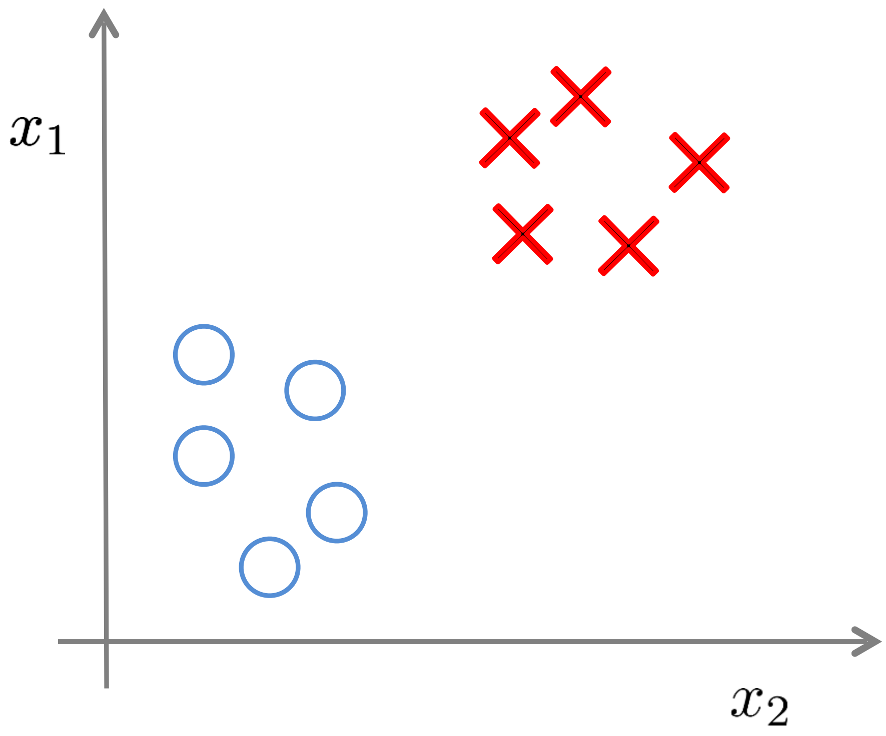
Applied
Machine Learning

Lecture 13
Unsupervised Learning
(Clustering)

Ekarat Rattagan, Ph.D.

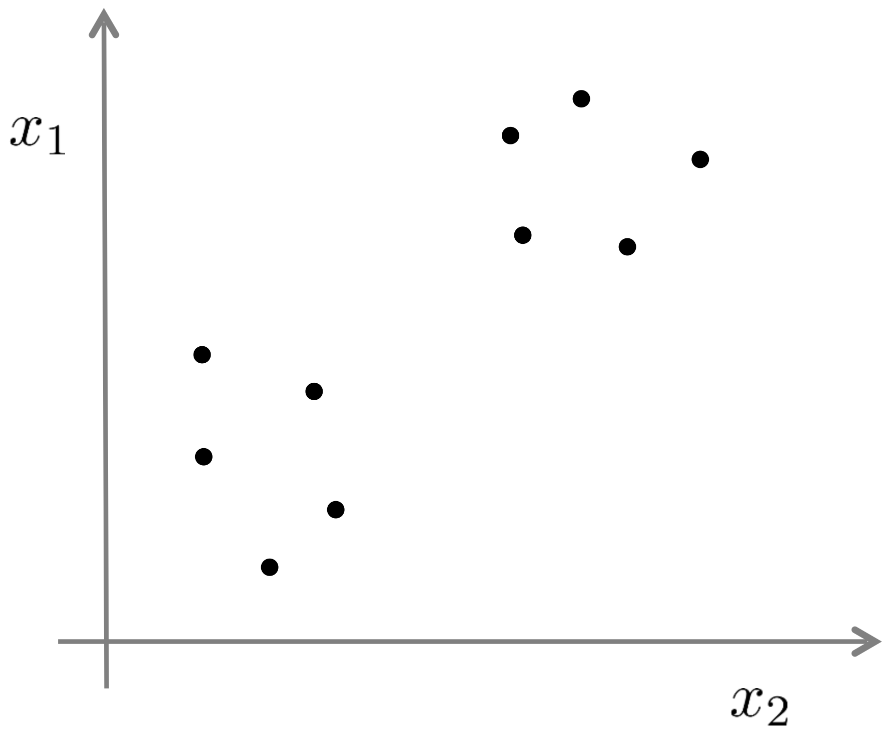
Slides adapted from Andrew NG

Supervised learning



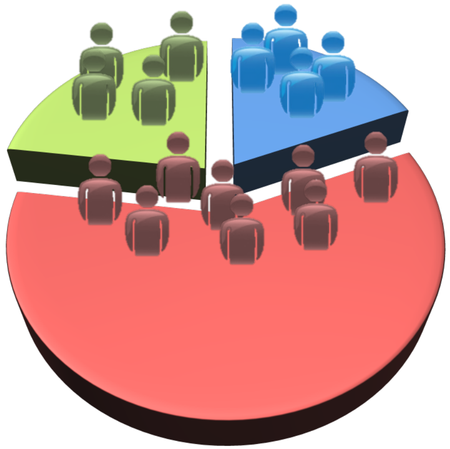
Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \ldots, (x^{(m)}, y^{(m)})\}$

Unsupervised learning

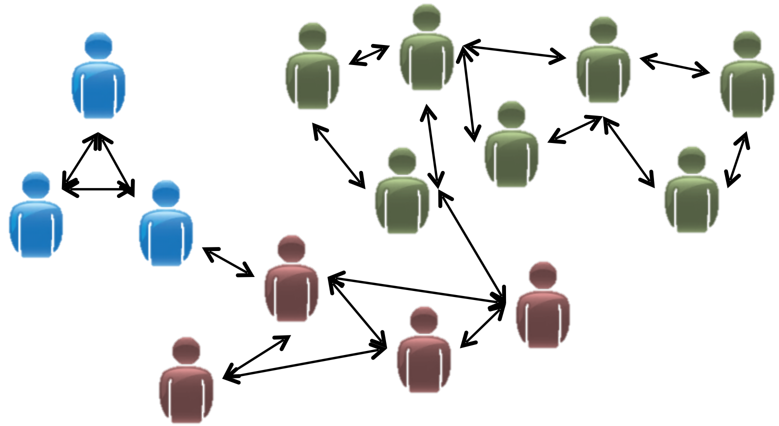


Training set: $\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(m)}\}$

Applications of clustering

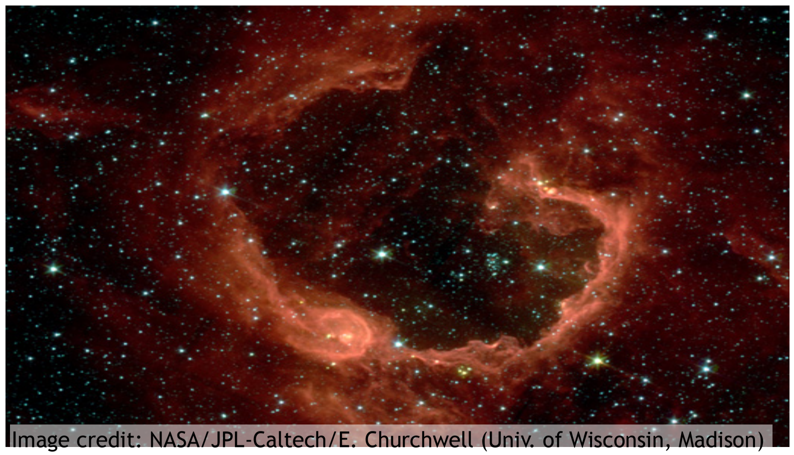


Market segmentation



Social network analysis

Organize computing clusters



Astronomical data analysis

K-means algorithm

- The KMeans algorithm clusters data by trying to separate samples in n groups of equal variance,

- Minimizing a criterion known as the inertia or within-cluster sum-of-squares (see below).

- This algorithm requires the number of clusters to be specified. The k-means algorithm divides a set of samples into disjoint clusters, each described by the mean of the samples in the cluster. The means are commonly called the cluster “centroids”; note that they are not, in general, points from , although they live in the same space.

