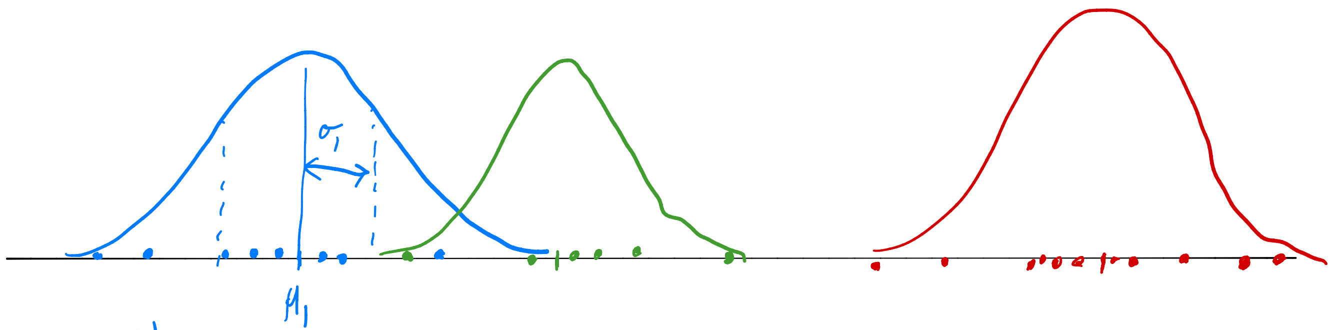
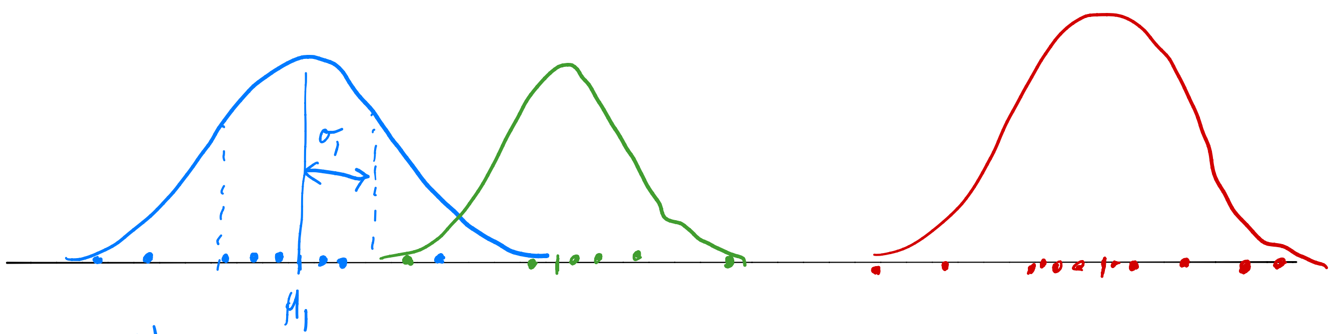
supervised setting (known labels)



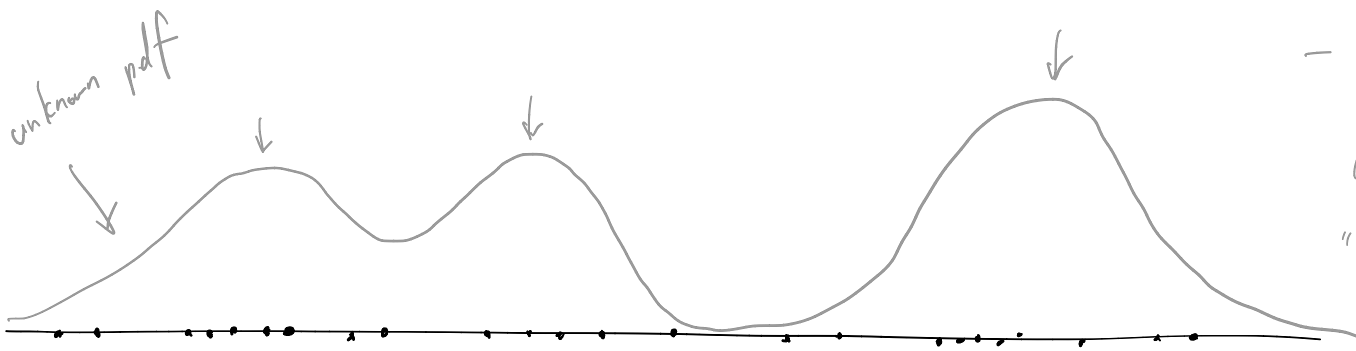
$x \sim N(\mu_1, \sigma_1^2)$

Supervised setting (known labels)



