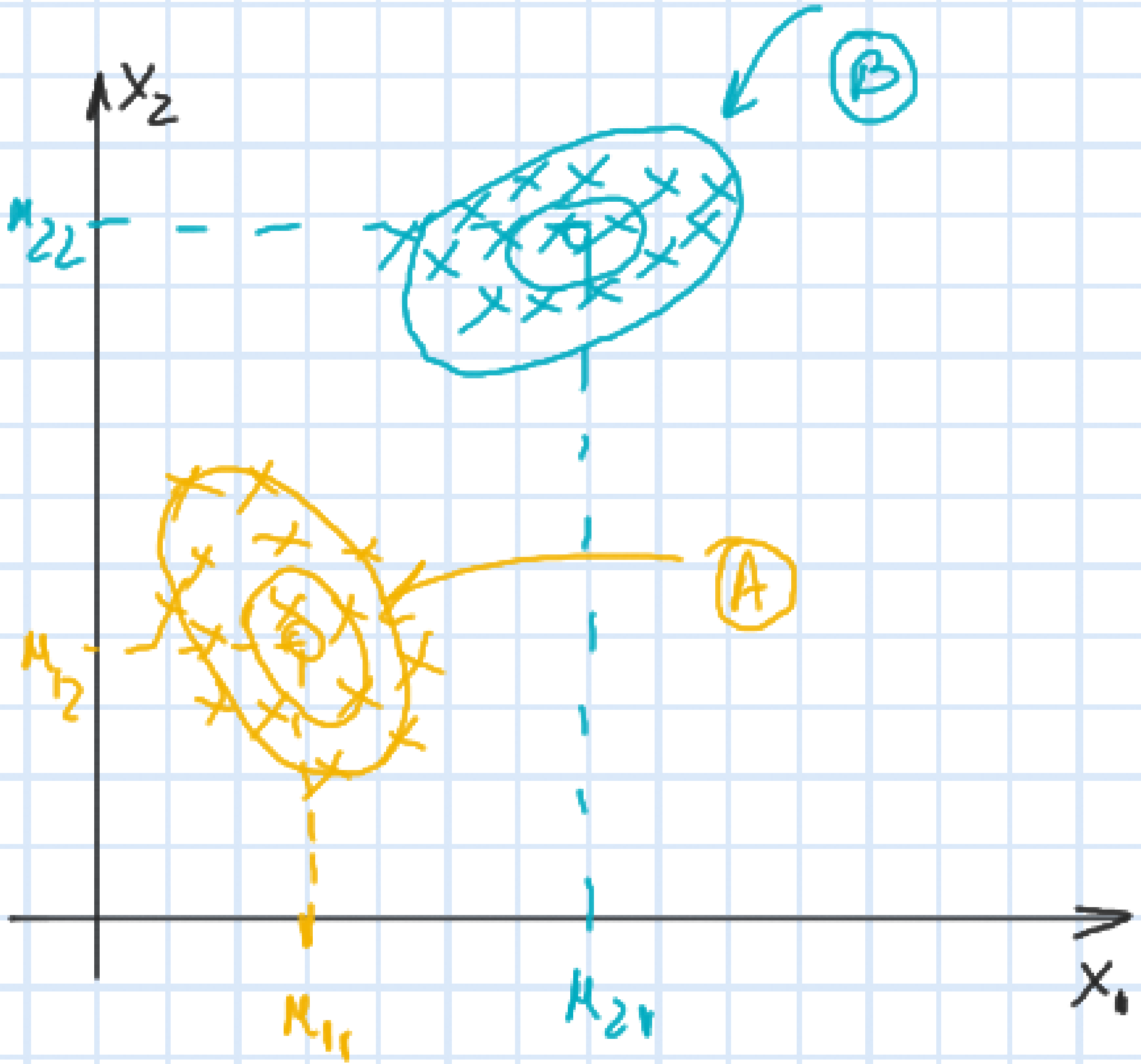


Gaussian Mixture Models

$$\mathcal{N}(\vec{x} \mid \vec{\mu}_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^K |\Sigma_i|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)\right)$$