- The K-means algorithm aims to choose centroids that minimise the inertia, or within-cluster sum-of-squares criterion:

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 {

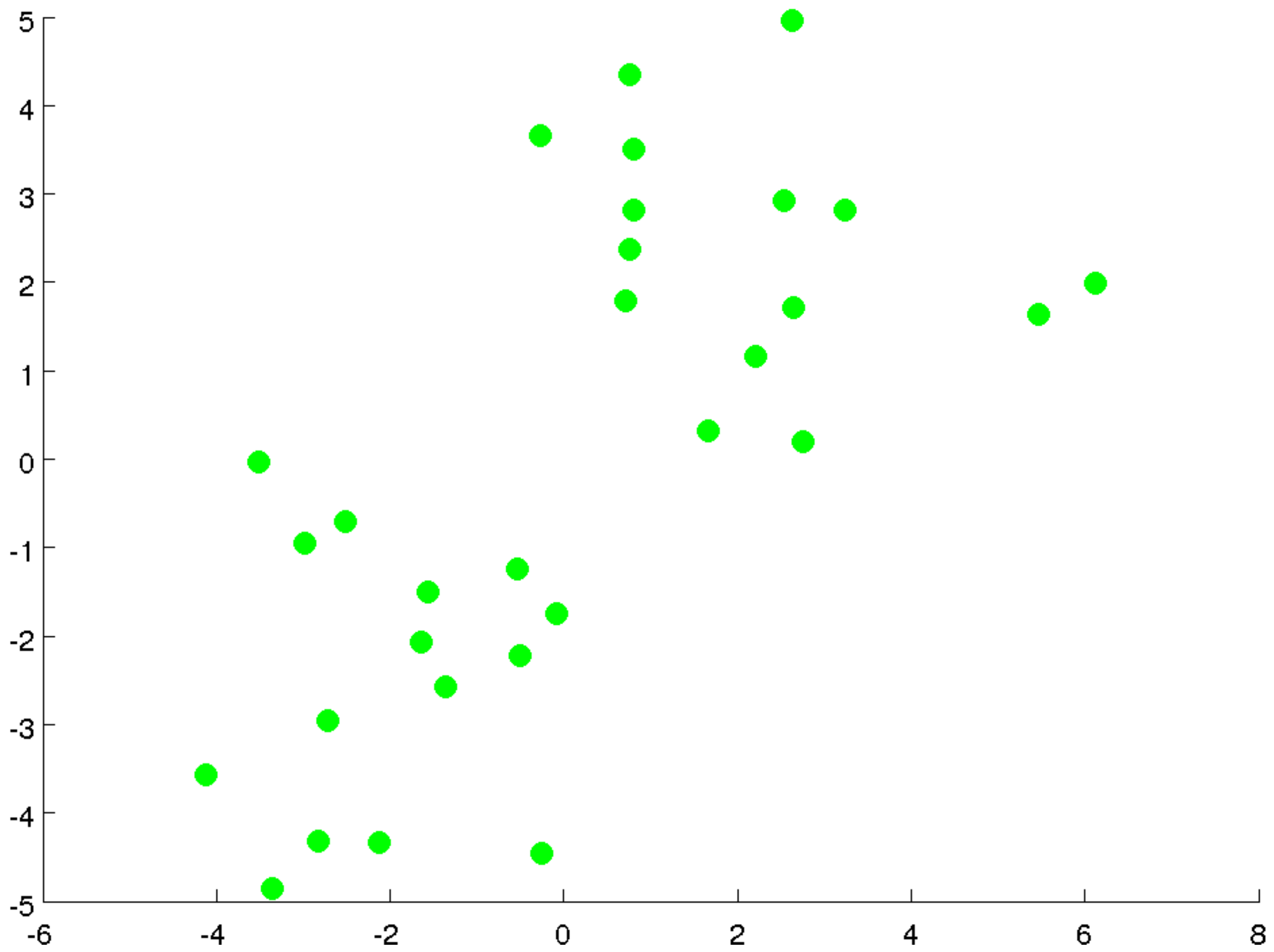
for i = 1 to m

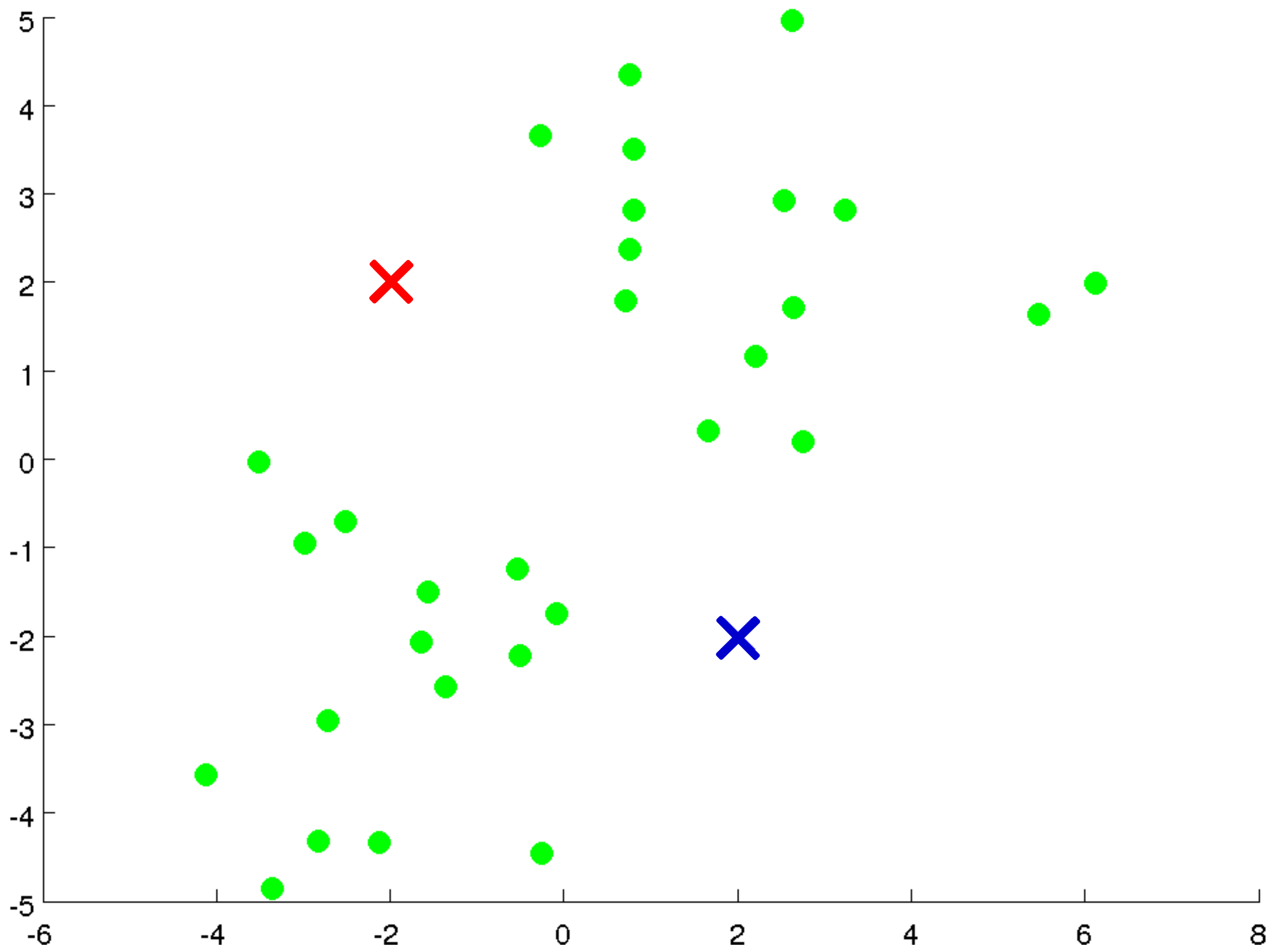
$c^{(i)}$ = index (from 1 to K) of cluster centroid closest to $x^{(i)}$

