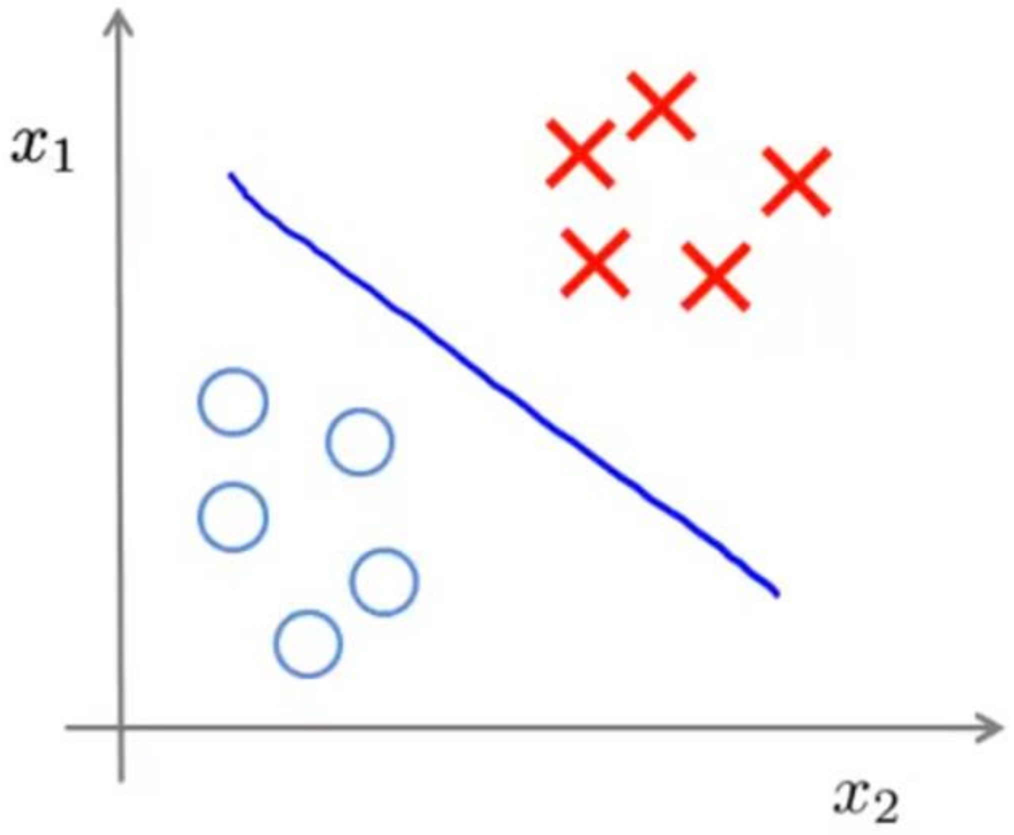
# Unsupervised Learning: Introduction

*Clustering*

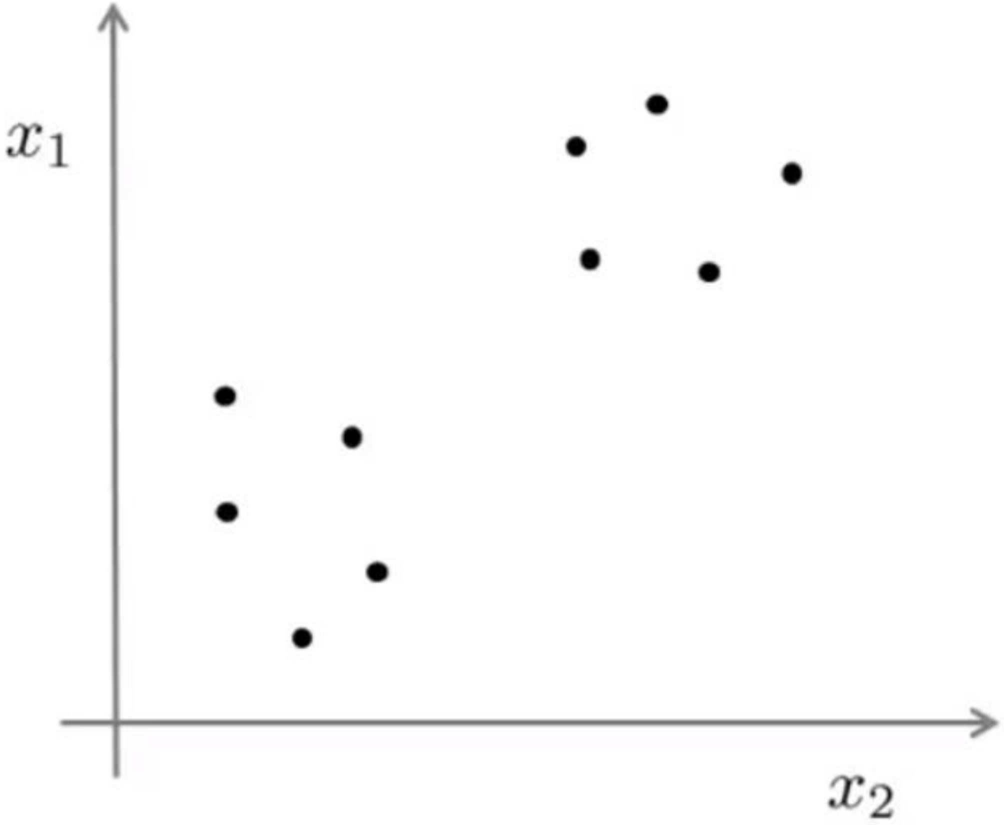
Unsupervised Learning

