L23: Mixture of Gaussians

soft clustering, EM, mean-shift clustering

April 6, 2026
FoDA
Jeff M. Phillips

Soft Clustering

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

Goal : each $x_i \in X$ will have partial assignment onto some clusters

each x_i, cluster j, weight w_{ij}

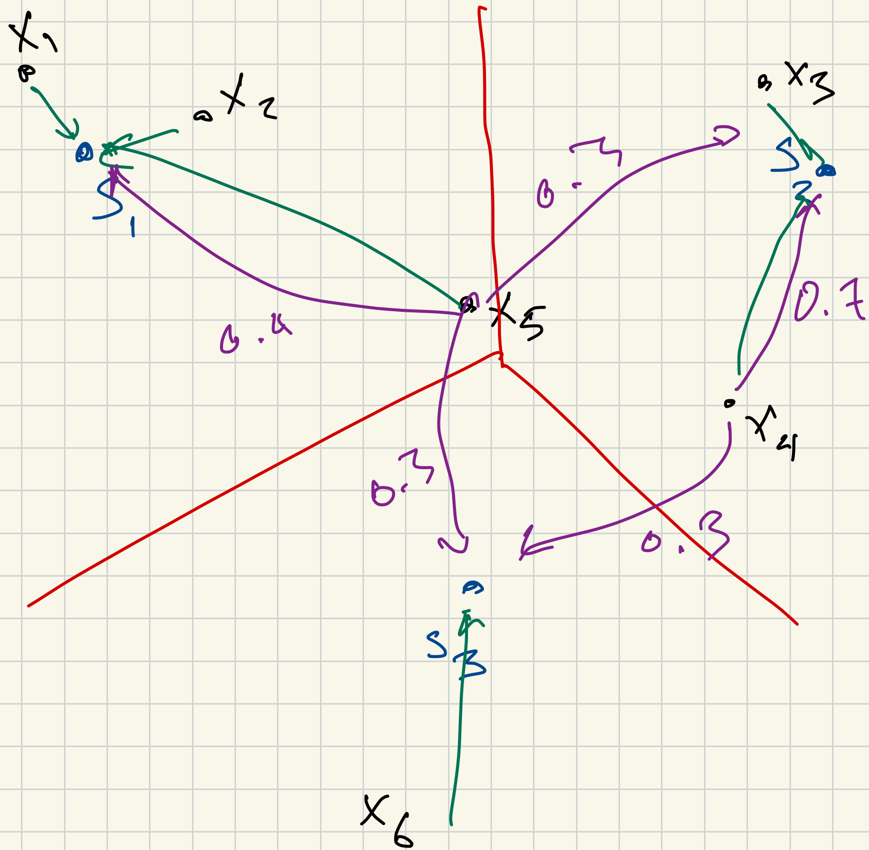
$$w_i = (w_{i1}, w_{i2}, \ldots w_{ik}) \subset \Delta_k \subset \mathbb{R}^k$$

