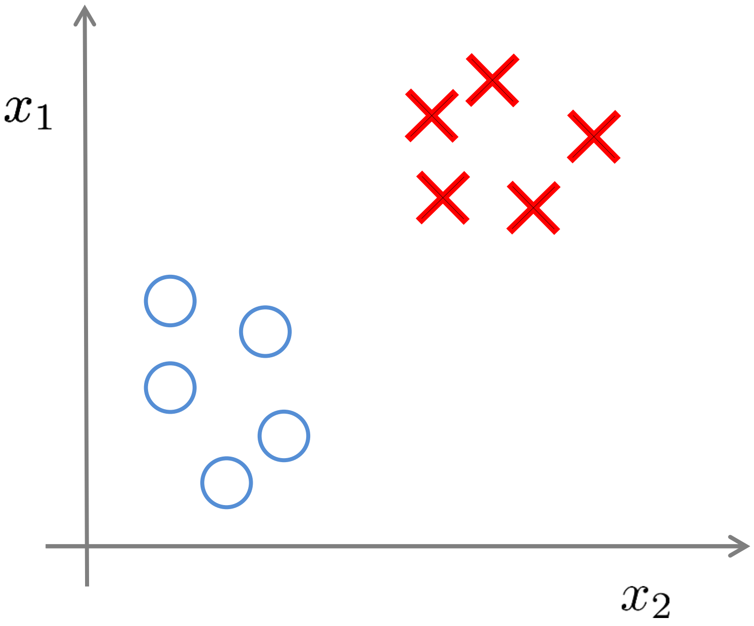
Internet Applications

Clustering

Unsupervised Learning

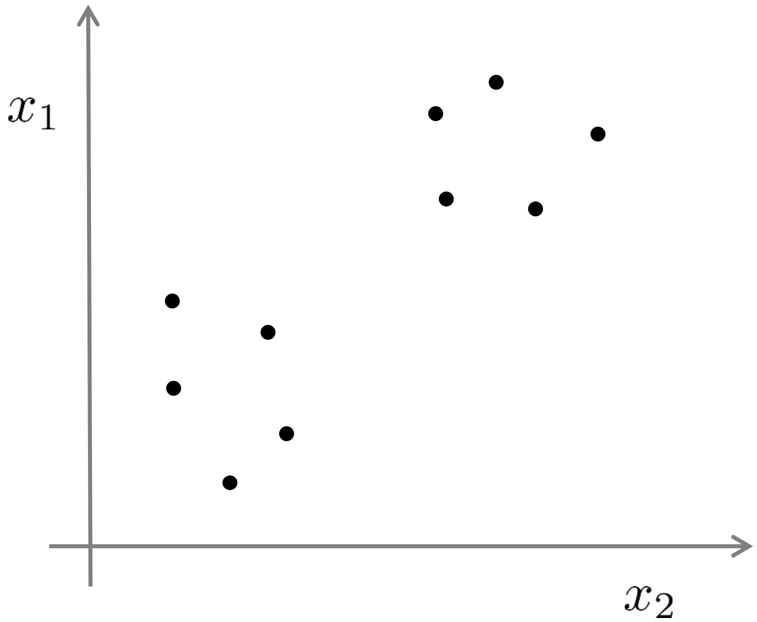
Supervised Learning



Training data:

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

Unsupervised learning

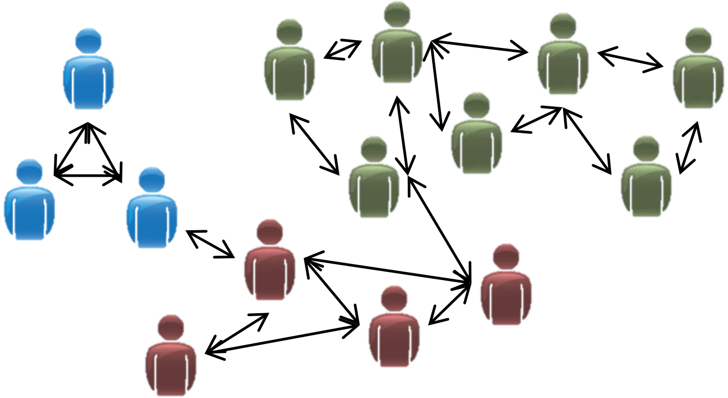


Training data:

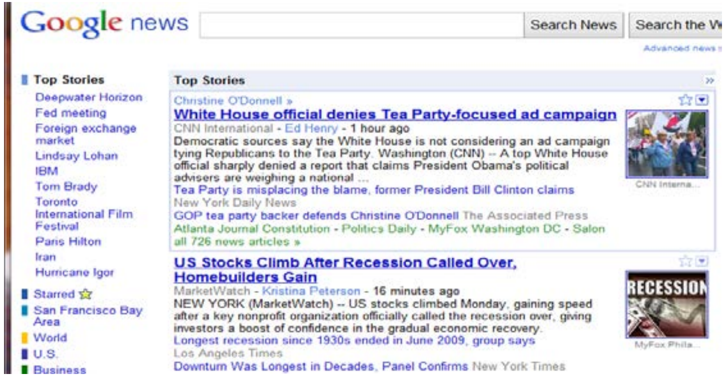
$$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(m)}\}$$

Examples of clustering applications

Market segmentation



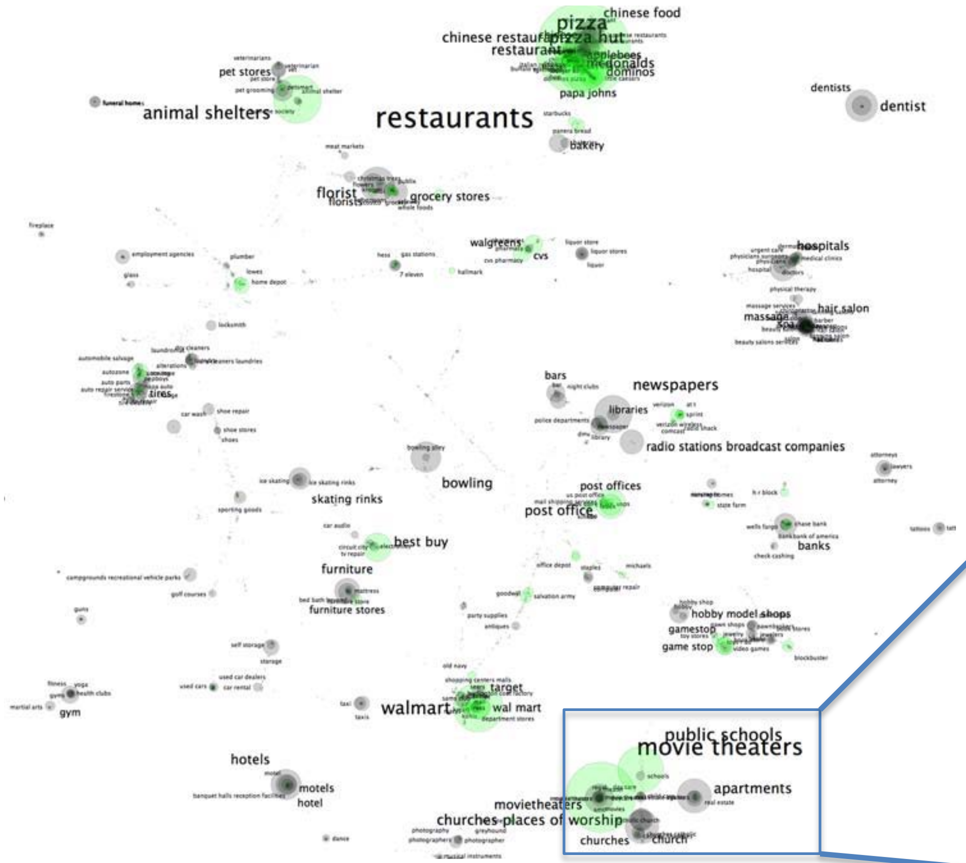
Social network analysis

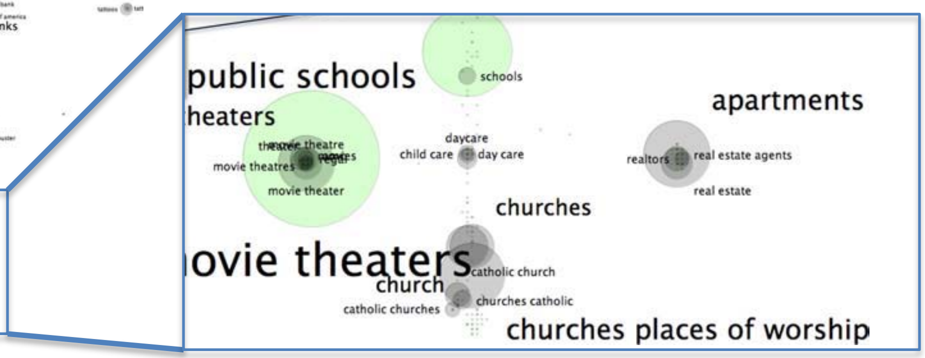


Group related elements
(News, products, websites,...)

