FoDA L24

Mixture of Gaussians

soft clustering, EM, mean-shift clustering

Nov 17, 2022

Soft Clustering

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

each $x_i \in X$ will have a partial assignment onto some clusters.

k-means

$\phi_S(x) = s_j$

$\hookrightarrow x \in X_j \subset X$

cluster

soft $w_i \in \Delta_k \subset \mathbb{R}^k$

$\sum_{j=1}^{k} w_{ij} = 1$

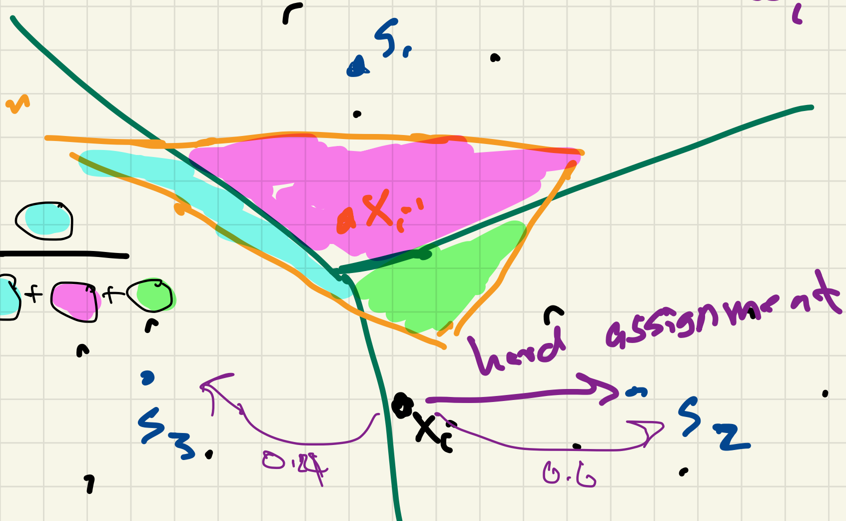
$w_{ij} \geq 0$

w_{i2}

$w_i = [0, .6, .4]$

Natural Neighbour Interpolation

$w_{i,3} =$



Soft Clustering

hard

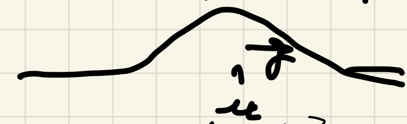
	1	2	3	4	5	6	7	
	1	1	1	0	0	1	0	1
	0	0	0	1	0	0	0	2
	0	0	0	0	1	0	1	3
	1	1	1	2	3	1	2	

Soft

1	0.9	1	0	0	0.4	0	1
0	0.1	0	1	0	0.3	0.5	2
0	0	0	0	1	0.3	0.5	3

Mixture of Gaussians

Normal



$e^{-\|x-\mu\|^2/\sigma^2}$

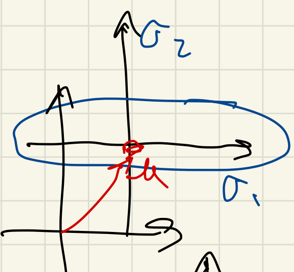
Input $X \subset \mathbb{R}^d$ # clusters $k \geq 1$

