MACHINE LEARNING

1211635

MAHASARAKHAM
UNIVERSITY

K-MEANS CLUSTERING

Clustering Algorithms

Olarik Surinta, PhD.
Lecturer

k-means clustering

- ***K-means*** *(MacQueen, 1976)* is one of the simplest unsupervised learning algorithms.
- ***K*-means clustering** aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
- The procedure follows a simple and easy way to classify a given data set through a certain number of clusters fixed a priori.

k-means clustering

- The main idea is to define *k* centroids, one for each cluster.
- These centroids should be placed in a cunning way because of different location causes different result.
- The better choice is to place them as much as far away from each other.

k-means clustering

- The next step is to take each point belonging to a given data set and associate it to the nearest centroid.
- At this point we need to re-calculate *k* new centroids as centers of the clusters resulting from the previous step.
- After we have these *k* new centroids, a new binding has to be done between the same data set points and the nearest new centroid.

k-means clustering

- A loop has been generated.
- As a result of this loop we may notice that the *k* centroids change their location step by step until no more changes are done. In other words centroids do not move any more.
- Finally, this algorithm aims at minimizing an objective function.

Objective function

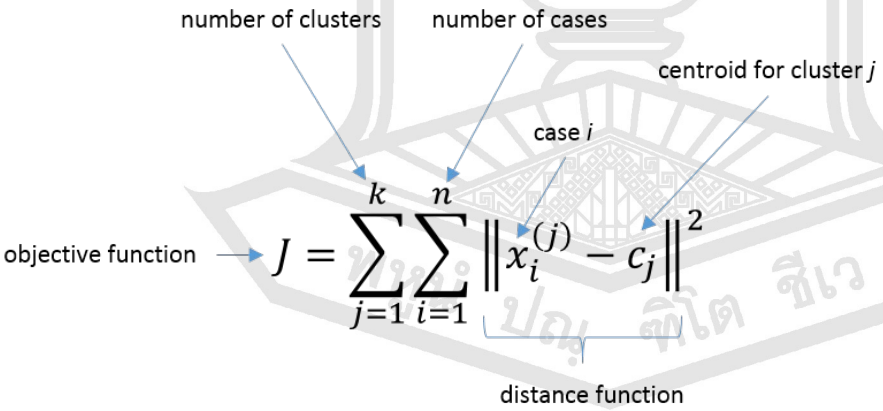
$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