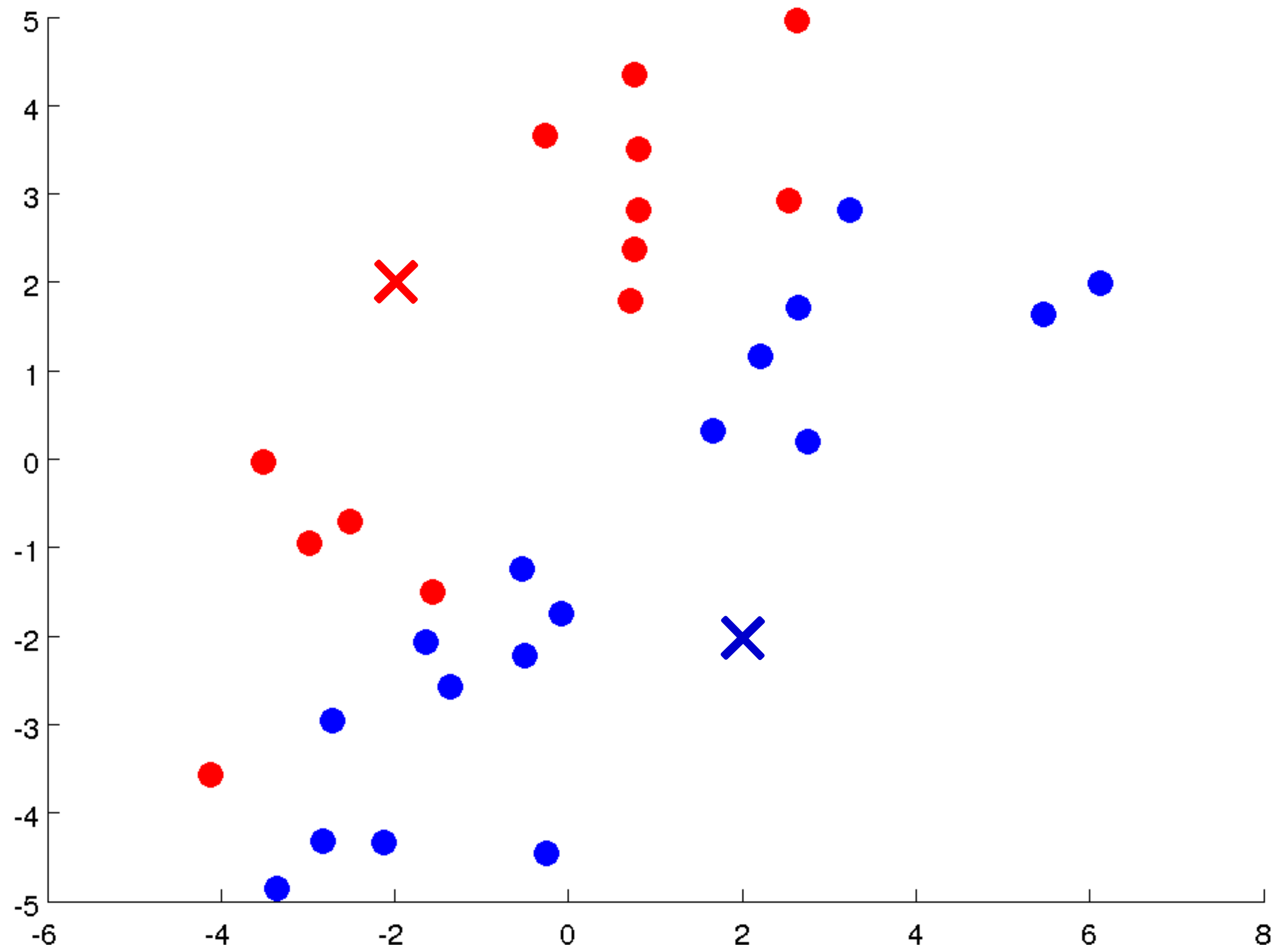
for k = 1 to K

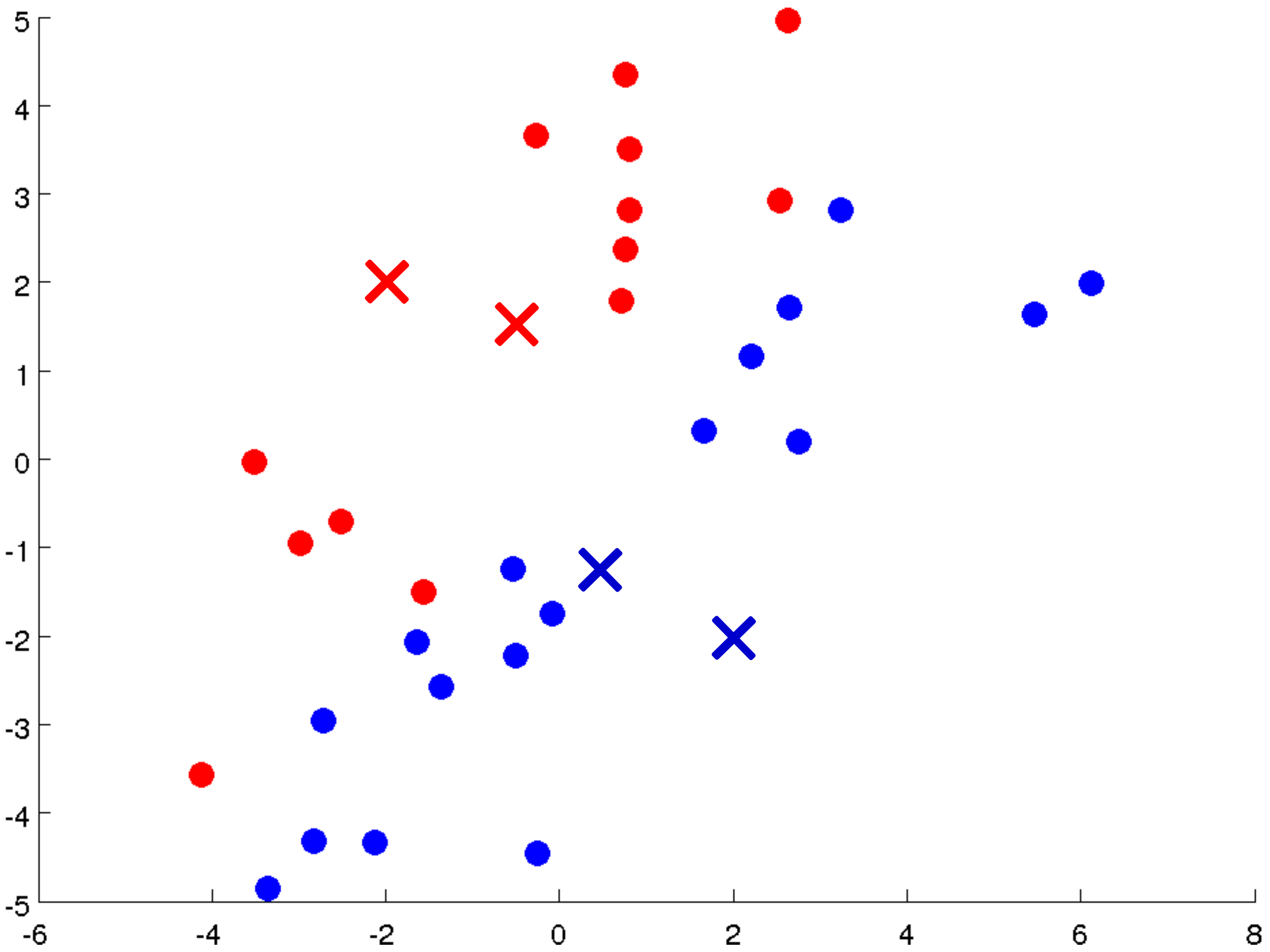
μ_k := average (mean) of points assigned to cluster k

