UNIVERSITY OF LEEDS

Class: Machine Learning

Unsupervised learning: k-means

Instructor: Matteo Leonetti

Learning outcomes

UNIVERSITY OF LEEDS

- Apply the k-means algorithm to a dataset

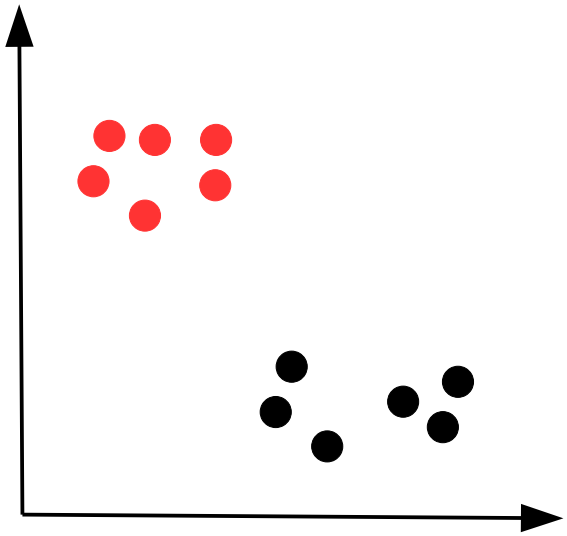
Clustering

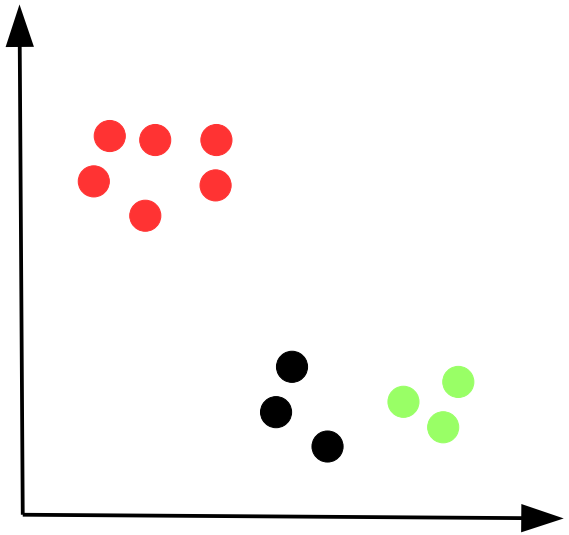
UNIVERSITY OF LEEDS

?

Scale

UNIVERSITY OF LEEDS

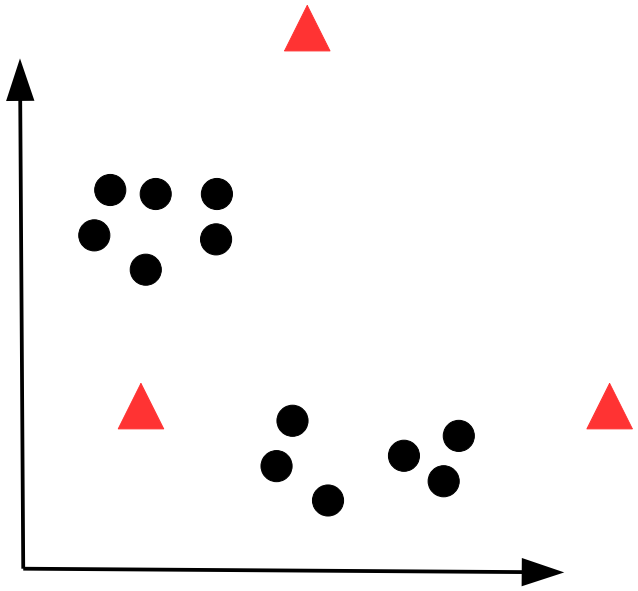




How many clusters do we have?