where,

||xi - vj|| is the Euclidean distance between *xi* and *vj*

ci is the number of data points in i*th* cluster

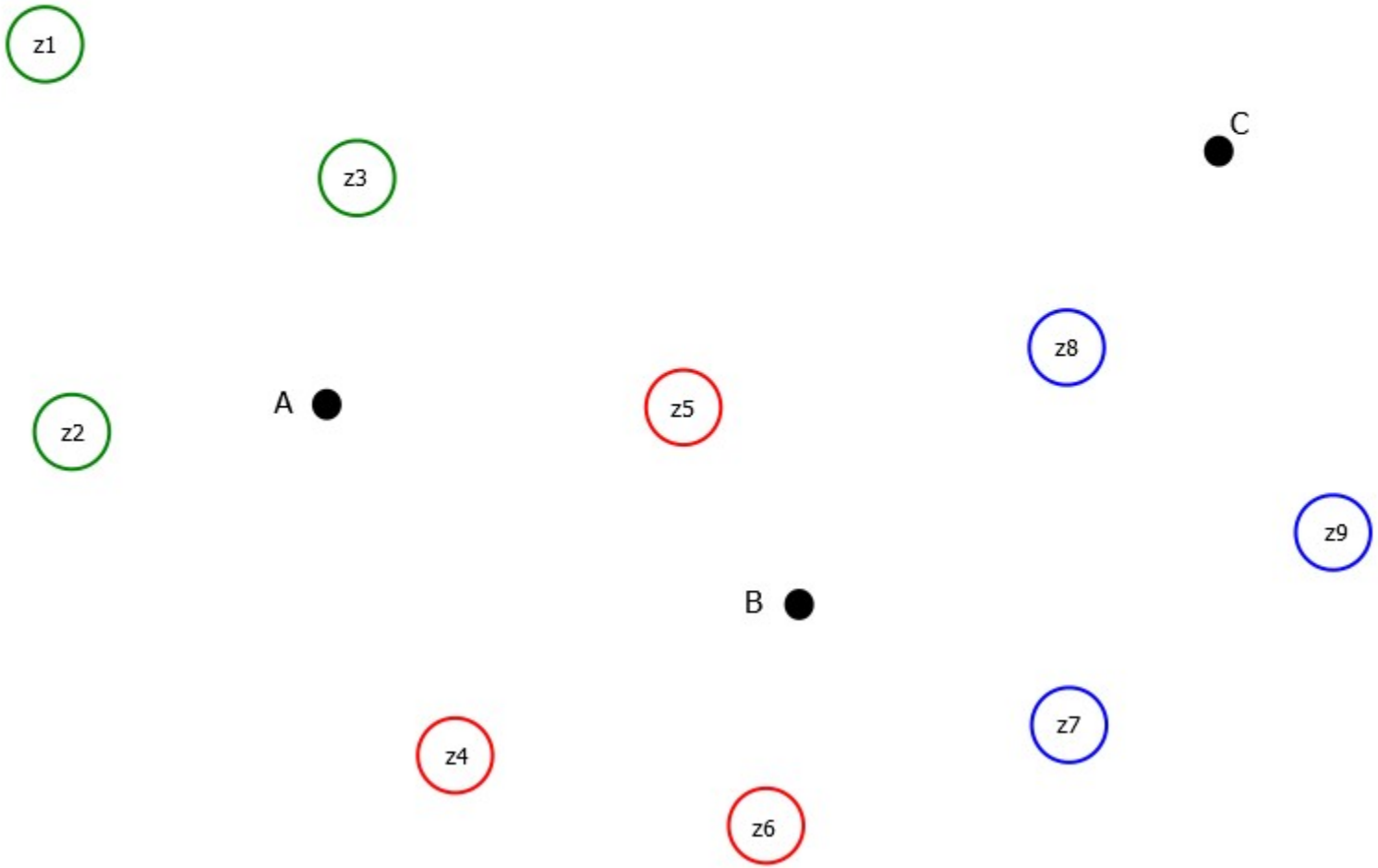
c is the number of cluster centers



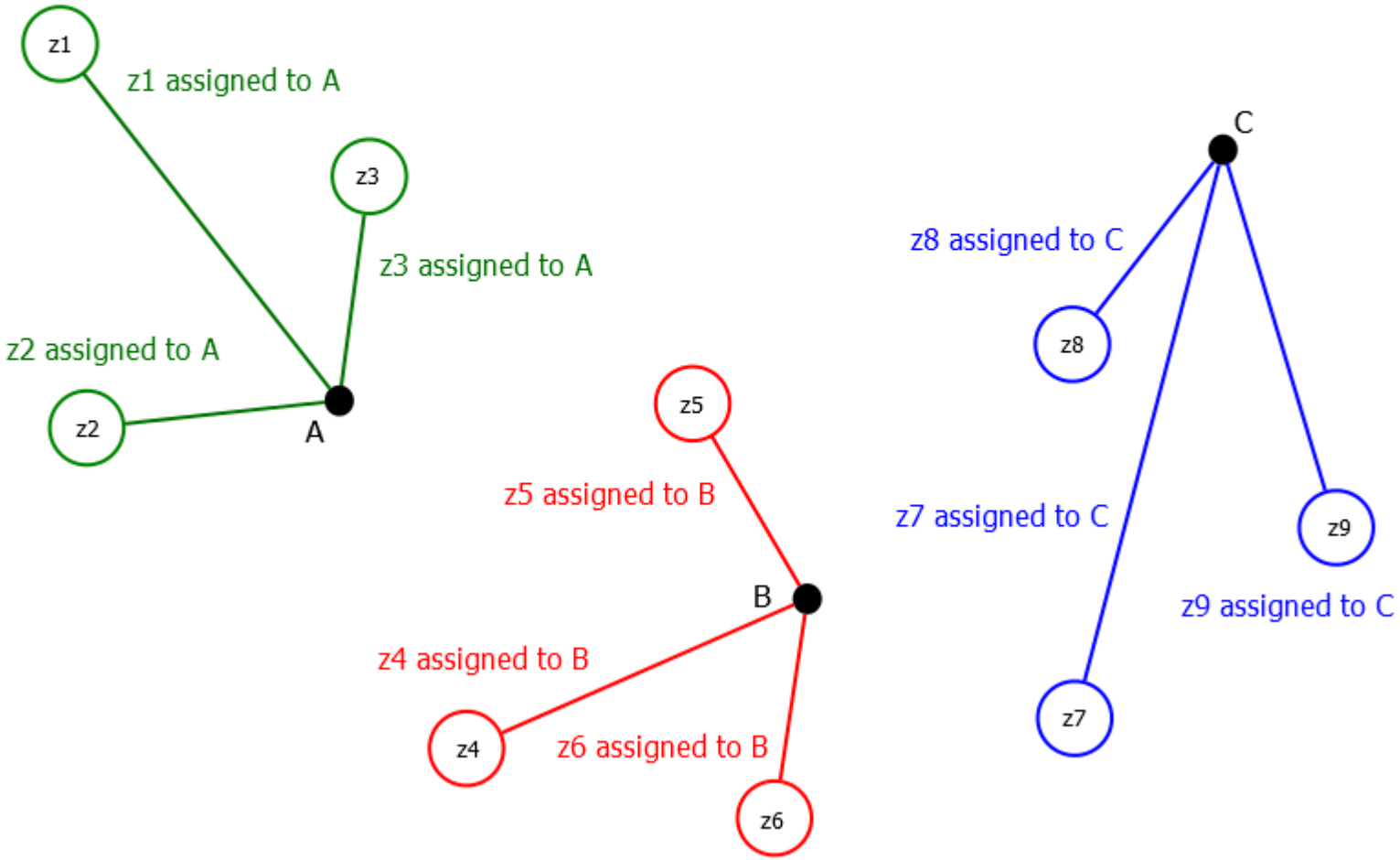
algorithm