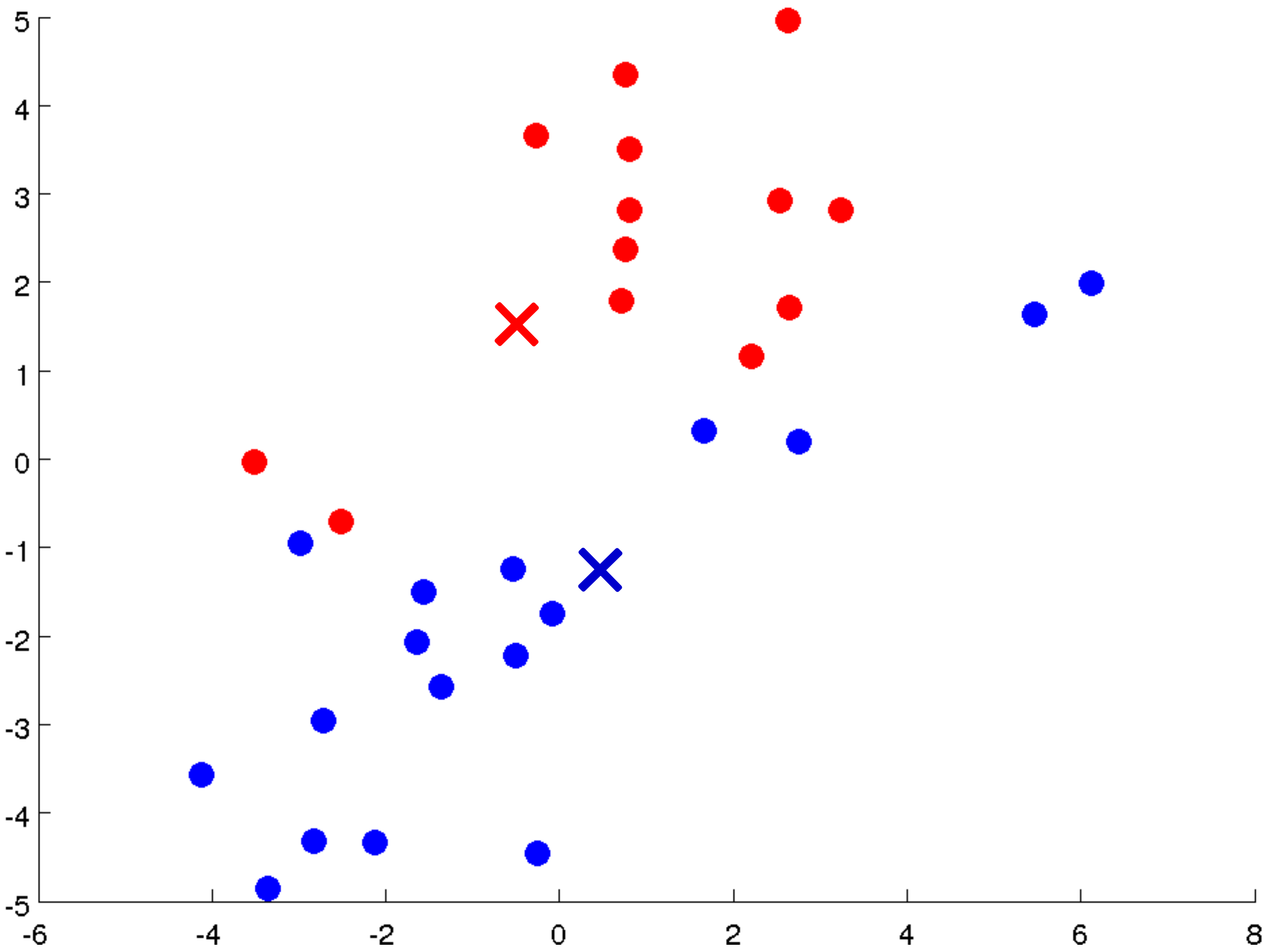
}


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

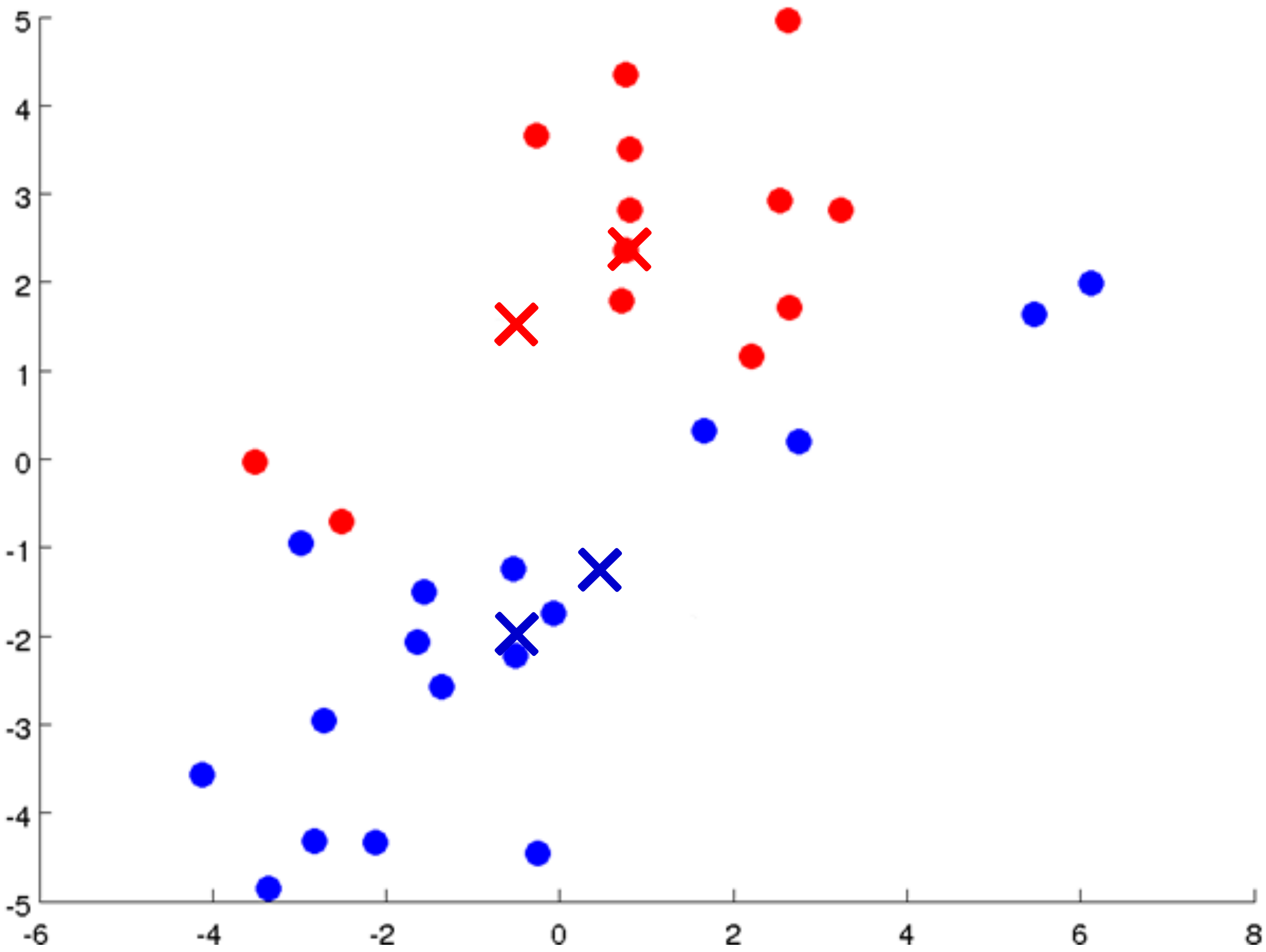

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

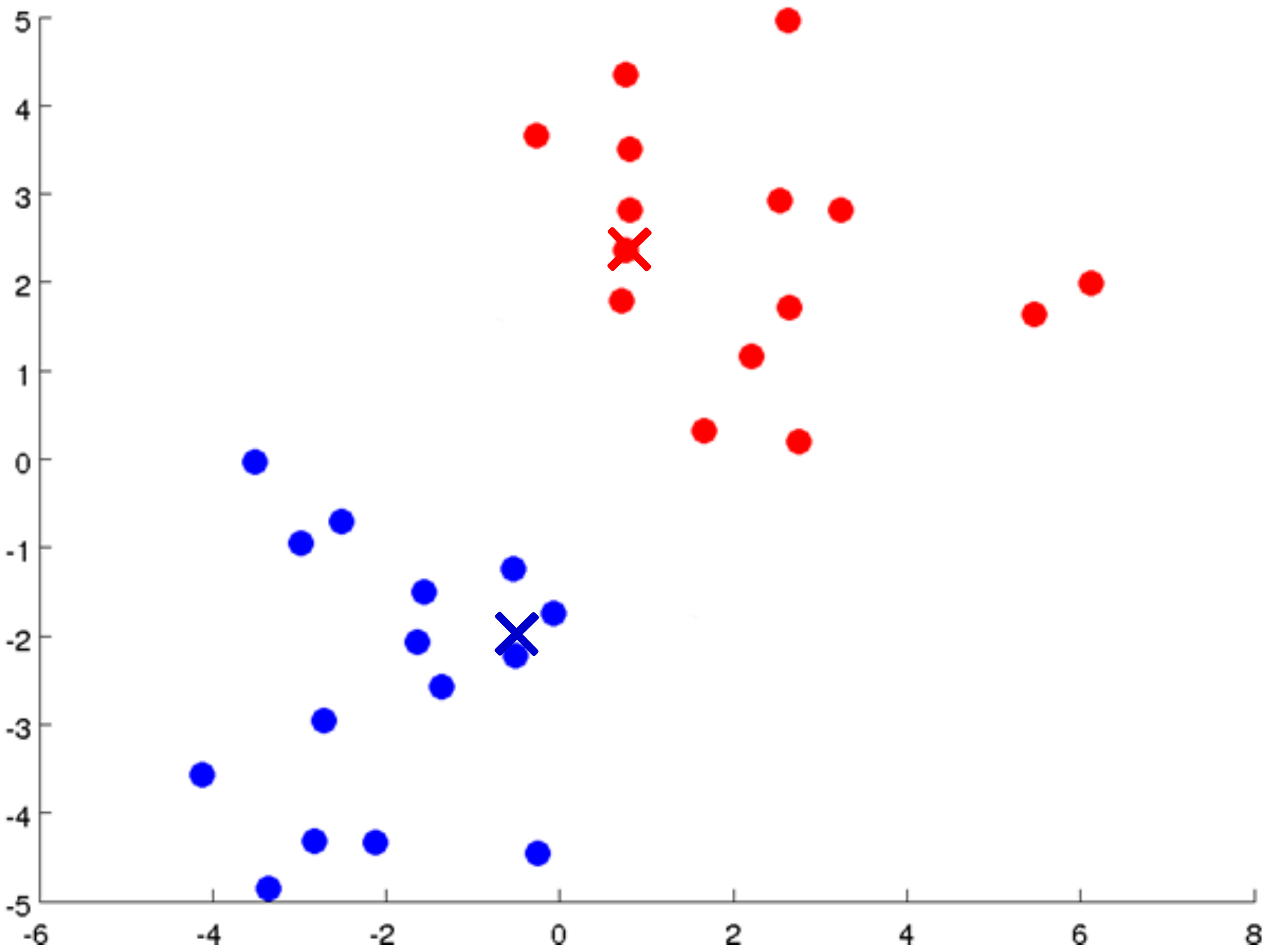

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8


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

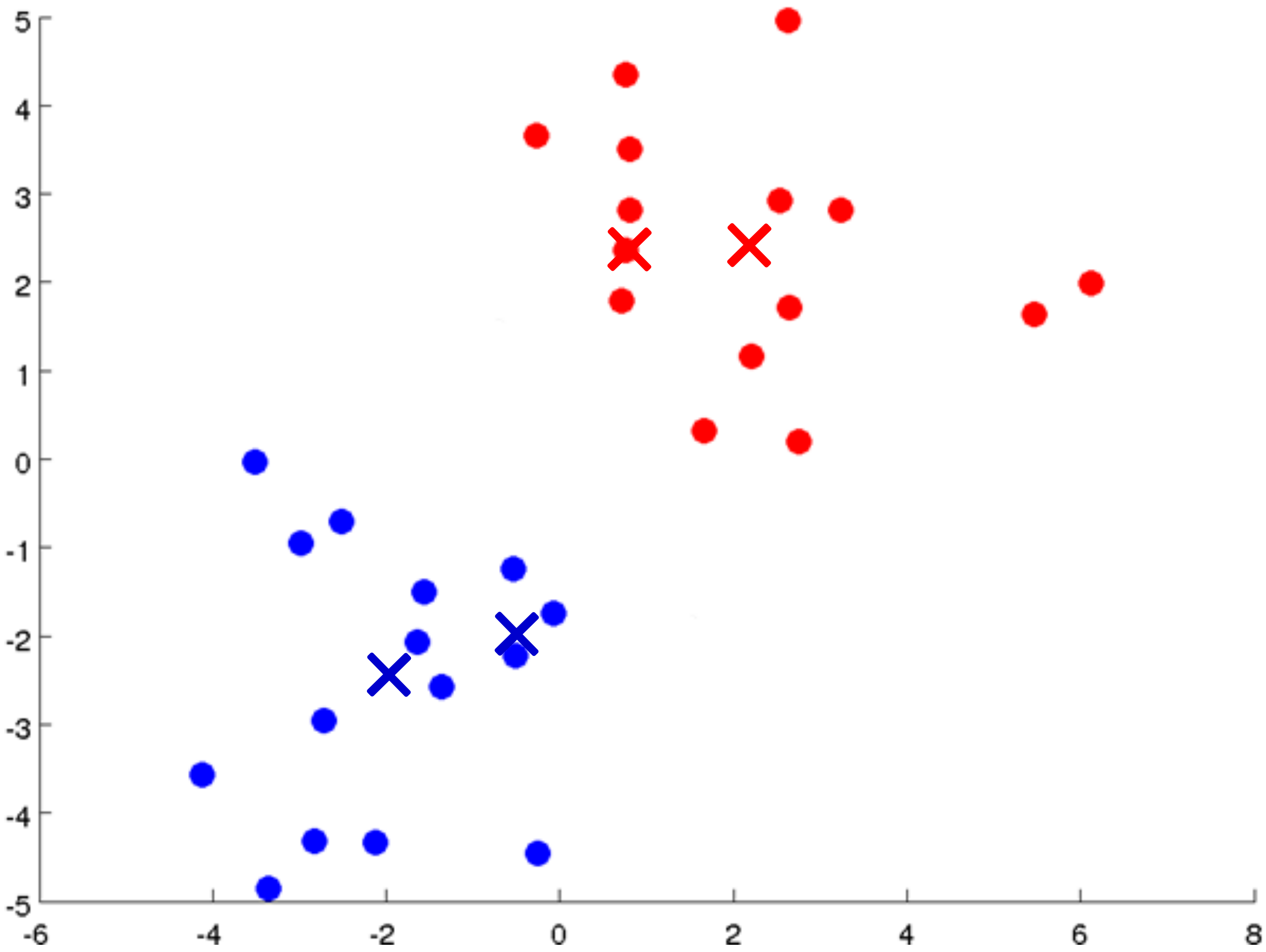


5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

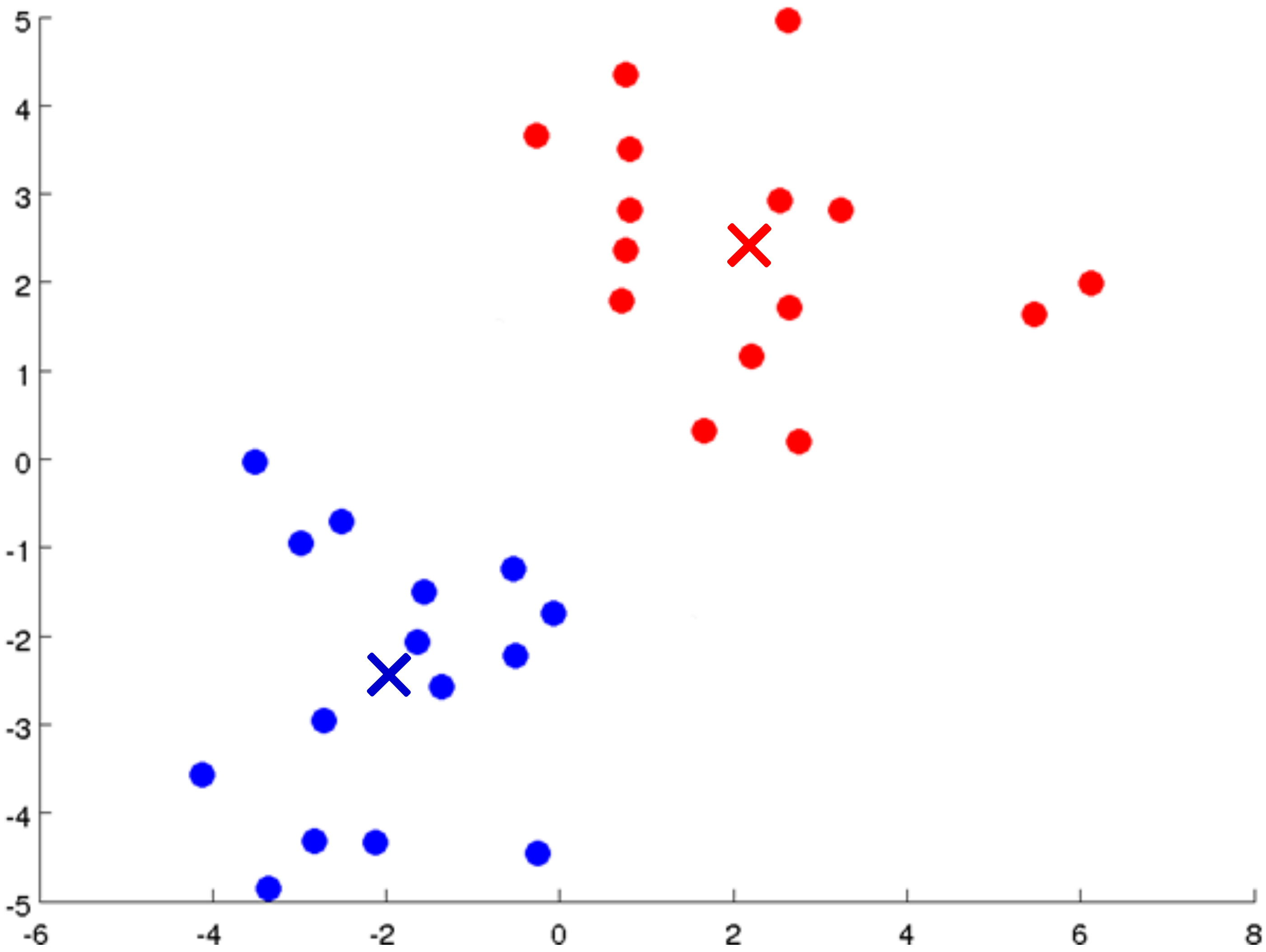

