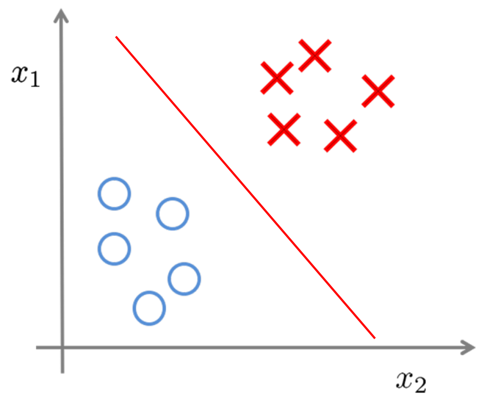
Machine Learning – Lab5

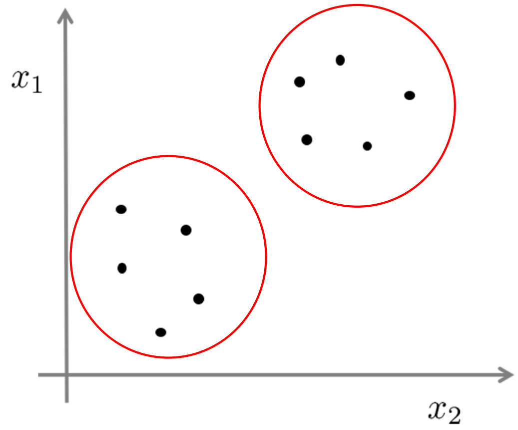
1. Unsupervised learning

Supervised vs Unsupervised Learning



Training set:

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \ldots, (x^{(m)}, y^{(m)})\}$$



Clustering
(but other types of unsupervised learning exist

Training set:

$$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(m)}\}$$

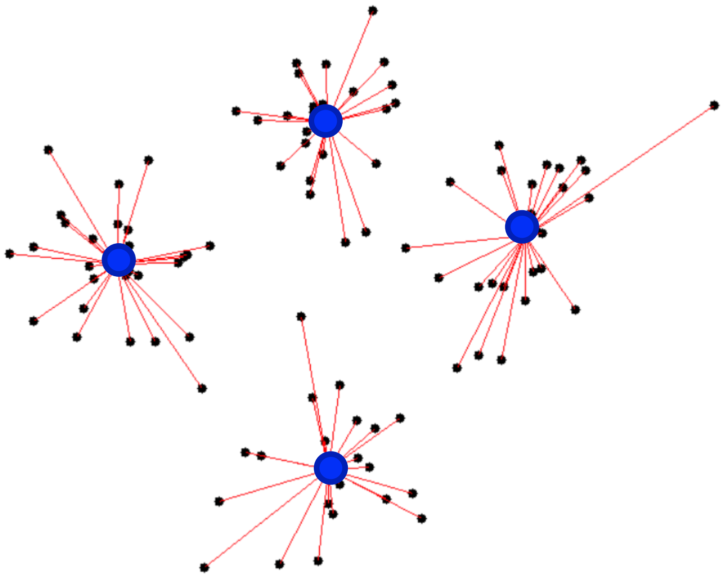
Find structure in the data...

