GMMs :

↳ Estimate data likelihood

↳ Perform clustering

Assuming we have samples $\{x_i\}_{i=1}^{N}$.

$$\mathcal{L} = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$

↳ We need to find parameters

$$\theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$$

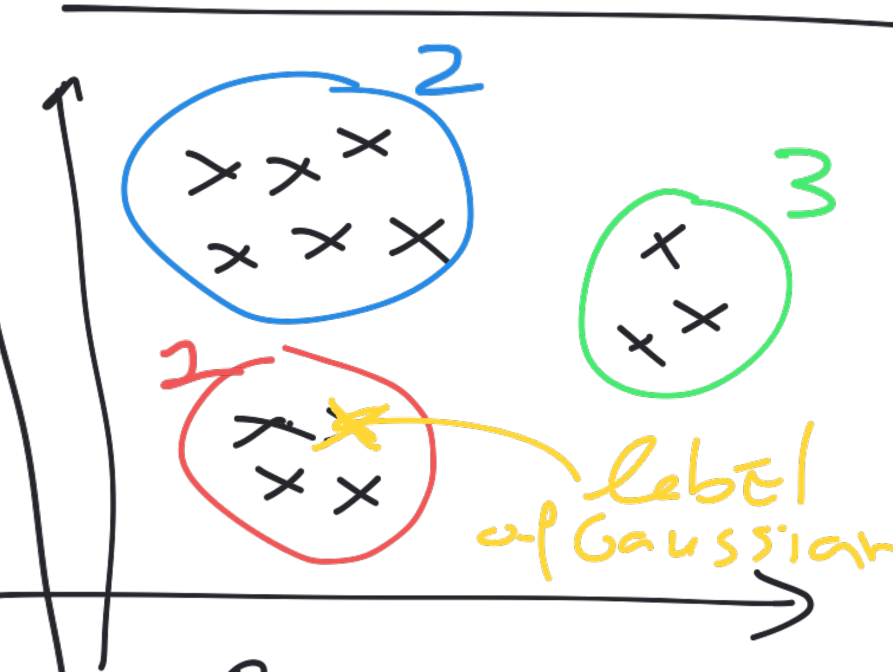
↳ MLE approach is "difficult" to solve for the parameters.

↳ Instead we use EM algorithm

EM Alg.

(1) Hidden variables

$z_i \equiv$ label of the Gaussian from which x_i was drawn from.



$$\mathcal{L}^o = \pi_1 \cdot N + \pi_2 \cdot N + \pi_3 \cdot N$$

$$z_i \in \{1, 2, 3\}$$

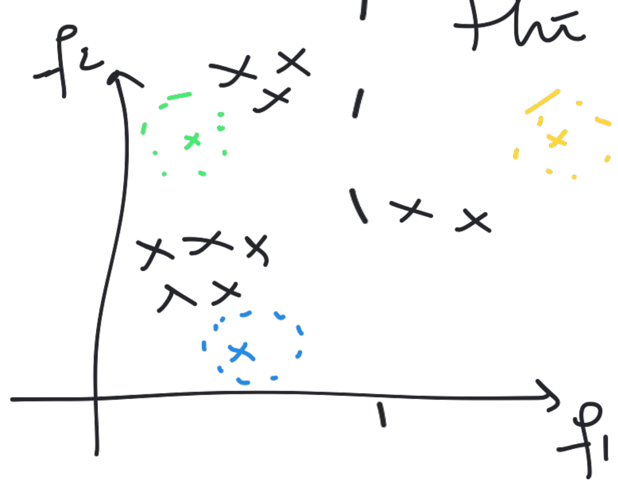
(2) Complete Data Likelihood

$$\mathcal{L}^c = \prod_{i=1}^{N} \pi_{z_i} \cdot \mathcal{N}(x_i \mid \mu_{z_i}, \Sigma_{z_i})$$

(3) Do not know z's, so EM will take the expected value of z's for the log-complete-likelihood:

$t=0$

(3.1) First randomly pick values for the parameters $\theta^t = \{\pi_k, \mu_k, \Sigma_k\}_k$



