FoDA : Clustering

L24 : Mixture of Gaussians

Clustering

Input: $X \subset \mathbb{R}^d$

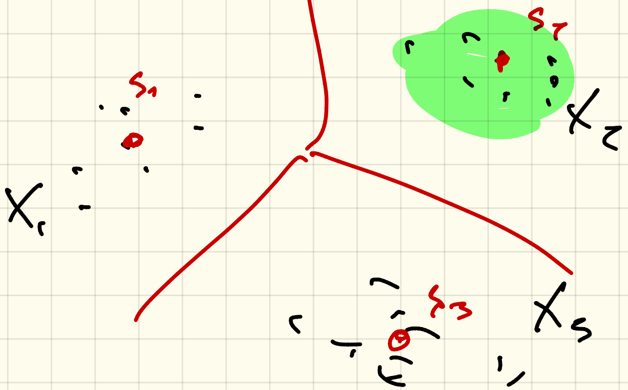
$k = \#$ clusters

$d : X \times X \to \mathbb{R}$

$d(x_1, x_2) = \|x_1 - x_2\|$

Output: set sites $S = \{s_1, s_2 \dots s_k\}$

set densities $f_1, f_2 \dots f_k$



Lloyd's Algo

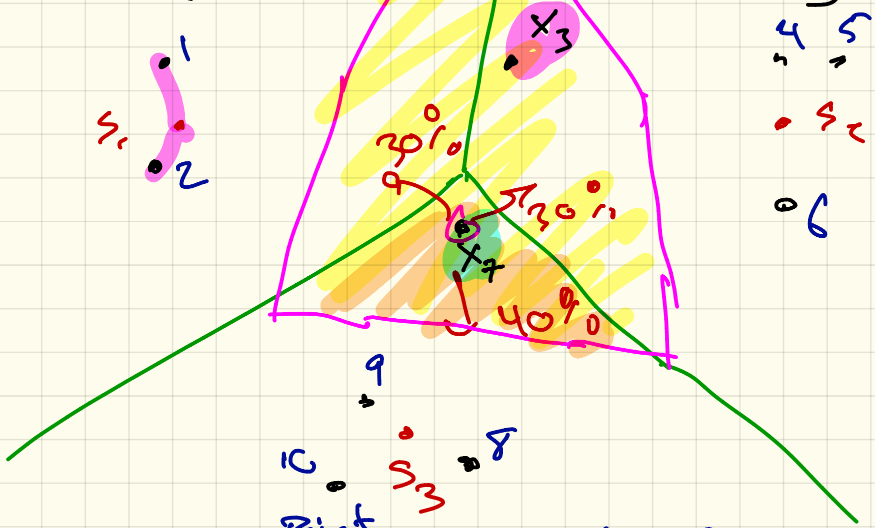
repeat

a. Assign each x_i to closest s_j

b. Center $s_j = \frac{1}{|X_j|} \sum_i x_i$

Soft Clustering

allows partial assignment of x_i onto sites s_j



Point

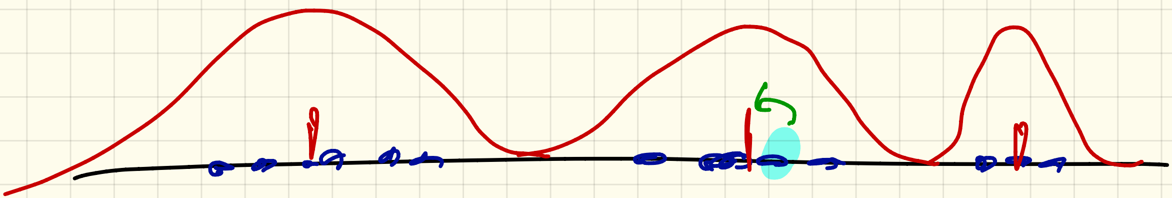
	1	2	3	4	5	6	7	8	9	10
cluster index	1	1	2	2	2	2	3	3	3	3

partial	1	2	3	4	5	6	7	8	9	10
prob 1	1	1	.4	0	0	0	.3	0	0	0
prob 2	0	0	.6	1	1	1	.3	0	0	0
prob 3	0	0	0	0	0	0	.4	1	1	1