K-Means Algorithm

Number of clusters: K

Training set: $\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(m)}\}$

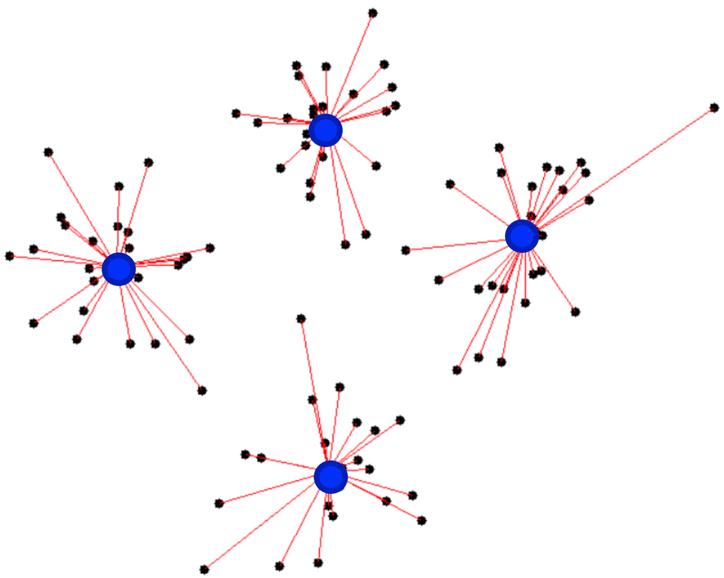
Centroids $\mu_1, \mu_2, \ldots, \mu_K$



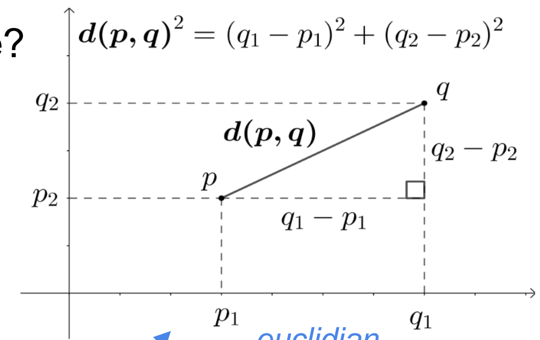
1. Define (identify) number of clusters
2. Find position of clusters (centroids) to minimize the average **distance** to cluster sets

http://shabal.in/visuals/kmeans

K-Means Algorithm



What is a distance?



euclidian distance

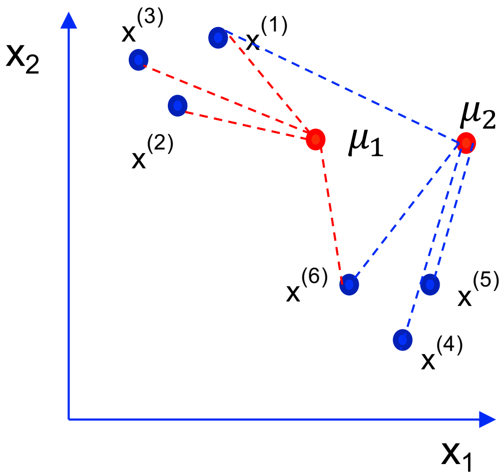
also called norm: $\| p(x_1, x_2) - q(x_1, x_2) \|^2$

n-dimensional norm $\| p(x_1, \ldots, x_n) - q(x_1, \ldots, x_n) \|^2$

$$= \sum_i (p(x_i) - q(x_i))^2$$

example

1) Find clusters

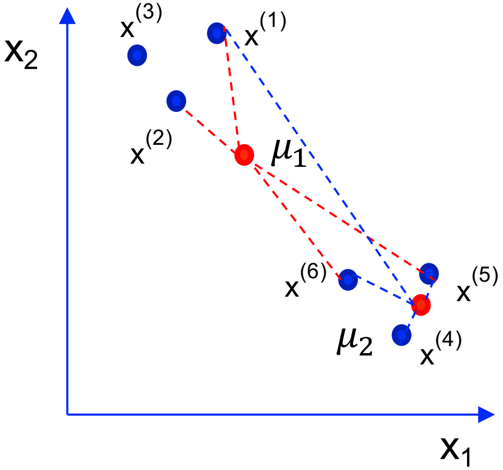


$$\begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \\ x^{(4)} \\ X^{(5)} \\ x^{(6)} \end{bmatrix} \quad \begin{array}{c} C \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} \end{array}$$

$$\min(\, \| x^{(i)} - \mu 1 \|, \| x^{(i)} - \mu 1 \| \,)$$

example

2) Move centroids



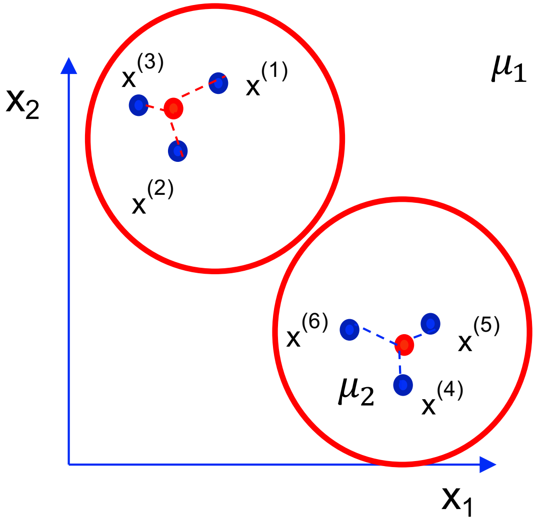
$$\begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \\ x^{(4)} \\ X^{(5)} \\ x^{(6)} \end{bmatrix} \quad \begin{array}{c} \text{C} \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} \end{array}$$

