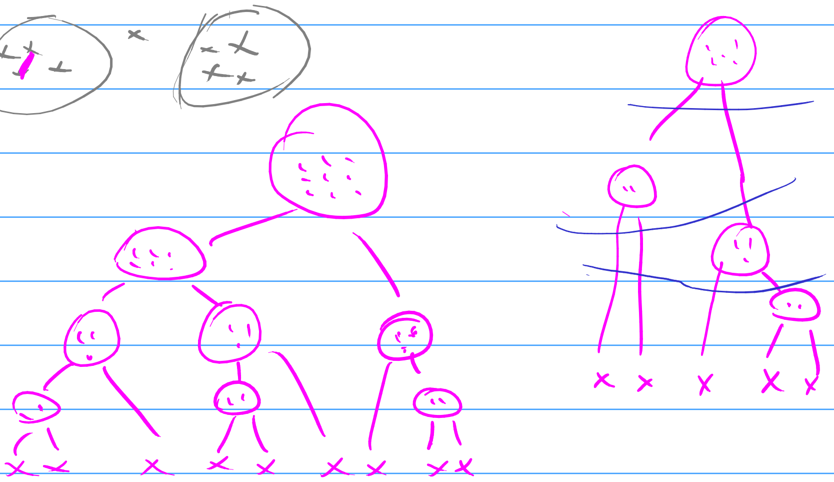


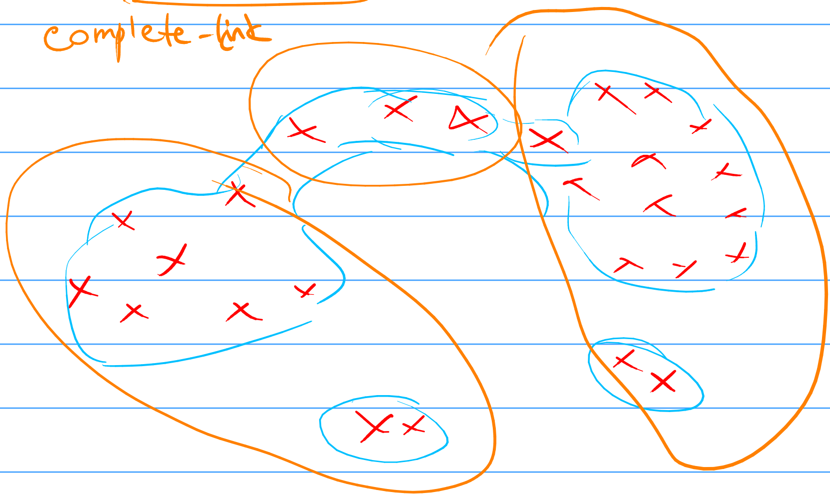
Hierachical clustering

single-link

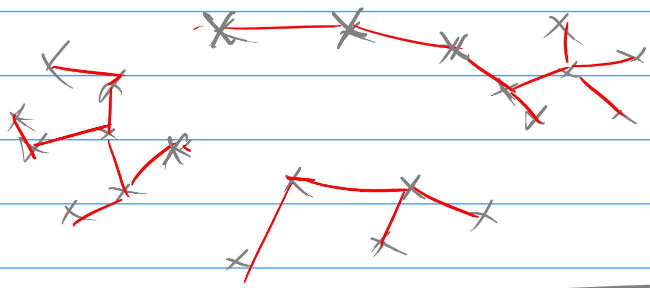
$$d(C_1, C_2) = \boxed{\min d(\bar{x}_1, \bar{x}_2)}$$
$$\text{avg } d(\_,\_)$$
$$\boxed{\max d(\_,\_)}$$

