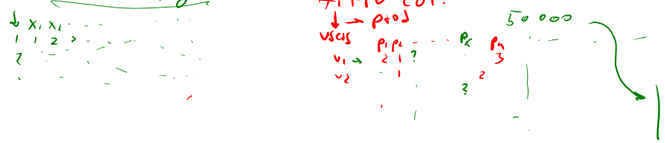
clustering ⟷ filtro col.



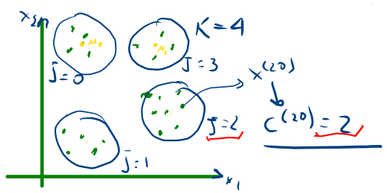
K means

$X_{(n,d)}$

Datos: $\{x^{(1)}, \dots, x^{(n)}\} \in \mathbb{R}^d$

$K = \#$ clusters $\qquad j = 1, \dots, K \;\; [0, K-1]$

encontrar asignación $c^{(i)} = j$ para cada $x^{(i)}$ $\mapsto$ punto $x^{(i)}$ esta en el cluster j

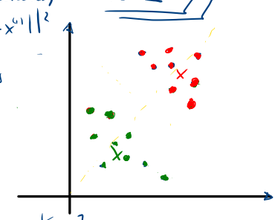


para cada cluster: μ_j

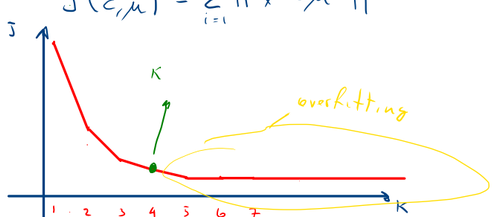
Algoritmo:

1. iniciar random (μ_j)
2. asignar $x^{(i)}$ a un cluster (más cercano)
   $$c^{(i)} = \underset{j=1 \dots K}{\arg\min} \, \|\mu^{(j)} - x^{(i)}\|^2$$
3. calcular **NUEVOS** centroides en base a los clusters
4. repetir 2.

$K=2$



K: Repetir Kmeans $K[K_0 \dots K_m]$

función de costo:

$$J(c, \mu) = \sum_{i=1}^{N} \| x^{(i)} - \mu^{(i)} \|^2$$