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8



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

Limitation

of K-Means

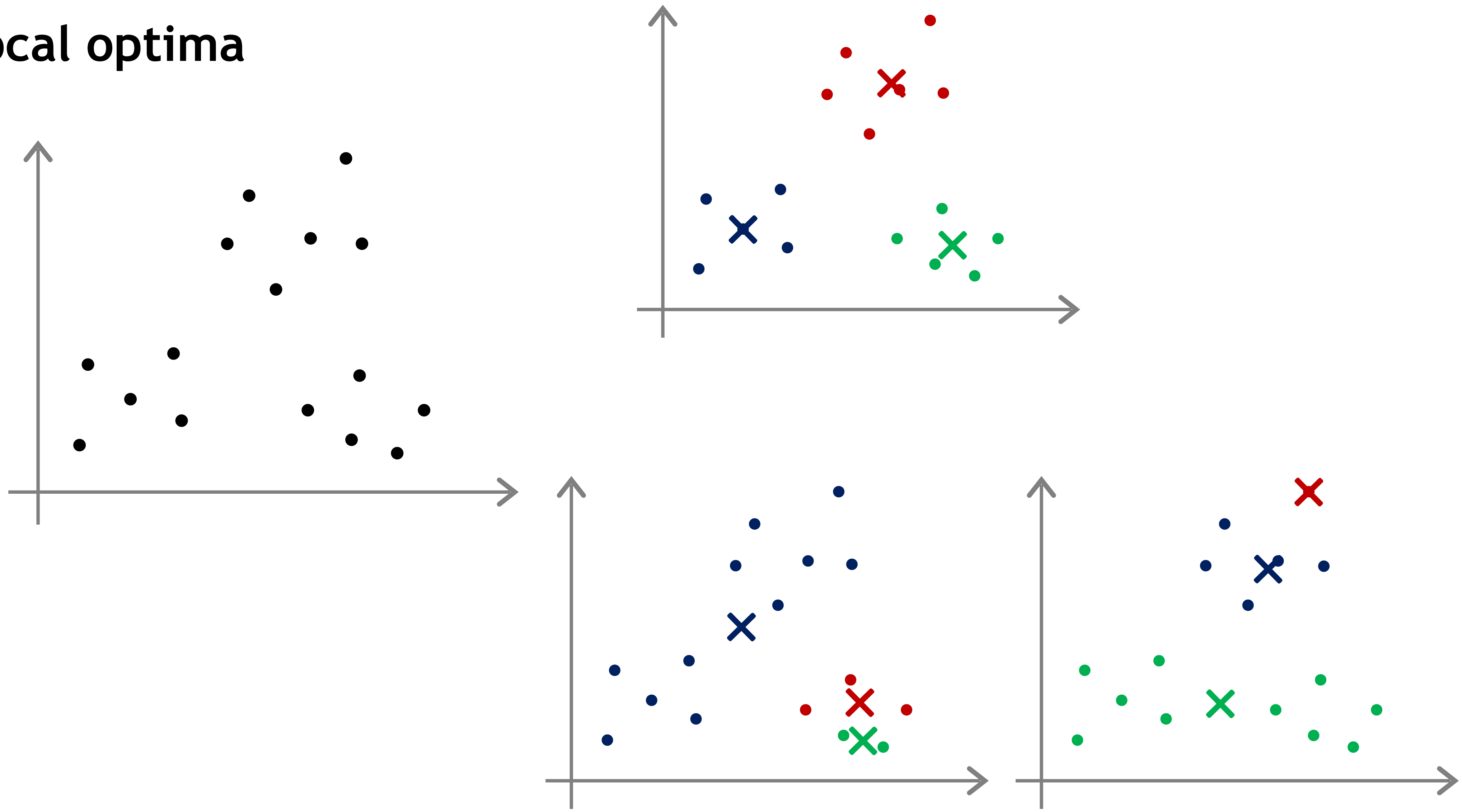
Limitation

1. Where to put initial centroids?
2. What is the right value of K?

Random initialization

Centroids

Local optima



The `k-means++` algorithm

We propose a specific way of choosing centers for the **`k-means`** algorithm. In particular, let $D(x)$ denote the shortest