Mixture of Gaussians

Input $X \subset \mathbb{R}^d$ $\quad k = \#$ Gaussians

Goal: Find the best distribution that X likely was drawn from, s.t. it is composed of k Gaussians



Gaussian Distribution → Σ Sigma

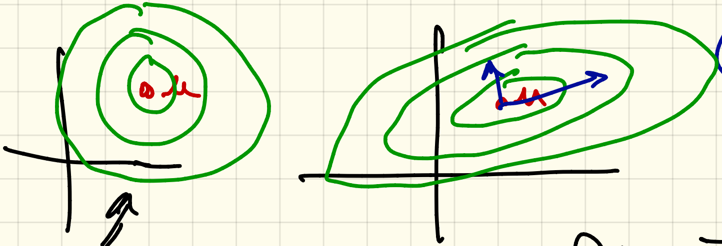
Σ sum

$$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$x \in \mathbb{R}^d$ $\quad$ $e^{\wedge()}$

mean $\mu \in \mathbb{R}^d$

$$\left(d_{\Sigma}(x,\mu)\right)^2$$



$$\Sigma = I \text{ (identity)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Sigma \in \mathbb{R}^{d\times d}$$

simple

$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \ \sigma_d \end{bmatrix}$$

$$\exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$$\exp\left(-\frac{1}{2}(x-\mu)^T (x-\mu)\right)$$

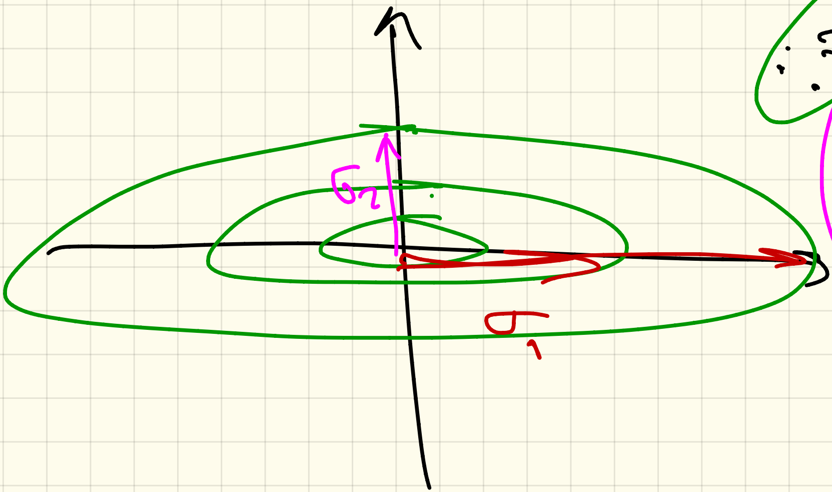
$$\exp\left(-\frac{1}{2}\langle x-\mu, x-\mu\rangle\right)$$

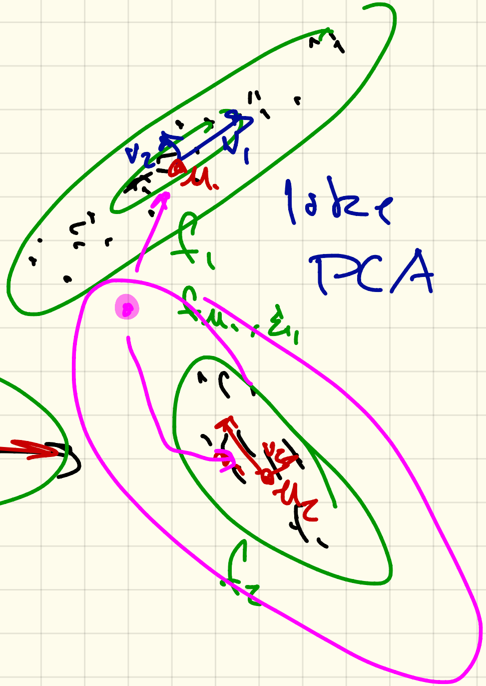
$$\exp\left(-\frac{1}{2}\|x-\mu\|^2\right) = \exp\left(-\frac{\|x-\mu\|^2}{2(\sigma=1)^2}\right)$$