K-means

UNIVERSITY OF LEEDS

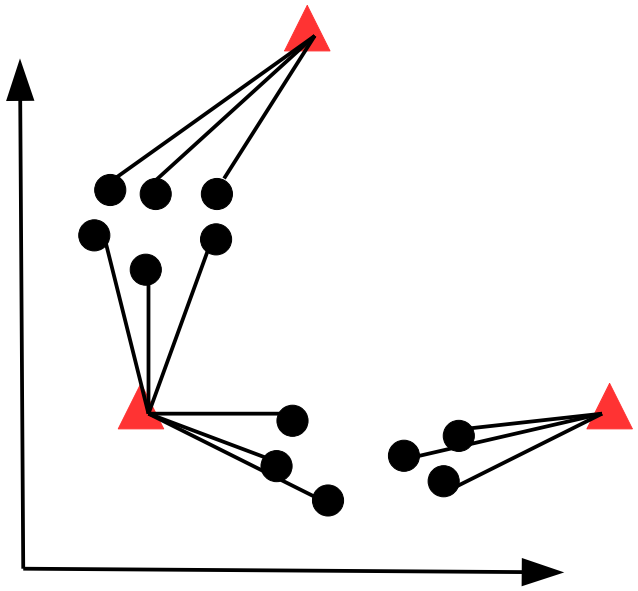


1. Choose the number of clusters (in the example: k=3)

2. Place k centroids randomly (the triangles)

K-means

UNIVERSITY OF LEEDS

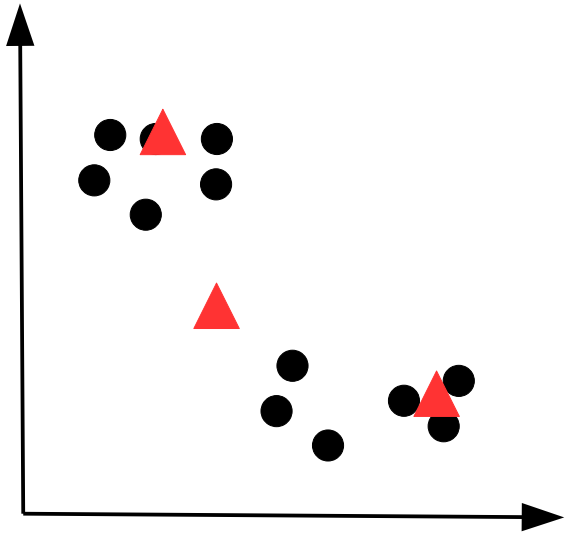


3. Identify the closest centroid to each point

4. Compute the new centroids for the clusters

K-means

UNIVERSITY OF LEEDS

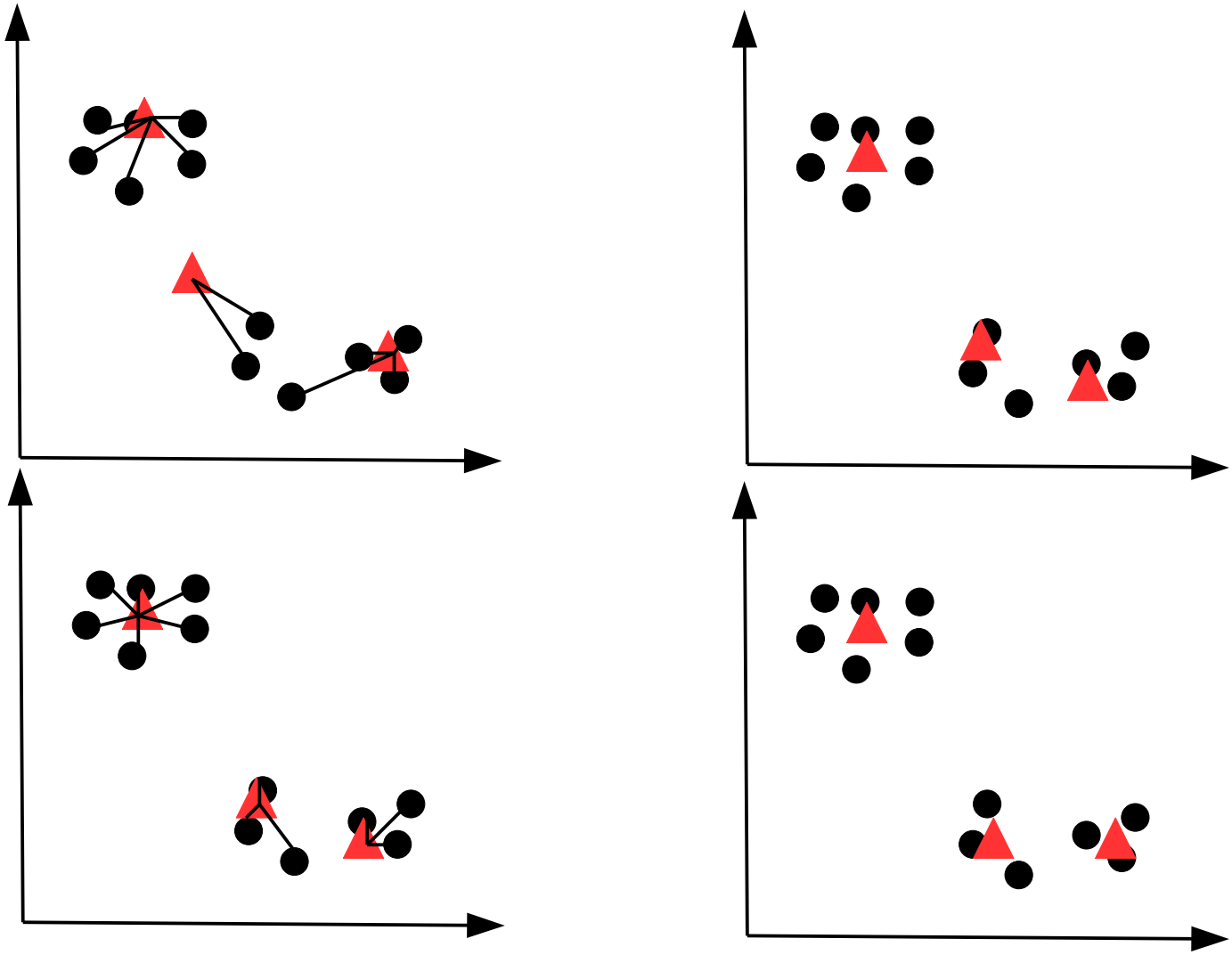


5. Repeat until the centroids do not move

