



K-MEANS ALGORITHM

SYRACUSE UNIVERSITY
School of Information Studies

CENTROID OF A CLUSTER

Centroid: The “gravity center” of a cluster, which is calculated as the average of all data examples in a cluster

For numeric variable, use the mean as the average.

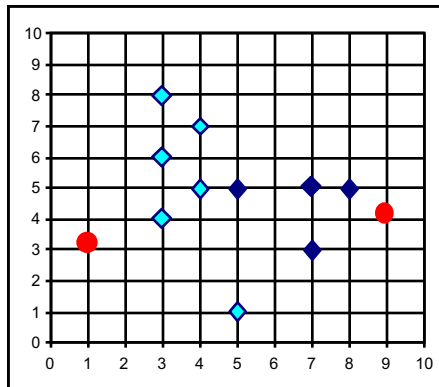
For nominal data, use the mode as the average.

THE K-MEANS CLUSTERING METHOD

-
- 1: Select K points as the initial centroids.
 - 2: **repeat**
 - 3: Form K clusters by assigning all points to the closest centroid.
 - 4: Recompute the centroid of each cluster.
 - 5: **until** The centroids don't change
-

THE K-MEANS CLUSTERING METHOD

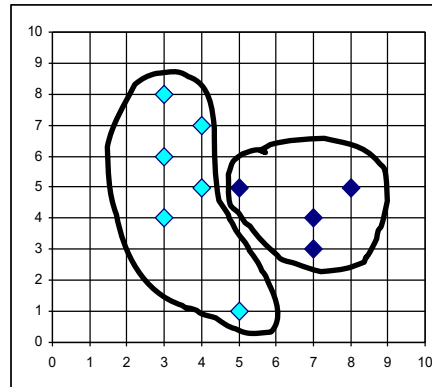
Example:



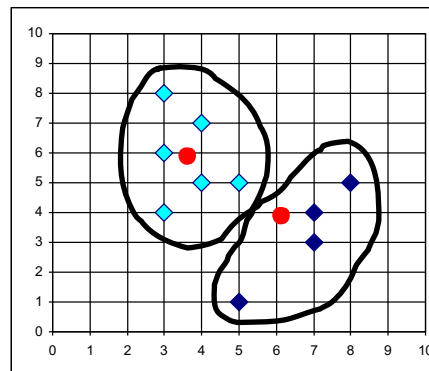
$k = 2$

Arbitrarily choose k object as initial cluster center.

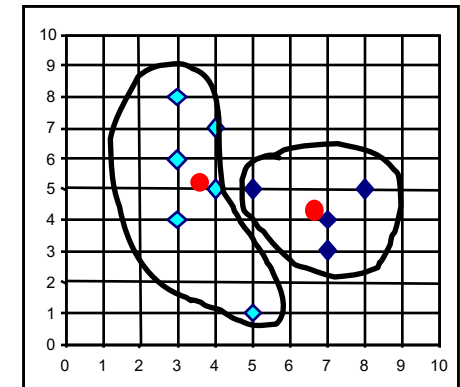
Assign each object to most similar center.



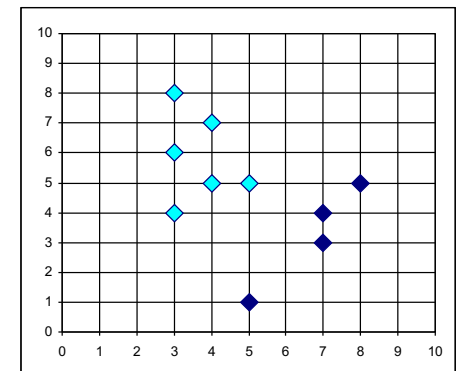
Reassign.



Update the cluster means.

