

K-Means Clustering

**INSTRUCTOR: HONGJIE CHEN
JUNE 16TH 2022**

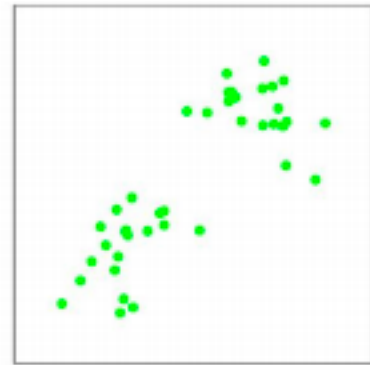
Clustering without Distribution Knowledge

- K-means:
 - Assume there are K clusters
 - Each node is assigned to the closest cluster

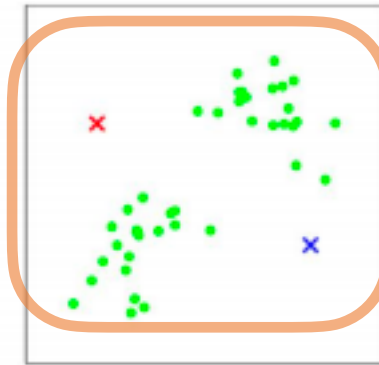
- Let's see an example

Example

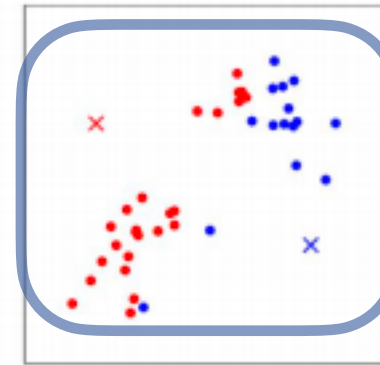
1. Random cluster initialization



(a)

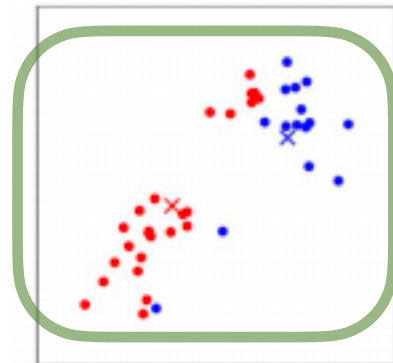


(b)

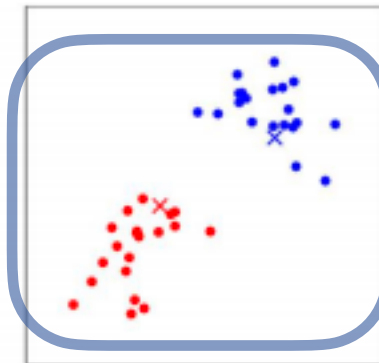


(c)

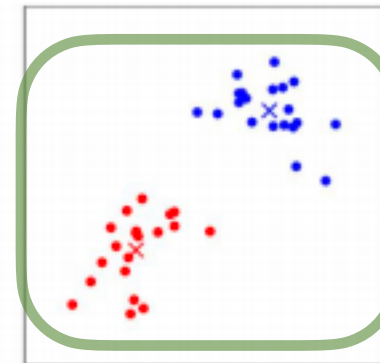
2. Node assignment



(d)



(e)



(f)

3. Update cluster

K-Means Algorithm

- Initialize K cluster centroids m_1, m_2, \dots, m_k
- Assign points to cluster, point x goes to cluster i if
 - $\|x - m_i\| < \|x - m_j\|, \forall j \neq i$
- Update cluster centroids, let S_i be the set of nodes in cluster i
 - $$m_i = \frac{1}{|S_i|} \sum_{x \in S_i} x$$
- Repeat Assignment and Update until converged (m_i doesn't changed)
- Demo: [links](#)

Best Clustering

- Objective

- $$\min \sum_i \text{distance}(x_i, m_{x_i})^2$$