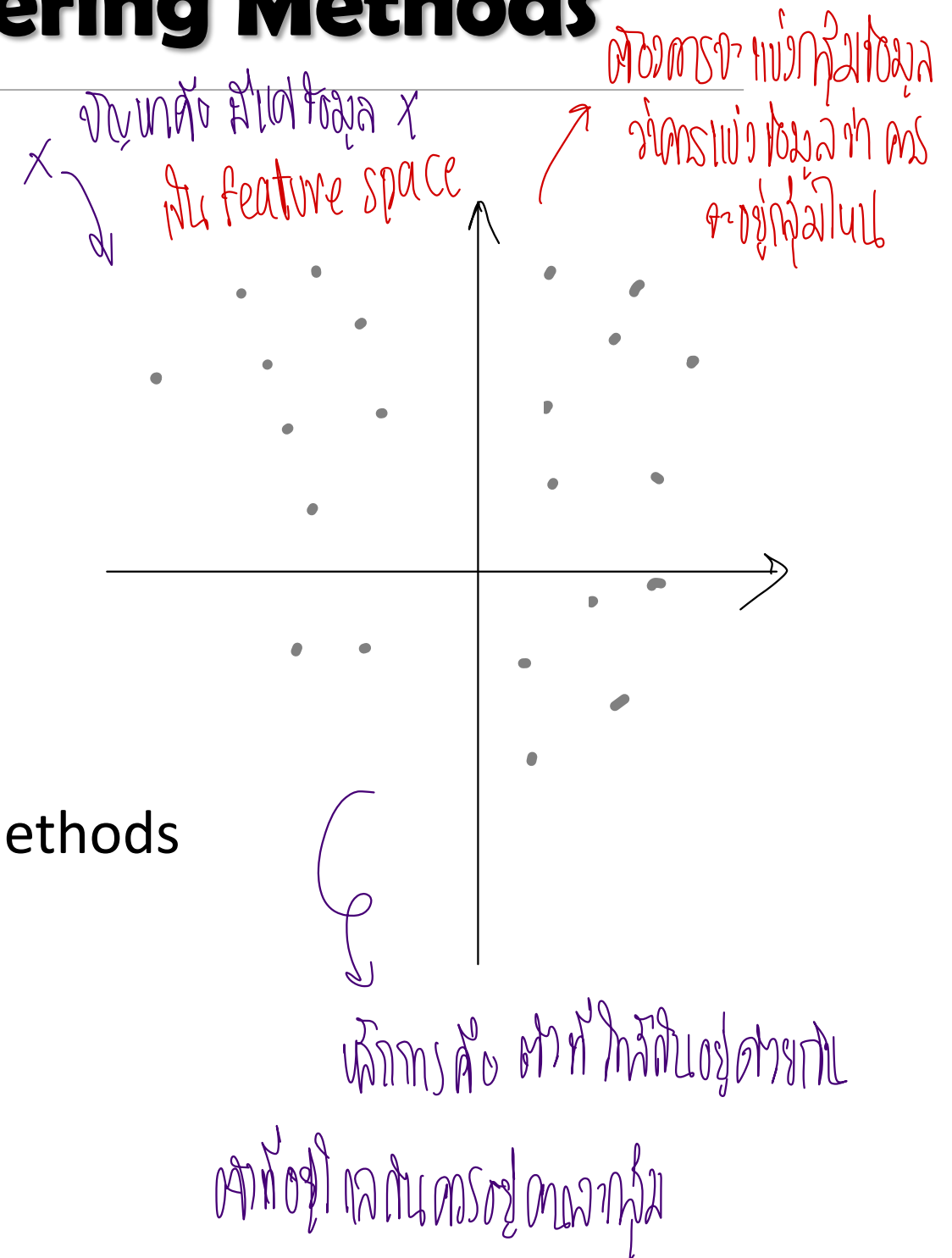


Partitioning-Based Clustering Methods

- 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



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.

အမှတ်အသား
↑ (parameter ကို)