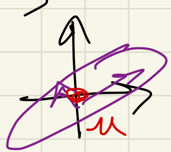
so integrates to 1

pdf Gaussian

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

$\mathbb{R}^d$ $\mathbb{R}^{d\times d}$ $-\frac{1}{2}\|x-\mu\|^2$



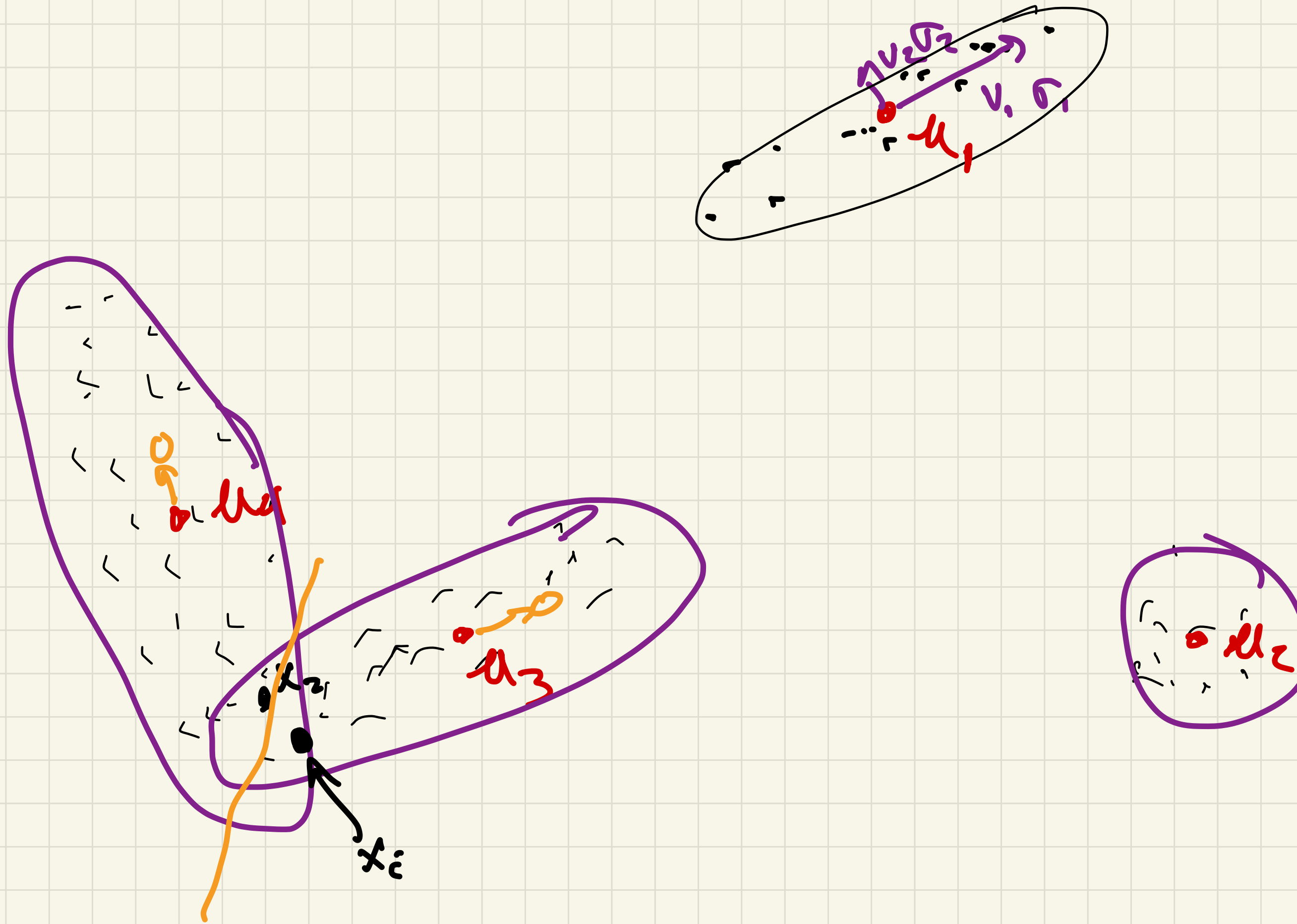


if $\Sigma_i = I_d$ ← identity

$$(x-\mu)^T\Sigma_i^{-1}(x-\mu) = (x-\mu)^T(x-\mu) = \langle x-\mu, x-\mu\rangle$$
$$= \|x-\mu\|^2$$

if $\Sigma_i = \sigma^2 \cdot I = \mathrm{diag}(\sigma^2, \sigma^2, \ldots, \sigma^2)$

$$(x-\mu)^T\Sigma_i^{-1}(x-\mu) = \frac{1}{\sigma^2}(x-\mu)^T(x-\mu) = \frac{\|x-\mu\|^2}{\sigma^2}$$



$\sqrt{v_2\sigma_2}$
$v_1\sigma_1$
μ_1
μ_2
μ_3
μ_4
x_2
x_i

Mixture of Gaussians

Input $x \in \mathbb{R}^d$ $\quad k \geq 1$

$\hookrightarrow$ fit k Gaussians

Output set (μ_j, Σ_j) $\quad j = 1 \ldots k$

assignment $w_i \in \Delta_k$ for all $x_i \in X$

define w_{ij} = prob. x_i from f_{μ_j, Σ_j}

$$= \frac{f_{\mu_j, \Sigma_j}(x_i)}{\sum_{j=1}^{k} f_{\mu_j, \Sigma_j}(x_i)}$$

$$\{\mu_j^*, \Sigma_j^*\} = \underset{\{\mu_j, \Sigma_j\}}{\arg\max} \prod_{x_i \in X} \sum_{j=1}^{k} w_{ij} f_{\mu_j, \Sigma_j}(x_i) = \underset{\{\mu_j, \Sigma_j\}}{\arg\min} -\sum_{x_i \in X} \sum_{j=1}^{k} w_{ij} \log(f_j(x_i))$$

$$-\ln\left(Z_{\Sigma_j}\exp\left(-(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)\right)\right)$$

$$= (x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j) + \ln(Z_{\Sigma_j})$$

$$\|x-\mu_j\|^2_{\Sigma_j}$$

if fix $\Sigma_j = I$

↳ k-means w/ soft assignment

set $S = \{1, 2, 5\}$

$f : S \to \mathbb{R}$

f $1 \to 10$
$2 \to 20$
$5 \to 5$

$s = \max_{s \in S} s$

$s = \arg\max_{s \in S} f(s) = 2$

EM Algo for MoG

0. Choose k pts $S \subset X$ $\quad s_j = \mu_j$

$\forall x_i \in X$ set $w_{ij} = 1$ if $\phi_S(x_i) = s_j$ o.w. $w_{ij} = 0$

otherwise

1. repeat

$f_j(x) = f_{\mu_j, \Sigma_j}(x)$

expectation update

for $j = 1 \ldots k$

Calculate $W_j = \sum_{i=1}^{n} w_{ij}$

Set $\mu_j = \frac{1}{W_j} \sum_{x_i \in X} w_{ij} x_i$

Set $\Sigma_j = \frac{1}{W_j} \sum_{x_i \in X} w_{ij} (x_i - \mu_j)(x_i - \mu_j)^T$

maximization assignment

for $x_i \in X$

for $j \in 1 \ldots k$ set $w_{ij} = \frac{f_j(x)}{\sum_{j=1}^{k} f_j(x)}$

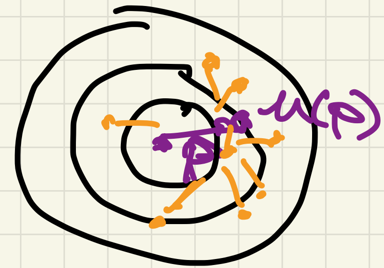
until (converged)

Mean Shift Clustering

input $X \subset \mathbb{R}^d$ kernel $K: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

$$K(p,q) = \exp\left(-\frac{\|p-q\|^2}{\sigma^2}\right)$$

$$\mu(p) = \frac{\sum_{x_i \in X} K(x_i, p)\, x_i}{\sum_{x_i \in X} K(x_i, p)}$$



repeat

$\forall p \in X$: calculate $\mu(p)$

$\forall p \in X$: set $p \to \mu(p)$

until (converged)