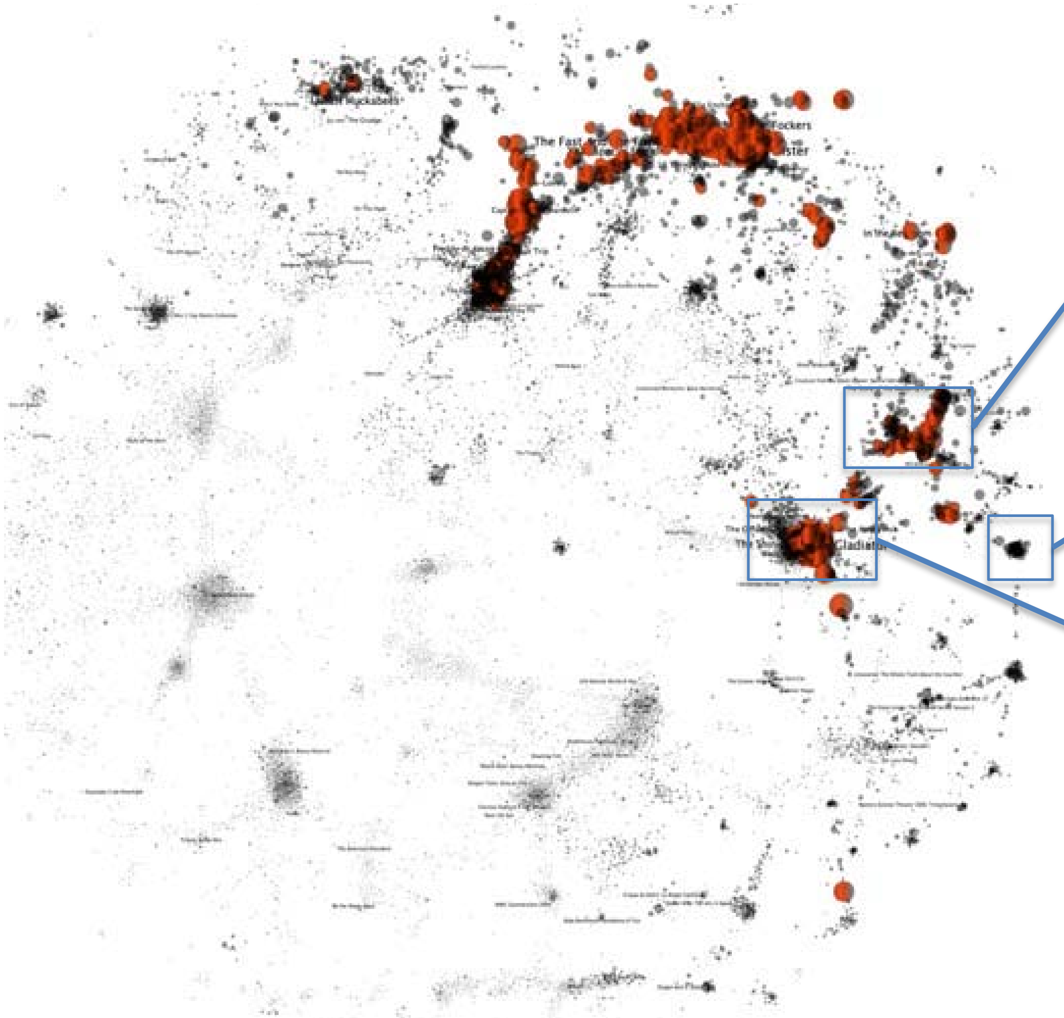
Analysis of astronomical data

Searches made on yellowpages.org organized by similarity (two searches are similar if, after the user has made it, she clicks on the same business category with high probability)