complete-link





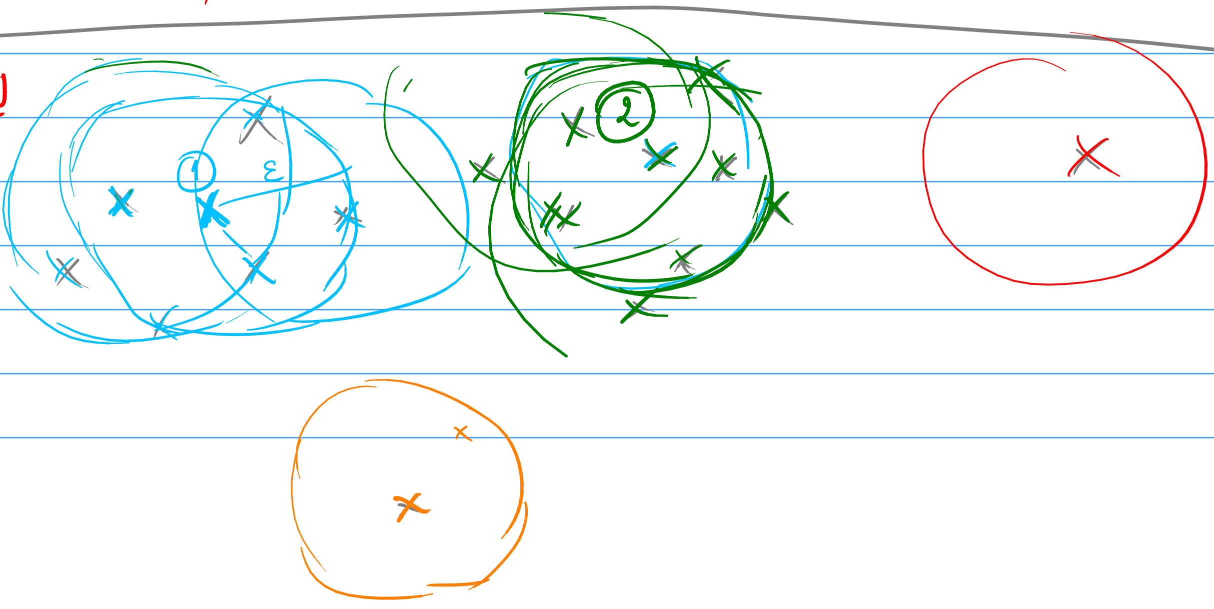
MST

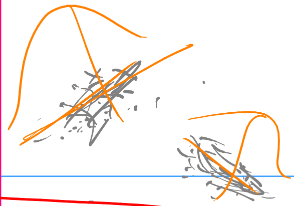


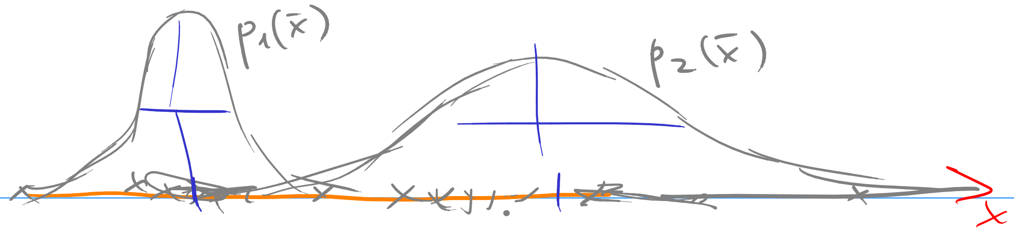
DBSCAN

ε - рад. окр

m - точек





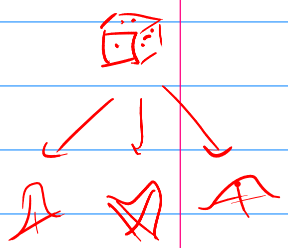


$$\boxed{p(\bar{x}) = \sum_{k=1}^{K} \alpha_k p_k(\bar{x}) \qquad \sum \alpha_k = 1}$$

$$p(\bar{x}) = \alpha\, p_1(\bar{x}) + (1-\alpha)\, p_2(\bar{x})$$

$$p(x \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \alpha) =$$

$$= \alpha \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2} + (1-\alpha)\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2}$$



$$p(t_n \mid \bar{x}_n, \bar{w}) = \sigma_n^{t_n}(1-\sigma_n)^{1-t_n} = t_n\sigma_n + (1-t_n)(1-\sigma_n)$$

$$p(X \mid \dots) = \prod_{n=1}^{N} p(x_n \mid -) =$$

$$= \prod_{n=1}^{N}\left(\frac{\alpha}{\sqrt{}} e^{\dots} + \frac{1-\alpha}{\sqrt{}} e^{-\dots}\right) \xrightarrow{\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha} \max$$

$$z_n = \begin{cases} 1, & x_n \in C_1 \\ 0, & x_n \in C_2 \end{cases} \qquad \alpha = p(z_n = 1)$$

$$p(X, Z \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \alpha) = \prod_{n=1}^{N}\left(\frac{\alpha}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}\right)^{z_n} \times$$

$$\times \left(\frac{1-\alpha}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2}\right)^{1-z_n}$$