$x \sim \mathcal{N}(\mu_1, \sigma_1^2)$

unsupervised setting (we lost the labels)

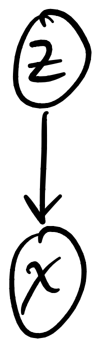


– mixture of Gaussians (MoG)

"Gaussian Mixture Model" (GMM)

GMM – 3 components

Graphical model / Generative model



Directed Acyclic Graph
(Graphical Model)

x : observable part (data) $\quad x \in \mathbb{R}^d$

z : latent variable $\quad z \in \{1, 2, \dots, K\}$

Generative Model

(1) Draw z from categorical distr. $\quad p(z=j) = \pi_j$

(2) Draw x from $p(x \mid z)$

Gaussian Mixture Model (GMM) - parameters

parameters $\theta = (\pi, \theta_1, \theta_2, \ldots, \theta_k)$

where $\pi \in \mathbb{R}^k$ with $\sum_{j=1}^{k} \pi_j = 1$ and $\pi_j \geq 0 \ \forall j$

and $\theta_j = (\mu_j, \Sigma_j)$ with $\mu_j \in \mathbb{R}^d$

and $\Sigma_j \succ 0$ $\Sigma_j \in \mathbb{R}^{d \times d}$

and symmetric

$(\Sigma_j = A_j A_j^T)$

Probability density function for GMM

$$p(x \mid \theta) = \sum_{j=1}^{K} p(x, z=j \mid \theta)$$

$$= \sum_{j=1}^{K} \underbrace{p(z=j \mid \theta)} \, p(x \mid z=j, \theta)$$

$$= \sum_{j=1}^{K} \pi_j \, p(x \mid z=j, \theta)$$

$$= \sum_{j=1}^{K} \pi_j \, p(x \mid \theta_j)$$

$$\frac{\exp\left(-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)\right)}{(2\pi)^{d/2} \, |\Sigma_j|^{1/2}}$$

irrational

Maximum Likelihood Estimation

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \underbrace{\log p(D \mid \theta)}$$

$$\log p(x_1, \ldots, x_n \mid \theta)$$

$$\overset{(iid)}{=} \sum_{i=1}^{n} \log p(x_i \mid \theta)$$

$$= \sum_{i=1}^{n} \log \left(\sum_{j=1}^{K} \pi_j \, p(x_i \mid \theta_j) \right)$$

Nothing wrong but annoying to optimize

Complete Data Log Likelihood (C.D.L.L.)

$$\log p(x_1, z_1, \ldots, x_n, z_n \mid \theta)$$

$$= \sum_{i=1}^{n} \log p(x_i, z_i \mid \theta)$$

$$= \sum_{i=1}^{n} \log\Big(p(z_i \mid \theta)\, p(x_i \mid z_i, \theta)\Big)$$

$$= \sum_{i=1}^{n} \log\left(\pi_{z_i} \cdot p(x_i \mid \theta_{z_i})\right)$$

$$= \sum_{i=1}^{n} \log \pi_{z_i} + \log p(x_i \mid \theta_{z_i})$$

Problematic to maximize b/c $z_1, \ldots, z_n$ are latent (hidden)

Expected C.D.L.L.