# Supervised learning



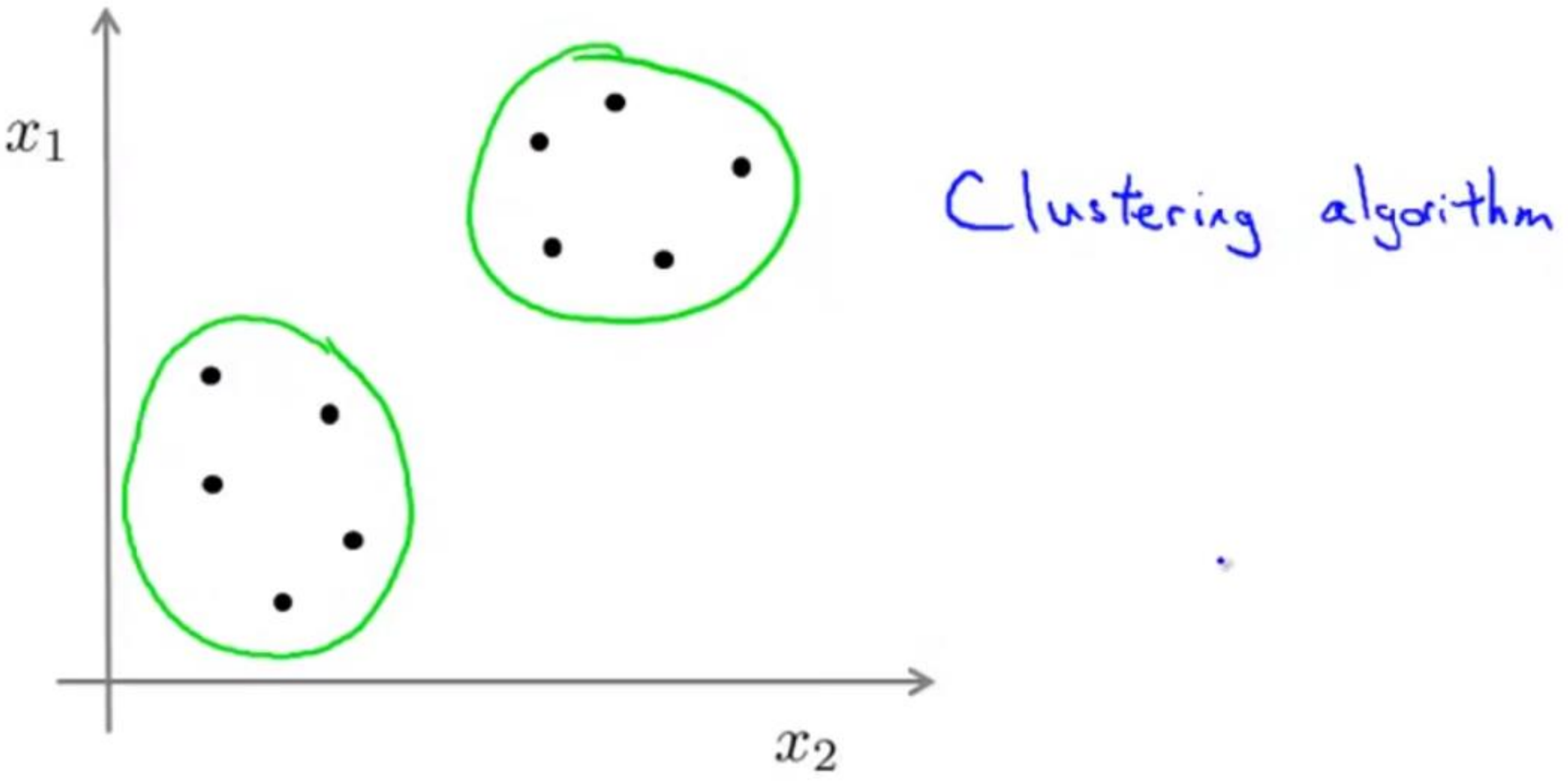

Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \ldots, (x^{(m)}, y^{(m)})\}$