Covariance

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & \boxed{0} \\ & \sigma_2 & & 0 \\ 0 & & \ddots & \sigma_d \end{bmatrix}$$

no off-diagonal covariance





like PCA

MLE (hard) Mix of Gaussians

Given X, k find $F = \{f_1, f_2 \dots f_k\} \subset$ Gaussians

$$\underset{F=\{f_1 \dots f_k\}}{\text{maximize}} \prod_{x \in X} \max_{f_i \in F} f_i(x)$$

MLE

if force $\Sigma_i = I$ $\quad f_i = f_{\mu_i, I}(x) = \exp\left(-\frac{\|x-\mu_i\|^2}{2}\right)$

$$\iff \underset{\mu_1 \dots \mu_k}{\text{minimize}} \sum_{x \in X} \min_{\mu_i \in F} -\ln(f_i(x))$$

$$= \underset{\mu_i}{\text{minimize}} \sum_{x \in X} \min_{\mu_i \in F} \frac{\|x-\mu_i\|^2}{2}$$

k-means objective

Sample Covariance for cluster $X_i \subset X$

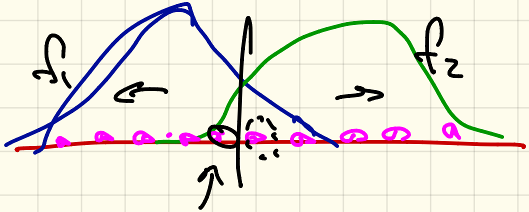
sample mean $\mu_i = \frac{1}{|X_i|} \sum_{x \in X_i} x$

sample covariance $\Sigma_i = \underset{x \in X_i}{\text{sum}} \underbrace{(x-\mu_i)(x-\mu_i)^T}_{\text{outer product}} \in \mathbb{R}^{d \times d}$

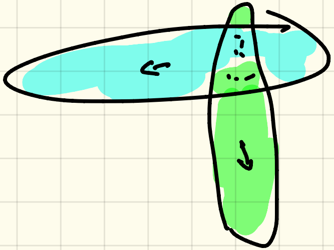
best estimate

$f_i = f_{\mu_i, \Sigma_i}$

Soft assignment



$f_1, \dots f_k \in F$



probability $x \sim f_i \equiv w_i(x)$

$$w_i(x) = \frac{f_i(x)}{\sum_{i=1}^{k} f_i(x)}$$

property $\forall x$ $\quad \sum_i w_i(x) = 1$

sample mean

weighted average

$$\mu_i = \frac{1}{W_i} \sum_{x \in X} w_i(x)\, x$$

$$W_i = \sum_{x \in X} w_i(x)$$

$$\Sigma_i = \frac{1}{W_i} \sum_{x \in X} w_i(x) (x-\mu_i)(x-\mu_i)^T$$

E-M Algorithms

0. Initialize $S = \{\mu_1, \dots, \mu_k\} \subset X \qquad \Sigma_i = I$

1. hard assign $x \in X \qquad w_i(x) = 1$ if $i = \phi_S(x)$

2. repeat

   b. for $i \in [1 \dots k]$ do

      - $W_i = \sum_{x \in X} w_i(x)$
      - $\mu_i = \frac{1}{W_i} \sum_{x \in X} w_i(x) \cdot x$
      - $\Sigma_i = \frac{1}{W_i} \sum_{x \in X} w_i(x) (x - \mu_i)(x - \mu_i)^T$

