# Lecture 18: K-means

Assume we have numeric features $x_i \in \mathbb{R}^p$

and define

$$D_{ii'} = \|x_i - x_{i'}\|^2$$

Want to choose $G_k$'s to minimize $W$

$$W = \sum_k \sum_{i,i' \in G_k} D_{ii'} = \sum_k \underbrace{\sum_{i \in G_k} \sum_{i' \in G_k} \|x_i - x_{i'}\|^2}$$

$$= \sum_k 2N_k \sum_{i \in G_k} \|x_i - \bar{x}_k\|^2$$

number in $G_k$

mean of pts in $G_k$

$$\textcircled{*} = \|x_i - x_{i'}\|^2 = \|x_i - \bar{x}_k + \bar{x}_k - x_{i'}\|^2$$

$$= ((x_i - \bar{x}_k) + (\bar{x}_k - x_{i'}))^T ((x_i - \bar{x}_k) + (\bar{x}_k - x_{i'}))$$

$$= (x_i - \bar{x}_k)^T (x_i - \bar{x}_k) \rightsquigarrow \|x_i - \bar{x}_k\|^2$$

$$+ (x_{i'} - \bar{x}_k)^T (x_{i'} - \bar{x}_k) \rightsquigarrow \|x_{i'} - \bar{x}_k\|^2$$

$$- 2(x_i - \bar{x}_k)^T (x_{i'} - \bar{x}_k)$$

$$\sum_{i \in G_k} \textcircled{*} = \sum_i \|x_i - \bar{x}_k\|^2 + \sum_i \|x_{i'} - \bar{x}_k\|^2 + 2(x_{i'} - \bar{x}_k)^T \underbrace{\sum_i (x_i - \bar{x}_k)}_{0}$$

$$\sum (y_i - \bar{y}) = 0$$

$$= \sum_i \|x_i - \bar{x}_k\|^2 + \|x_{i'} - \bar{x}_k\|^2 N_k$$

$$\sum_{i'} \sum_{i} (*) = \sum_{i'} \underbrace{\sum_{i} \|x_i - \bar{x}_k\|^2}_{\text{no } i'} + N_k \sum_{i'} \|x_{i'} - \bar{x}_k\|^2$$

$$= N_k \sum_i \|x_i - \bar{x}_k\|^2 + N_k \sum_{i'} \|x_{i'} - \bar{x}_k\|^2$$

$$= 2N_k \sum_{i \in G_k} \|x_i - \bar{x}_k\|^2$$

$$\boxed{W = \sum_k 2N_k \sum_{i \in G_k} \|x_i - \bar{x}_k\|^2.}$$

try to minimize.

## Lloyd's Algorithm

0 **Initialize Step**:

make some random initializations of cluster means

$$\mu_1^{(0)}, \ldots, \mu_K^{(0)}$$

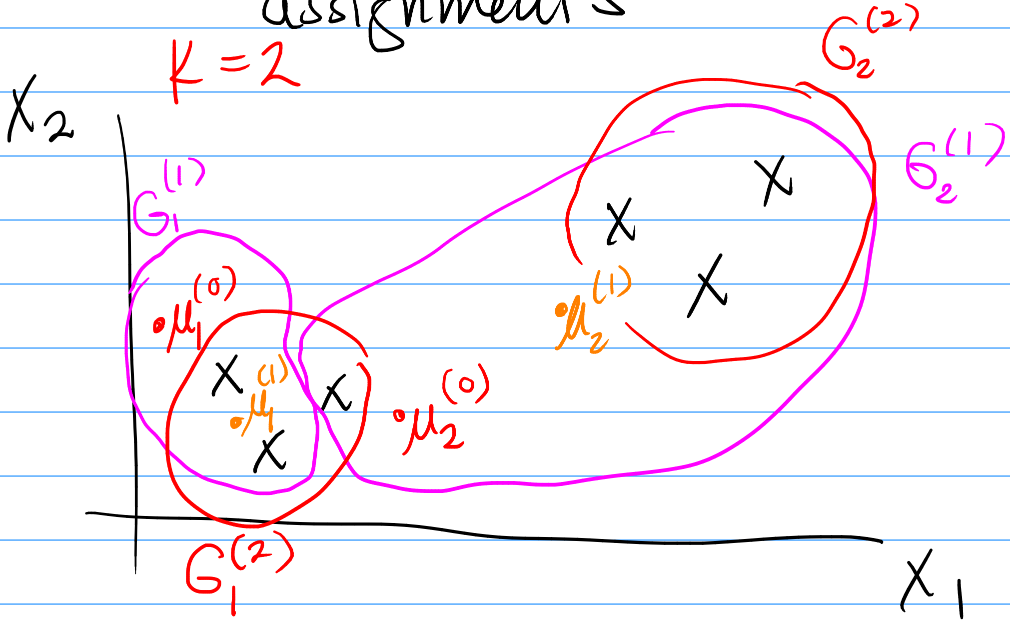
For $t = 1, 2, 3, \ldots$

at the $t^{th}$ iteration

(1) Assignment Step:

assign $i$ or $x_i$ to cluster $G_k$ w/ the closest mean

(2) Update Step: re-compute the $\mu$s as the means of the (new) cluster assignments




Why is this reasonable?