$$\mathbb{E}\left[\, C.D.L.L. \,\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} \log\left(\pi_{z_i}\, p(x \mid \theta_{z_i})\right) \;\middle|\; x_1, \ldots, x_n, \theta^{(t-1)} \right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} \log\left(\prod_{j=1}^{K} \left(\pi_j\, p(x \mid \theta_j)\right)^{\mathbb{1}[j = z_i]} \right) \;\middle|\; x_1 \ldots x_n, \theta^{(t-1)} \right]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{K} \mathbb{E}\left[\mathbb{1}[j = z_i] \log\left(\pi_j\, p(x_i \mid \theta_j)\right) \;\middle|\; x_1 \ldots x_n, \theta^{(t-1)} \right]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{K} \underbrace{p\left(z_i = j \mid x_i, \theta^{(t-1)}\right)}_{r_{ij} \text{ – responsibility of component } j \text{ for data point } x_i} \log\left(\pi_j\, p(x_i \mid \theta_j)\right)$$

How to update model parameters θ?

(1) Given $\theta^{(t-1)}$, update responsibilities r_{ij} for all i, j Expectation step (E-step)

(2) Given responsibilities, update $\theta^{(t-1)}$ to $\theta^{(t)}$ Maximization step (M-step)

How to update model parameters θ?

(1) Given $\theta^{(t-1)}$, update responsibilities r_{ij} for all i, j — Expectation step (E-step)

(2) Given responsibilities, update $\theta^{(t-1)}$ to $\theta^{(t)}$ — Maximization step (M-step)

E-M analogy with Lloyd's Algorithm

GMM (EM)	Clustering (Lloyd's Algorithm)
E-step: updating responsibilities (soft clustering)	assigning points to clustering (hard clustering)
M-step: Update $\theta = \left[\pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_K \end{pmatrix}, \theta_1, \ldots, \theta_K\right]$	Update $\theta = (\mu_1, \ldots, \mu_K)$

E-step

Compute responsibilities r_{ij} , given $\theta^{(t-1)}$

$$r_{ij} = p(z_i = j \mid x_i, \theta^{(t-1)})$$

$$= \frac{p(x_i, z_i = j \mid \theta^{(t-1)})}{p(x_i \mid \theta^{(t-1)})}$$

$$= \frac{p(z_i = j \mid \theta^{(t-1)})\, p(x_i \mid z_i = j, \theta^{(t-1)})}{\sum_{k=1}^{K} p(z_i = k \mid \theta^{(t-1)})\, p(x_i \mid z_i = k, \theta^{(t-1)})} = \frac{\pi_j^{(t-1)}\, p(x_i \mid \theta_j^{(t-1)})}{\sum_{k=1}^{K} \pi_k^{(t-1)}\, p(x_i \mid \theta_k^{(t-1)})}$$

M-step: update π and $\theta_1, \ldots, \theta_K$

$$\max_{\pi} \sum_{i=1}^{n} \sum_{j=1}^{k} r_{ij} \Big(\log \pi_j + \underbrace{\log p(x_i \mid \theta_j)}_{\text{ignore :)}} \Big)$$

$$\equiv \max_{\pi} \sum_{j=1}^{k} \underbrace{\left(\sum_{i=1}^{n} r_{ij} \right)}_{} \log \pi_j$$

r_j - general responsibility of comp. j for data

$$\equiv \max_{\pi} \exp\left(\sum_{j=1}^{k} \log \pi_j^{r_j} \right)$$

"counts"

$$\equiv \max_{\pi} \prod_{j=1}^{k} \pi_j^{r_j} \longrightarrow \pi_j = \frac{r_j}{\sum_{k=1}^{k} r_k} = \frac{r_j}{n}$$

M-step: update π and $\theta_1, \ldots, \theta_K$

$$\max_{\theta_j} \sum_{i=1}^{n} \sum_{j'=1}^{K} r_{ij'} \left(\underbrace{\log \pi_{j'}}_{\text{ignore :)}} + \log p(x_i \mid \theta_{j'}) \right)$$

$$\equiv \max_{\theta_j} \sum_{i=1}^{n} r_{ij} \log p(x_i \mid \theta_j)$$

$$\mu_j = \frac{\sum_{i=1}^{n} r_{ij} \cdot x_i}{\sum_{i=1}^{n} r_{ij}} \qquad \Sigma_j = \frac{\sum_{i=1}^{n} r_{ij} (x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{n} r_{ij}}$$