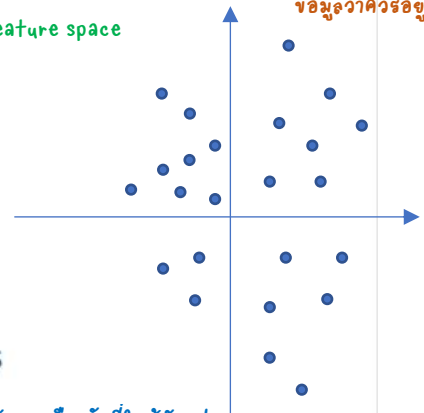


## Partitioning-Based Clustering Methods

- ❑ Basic Concepts of Partitioning Algorithms
- ❑ The K-Means Clustering Method
- ❑ Initialization of K-Means Clustering
- ❑ The K-Medoids Clustering Method
- ❑ The K-Medians and K-Modes Clustering Methods
- ❑ The Kernel K-Means Clustering Method

ปัญหาคือ มีแค่ข้อมูล  $x$   
เป็น feature space



ต้องการจะแบ่งกลุ่ม  
ข้อมูลว่าควรอยู่กลุ่มไหน

หลักการ คือ ตัวที่ใกล้กันอยู่  
ด้วยกัน ตัวที่อยู่ไกลกันควรอยู่  
คนละกลุ่มกัน

11

## Partitioning Algorithms: Basic Concepts

- ❑ Partitioning method: Discovering the groupings in the data by optimizing a specific objective function and iteratively improving the quality of partitions
- ❑ K-partitioning method: Partitioning a dataset  $D$  of  $n$  objects into a set of  $K$  clusters so that an objective function is optimized (e.g., the sum of squared distances is minimized, where  $c_k$  is the centroid or medoid of cluster  $C_k$ )
  - ❑ A typical objective function: **Sum of Squared Errors (SSE)**

$$SSE(C) = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - c_k\|^2$$
- ❑ Problem definition: Given  $K$ , find a partition of  $K$  clusters that optimizes the chosen partitioning criterion
  - ❑ Global optimal: Needs to exhaustively enumerate all partitions
  - ❑ Heuristic methods (i.e., greedy algorithms): *K-Means*, *K-Medians*, *K-Medoids*, etc.

12

# The *K-Means* Clustering Method

เป็นวิธีที่ง่ายที่สุด

□ *K-Means* (MacQueen'67, Lloyd'57/'82)

□ Each cluster is represented by the center of the cluster

□ Given  $K$ , the number of clusters, the *K-Means* clustering algorithm is outlined as follows  $K$  สามารถกำหนดค่าเองได้

□ Select  $K$  points as initial centroids

□ Repeat

□ Form  $K$  clusters by assigning each point to its closest centroid

□ Re-compute the centroids (i.e., *mean point*) of each cluster

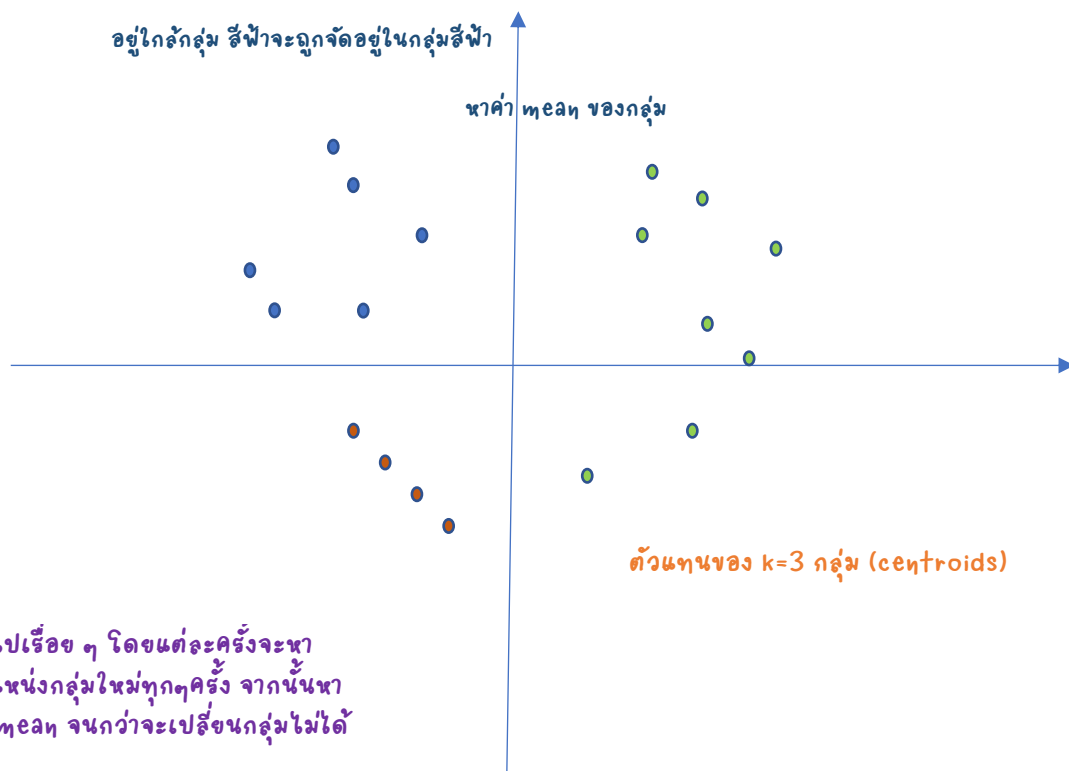
□ Until convergence criterion is satisfied

□ Different kinds of measures can be used

ทำการสุ่มจุด จะเข้าไปอยู่ทุกตัวว่า  
ตัวไหนอยู่ใกล้ centroids ไหน

□ Manhattan distance ( $L_1$  norm), Euclidean distance ( $L_2$  norm), Cosine similarity

13



กำหนดกลุ่ม 2 กลุ่ม

## Example: *K*-Means Clustering