distance from a data point to the closest center we have already chosen. Then, we define the following algorithm, which we call **`k-means++`**.

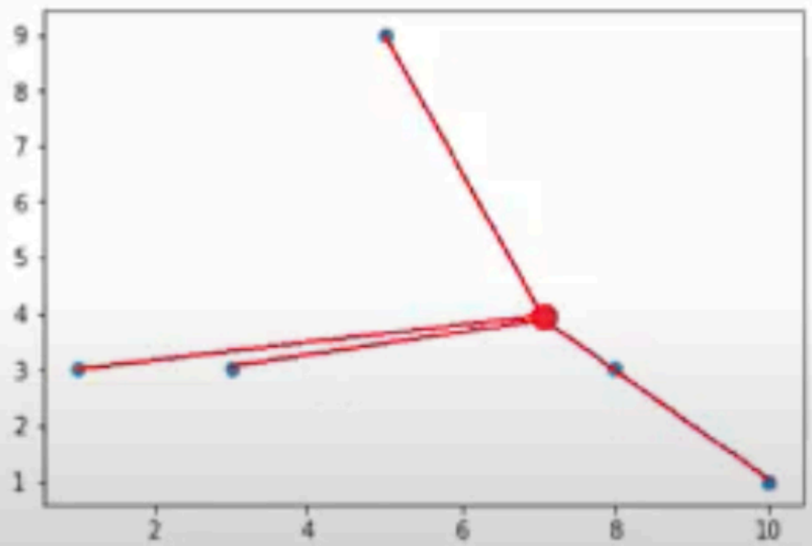
1a. Take one center c_1, chosen uniformly at random from $\mathcal{X}$.
1b. Take a new center c_i, choosing $x \in \mathcal{X}$ with probability $\frac{D(x)^2}{\sum_{x \in \mathcal{X}} D(x)^2}$. (Assign probability to each x)
1c. Repeat Step 1b. until we have taken k centers altogether.