- The algorithm is composed of the following steps:
  1) Place *k* points into the space represented by the objects that are being clustered. These points represent initial group centroids.
  2) Assign each object to the group that has the closest centroid.
  3) When all objects have been assigned, recalculate the positions of the *k* centroids.
  4) Repeat steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calcuated.

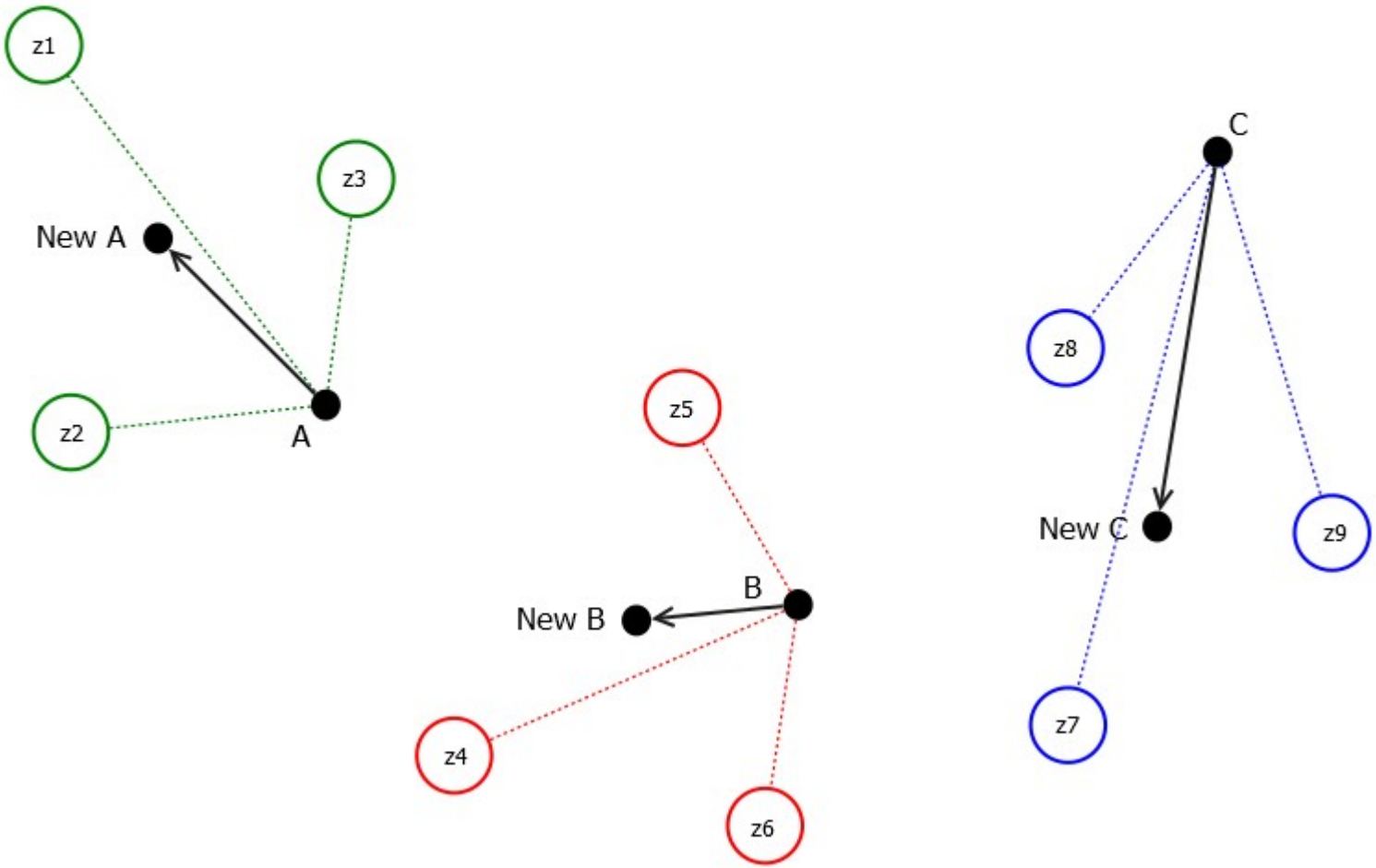
K-Means Clustering
Step One - Initialization