Cluster new points with the closest cluster centre (centroid)

K-means

UNIVERSITY OF LEEDS



K-means - characteristics

UNIVERSITY OF LEEDS

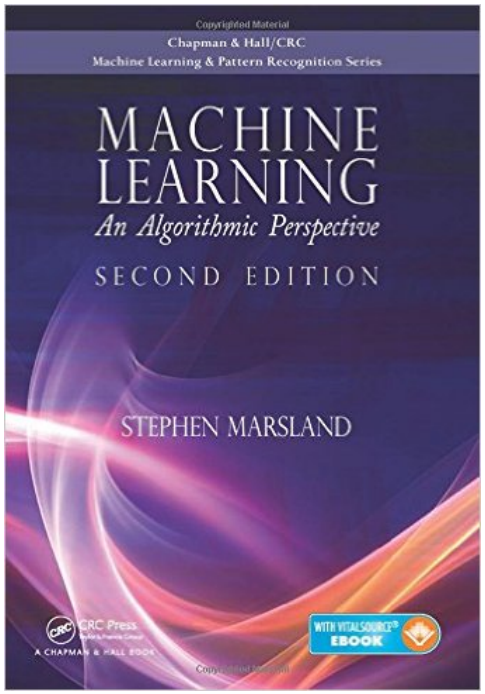
Very easy to implement 

You have to choose the number of clusters 

Subject to local minima (clusters depend on initial positions of the centroids) 

Yet, a popular first thing to try!

UNIVERSITY OF LEEDS



Chapter 14 (intro)
Chapter 14.1