http://abeautifulwww.com



Bridget Jones Diary
When Harry Met Sally

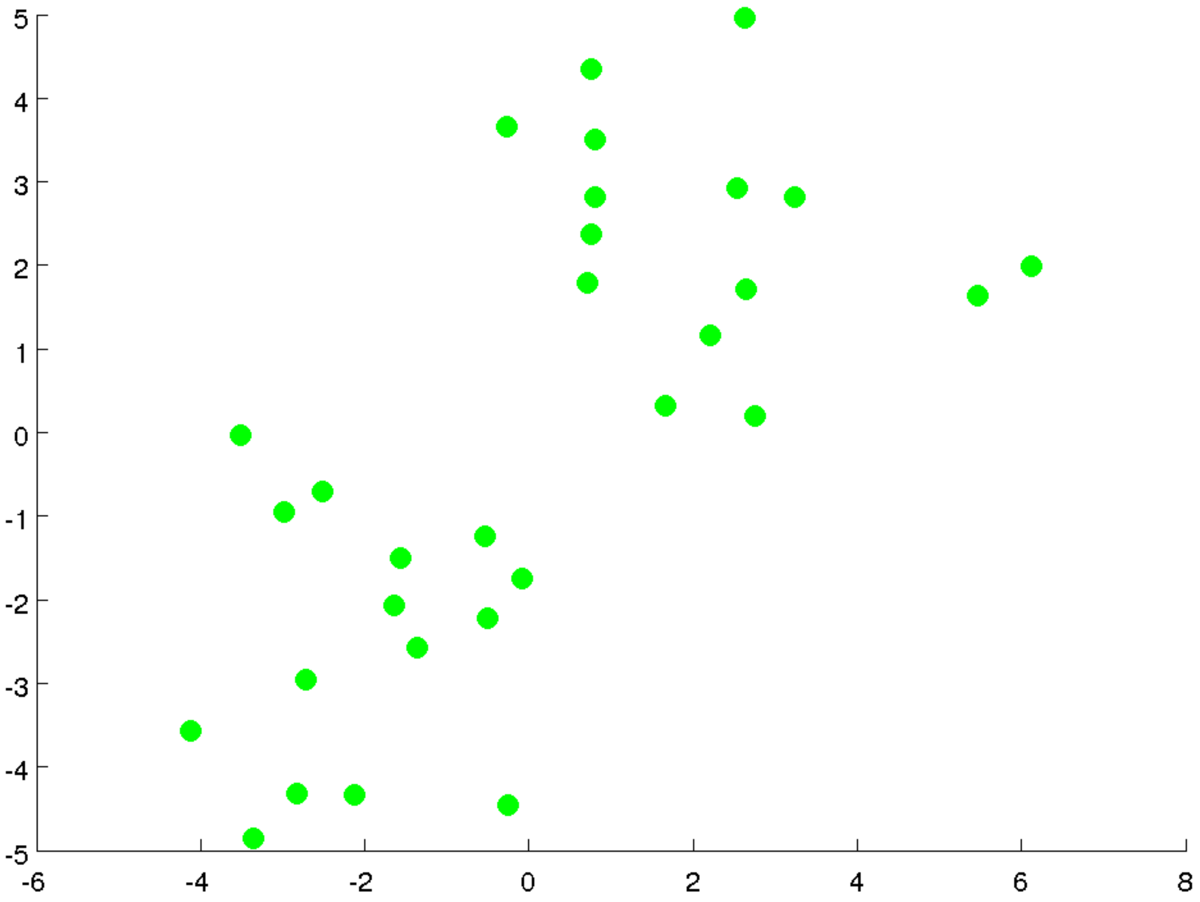
Star Trek Movies
Dune
X Files

Raiders of the Lost Ark
The Big Lebowsky

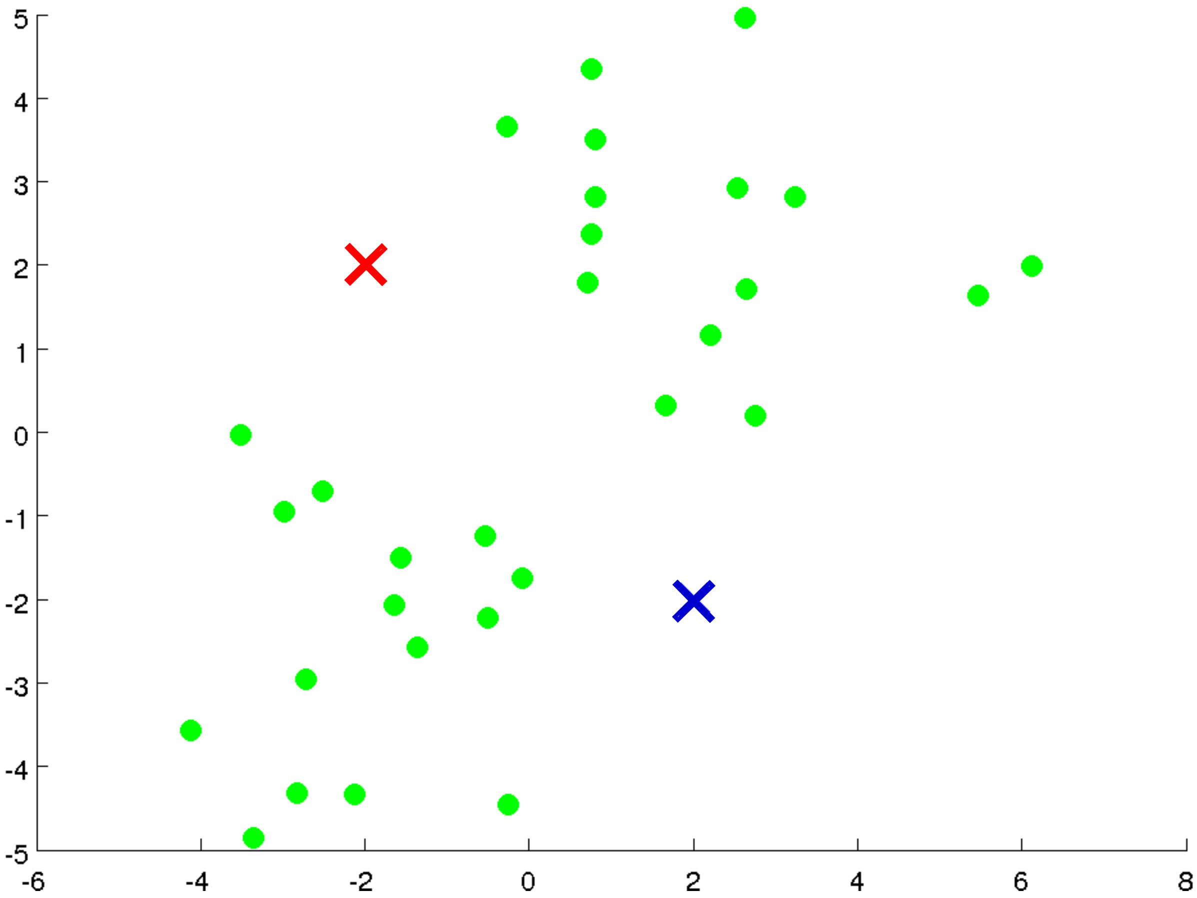
Clustering

K-means algorithm

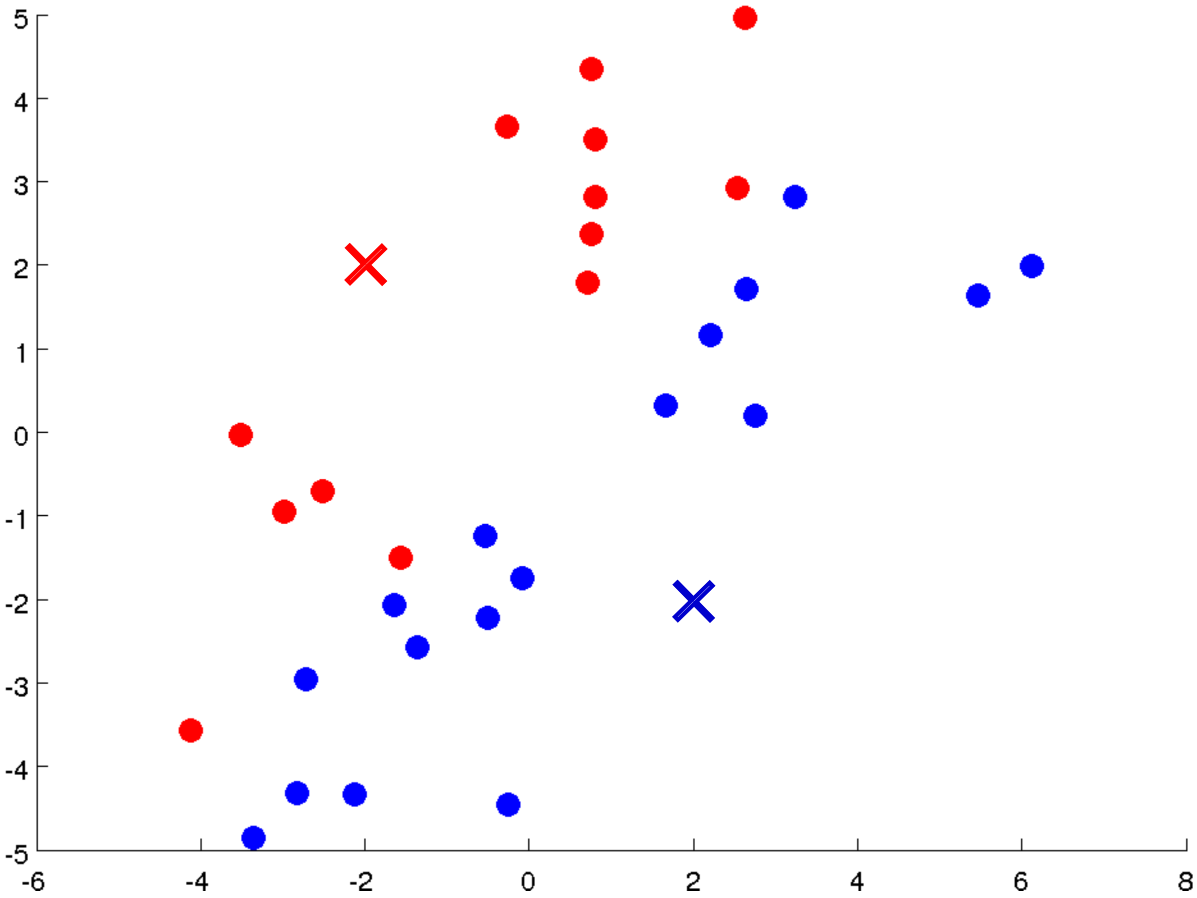
Aplicaciones en Internet



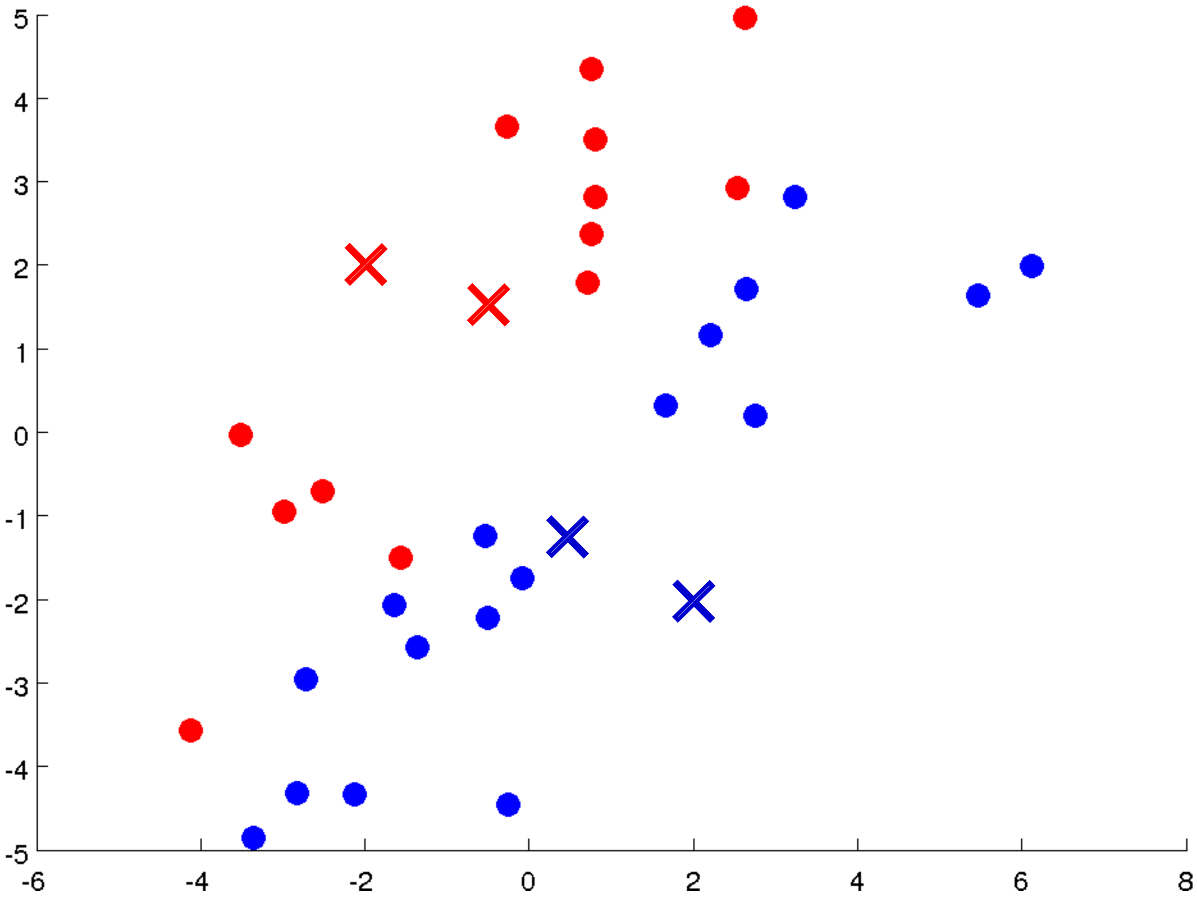
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8



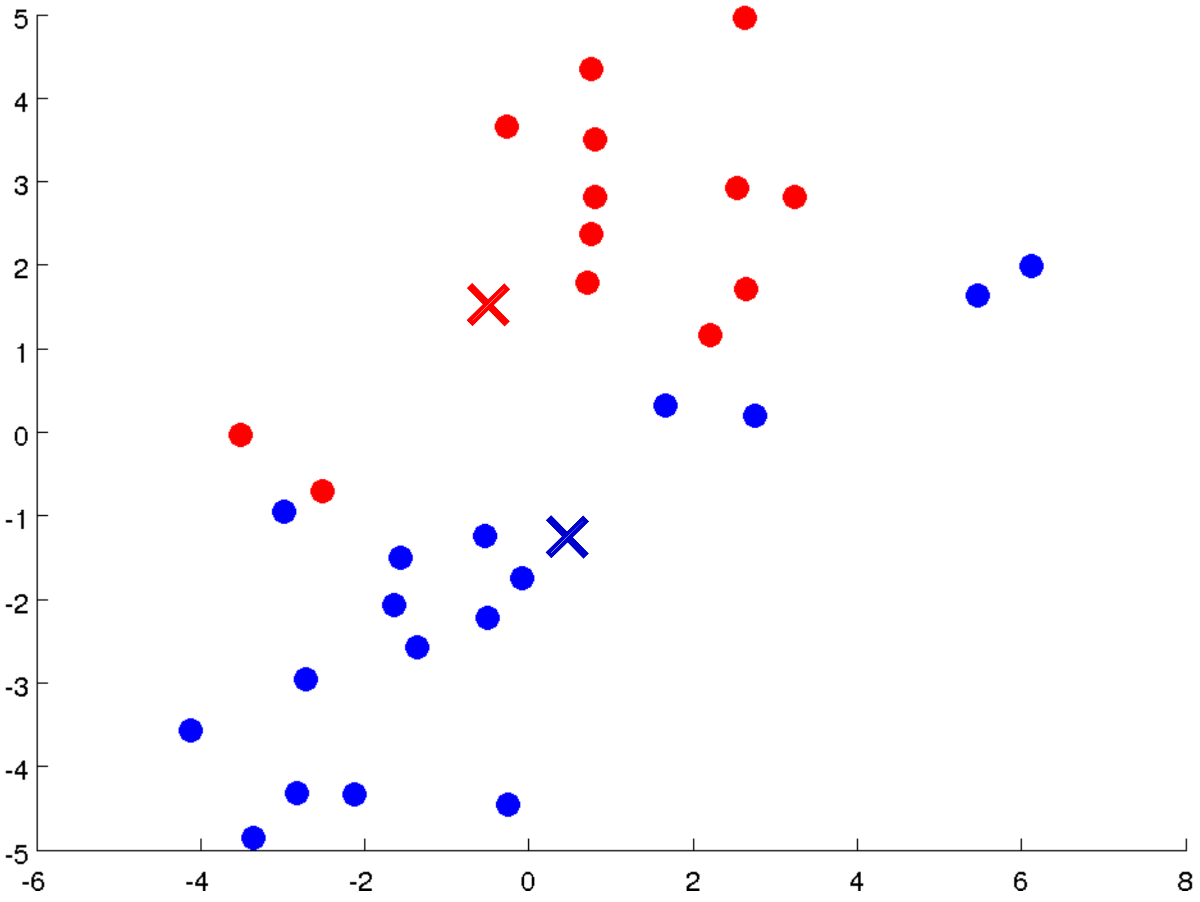
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