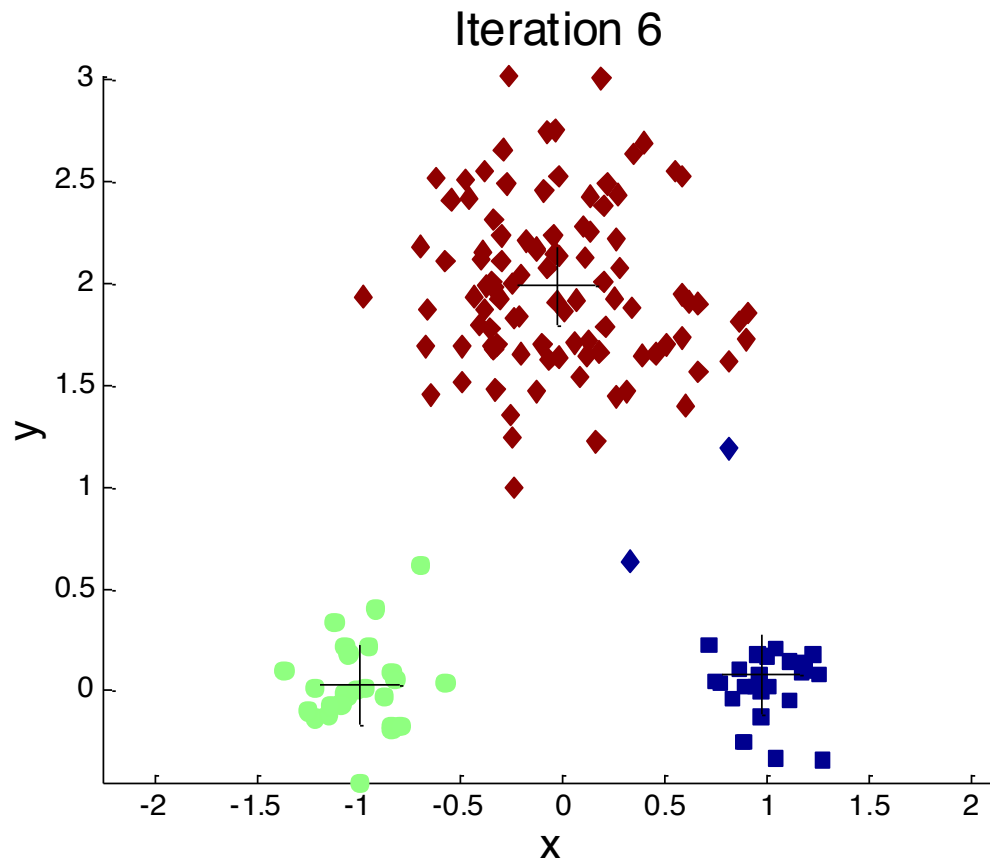


Reassign.



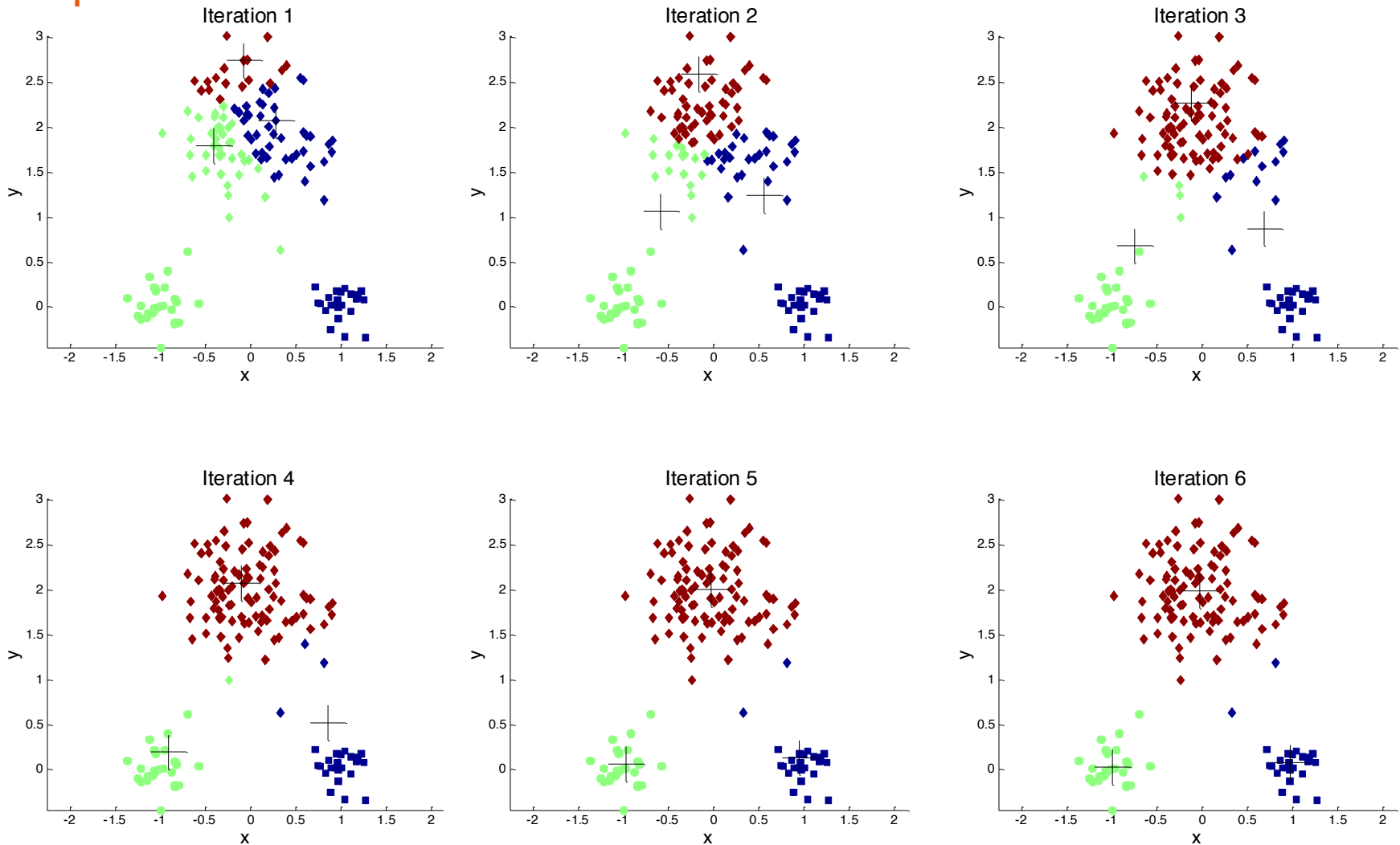
Update the cluster means.