Arthur, David, and Sergei Vassilvitskii. "k-means++: The Advantages of Careful Seeding."

Suppose we have the small dataset [(7,4),(8,3),(5,9),(3,3),(1,3),(10,1)] to which we wish to assign 3 clusters.

We begin by randomly selecting (7,4) to be a cluster center.

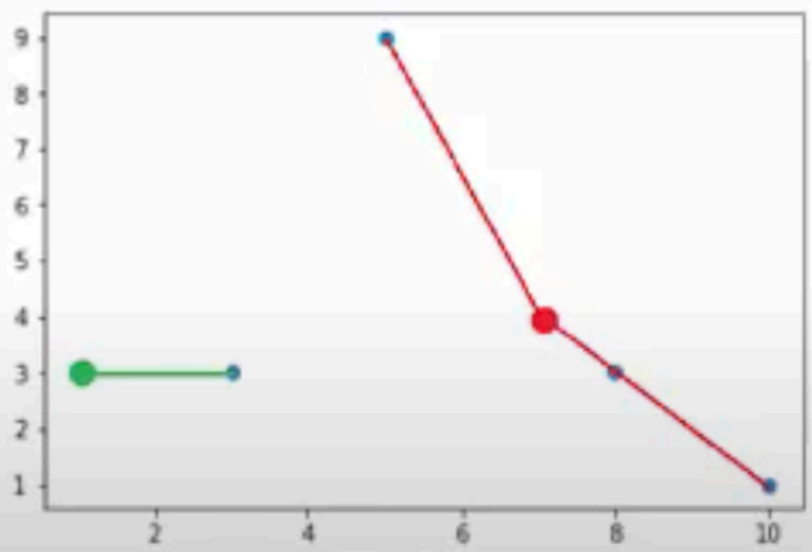
x	prob
(7,4)	-
(8,3)	2/103
(5,9)	29/103
(3,3)	17/103
(1,3)	37/103
(10,1)	18/103



Suppose we have the small dataset [(7,4),(8,3),(5,9),(3,3),(1,3),(10,1)] to which we wish to assign 3 clusters.

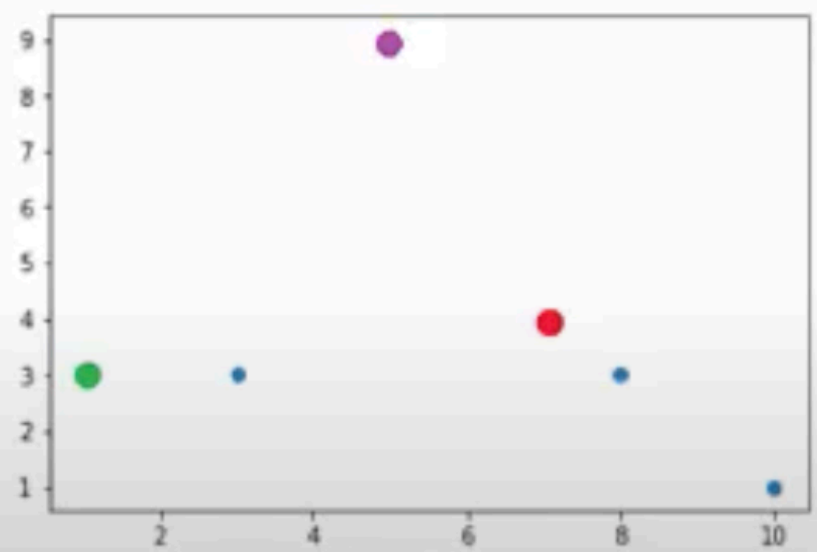
We add (1,3) to the list of cluster centers.

x	prob
(7,4)	-
(8,3)	2/53
(5,9)	29/53
(3,3)	4/53
(1,3)	-
(10,1)	18/53



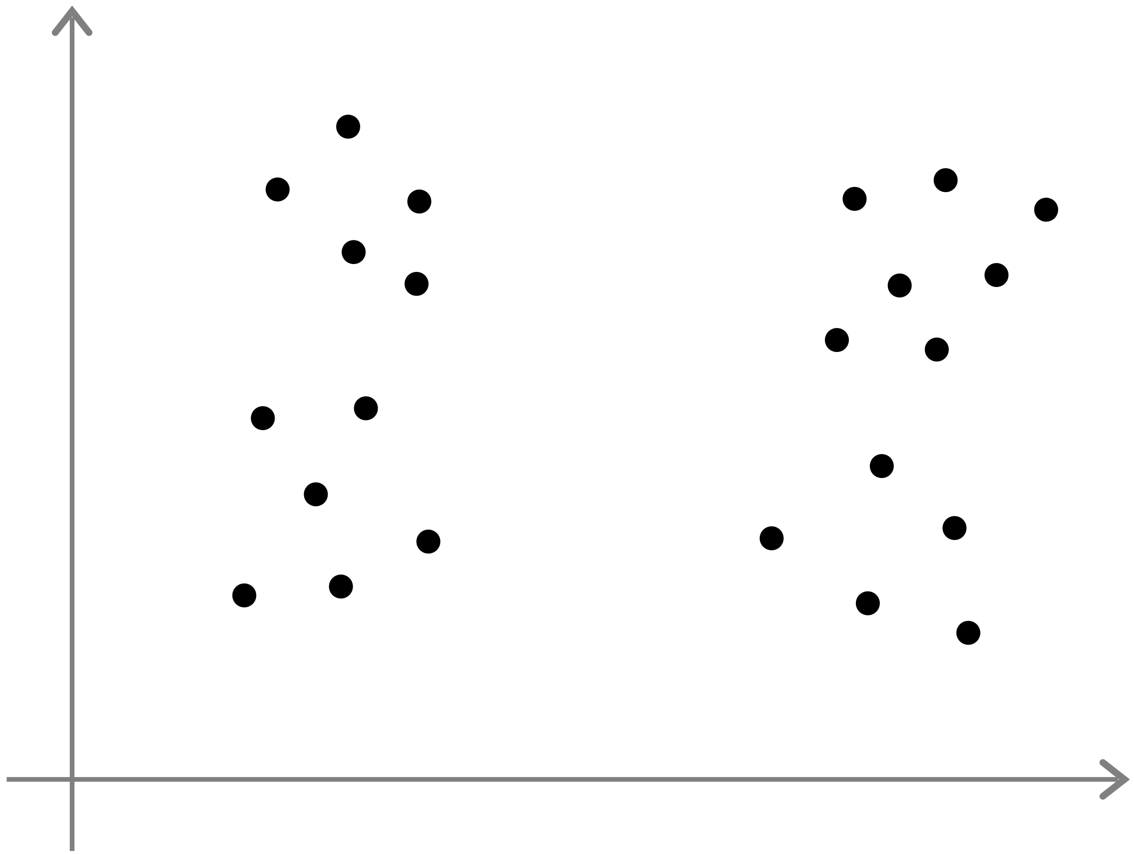
Suppose we have the small dataset [(7,4),(8,3),(5,9),(3,3),(1,3),(10,1)] to which we wish to assign 3 clusters.
We add (5,9) to the list of cluster centers.

x	prob
(7,4)	-
(8,3)	
(5,9)	-
(3,3)	
(1,3)	-
(10,1)	