(3.1.1) Find exp. value for z's, while holding θ^t fixed. E-STEP

(3.1.2) For the z's found, update the parameters θ. M-STEP

optimizing the Q function

Repeat until convergence
- # iterations
- ERROR MEASURE

EM algorithm is called an Alternating Optimization algorithm

$$Q(\theta,\theta^t) = E[\ln \mathcal{L}^c \mid X, \theta^t]$$

$$= \sum_{z_i=1}^{K} \ln \mathcal{L}^c \cdot P(z_i \mid x_i, \theta^t)$$

$$= \sum_{k=1}^{K} \underbrace{\ln\left(\prod_{i=1}^{N} \pi_k^t \cdot \mathcal{N}(x_i \mid \mu_k^t, \Sigma_k^t)\right)}_{\text{E-STEP this term is constant}} \cdot \overbrace{\underbrace{P(z_i=k \mid x_i, \theta^t)}_{\text{M-STEP this term is constant}}}^{\text{"} C_{ik}}$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{N} \left(\ln(\pi_k) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_k^2) - \frac{1}{2\sigma_k^2}\|x_i - \mu_k\|_2^2 \right) \cdot C_{ik}$$

$p(z_1 \mid x_1, \theta^t)$

C_{ik} is a matrix $N \times K$

	1	2	...	K
x_1	.8	0.2	...	0
x_2				
x_3				
$\vdots$				
x_N				

E-STEP: Comput C_{ik}, membership OR Responsabilities of point x_i in Gaussian k.

$$\boxed{\sum_k C_{ik} = 1}$$

M - STEP:

① $$\frac{\partial Q}{\partial \mu_k} = 0$$

$$\Longleftrightarrow \mu_k^{t+1} = \frac{\sum_{i=1}^{N} x_i \cdot c_{ik}}{\sum_{i=1}^{N} c_{ik}}$$

, weighted mean for each point belonging component k

② $$\underline{\underline{\Sigma_k}} = \underline{\underline{\sigma_k^2}} \cdot I$$

$$\frac{\partial Q}{\partial \sigma_k^2} = 0 \Longleftrightarrow \sigma_k^2 = \frac{\sum_{i=1}^{N} \|x_i - \mu_k\|_2^2 \cdot c_{ik}}{\sum_{i=1}^{N} c_{ik}}$$

(3) $$\frac{\partial Q}{\partial \pi_k} = 0$$

$$\boxed{\sum_{k=1}^{K} \pi_k = 1}, \quad 0 \leq \pi_k \leq 1 \implies \boxed{1 - \sum_{k=1}^{K} \pi_k = 0}$$

Equality Constraint

$$Q_\pi(\theta, \theta^t) = Q(\theta, \theta^t) + \lambda \left(1 - \sum_{k=1}^{K} \pi_k\right)$$

Lagrange Multipliers

Lagrange multiplier

$$\frac{\partial Q_\pi}{\partial \pi_k} = 0 \iff \boxed{\sum_{i=1}^{N} \frac{1}{\pi_k} c_{ik} - \lambda = 0} \iff \pi_k = \frac{\sum_{i=1}^{N} c_{ik}}{\lambda}$$

(*)