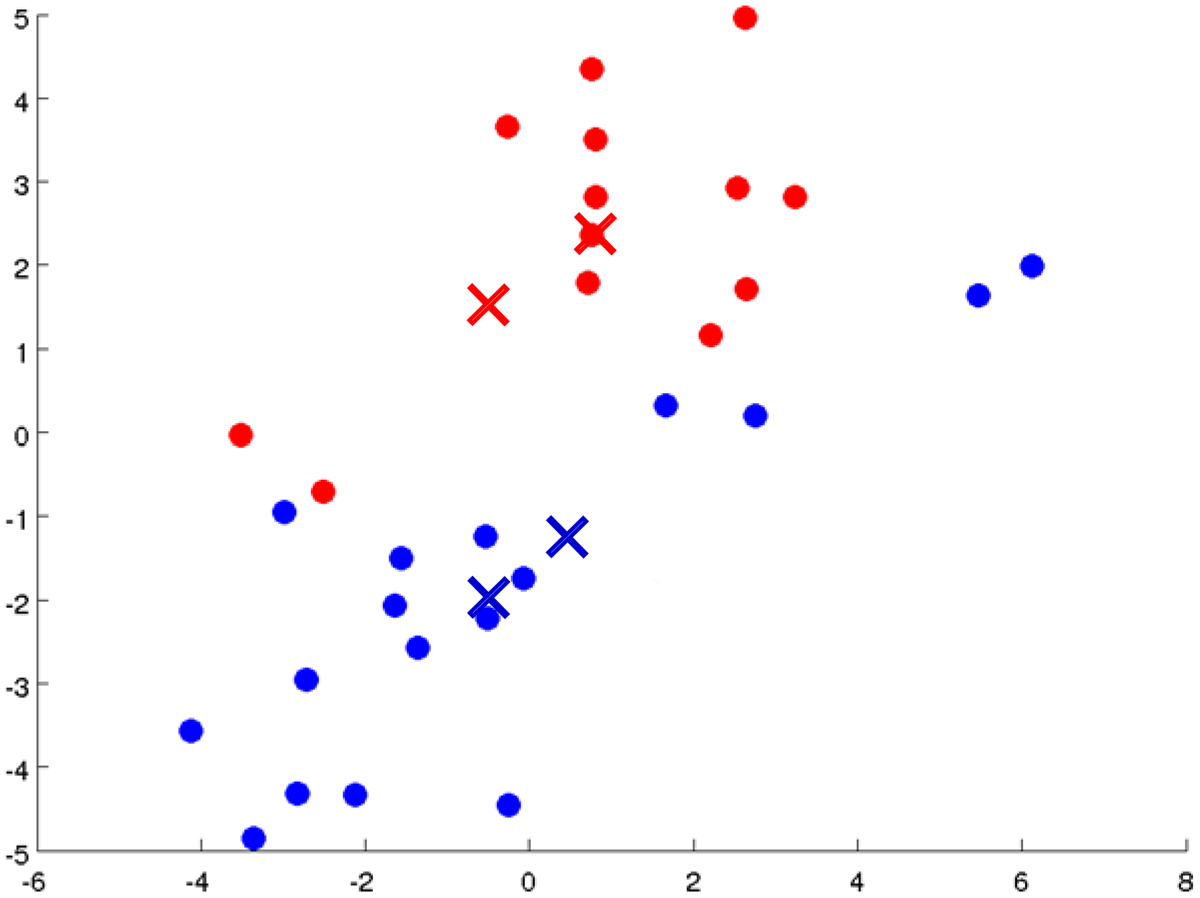


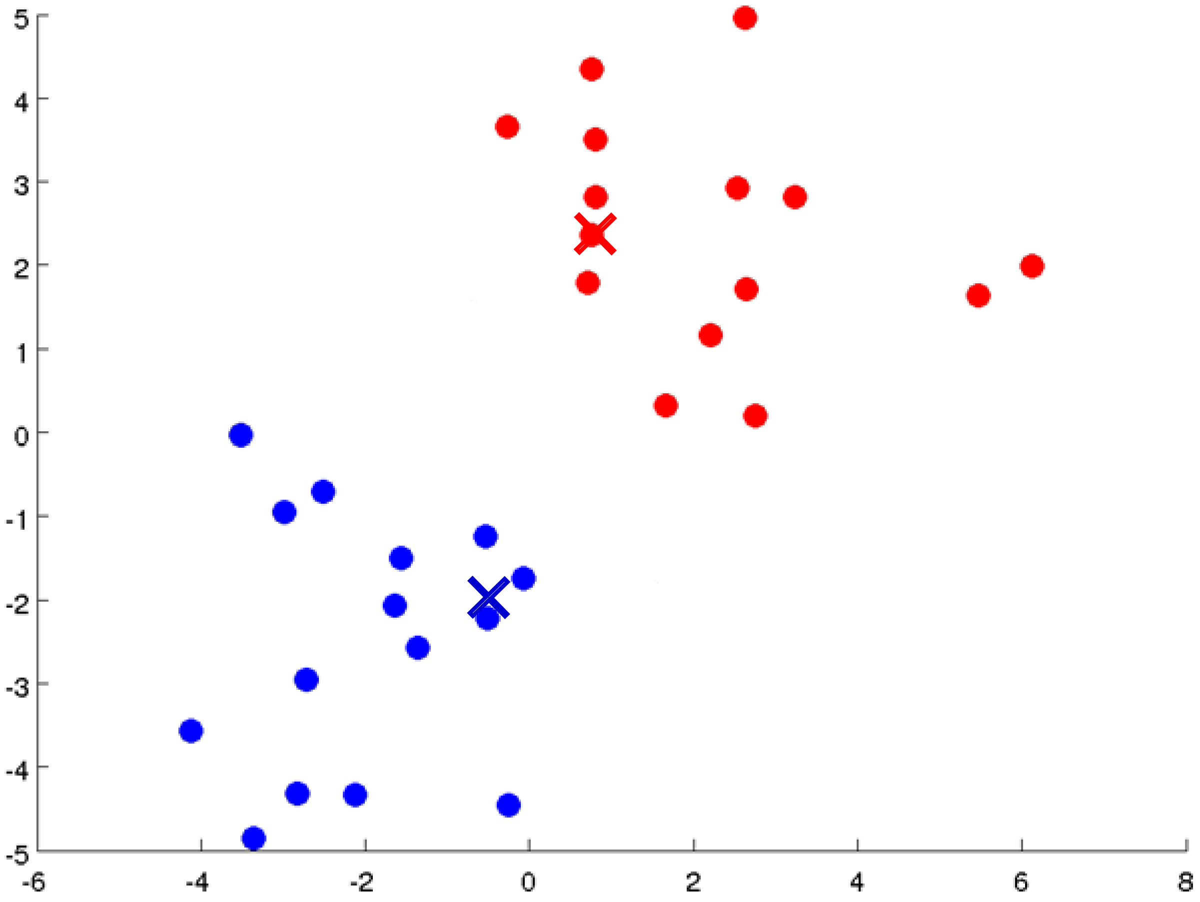
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