$$\mu_1 = \frac{1}{k1} \sum_{k1} x^{(k)}$$

$$\mu_2 = \frac{1}{k2} \sum_{k2} x^{(k)}$$

example

Find new clusters

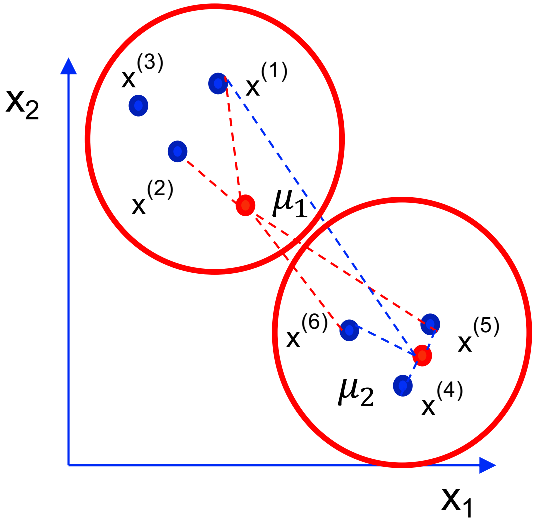


$$\begin{array}{cc} & C \\ \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \\ x^{(4)} \\ X^{(5)} \\ x^{(6)} \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \end{bmatrix} \end{array}$$

$\min(\| x^{(i)} - \mu 1 \|, \| x(^{i)} - \mu 1 \|)$

example

Move centroids



$$\begin{array}{cc} & C \\ \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \\ x^{(4)} \\ X^{(5)} \\ x^{(6)} \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \end{bmatrix} \end{array}$$

$$\mu_1 = \frac{1}{k1} \sum_{k1} x^{(k)}$$

$$\mu_2 = \frac{1}{k2} \sum_{k2} x^{(k)}$$

K-Means Algorithm

Randomly initialize K cluster centroids $\mu_1, \mu_2, \ldots, \mu_K \in \mathbb{R}^n$

Repeat {

for i = 1 to m

$c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$

for k = 1 to K

μ_k := average (mean) of points assigned to cluster k

}

Cluster assignment step, minimise J with respect to $c^{(i)}$ *while holding* μ_k

$$\min_{\substack{c^{(1)},\ldots,c^{(m)},\\ \mu_1,\ldots,\mu_K}} J$$

Move centroid step, minimise J with respect to μ_k *while holding* $c^{(i)}$

Optimization Objective

What is the cost function that K-Means is minimizing?

$c^{(i)}$ = index of cluster (1,2,…,K) to which example $x^{(i)}$ is currently assigned

μ_k = cluster centroid k ($\mu_k \in \mathbb{R}^n$)

$\mu_{c^{(i)}}$ = cluster centroid of cluster to which example $x^{(i)}$ has been assigned

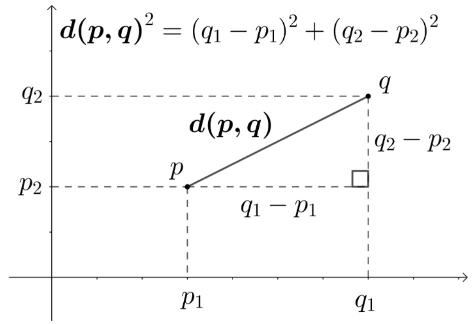
Optimization objective:

$$J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K) = \frac{1}{m} \sum_{i=1}^{m} ||x^{(i)} - \mu_{c^{(i)}}||^2$$

$$\min_{\substack{c^{(1)}, \dots, c^{(m)}, \\ \mu_1, \dots, \mu_K}} J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K)$$