# Unsupervised learning




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

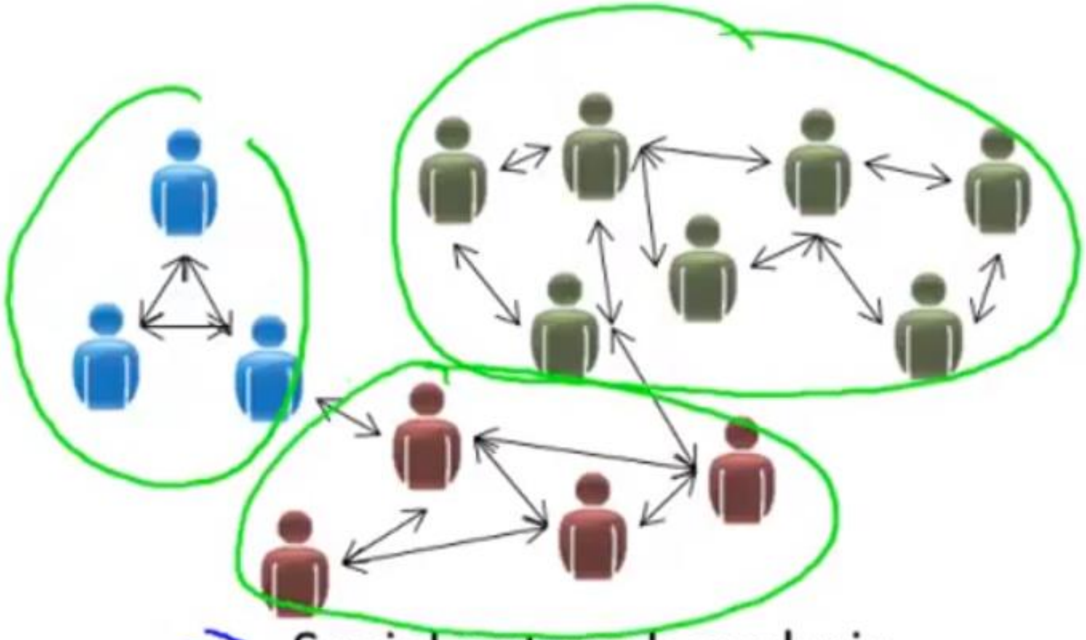
# Unsupervised learning




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