$$\xrightarrow{\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha} \max$$

$$\ln p(X, Z \mid \dots) = \sum_n \Big[z_n\big(\ln\alpha + \overbrace{\ln p_1(x_n)}^{\mu_1, \sigma_1}\big) + (1-z_n)\big(\ln(1-\alpha) + \overbrace{\ln p_2(x_n)}^{\mu_2, \sigma_2}\big)\Big]$$

$$= \sum_n z_n \ln p_1(x_n) + \overbrace{\sum_n (1-z_n)\ln p_2(x_n)}^{n:\, z_n = 0} +$$

$$+ \sum_n \big[z_n \ln\alpha + (1-z_n)\ln(1-\alpha)\big]$$

$$\sum_{n:\, z_n = 1} \ln p_1(x_n)$$

$$\alpha_{ML} = \frac{\sum z_n}{N}$$

EM-algorithm:

$$p(X|\bar\theta) \to \max$$

$$\boxed{p(X,Z|\bar\theta)}$$

$z_n \in \{0,1\}$

$$p(X|\bar\theta) = \int p(X,Z|\bar\theta)\,dZ$$

expectation

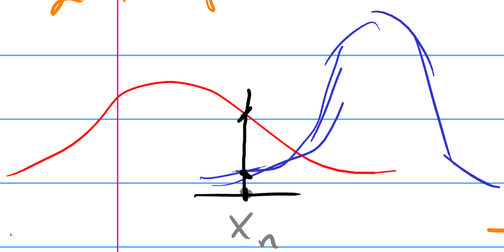
$$\text{E-step:}\quad \overbrace{\mathbb{E} Z_n^{(m)}}^{\in[0,1]} = \mathbb{E}_{\bar\theta^{(m)}}[z_n]$$

$$\text{M-step:}\quad \bar\theta^{(m+1)} = \arg\max_{\theta}\ p(X, Z^{(m)} | \bar\theta)$$

maximization

Clustering: 2 кластера

E-step:
$$\mathbb{E} z_n^{(m)} := p\left(C_1 | x_n, \mu_1^{(m)}, \mu_2^{(m)}, \sigma_1^{(m)}, \sigma_2^{(m)}, \alpha^{(m)}\right) = \frac{p(x_n|C_1,\theta)\,p(C_1|\theta)}{p(x_n|C_1,\theta)\,p(C_1|\theta) + \dots C_2 \dots} =$$

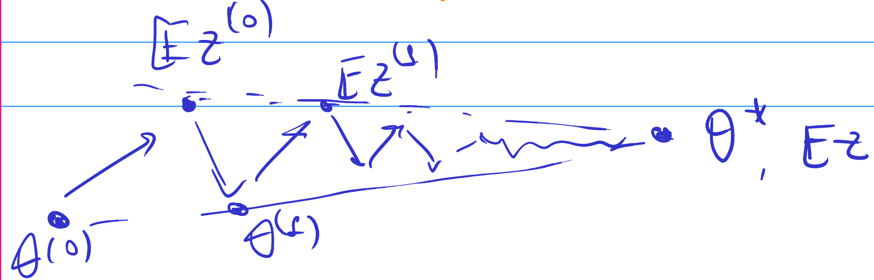


$$= \frac{\alpha \cdot \mathcal{N}(x_n|\mu_1,\sigma_1^2)}{\alpha \cdot \mathcal{N}(x_n|\mu_1,\sigma_1^2) + (1-\alpha)\,\mathcal{N}(x_n|\mu_2,\sigma_2^2)}$$

M-step:

$$\ln p(X, \mathbb{E}Z^{(m)}|\theta) = \sum_n \mathbb{E} z_n^{(m)} \ln \mathcal{N}(x_n|\mu_1,\sigma_1^2) +$$
$$+ \sum_n \left(1 - \mathbb{E} z_n^{(m)}\right) \ln \mathcal{N}(x_n|\mu_2,\sigma_2^2)$$
$$+ \sum_n \left[\mathbb{E} z_n^{(m)} \ln\alpha + \left(1-\mathbb{E} z_n^{(m)}\right)\ln(1-\alpha)\right] \xrightarrow[\theta]{} \max$$

$$\alpha^{(m+1)} = \frac{\sum \mathbb{E} z_n^{(m)}}{N}, \quad \mu_1^{(m+1)} = \frac{\sum_n \mathbb{E} z_n^{(m)} \cdot x_n}{\sum_n \mathbb{E} z_n^{(m)}}$$



k-means