K-Means Revisión

$$\{x^{(1)}, x^{(2)}, \dots, x^{(n)}\} \longrightarrow k \text{ clusters}$$

$$x^{(i)} \in \mathbb{R}^m$$

1. Inicializar centroides $\mu_1, \mu_2, \dots \mu_k \in \mathbb{R}^m$ aleatoriamente
2. a. Para cada $i \longrightarrow c^{(i)} := \arg\min_j \|x^{(i)} - \mu_j\|^2$ } E

   b. Para cada $j \longrightarrow \mu_j := \dfrac{\sum_{i=1}^{N} 1\{c^{(i)}=j\}\, x^{(i)}}{\sum_{i=1}^{N} 1\{c^{(i)}=j\}}$ } M

Distortion Function

$$J(c,\mu) = \underbrace{\sum_{i=1}^{N} \|x^{(i)} - \mu_{c^{(i)}}\|^2}_{\text{Coordinate Descent}}$$

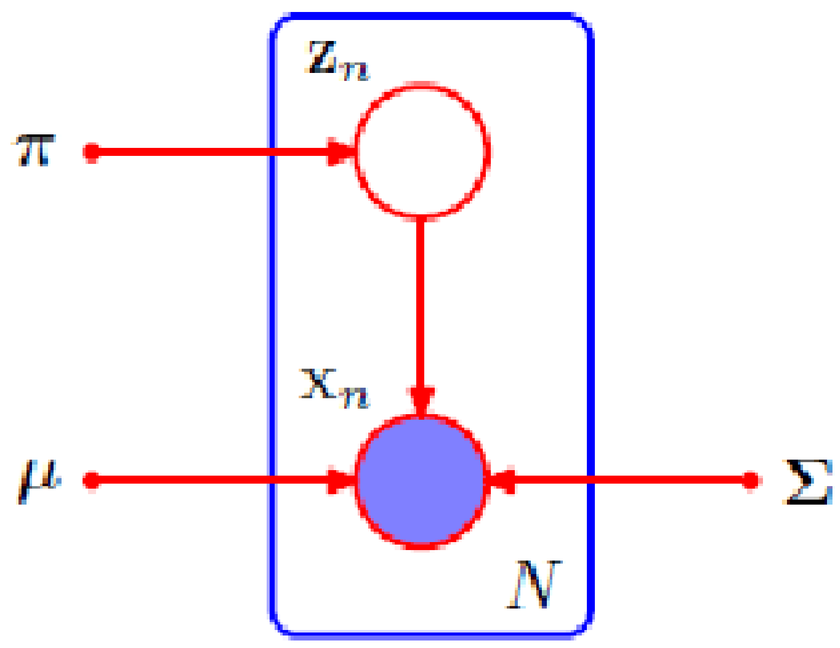
Gaussian Mixture Model - EM

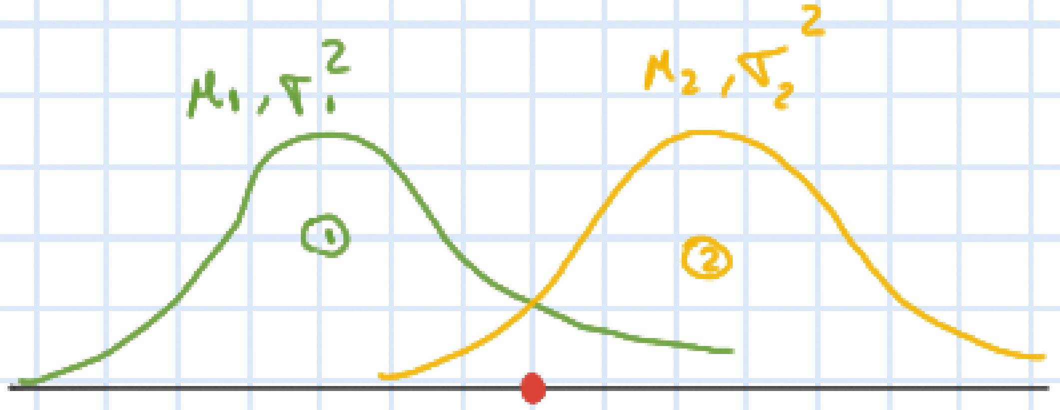
$$p(x^{(i)}, z^{(i)}) = p(x^{(i)} | z^{(i)})\, p(z^{(i)})$$

$$\text{prior} \begin{cases} z^{(i)} \sim \text{Multinomial}(\pi) \longrightarrow \\ \pi_j \longrightarrow p(z^{(i)} = j) \end{cases}$$

- $\pi_j \geq 0$
- $\sum_{j=1}^{k} \pi_j = 1$

$$x^{(i)} | z^{(i)} = j \sim \mathcal{N}(\mu_j, \Sigma_j)$$





$\longrightarrow$ ¿Viene de 1 o de 2?