အမှတ်အသား
↑

The *K-Means* Clustering Method

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

- Each cluster is represented by the center of the cluster

→ အကယ်၍ $k=3$

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

- Select K points as initial centroids

အစဉ်အဆက်

□ Repeat

- Form K clusters by assigning each point to its closest centroid

အနီးဆုံး ချုပ်နှောင်မှု

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

အမှတ်အသား centroids နှစ်

- Until convergence criterion is satisfied

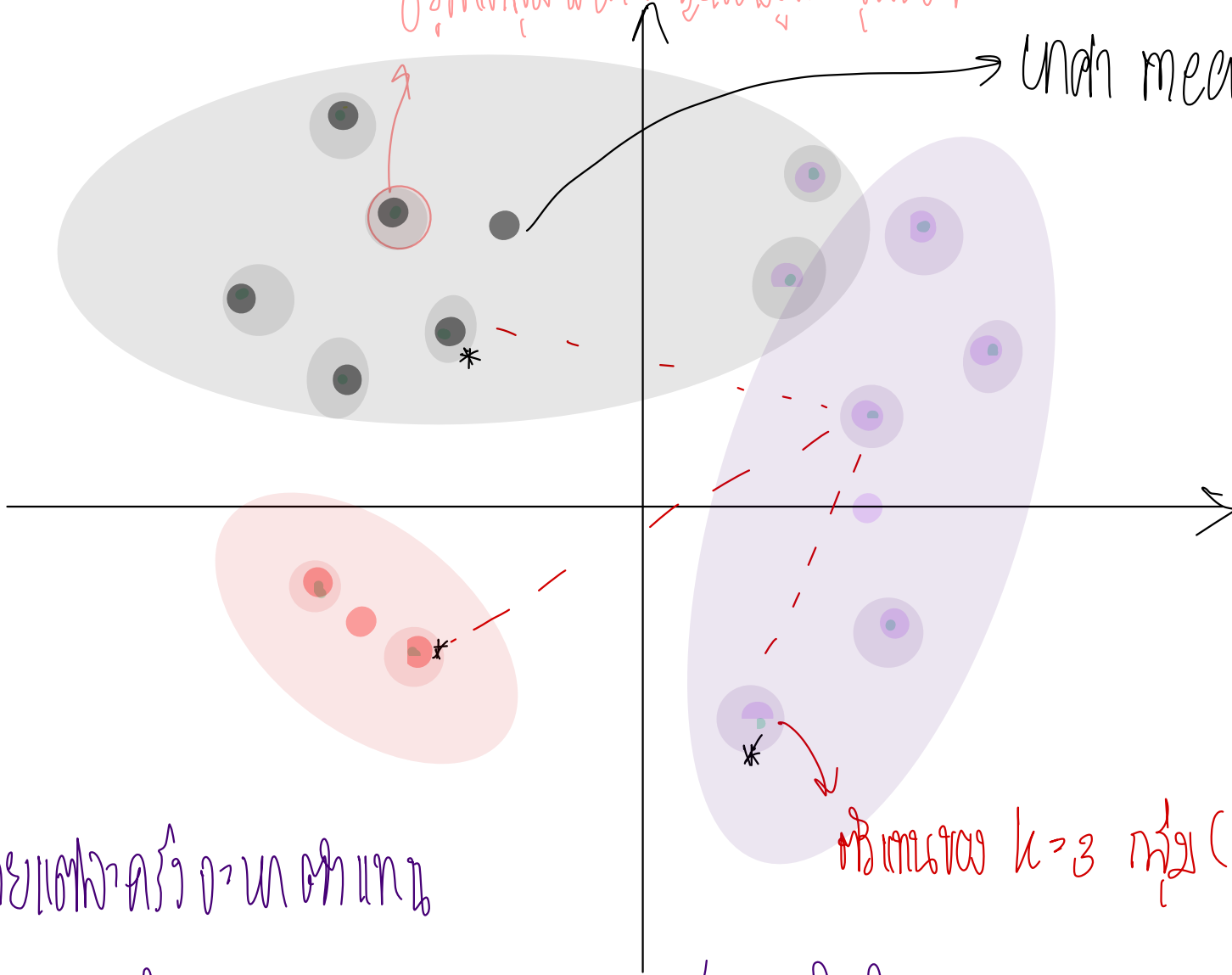
- Different kinds of measures can be used

- Manhattan distance (L_1 norm), Euclidean distance (L_2 norm), Cosine similarity

$$k = 3$$

ចំណុចទី៣ ៖ ពិចារណា ៖ ប្រសិនបើ n គឺជាចំនួនគត់វិជ្ជមាន តើ $n!$ ចែកចេញដោយ 2^n ដែរឬទេ?

• Und mean \vec{v}_{n2}

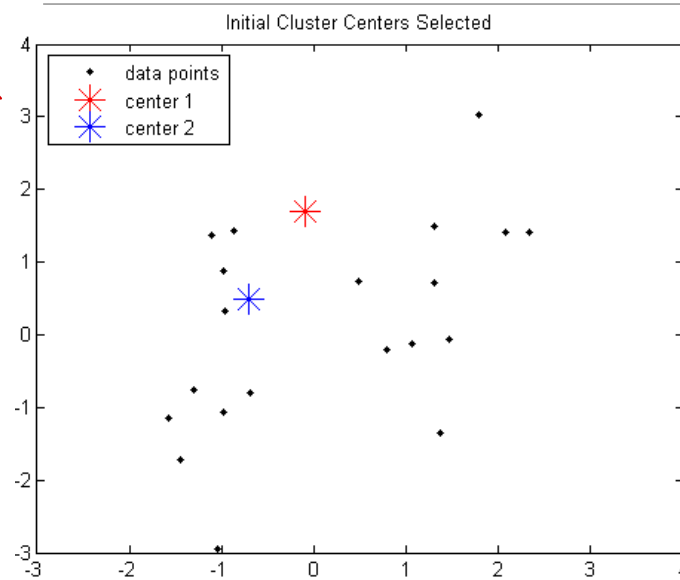


Number of clusters $k=3$ n_j (centroids)