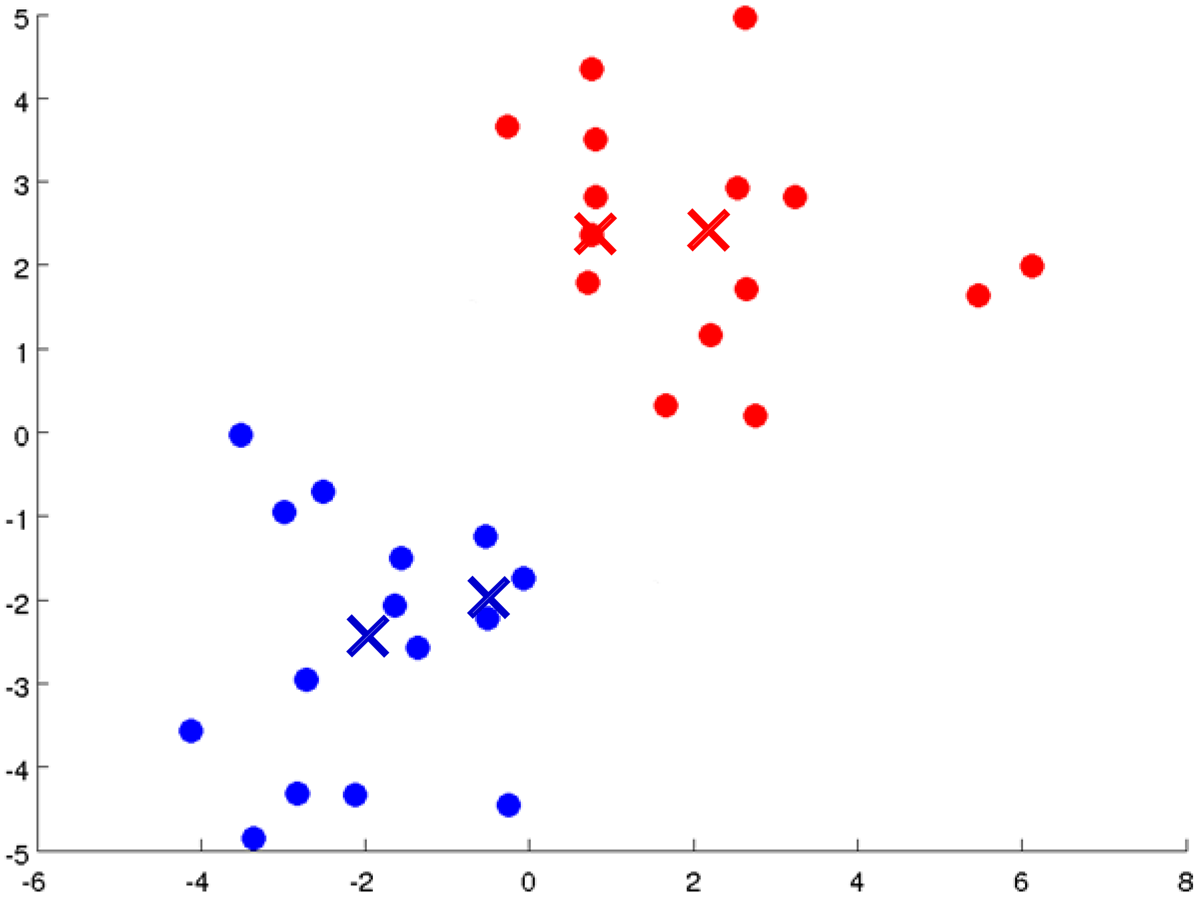


5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

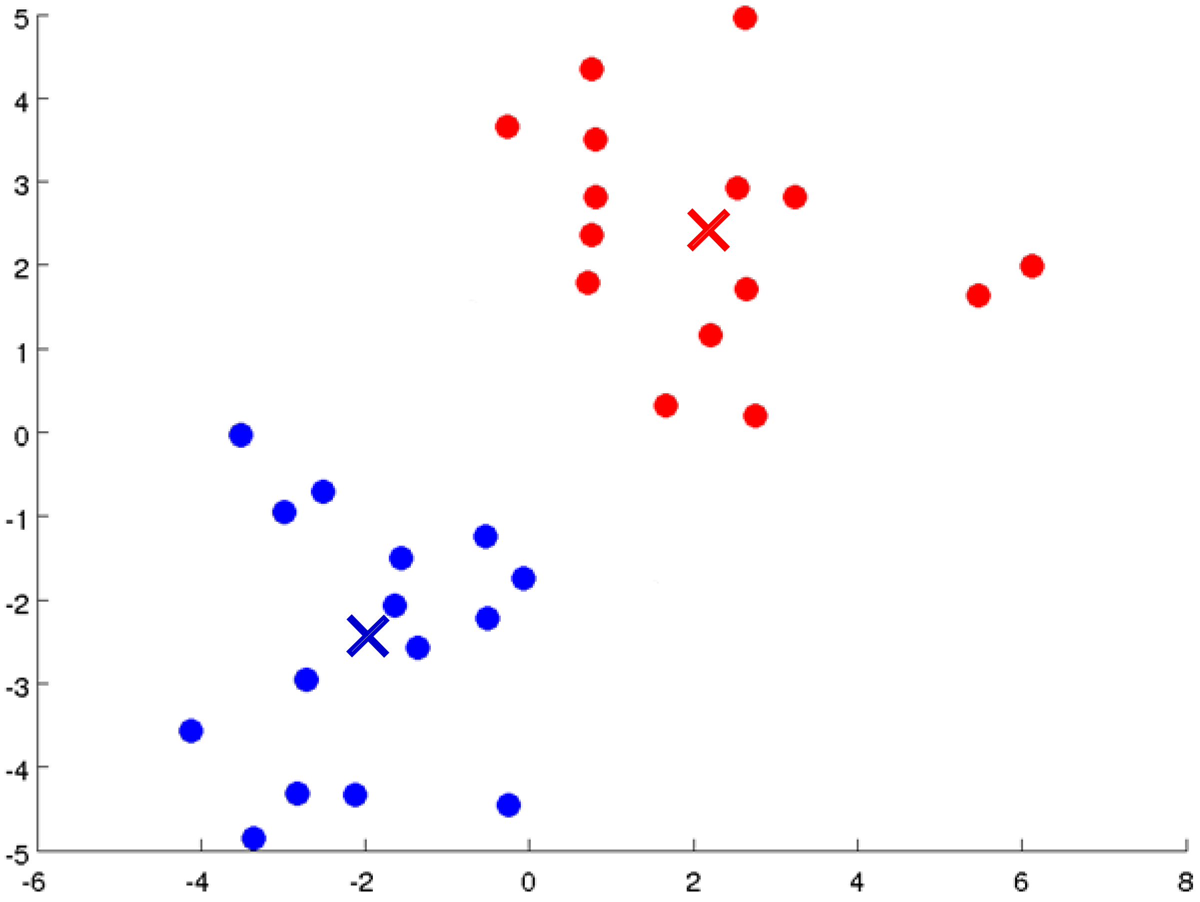

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8


5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8


5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8



5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8



5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

K-means algorithm

Input:

- K (number of clusters)
- Training set $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

$x^{(i)} \in \mathbb{R}^n$ (drop $x_0 = 1$ convention)

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 $x^{(i)}$

end for

for k = 1 to K

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

end for

}

K-means algorithm

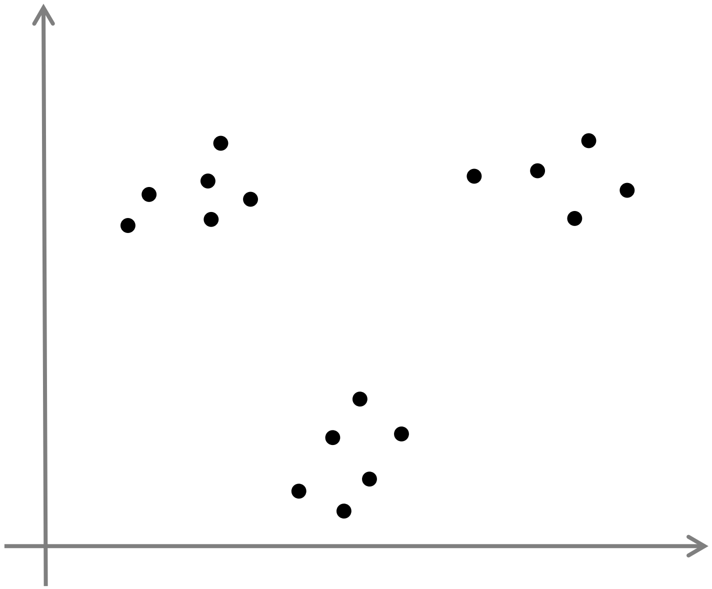
Example of index allocation

$$c^{(i)} := \arg\min_k \left\| x^{(i)} - \mu_k \right\|^2$$

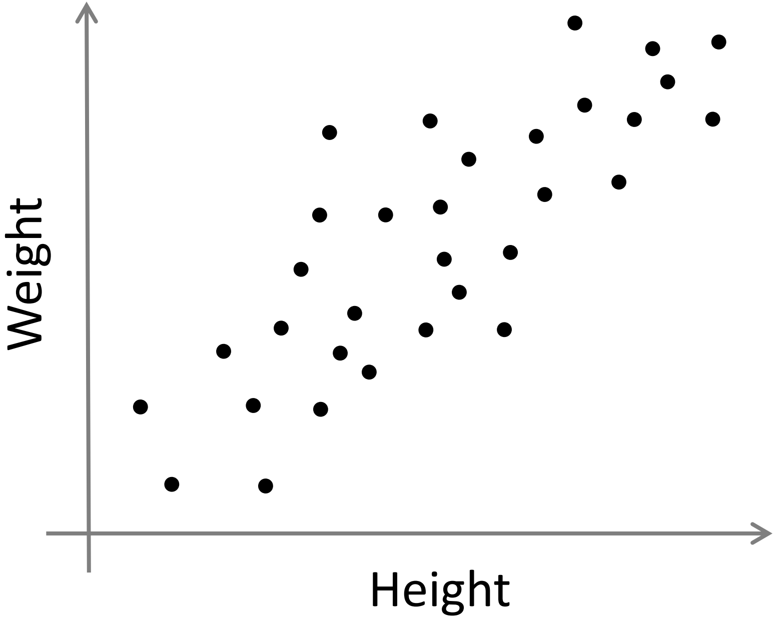
Example μ_k computation

$$\mu_k := \frac{1}{|C_k|} \sum_{i \in C_k} x^{(i)}$$

K-means for non-separated clusters



T-shirt sizing



K-means optimization objective

$c^{(i)}$ = 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:

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

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

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

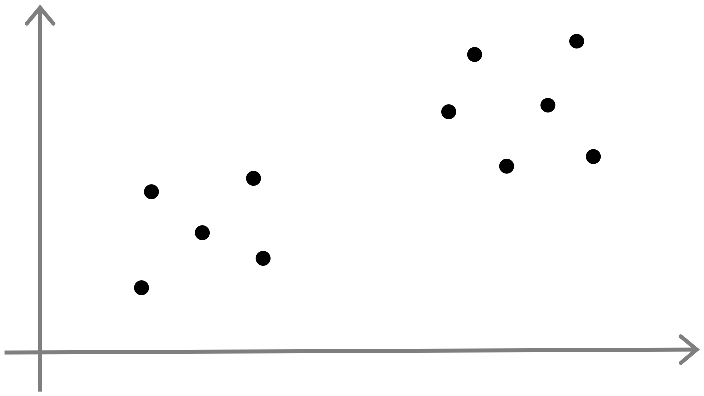
}

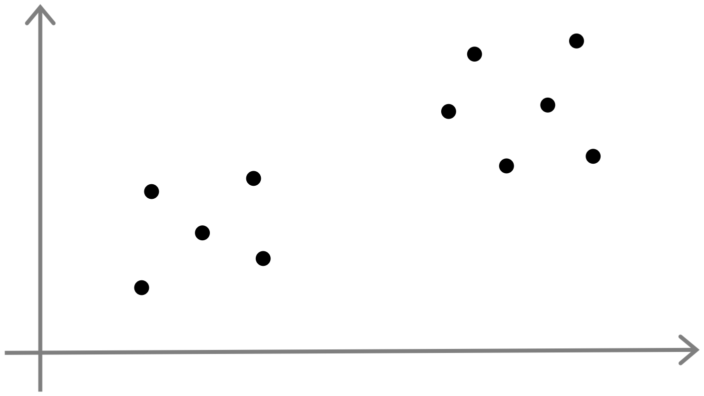
Random initialization

Should have $K < m$

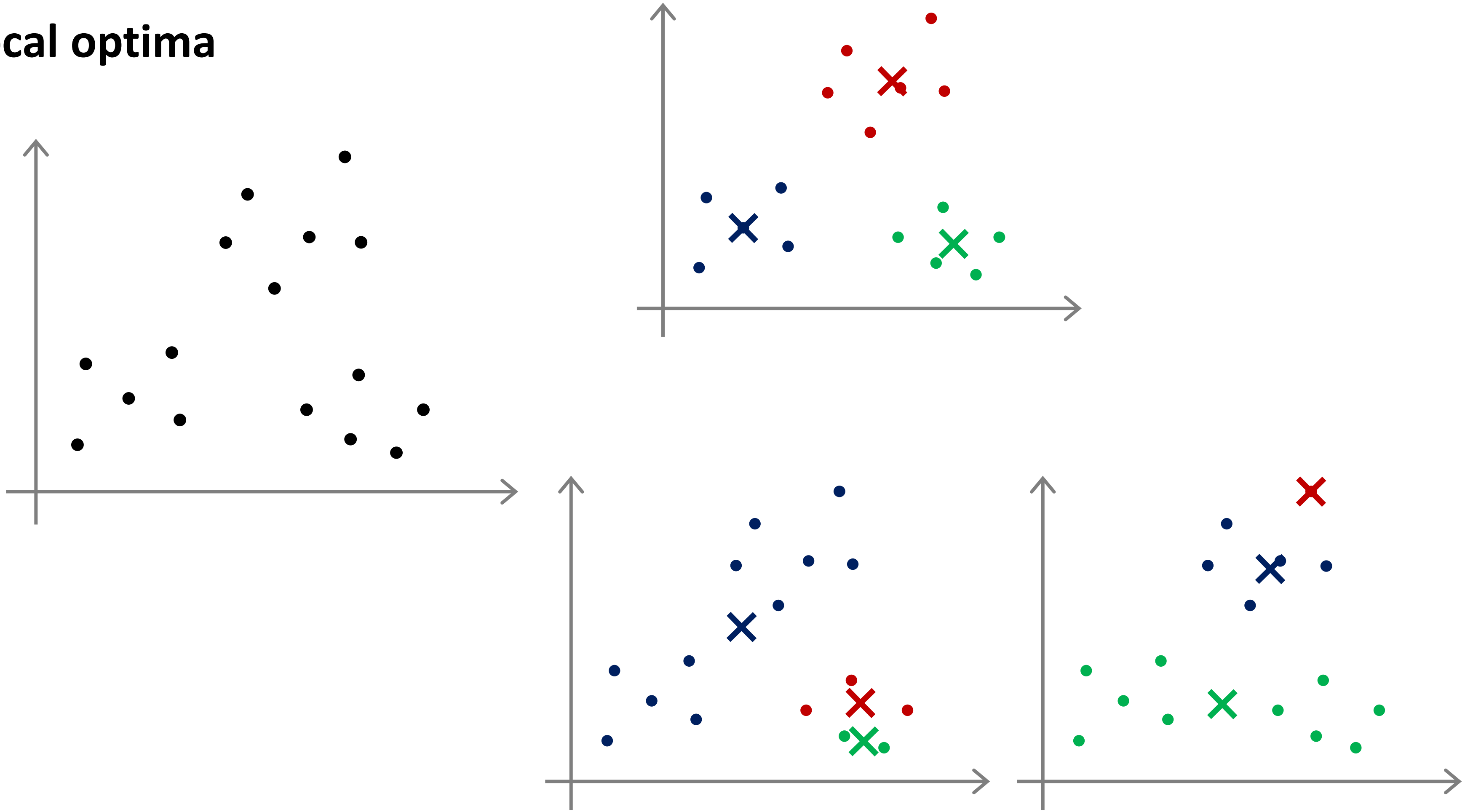
Randomly pick K training examples.

Set $\mu_1, \ldots, \mu_K$ equal to these K examples.





Local optima



Random initialization

For i = 1 to 100 {

Randomly initialize K-means.

Run K-means. Get $c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K$.

Compute cost function (distortion)

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

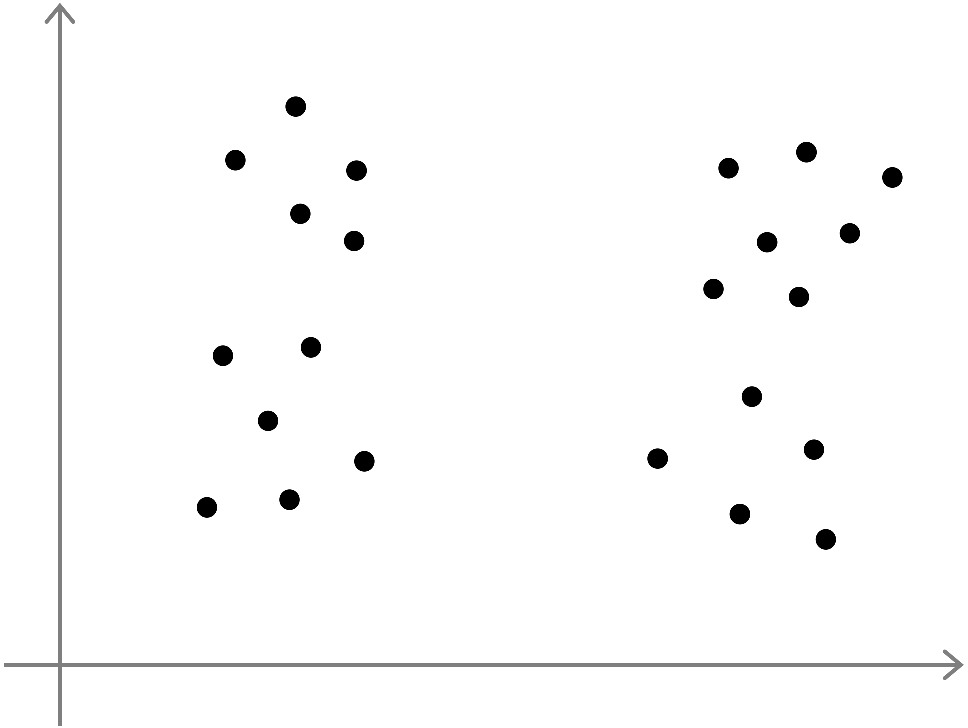
}

Pick clustering that gave lowest cost $J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K)$

Which of the following methods is suitable for initializing k-means?

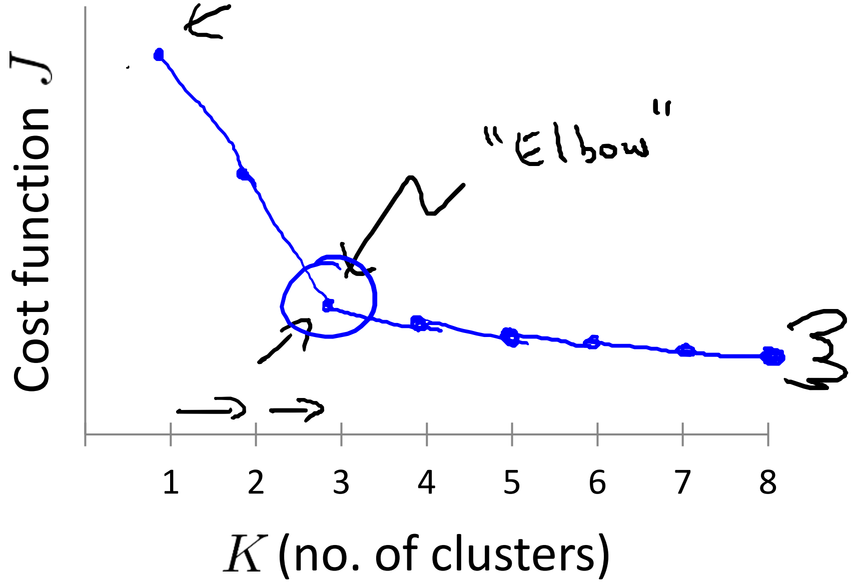
- ❑ Choose an integer i randomly from {1,…,K} and set $\mu_1=\mu_2=\ldots\mu_K=x^{(i)}$
- ❑ Choose K integers, i1, i2, ..., iK randomly from {1,…,K} and set $\mu_1=x^{(i1)}$, $\mu_2=x^{(i2)}$,…, $\mu_K=x^{(iK)}$.
- ❑ Choose K integers, i1, i2, ..., iK randomly from {1,…,m} and set $\mu_1=x^{(i1)}$, $\mu_2=x^{(i2)}$,…, $\mu_K=x^{(iK)}$.
- ❑ Choose m integers, i1, i2, ..., im randomly from {1,…,m} and set $\mu_1=x^{(i1)}$, $\mu_2=x^{(i2)}$,…, $\mu_K=x^{(im)}$.

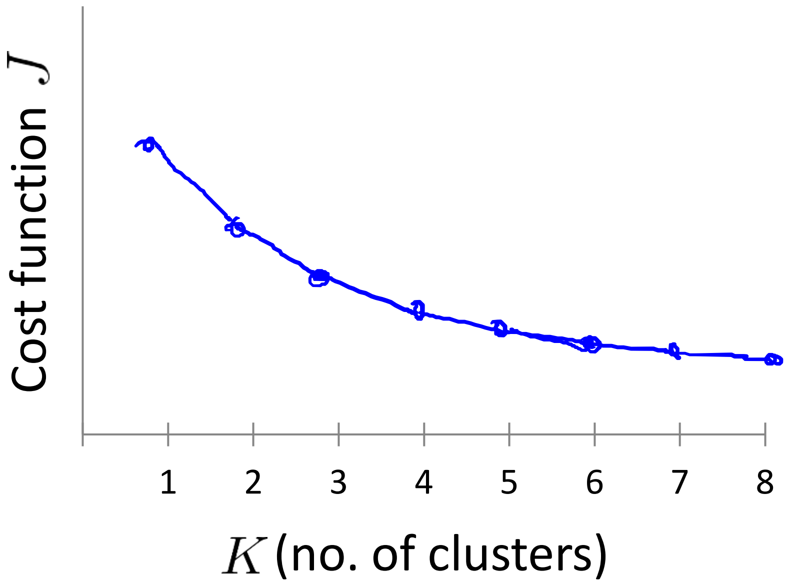
What is the right value of K?



