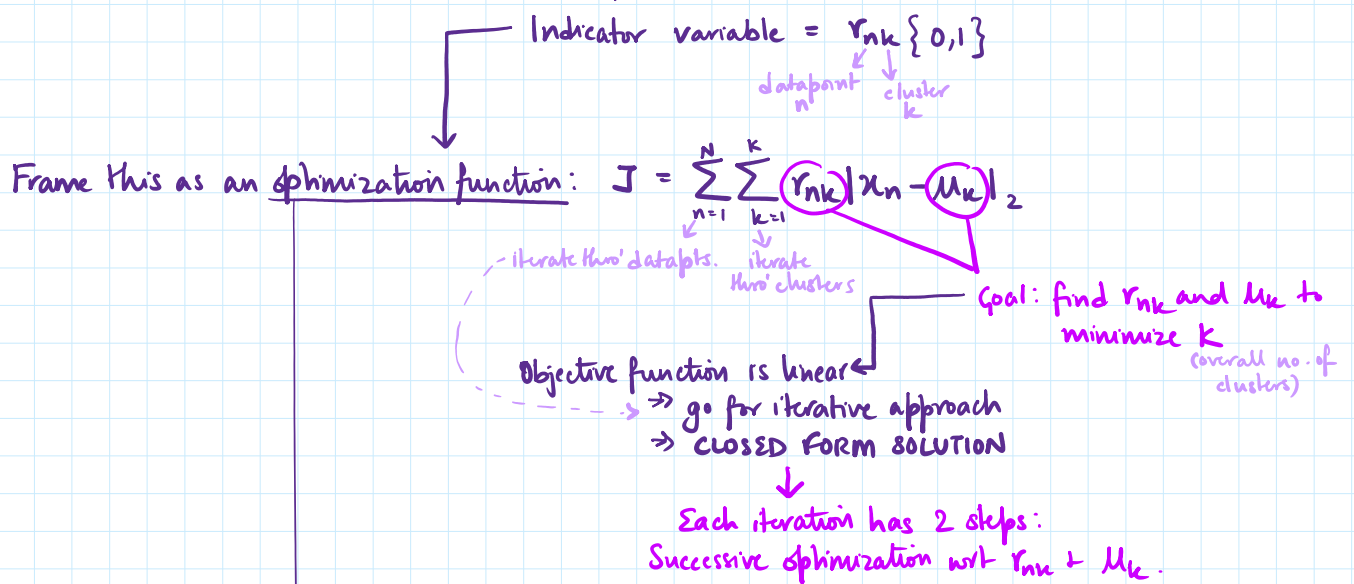


You have a dataset  $\{x_i\}_{i=1,2,\dots,n}$  in a 'D' dimensional space.  $\rightarrow$  partition into  $k$  clusters.



Optimization:

- ① Minimize  $J$  wrt  $r_{nk}$  keeping  $\mu_k$  fixed  $\rightarrow$  expected class? Expectation step
- ② Minimize  $J$  wrt  $\mu_k$  keeping  $r_{nk}$  fixed  $\rightarrow$  MLE of mean? Maximization step

ic: EM approach!

**E Step** — determine  $r_{nk}$  — keep  $\mu_k$  fixed

$J = \sum_{n=1}^N \sum_{k=1}^k r_{nk} \|x_n - \mu_k\|_2^2$  — linear fn of  $r_{nk} \rightarrow$  closed form solution.

terms  $x_n$  are independent  $\rightarrow$  optimize for each  $x_n$  separately.

$$r_{nk} = \begin{cases} 1 & \text{if } k = \operatorname{argmin}_k \|x_n - \mu_k\|_2 \\ 0 & \text{otherwise} \end{cases} \rightarrow \text{choose } r_{nk}=1 \text{ for a value } k, n \text{ that gives minimum value } \|x_n - \mu_k\|_2$$

Basically, assign  $x_n$  to a cluster  $k$  if the mean  $\mu_k$  is the closest (we'd fixed  $\mu_k$  — don't forget!)

**M Step** — optimize  $\mu_k$  — keep  $r_{nk}$  fixed

$$J = \sum_{n=1}^N \sum_{k=1}^k r_{nk} \|x_n - \mu_k\|_2^2$$

$$\frac{\partial J}{\partial \mu_k} = 0 \Rightarrow 2 \sum_{n=1}^N r_{nk} (x_n - \mu_k) = 0$$

$$\mu_k = \frac{\sum_{n=1}^N r_{nk} x_n}{\sum_{n=1}^N r_{nk}} \rightarrow \text{no. of points assigned to cluster } k.$$

Basically, set  $\mu_k$  = mean of all data points assigned to cluster  $k$  holding  $r_{nk}$  fixed.