IMPORTANCE OF CHOOSING INITIAL CENTROIDS

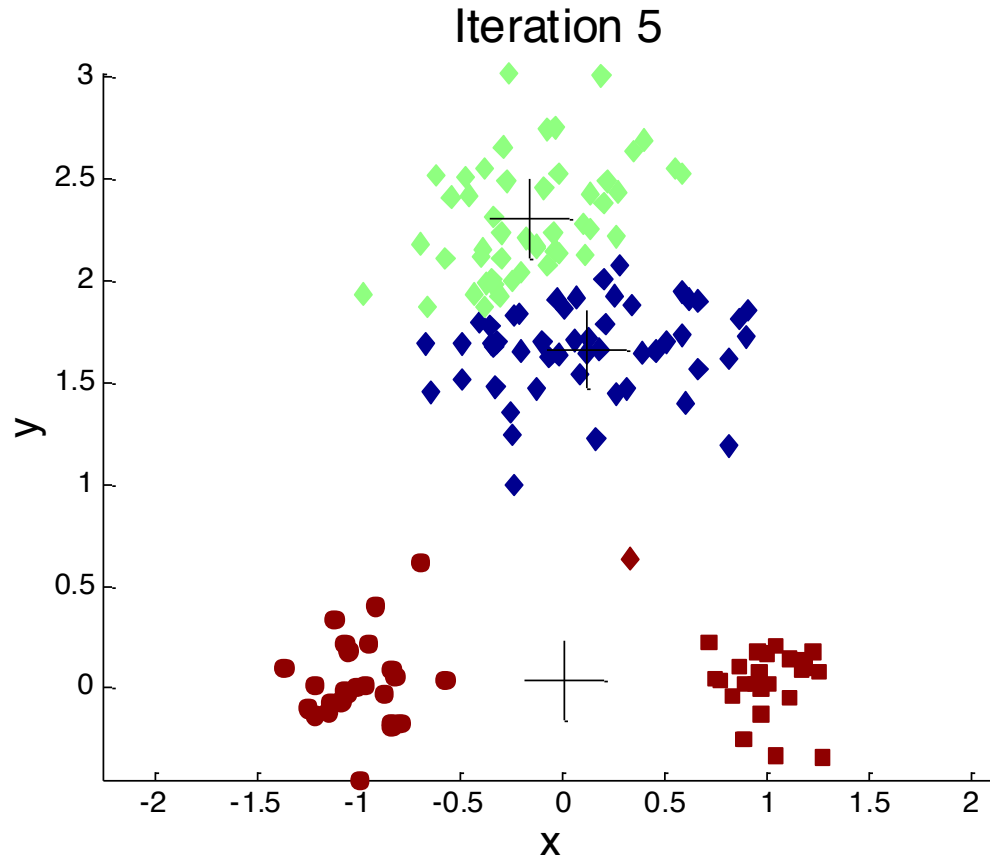


A good clustering result

IMPORTANCE OF CHOOSING INITIAL CENTROIDS



IMPORTANCE OF CHOOSING INITIAL CENTROIDS



A less meaningful result

IMPORTANCE OF CHOOSING INITIAL CENTROIDS