Choosing the value of K

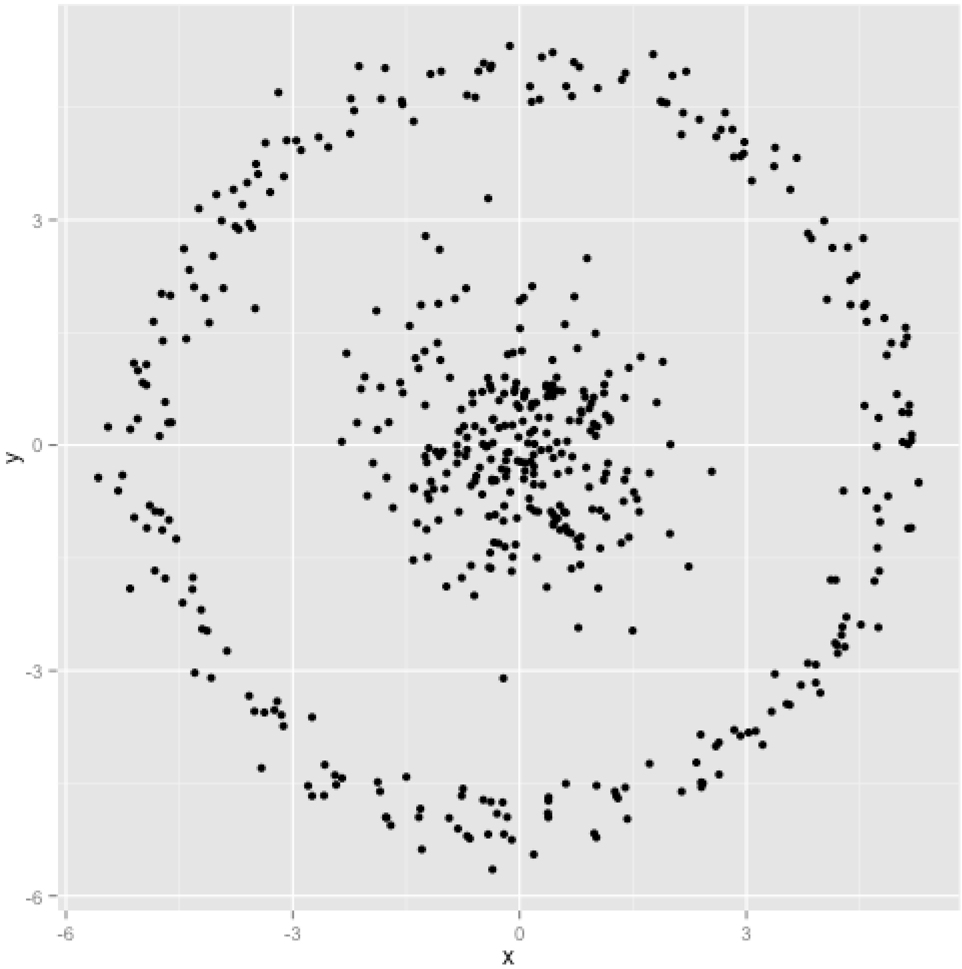
Elbow method:



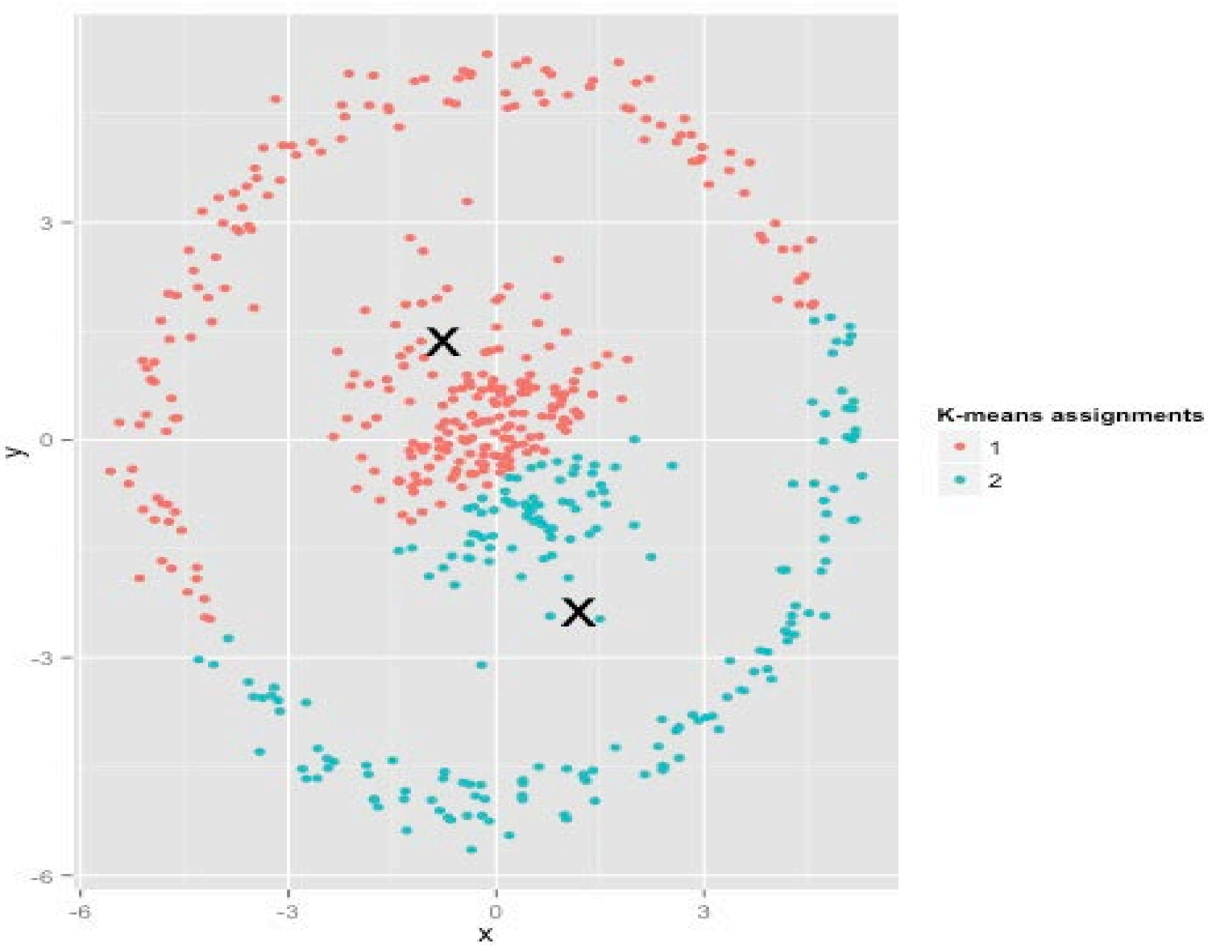


Suppose we run k-means using K=3 and K=5 and check that the cost function J is greater for K=5 than for K=3. What can we deduce from this?

- ❑ This is mathematically impossible. There must be a mistake in the code.
- ❑ That the correct number of clusters is K=3.
- ❑ In the execution with K=5, k-means remained at a non-optimal local minimum. It is convenient to re-run with a larger number of random initials.
- ❑ In the execution with K=3, k-means was lucky. It is advisable to increase the random initializations for K=3 until it is no longer better than K=5.