Dataset con Personas

Como siempre, planteamos M.V. para conocer θ

$$\ell(\pi, \mu, \Sigma) = \sum_{i=1}^{n} \log p(x^{(i)}; \pi, \mu, \Sigma)$$

$$= \sum_{i=1}^{n} \log \sum_{z^{(i)}=1}^{k} p(x^{(i)} | z^{(i)}; \mu, \Sigma)\, p(z^{(i)}; \pi)$$

$$\Sigma_j = \frac{\sum_{i=1}^{N} 1\{z^{(i)} = j\}(x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{n} 1 \cdot \{z^{(i)} = j\}}$$

Asumamos que conocemos z

$$\ell(\pi, \mu, \Sigma) = \sum_{i=1}^{N} \log p(x^{(i)} | z^{(i)}; \mu, \Sigma) + \log p(z^{(i)}; \pi)$$

Maximizando respecto de π_j, μ_j, Σ_j

$$\mu_j = \frac{\sum_{i=1}^{N} 1\{z^{(i)} = j\}\, x^{(i)}}{\sum_{i=1}^{n} 1 \cdot \{z^{(i)} = j\}}$$

$$\pi_j = \frac{1}{N} \sum_{i=1}^{N} 1\{z^{(i)} = j\}$$

¡ NO CONOCEMOS $z^{(i)}$!

$\downarrow$

Expectation Maximization $\longrightarrow$ K - Means

Vamos a estimar los $z^{(i)}$

(1) $P(z_i = j) = \frac{1}{k}$

(2) $\Sigma_j = \sigma^2 I$

(3) $\lim_{\sigma^2 \to 0}$

E - step

- Para cada i, j: $w_j^{(i)} = p(z^{(i)} = j \mid x^{(i)}; \pi, \mu, \Sigma)$

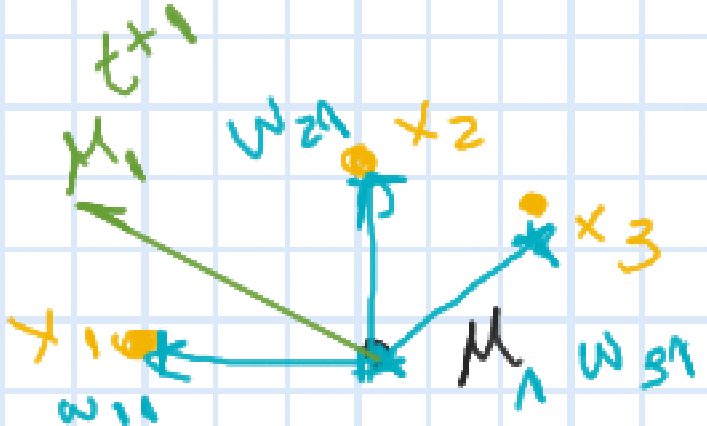
Usando Bayes tenemos:

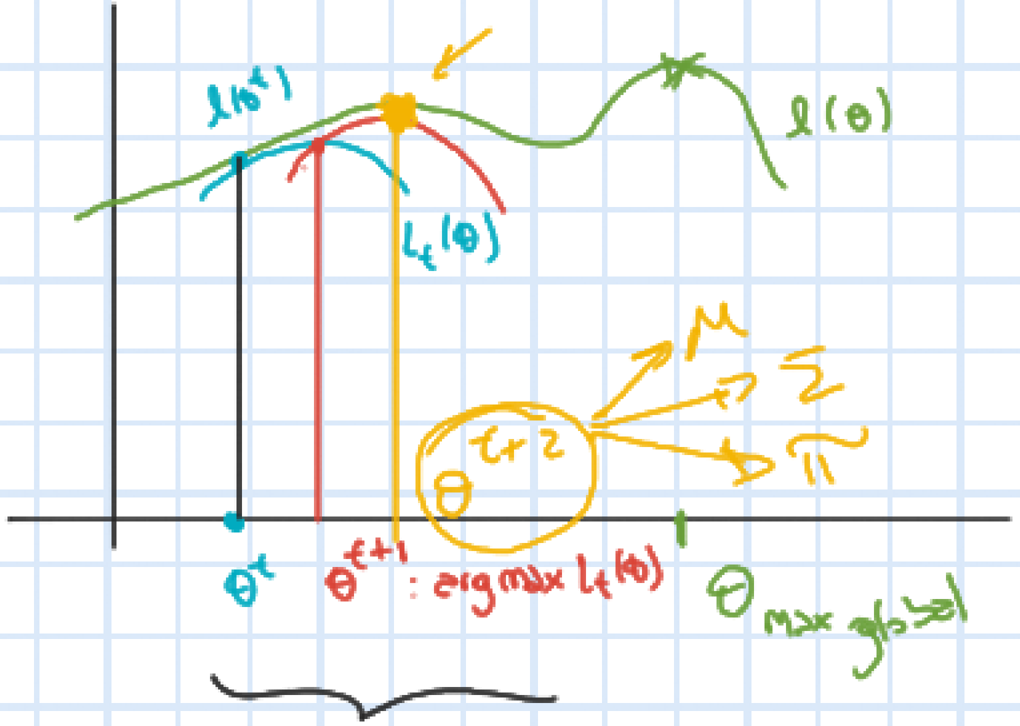
$$\underbrace{p(z^{(i)} = j \mid x^{(i)}; \pi, \mu, \Sigma)}_{w_j} = \frac{p(x^{(i)} \mid z^{(i)} = j; \mu, \Sigma)\, p(z^{(i)} = j; \pi)}{\sum_{l=1}^{k} p(x^{(i)} \mid z^{(i)} = l; \mu, \Sigma)\, p(z^{(i)} = l; \pi)}$$

- M - step

$$\pi_j = \frac{1}{N} \sum_{i=1}^{N} w_j^{(i)} \qquad \Sigma_j := \frac{\sum_{i}^{N} w_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{N} w_j^{(i)}}$$

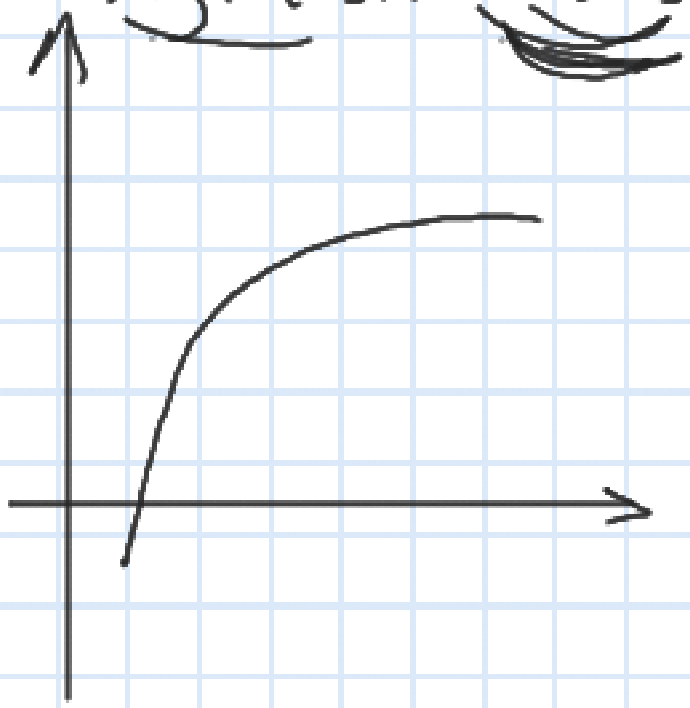
$$\mu_j = \frac{\sum_{i}^{N} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{N} w_j^{(i)}}$$





Expectation Maximization

$\log(\mathbb{E}(x)) \geq \mathbb{E}(\log x)$



1. $L_t(\theta) \leq \ell(\theta)$ lowerbound
2. $L_t(\theta^t) = \ell(\theta^t)$ Tight

E-step: Encontrar $L_t(\theta)$ dado $\theta^{(t)}$

M-step: $\theta^{t+1} = \underset{\theta}{\arg\max}\, L_t(\theta)$