$$w_{ij} \in [0,1]$$

treat w_{ij} as probability

$$\sum_j w_{ij} = 1$$

$$\Pr[x_i \in \text{cluster } j] = w_{ij}$$



hard clustering

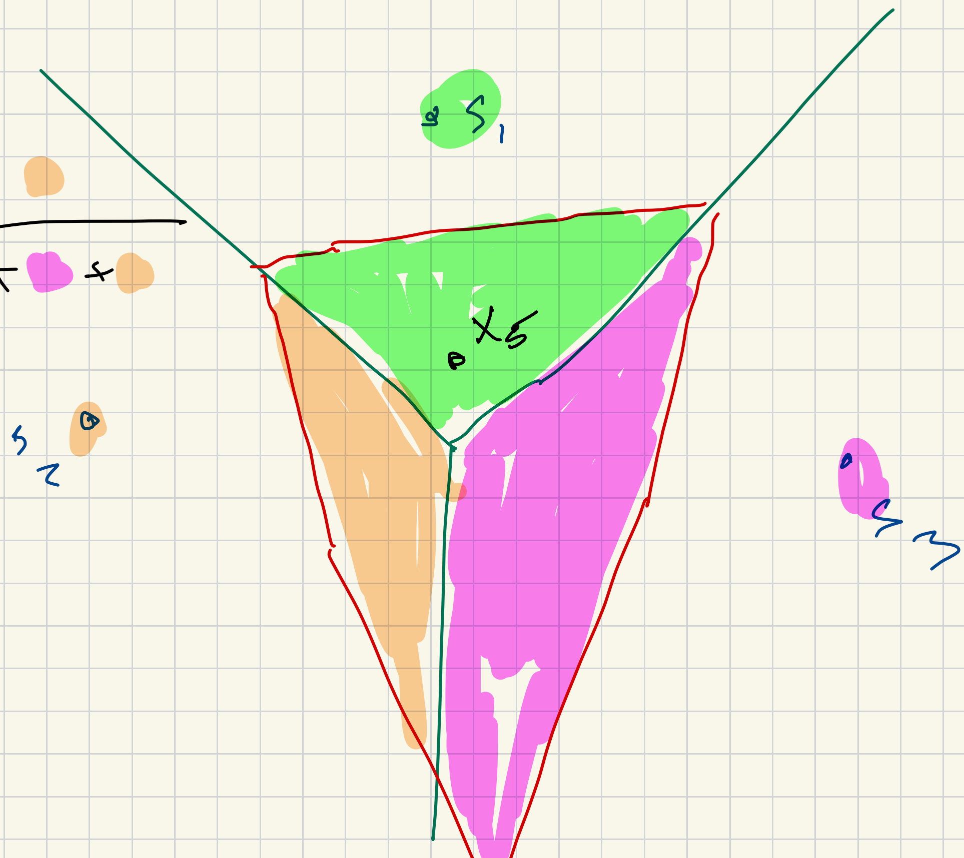
	x_1	x_2	x_3	x_4	x_5	x_6
S_1	1	1	0	0	1	0
S_2	0	0	1	1	0	0
S_3	0	0	0	0	0	1

soft clustering

	x_1	x_2	x_3	x_4	x_5	x_6
S_1	1	1	0	0	0.4	0
S_2	0	0	1	0.7	0.3	0
S_3	0	0	0	0.3	0.3	

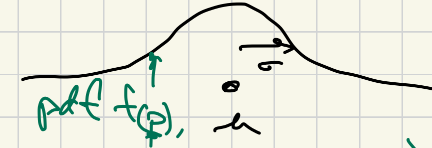
Natural Neighbor Interpolation

$$w_{5,2} = \frac{}{ + + }$$



Mixture of Gaussians

Normal



$$f(p) = Z \cdot \exp\left(-\frac{\|\mu - z\|}{2\sigma^2}\right)$$

pdf Gaussian

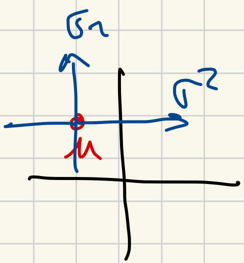
so integrate to 1

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

$\mathbb{R}^d$ $\quad$ $\mathbb{R}^{d\times d}$

if $\Sigma = I$ identity $= \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$

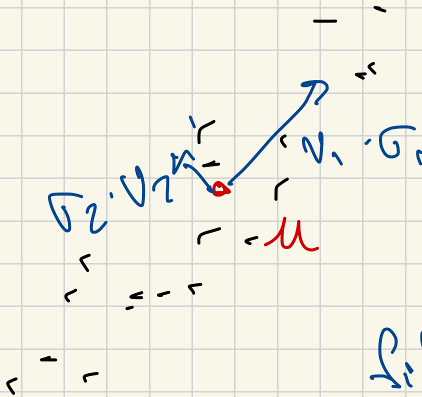
$$(x-\mu)^T \Sigma^{-1} (x-\mu) = (x-\mu)^T (x-\mu) = \langle x-\mu, x-\mu \rangle = \|x-\mu\|^2$$