   model building

   centering

   expectation step

   partial assignment

   a. for $x \in X$

      for $i = 1 \dots k \qquad w_i(x) = \frac{f_i(x)}{\sum_i f_i(x)}$

   maximization step

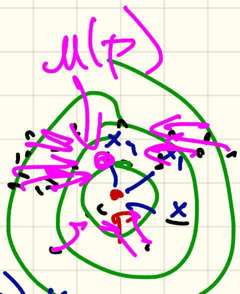
until ("converged")

Mean Shift Algo

Input $X \subset \mathbb{R}^d$, Kernel $K(x_1, x_2) = \exp\left(\frac{-\|x_1 - x_2\|^2}{2\sigma^2}\right)$

σ given

center of "mass" $\mu(p) = \dfrac{\sum_{x \in X}' K(x, p)\, x}{\sum_{x \in X}' K(x, p)}$



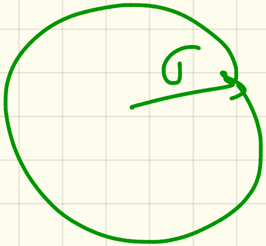
---

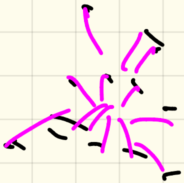
repeat

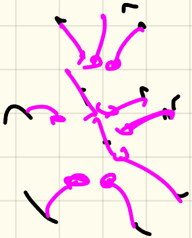
for all $p \in X$: calc $\mu(p) = \dfrac{\sum_x' K(x, p)\, x}{\sum_x' K(x, p)}$

for all $p \in X$: set $p \leftarrow \mu(p)$

until ("converged")







does not need R

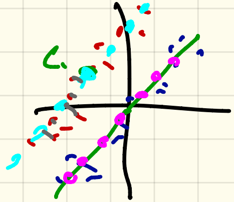
$$B = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$$

$$B v_3 = \left(\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T\right) v_3$$

$$= \sigma_1 u_1 v_1^T v_3 + \sigma_2 u_2 v_2^T v_3$$

$$\neq \sigma_1 v_1^T u_1 v_3 + \sigma_2 v_2^T u_2 v_3$$

$$\| \underset{n\times d}{A} - \underset{n\times d}{\pi_B(A)} \|^2$$



c 1. Find center $c \in \mathbb{R}^d$ of A.

2. center $A \to \tilde{A}$ $\quad [\tilde{A}]_{ij} = A_{ij} - c_j$

3. $\mathrm{svd}(\tilde{A}) = U S V^T$

4. best rank-k $\quad \tilde{A}_k \in \mathbb{R}^{n\times d}$

5. uncenter $\quad \pi_B(A) \quad \underset{\in \mathbb{R}^{n\times d}}{[\pi_B(A)]_{ij}} = \hat{A}_k + c_j$

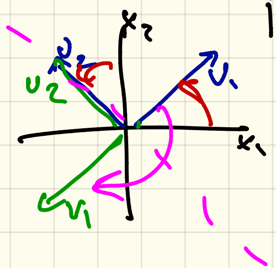
6. $A - \pi_B(A)$

$$\text{svd}(A) = U,\ S,\ V^T$$

$$\|Ax\| = 4 \qquad \underset{4\times 2}{\underline{n \times d}}$$

x unit vector $\qquad \|x\| = 1$

$\mathbb{R}^2$



$$\|V^T x\| = \|x\| = 1$$

$$S = \begin{bmatrix} 7 & 6 \\ 0 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\|S V^T x\| = \|U S V^T x\| = \|Ax\| = 4$$

$$\|\underbrace{U S V^T}\, x\| = \|Ax\| = 4$$

$$\alpha^* = (X^T X)^{-1} X^T y$$

optimal for SSE

$$\nabla f(\alpha^*) = 0$$