$$\hat{G}_1, \dots, \hat{G}_K = \underset{G_1, \dots, G_K}{\operatorname{argmin}} \sum_k N_k \sum_{i \in G_k} \|x_i - \bar{x}_k\|^2$$

Generalize:

$$\hat{G}_1, \dots, \hat{G}_K, \hat{m}_1, \dots, \hat{m}_K = \underset{G_i, m_i}{\operatorname{argmin}} \overbrace{\sum_k N_k \sum_{i \in G_k} \|x_i - m_i\|^2}^{W}$$

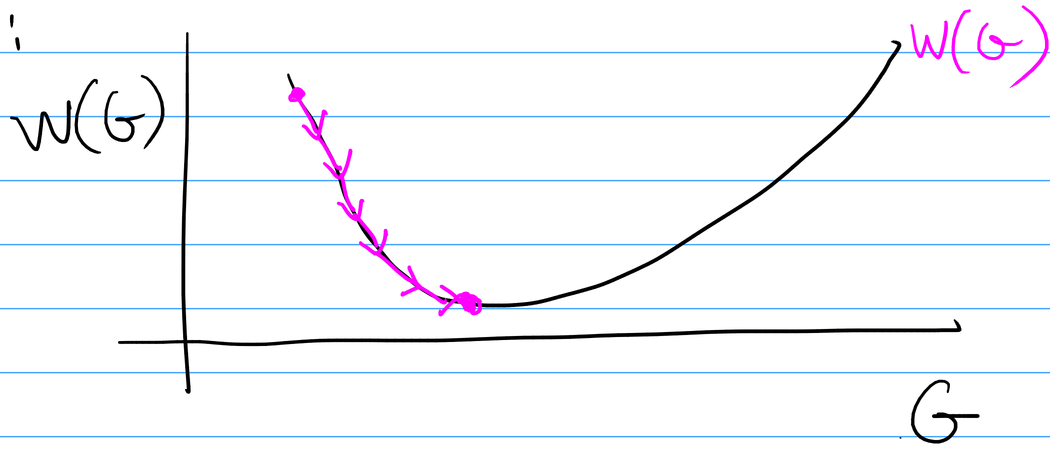
Fact: (1) given fixed $G_k$s best $m_k = \bar{X}_k$

$$\bar{y} = \underset{m}{\operatorname{argmin}} \|y_i - m\|^2$$

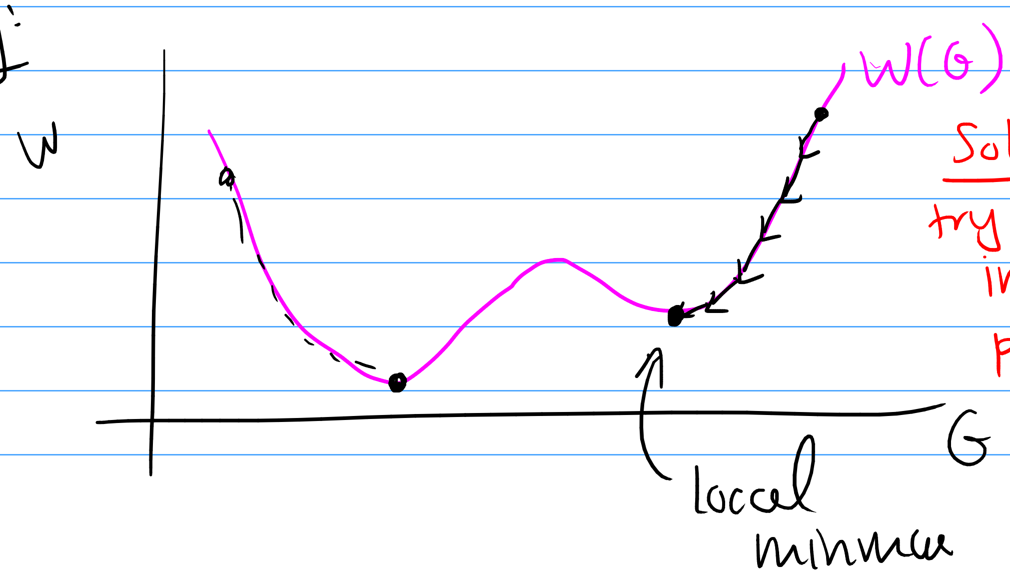
Fact: (2) given fixed $m_k$ the best $G_k$ is to assign pts to cluster w/ closest $m_k$.

B/c of these two facts, at each step Lloyd's algorithm makes $W$ smaller. (at least not larger)

ideally:




reality:




Soln: try different initializations, pick one w/ smallest $W$

## How do we choose K?

As I increase K, W will decrease.


---

What about non-numeric data or non-euclidean distance?

All we need is D.

## K-mediods

**Step 0: initialization**

Choose some random pts $u_k^*$, $k=1,\dots,K$
as "representative" of some clusters
— called mediods

For $t=1,2,3,\dots$
For each val of $t$

**Step 1: assignment**

assign each $i$ to group $G_k$ if its dissim w/ $i_k^*$ is smallest
all possible mediods

assign $i$ to $G_k$ if $D_{i i_k^*} \leq D_{i i_{k'}^*} \ \forall k'$

## Step 2: Update:

Choose new representatives for each group as the point w/ least total dissim to all other pts in group

$$i_k^* = \underset{i \in G_k}{\operatorname{argmin}} \sum_{i' \in G_k} D_{ii'}$$

---