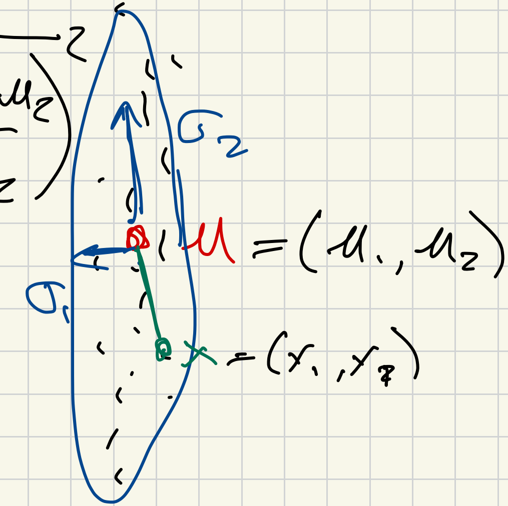


if $\Sigma = \sigma^2 I = \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & \cdots & \sigma^2 \end{bmatrix} = \mathrm{diag}(\sigma^2, \sigma^2, \dots, \sigma^2)$

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

$$d_{\Sigma_1}(x,\mu) = \|x-\mu\|_{\Sigma_1} = \sqrt{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}$$

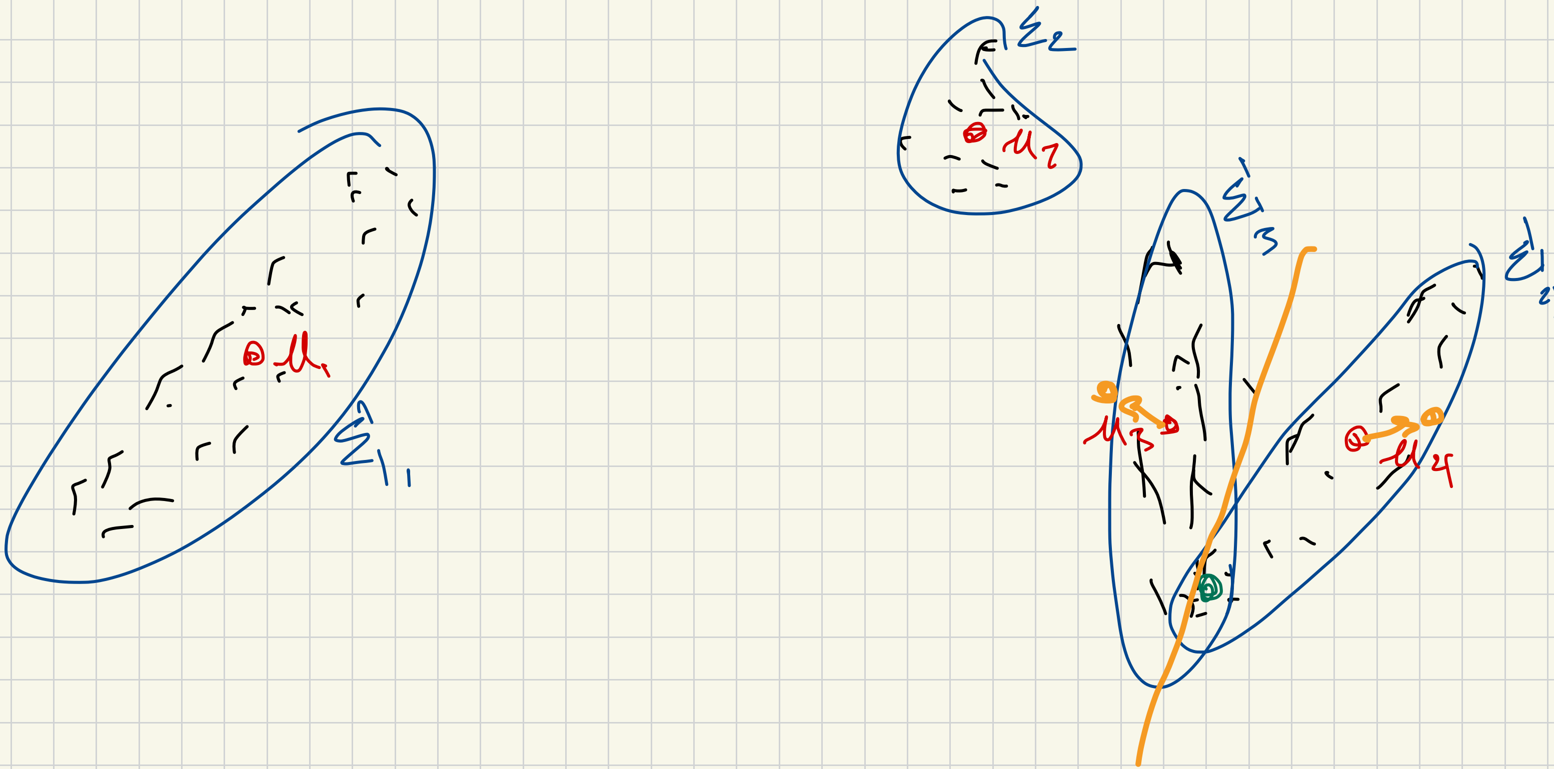




fit w/ $\Sigma_1 \in \mathbb{R}^2$

Mixture of Gaussians

Input $X \subset \mathbb{R}^d$ # Gaussians $= k = 4$



MoG Goal

Output

- set $\{(\mu_j, \Sigma_j)\}_j \quad j \in 1 \ldots k$

$$f_{\mu_j, \Sigma_j} = f_j$$

- assignment $\forall x_i \in X$

$$w_i \in \Delta_k \qquad w_{ij} = \Pr[x_i \in C_j]$$

$$w_{ij} = \frac{f_j(x_i)}{\sum_{j=1}^{k} f_j(x_i)} \qquad f_j(x) = f_{\mu_j, \Sigma_j}(x)$$

Goal

Maximum Likelihood Estimate

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

$$= \underset{\{\mu_j, \Sigma_j\}_j}{\arg\min} - \sum_{x_i \in X} \sum_{j=1}^{k} w_{ij} \ln\left(f_{\mu_j, \Sigma_j}(x_i)\right)$$

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