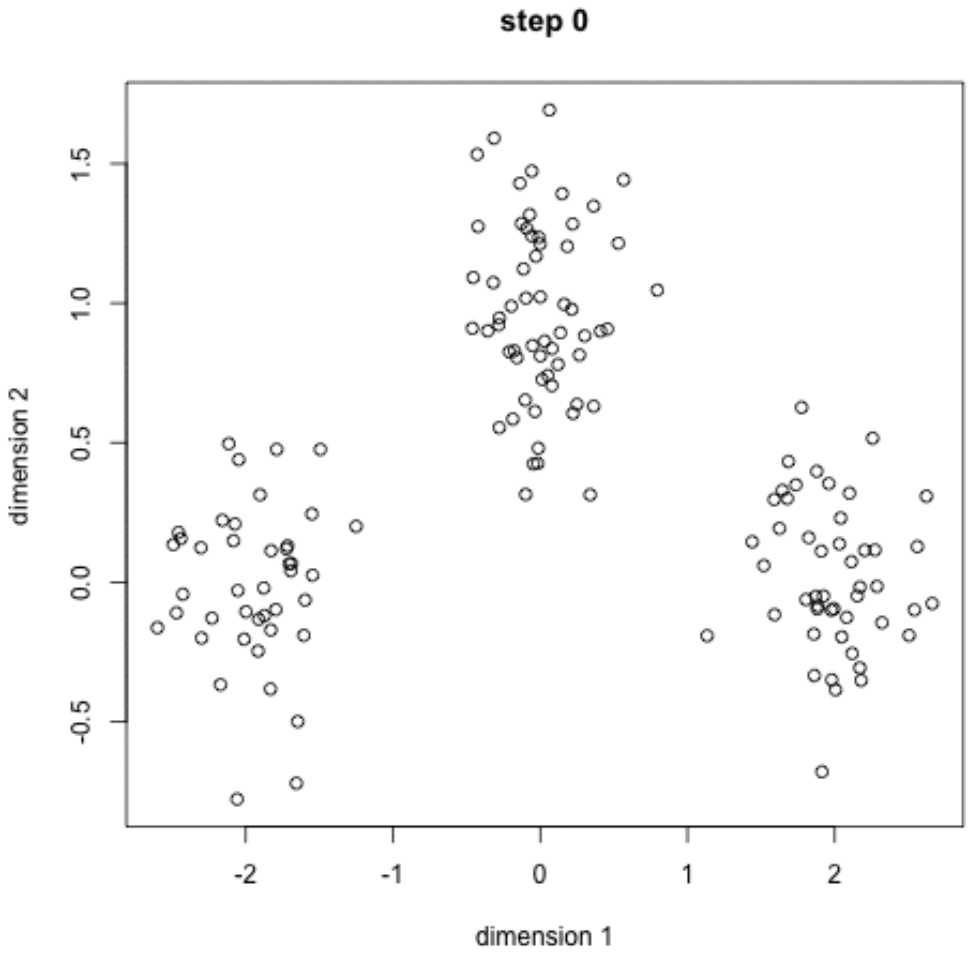


K-Means Clustering
Step Two - Assignment

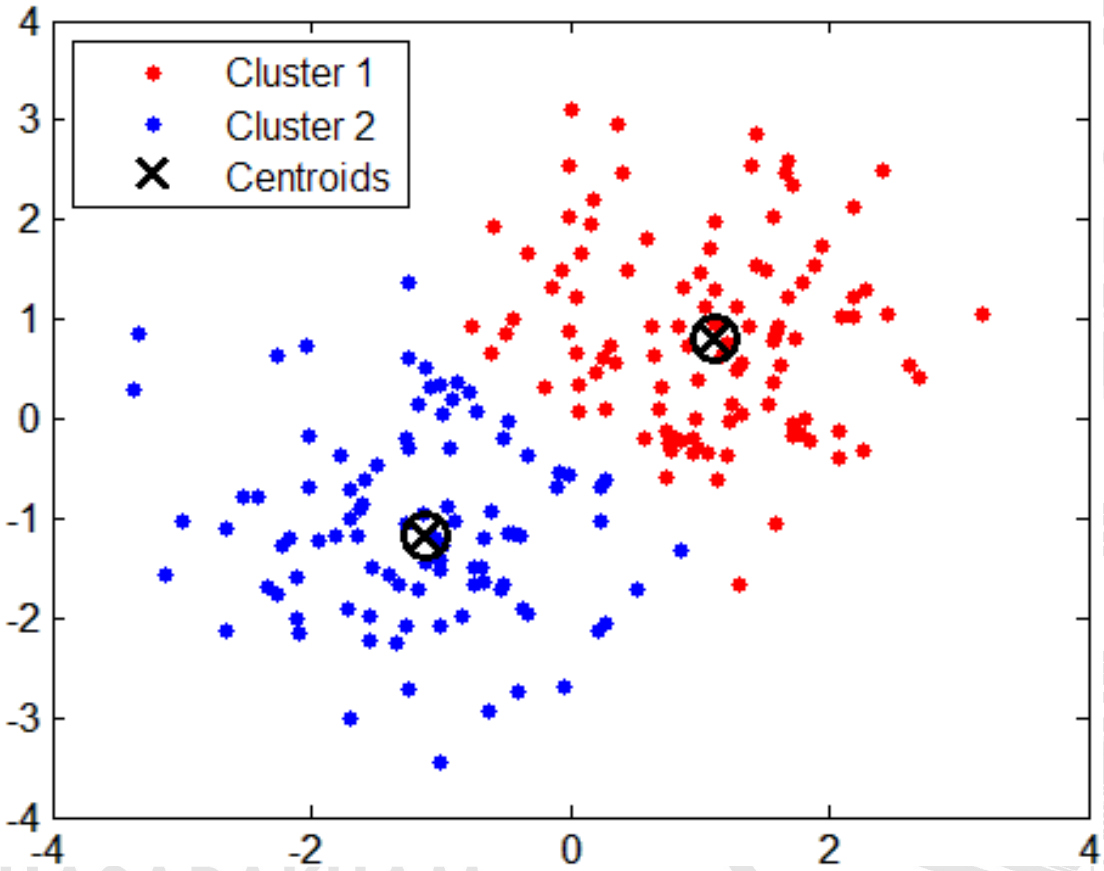


K-Means Clustering
Step Three - Updates





Click at the link to see the animation:
http://www.turingfinance.com/clustering-countries-real-gdp-growth-part2/



Cr. http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/mvoget/cluster/cluster.html

References

- https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html
- http://mnemstudio.org/clustering-k-means-example-1.htm
- https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm
- http://www.turingfinance.com/clustering-countries-real-gdp-growth-part2/

References

- http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/sphilip/kmeans.html
-