

# ④ K-Means - Understanding Loss, Coordinate Ascent

## K Means :

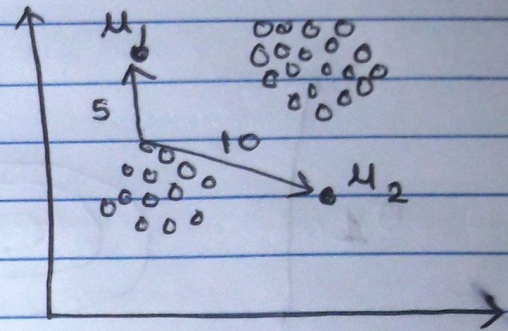
The maths behind

(Data points)

$$X = \{x_1, x_2, \dots, x_N\} \quad \text{Total } N$$

cluster centre

$$\mu = \{\mu_1, \mu_2, \dots, \mu_K\} \quad K$$



Loss / Inertia function = J

$$J = \sum_{n=1}^N \sum_{k=1}^K x_{nk} (x_n - \mu_k)^2$$

$\uparrow$        $\uparrow$        $\uparrow$        $\uparrow$   
 all      all      for      squared  
 points      clusters      taking      distance  
                                  min  
                                  square  
                                  distance  
                                  (5 Here)  
                                  = square  
                                  distance

$x_{nk}$  = boolean indicator

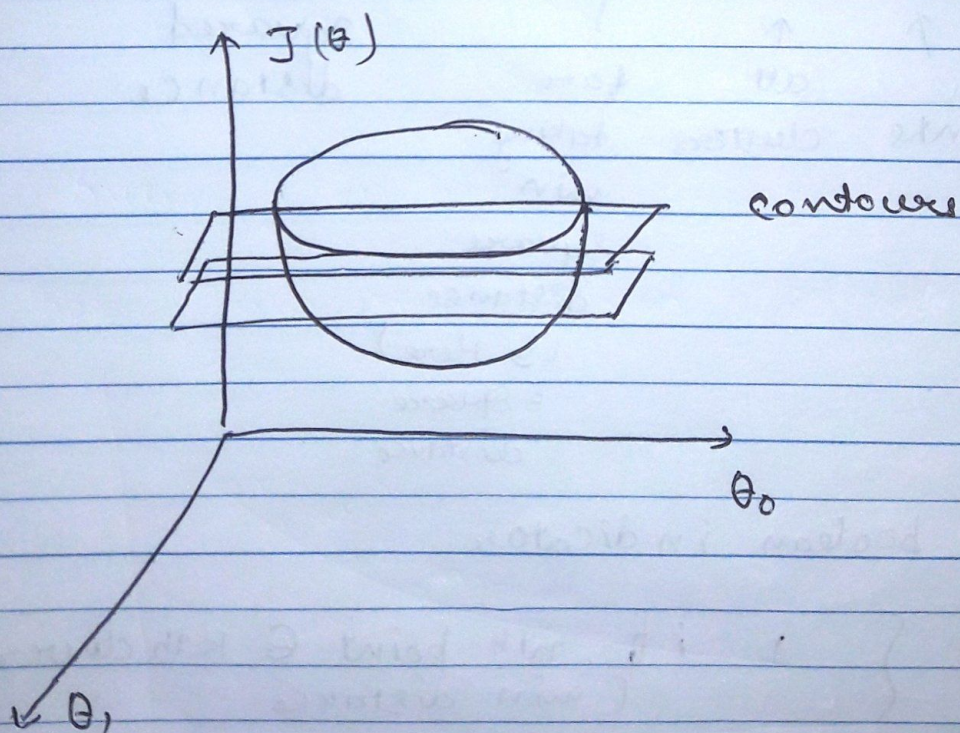
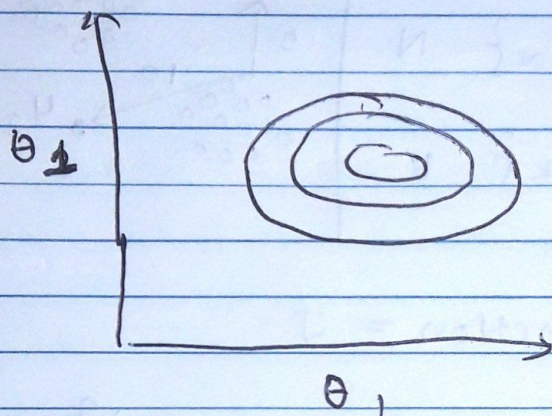
$$= \begin{cases} 1 & \text{if } n^{\text{th}} \text{ point } \in K^{\text{th}} \text{ cluster} \\ & \text{[min distance]} \\ 0 & \text{otherwise} \end{cases}$$



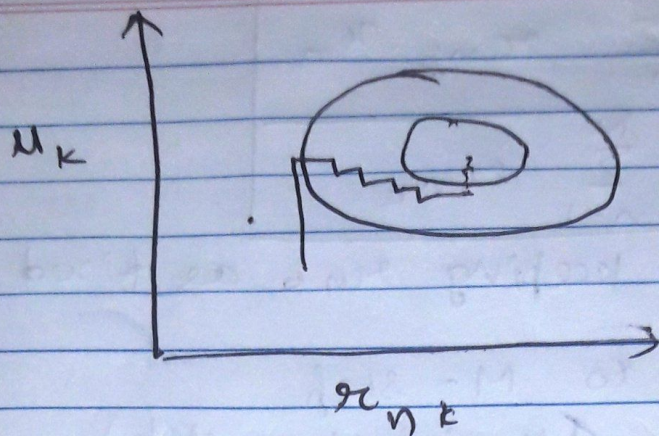
Learning parameters:

$x_{nk}$  and  $\mu_k$

Coordinate Descent:





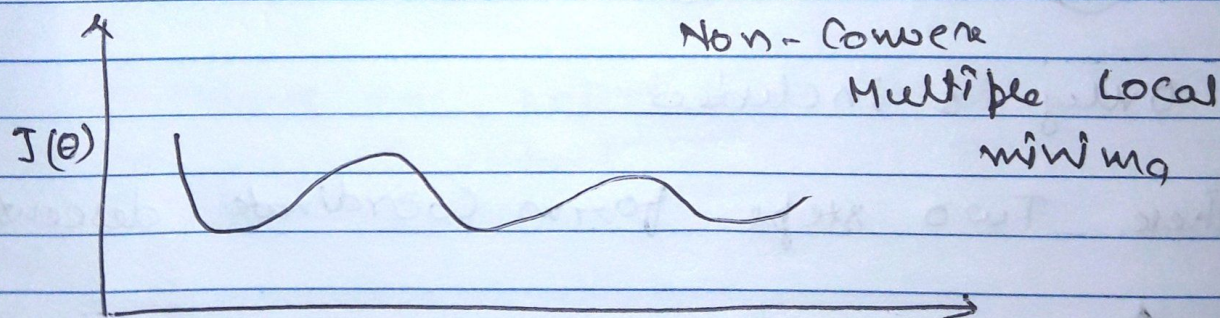


Step:  
finding  $x_{nk}$   
keeping  $\mu_k$   
as fixed

Coordinate Descent : wikipedia

$$x_{nk} = \begin{cases} 1 & \text{if } k = \arg \min_j \|x_n - \mu_j\|^2 \\ 0 & \text{otherwise} \end{cases}$$

$$J = \sum_n \sum_k x_{nk} (x_n - \mu_k)^2$$



if  $x_{nk}$  = fixed

⇒ Derivative

$$\Rightarrow \sum_{n=1}^N x_{nk} (x_n - \mu_k) = 0$$

$$\Rightarrow \mu_k \sum_{n=1}^N x_{nk} x_n - \mu_k \sum_{n=1}^N x_{nk} = 0$$



$$\Rightarrow \mu_k = \frac{\sum_{n=1}^N x_{nk} x_n}{\sum_{n=1}^N x_{nk}}$$

↓ find  $\mu_k$  keeping  $x_{nk}$  as fixed

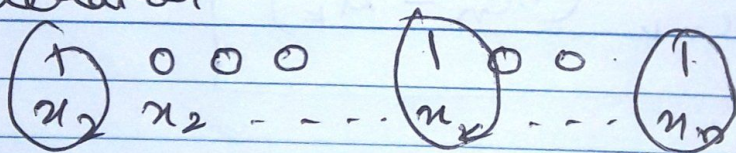
This is similar to M-step  
(Maximization step)

E-step = Expectation - step

(Denominator)

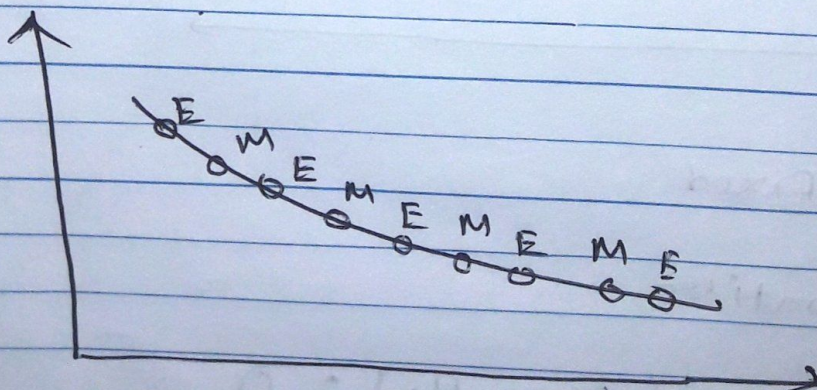
$x_{nk}$  = No. of points in cluster

Numerator

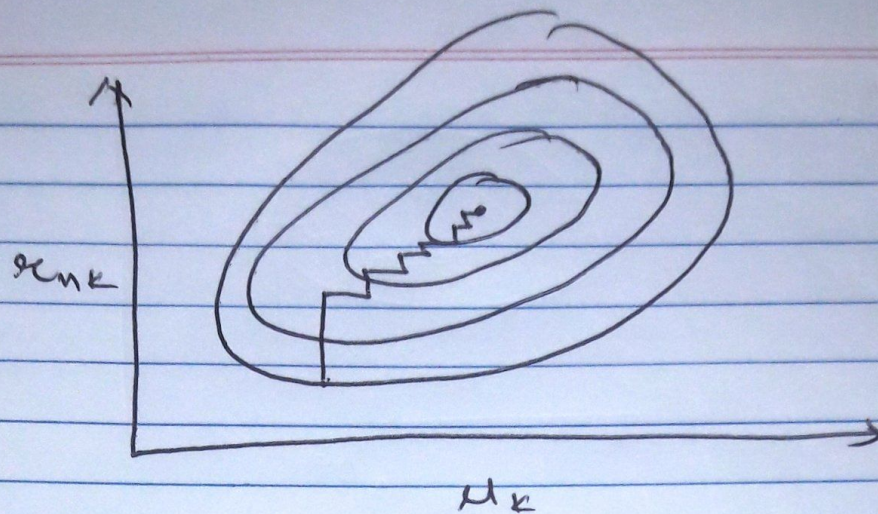


Only 1 included

- These Two steps form Coordinate descent,







Converges to local minima but not ~~mean~~ not global minima.

Because multiple local minima is present.

- We can have a problem where we can stuck in local minima

So choosing initial points is crucial.

K-Means ++ solve that problems to some extent.