$$= (x-\mu_j)^T \Sigma_j^{-1} (x-\mu) - \cancel{\ln\left(Z_{\Sigma_j}\right)} \quad \text{if } \Sigma_j = I$$

$$= \|x-\mu_j\|^2 \implies \text{soft } k\text{-means.}$$

EM Algo for MoG

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

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

1. repeat

Expectation

(a) for $j = 1$ to k

- calculate $W_j = \sum_{i=1}^{n} w_{ij}$
- set $\mu_j = \frac{1}{W_j} \sum_{x_i \in X} w_{ij} \cdot x_i$ $\leftarrow$ weighted average
- set $\Sigma_j = \frac{1}{W_j} \sum_{x_i \in X} w_{ij} (x_i - \mu_j)(x_i - \mu_j)^T$

$f_j = f_{\mu_j, \Sigma_j}$

Maximization

(b) for $x_i \in X$

for $j = 1$ to k set $w_{ij} = \frac{f_j(x_i)}{\sum_{j'} f_{j'}(x_i)}$

until converge

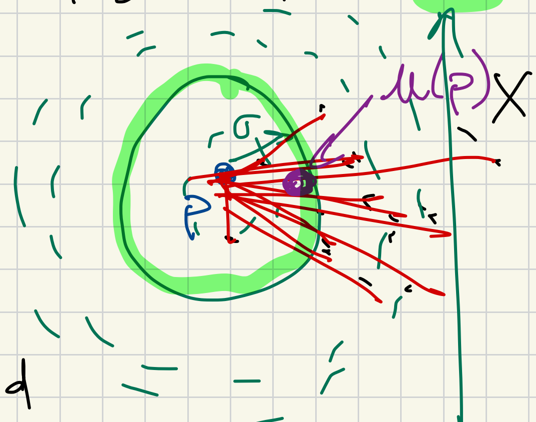
Mean Shift Clustering

<u>Input</u> $X \subset \mathbb{R}^d$

kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{>0}$

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

still a parameter



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

<u>repeat</u>

$\forall p_i \in X$: calculate $\mu(p_i) \in \mathbb{R}^d$

$\forall p_i \in X$: set $p_i \to \mu(p_i) = x_i$

<u>until converged</u>