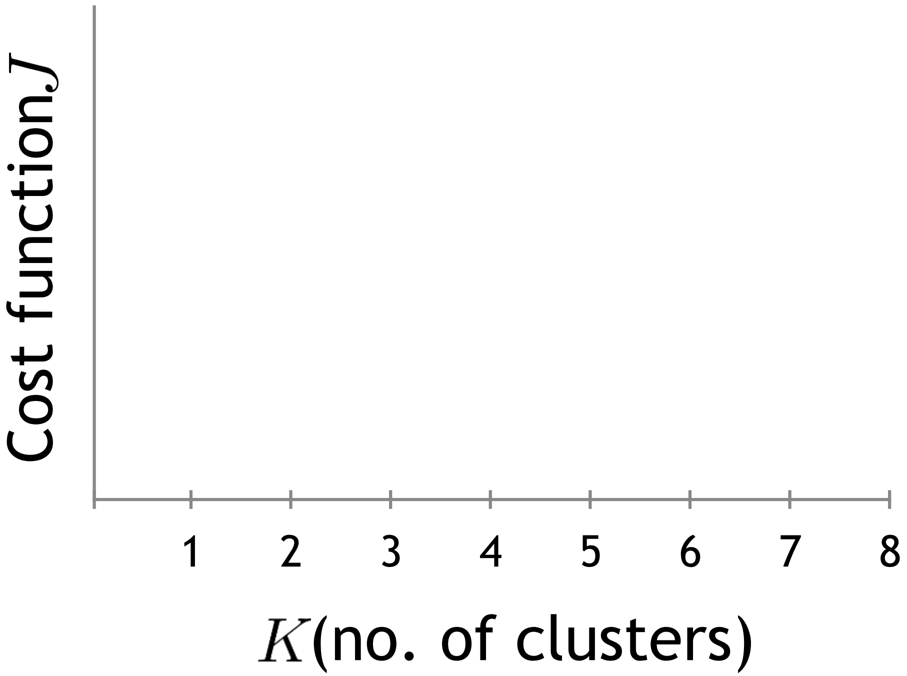
Choosing the number of clusters

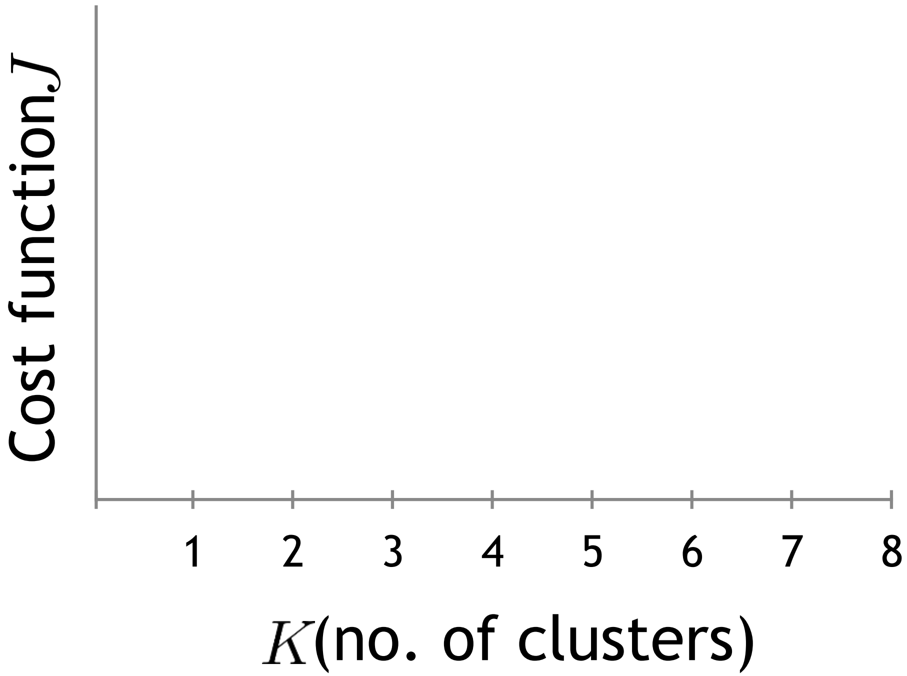
What is the right value of K?



Choosing the value of K

Elbow method:





Choosing the value of K

Silhouette analysis:

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & \text{if } a(i) < b(i) \\ 0 & \text{if } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1 & \text{if } a(i) > b(i) \end{cases}$$

a(i) : the average distance between 'i' and all other data within the same cluster

b(i) : the lowest average distance of 'i' to all points in any other clusters, of which 'i' is not a member

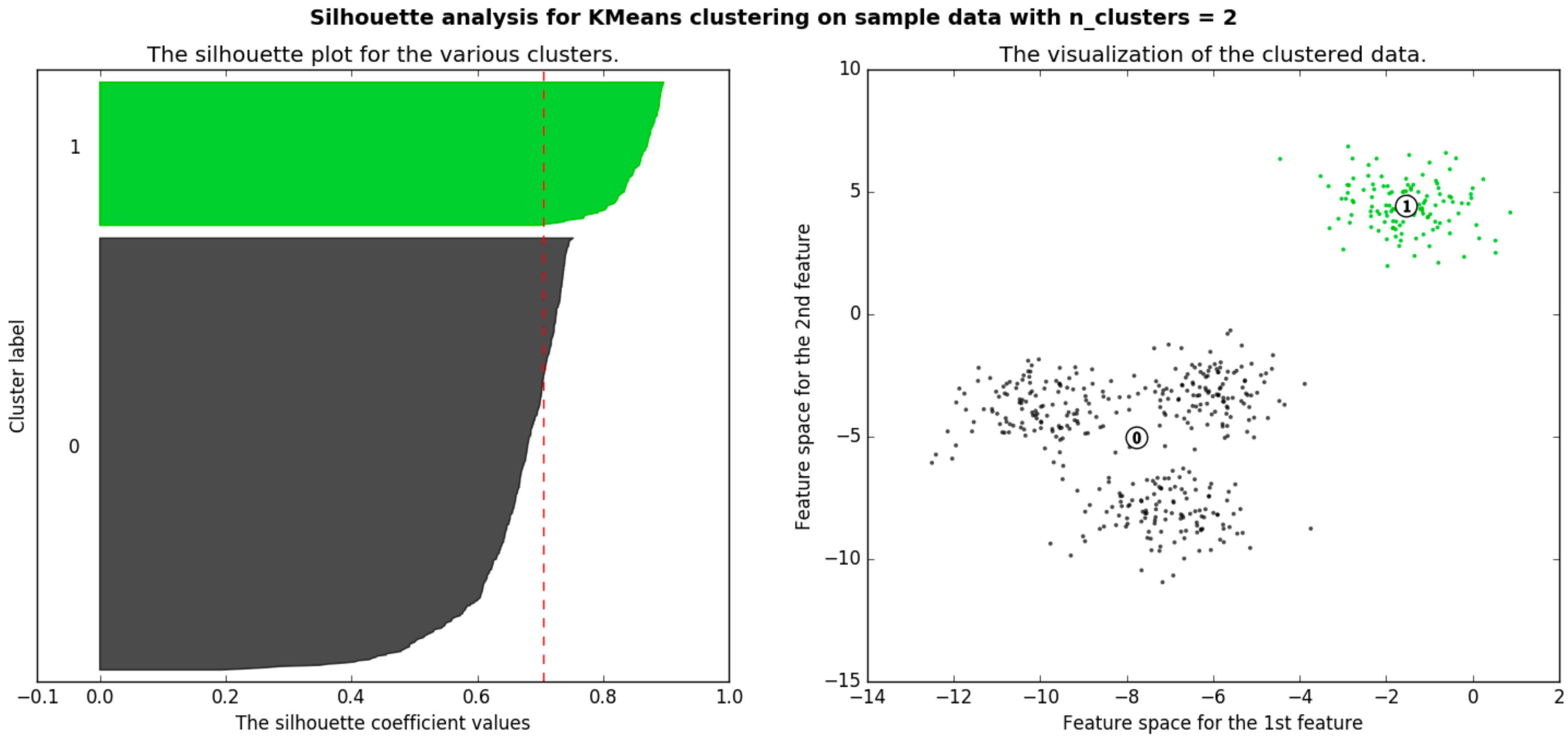
This metric ranges from -1 to 1 for each observation in your data and can be interpreted as follows:

- Values close to 1 suggest that the observation is well matched to the assigned cluster
- Values close to 0 suggest that the observation is borderline matched between two clusters
- Values close to -1 suggest that the observations may be assigned to the wrong cluster

https://campus.datacamp.com/courses/cluster-analysis-in-r/k-means-clustering?ex=9

Choosing the value of K

Silhouette analysis:



https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html