

# K-mean Algorithm (1)

[https://en.wikipedia.org/wiki/K-means\\_clustering](https://en.wikipedia.org/wiki/K-means_clustering)

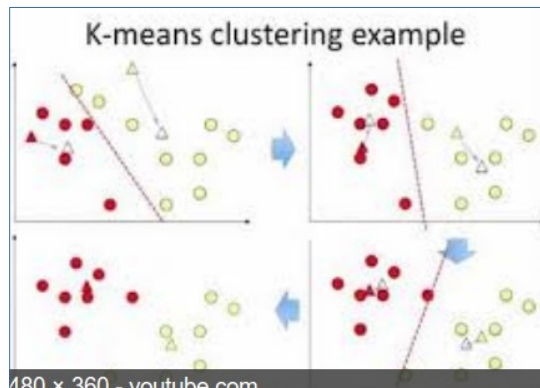
Given a set of observations ( $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ ), where each observation is a  $d$ -dimensional real vector,  $k$ -means clustering aims to partition the  $n$  observations into  $k$  ( $\leq n$ ) sets  $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$  so as to minimize the within-cluster sum of squares (WCSS) (i.e. **variance**). Formally, the objective is to find:

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg \min_{\mathbf{S}} \sum_{i=1}^k |S_i| \text{Var } S_i$$

Example:



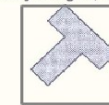
Form  
feature  
vectors



Cluster  
Seekingg

## Example On Simple Pattern Recognition

Given two binary images, derived from two objects, T and O, design a technique to identify them



Example: Computation of  
(1) Area (size);  
(2) X-bar;  
(3) Y-bar;  
(4) Orientation, theta angle  
(5) Perimeter of an object

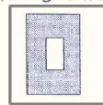


Fig1(a),(b)

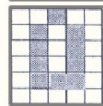


Fig2(a),(b)

Good continuation or noise? What to do with this noise?

Feature Vector		Size	X-bar	Y-bar	Orientation	Perimeter	
V_1(v1,..., v5)	T	v11	v12	v13	v14	v15	From Fig1(b)
V_2(v1,..., v5)	L	v21	v22	v23	v24	v25	From Fig2(b)

Harry Li, PhD SJSU

# K-mean Algorithm (2)

[https://en.wikipedia.org/wiki/K-means\\_clustering](https://en.wikipedia.org/wiki/K-means_clustering)

**Assignment step:** Assign each observation to the cluster whose mean has the least squared **Euclidean distance**, this is intuitively the "nearest" mean.<sup>[7]</sup> (Mathematically, this means partitioning the observations according to the **Voronoi diagram** generated by the means).

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\},$$

**Update step:** Calculate the new means to be the **centroids**

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$