But $\sum_{k=1}^{K} \pi_k = 1$

$$\Longleftrightarrow \sum_{k=1}^{K} \frac{\sum_{i=1}^{N} c_{ik}}{\lambda} = 1$$

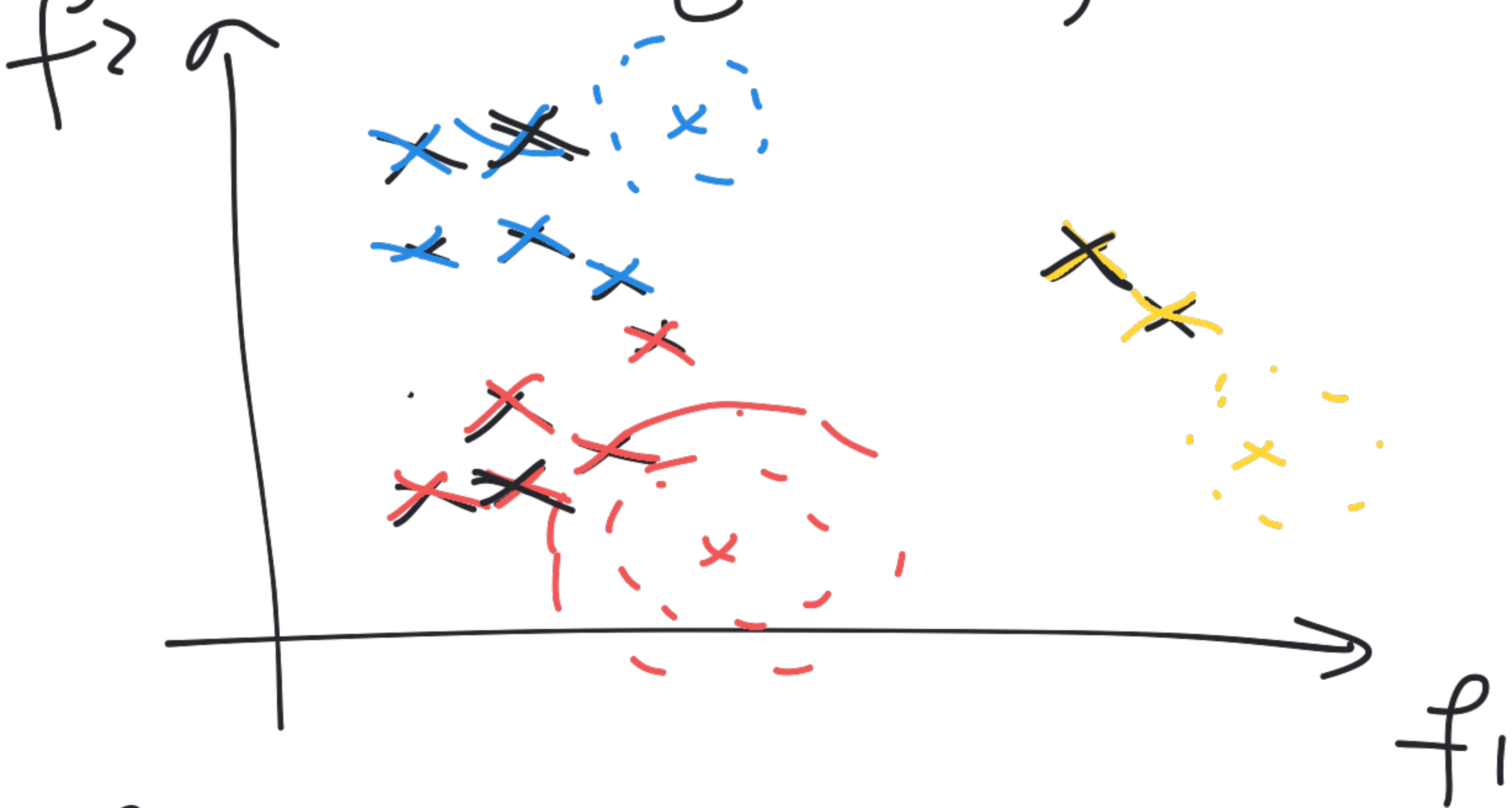
$$\Longleftrightarrow \lambda = \sum_{i=1}^{N} \sum_{k=1}^{K} c_{ik}$$

Plugging λ in $*$: $\pi_k = \dfrac{\sum_{i=1}^{N} c_{ik}}{\sum_{i=1}^{N} \underbrace{\sum_{k=1}^{K} c_{ik}}_{=1}}$