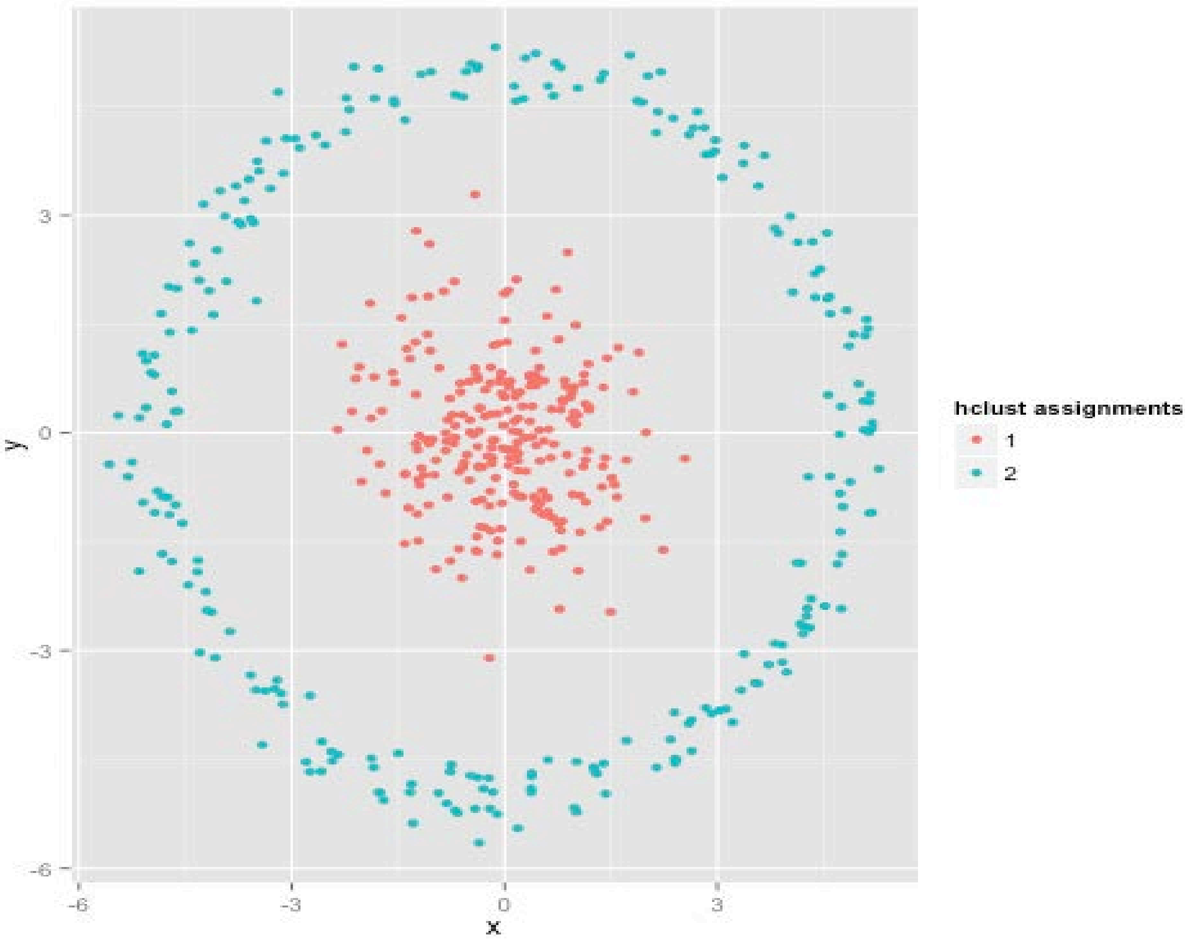


3
0
-3
-6
y
-6
-3
0
3
x

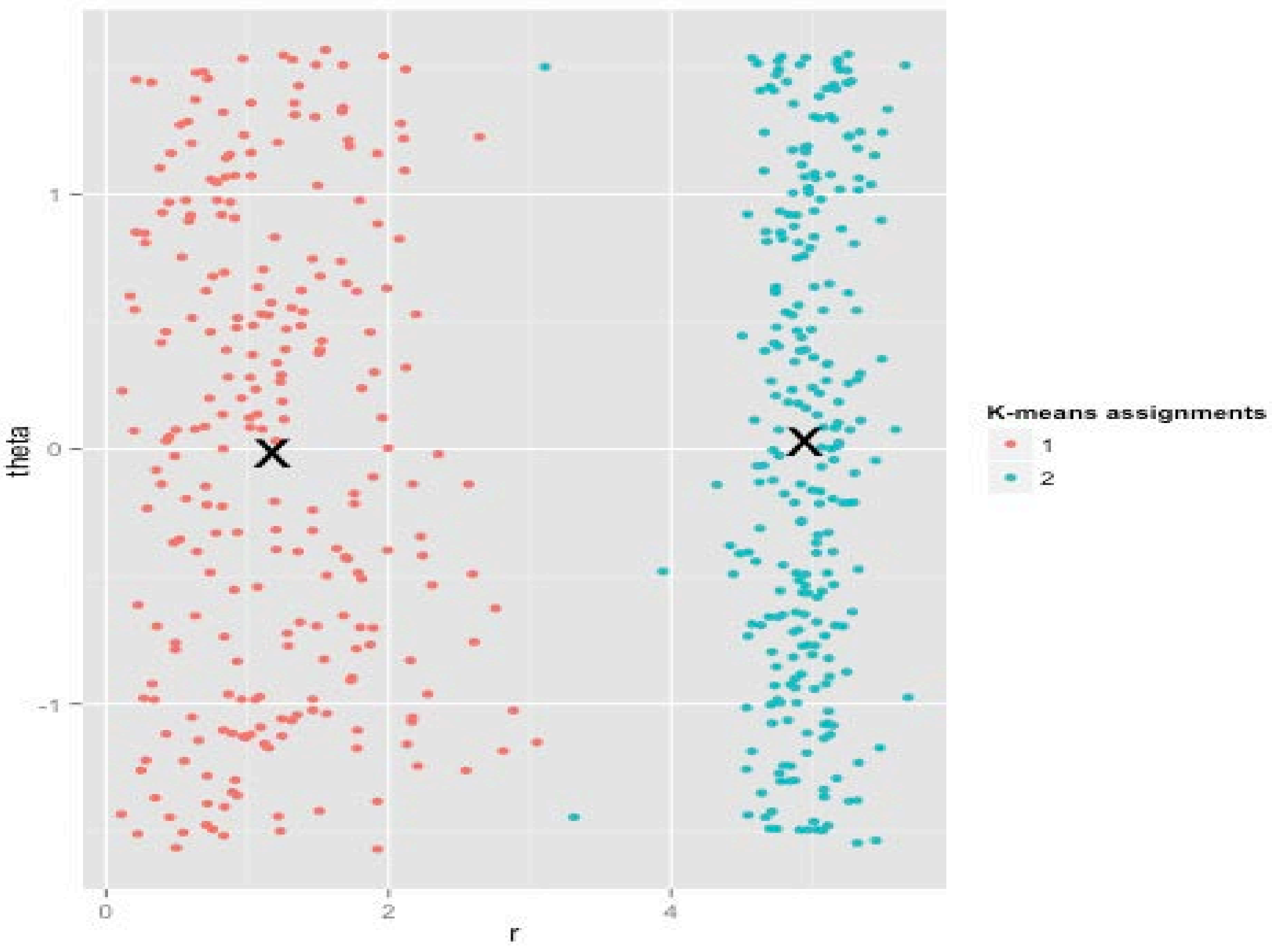


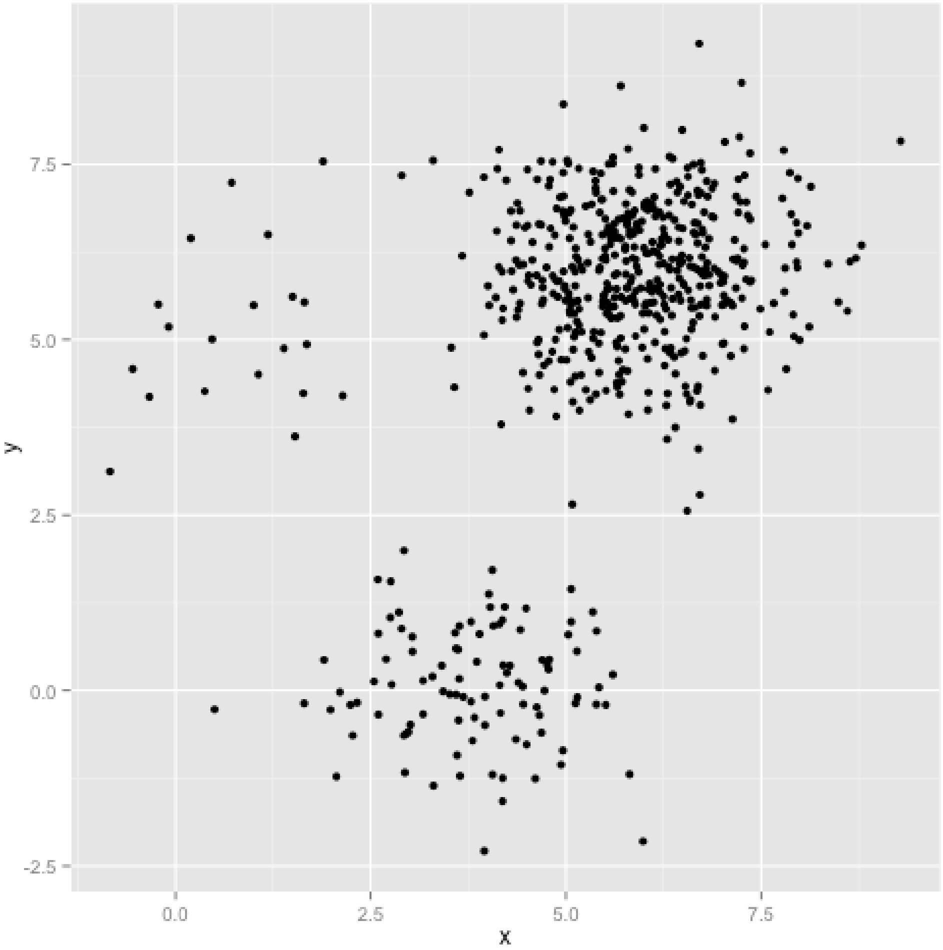
K-means assignments
1
2
y
x
3
0
-3
-6

Single-linkage hierachical clustering

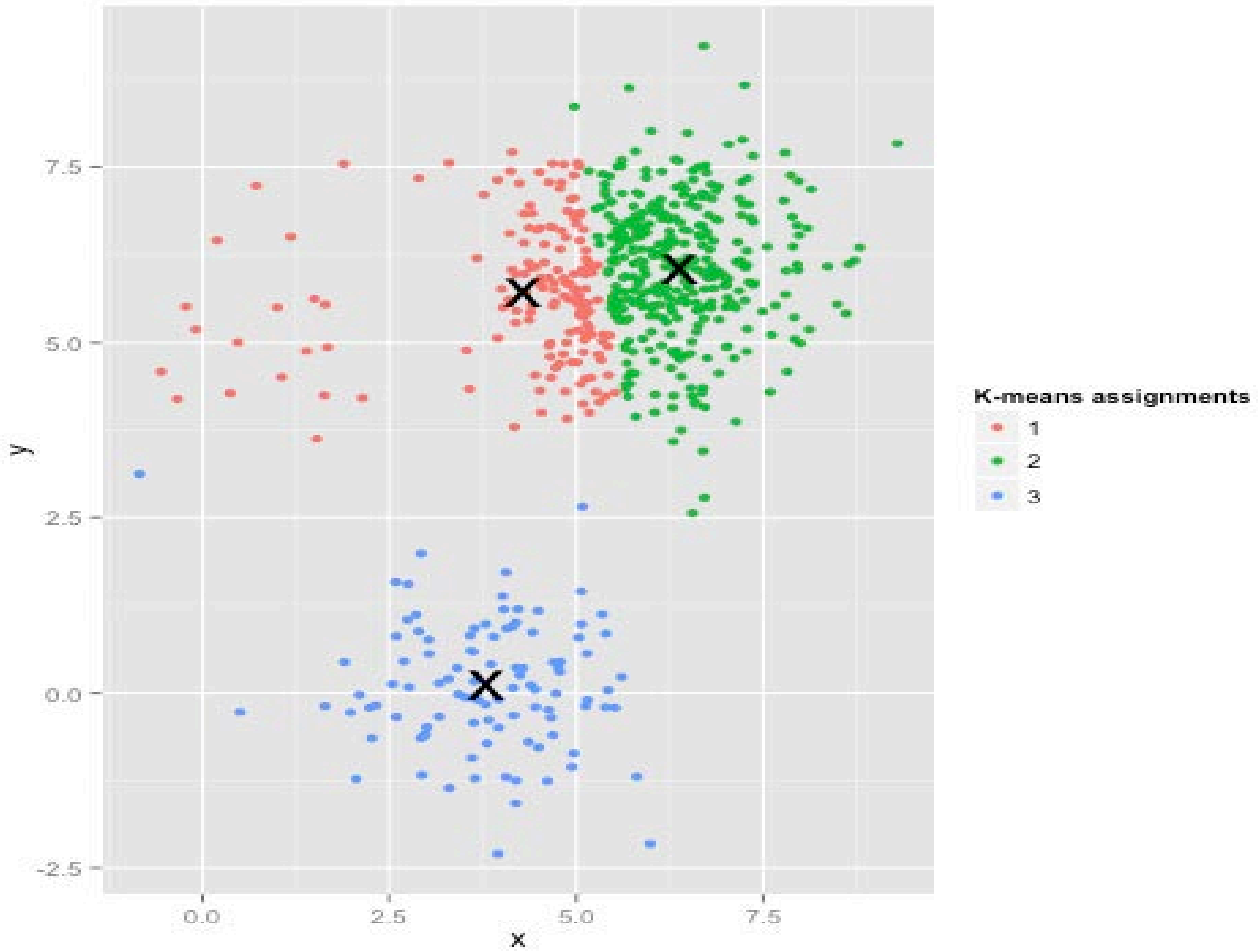


K-means using polar coordinates





7.5
5.0
2.5
0.0
-2.5
y
0.0
2.5
5.0
7.5
x



K-means assignments
1
2
3
7.5
5.0
2.5
0.0
-2.5
0.0
2.5
5.0
7.5
x
y