ការវិវឌ្ឍន៍ តាមគោលការណ៍ ទូទៅ គឺជា
 ក្នុងការវិវឌ្ឍន៍ គឺជា ការវិវឌ្ឍន៍ ទូទៅ គឺជា

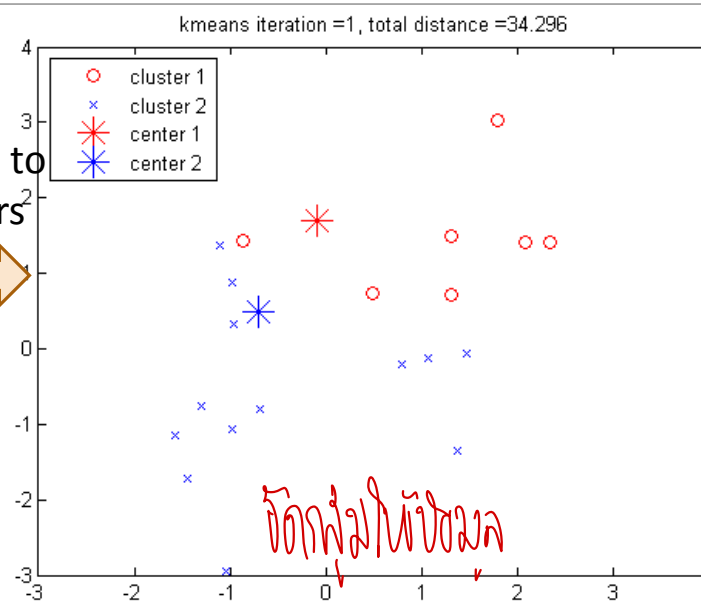
เลือกจุด 2 จุด

Example: *K-Means* Clustering

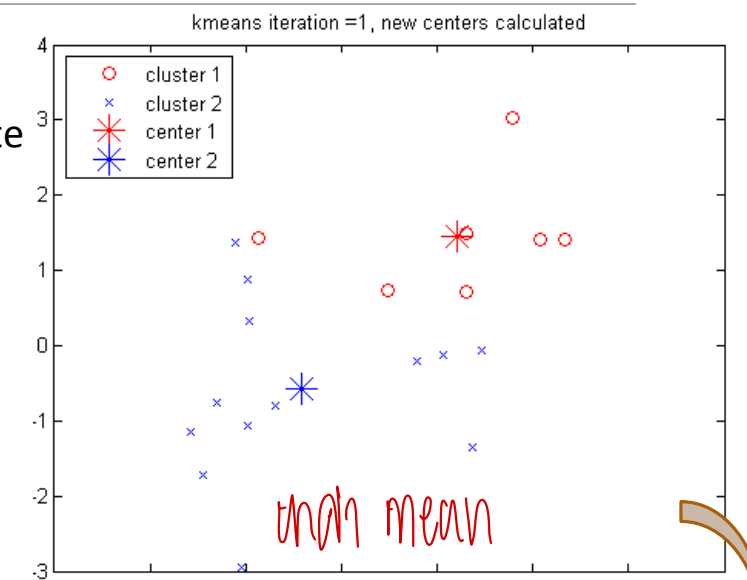


The original data points & randomly select $K = 2$ centroids

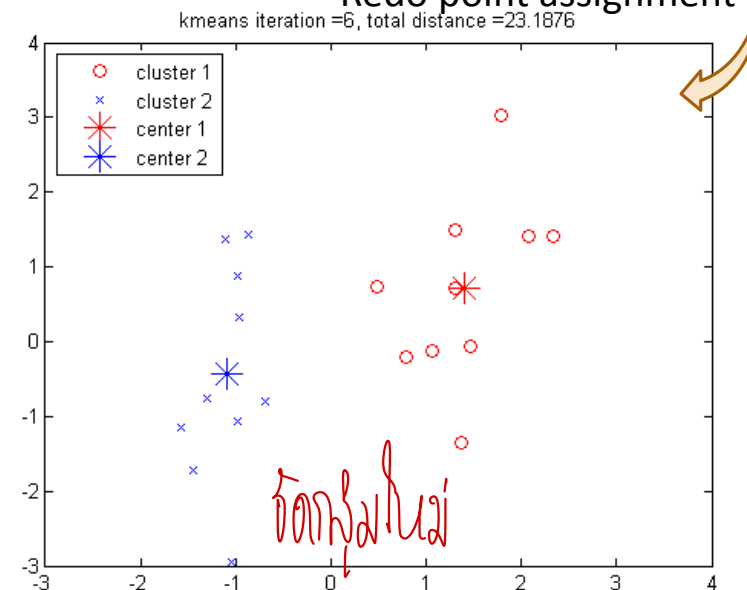
Assign points to clusters



Recompute cluster centers



Redo point assignment



Execution of the *K-Means* Clustering Algorithm

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