$$\Longleftrightarrow \boxed{\pi_k = \frac{\sum_{i=1}^{N} c_{ik}}{N}}$$

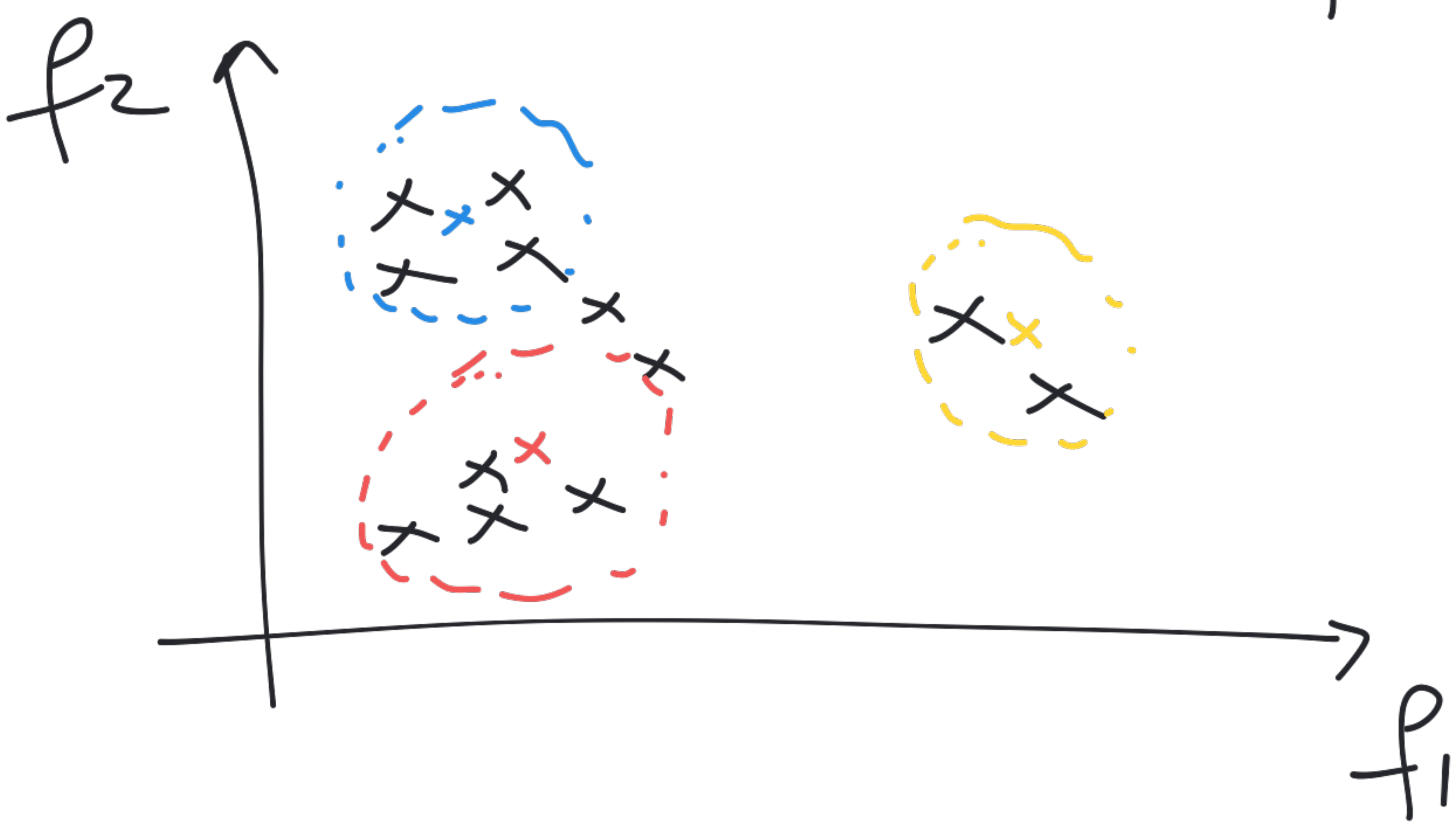
Sums up responsibilities of all points in component k.

$t = 0, \theta^t$



E-STEP:

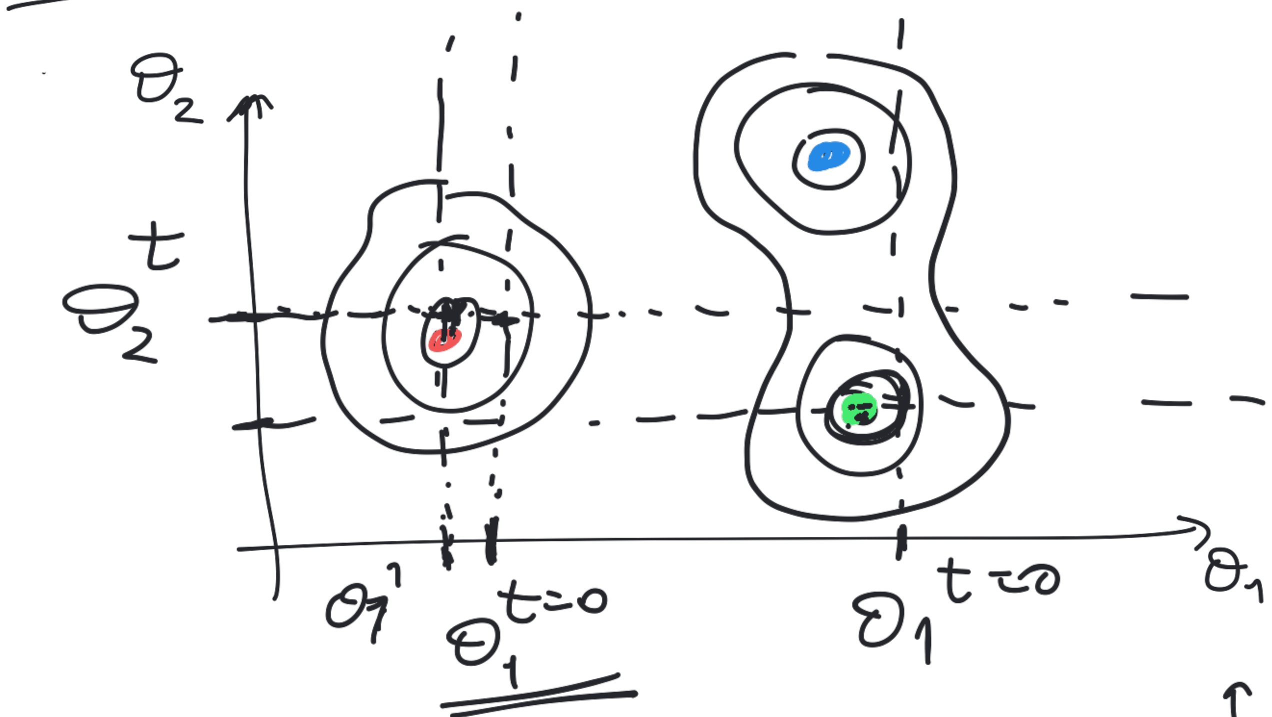
Assign prob./ responsabilities of each sample in component k



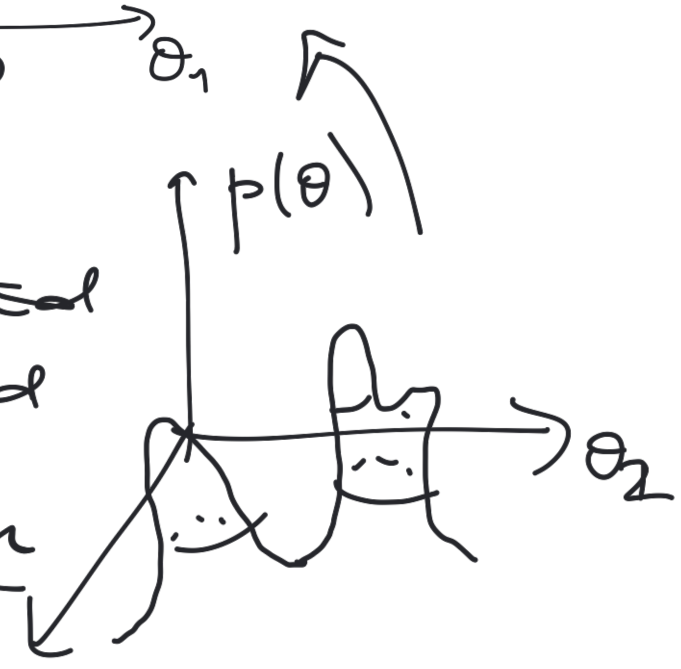
M-STEP:

based on these responsabilities, update parameters

EM uses alternating opt. approach:



→ The final converged solution will depend on the initialization of the parameters.



STRATEGY to initialize the parameters:

For the means, consider one of the samples.