find the parameters that minimise the cost functions

Euclidean Distance

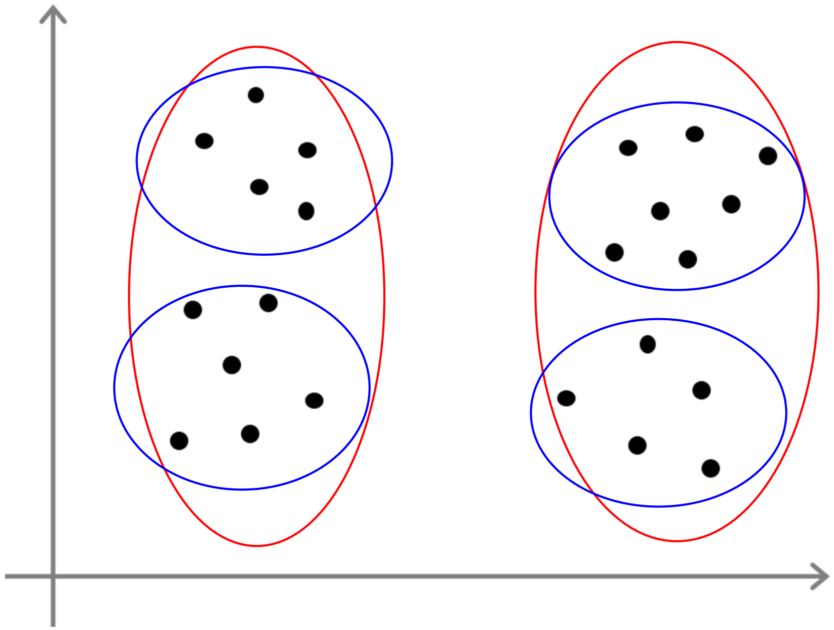


J is also called *distortion function*

https://www.coursera.org/learn/machine-learning

Choosing the number of Clusters

What is the right value of K?

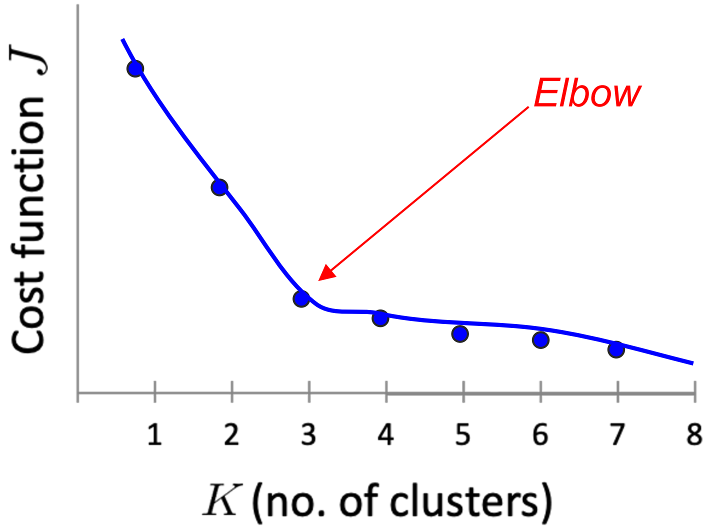


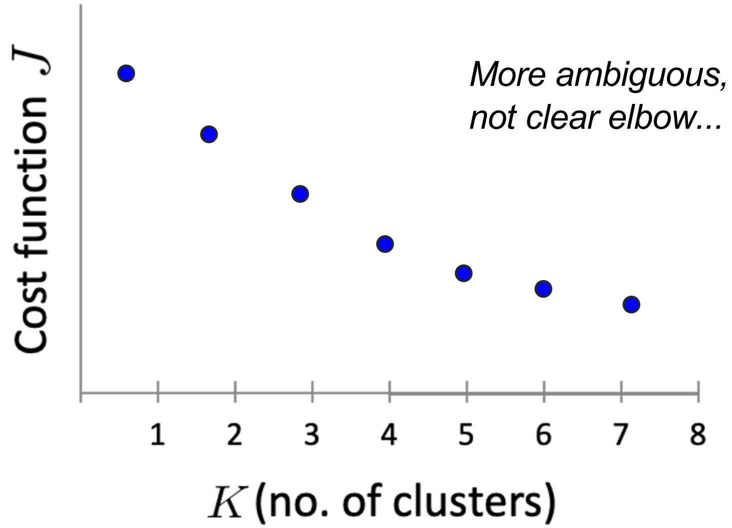
K=2 or K=4 ?

Number of cluster can be ambiguous

Choosing the number of Clusters

The Elbow Method





https://www.coursera.org/learn/machine-learning