# Applications of clustering

Market segmentation



Social network analysis

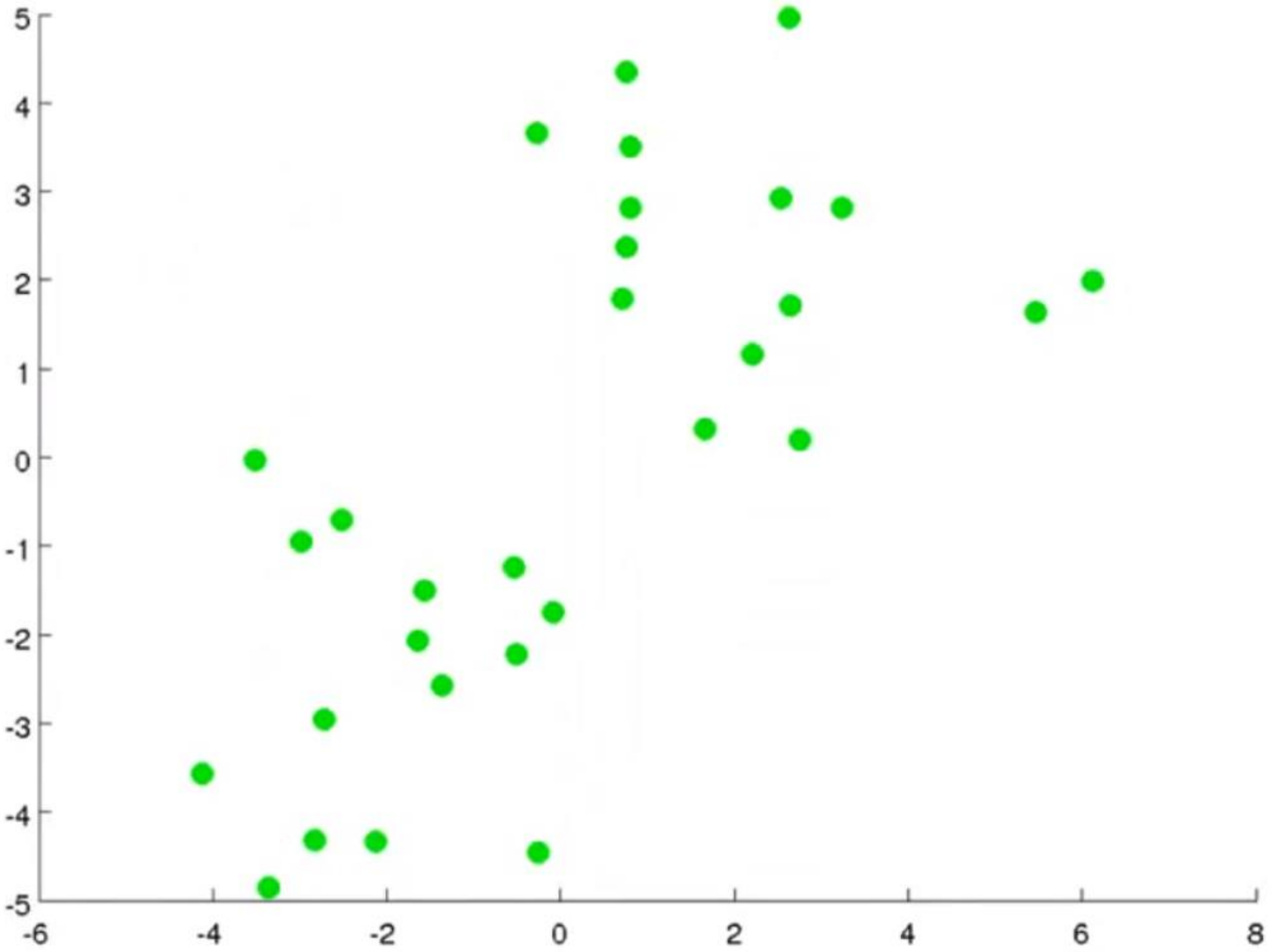
Organize computing clusters


Astronomical data analysis

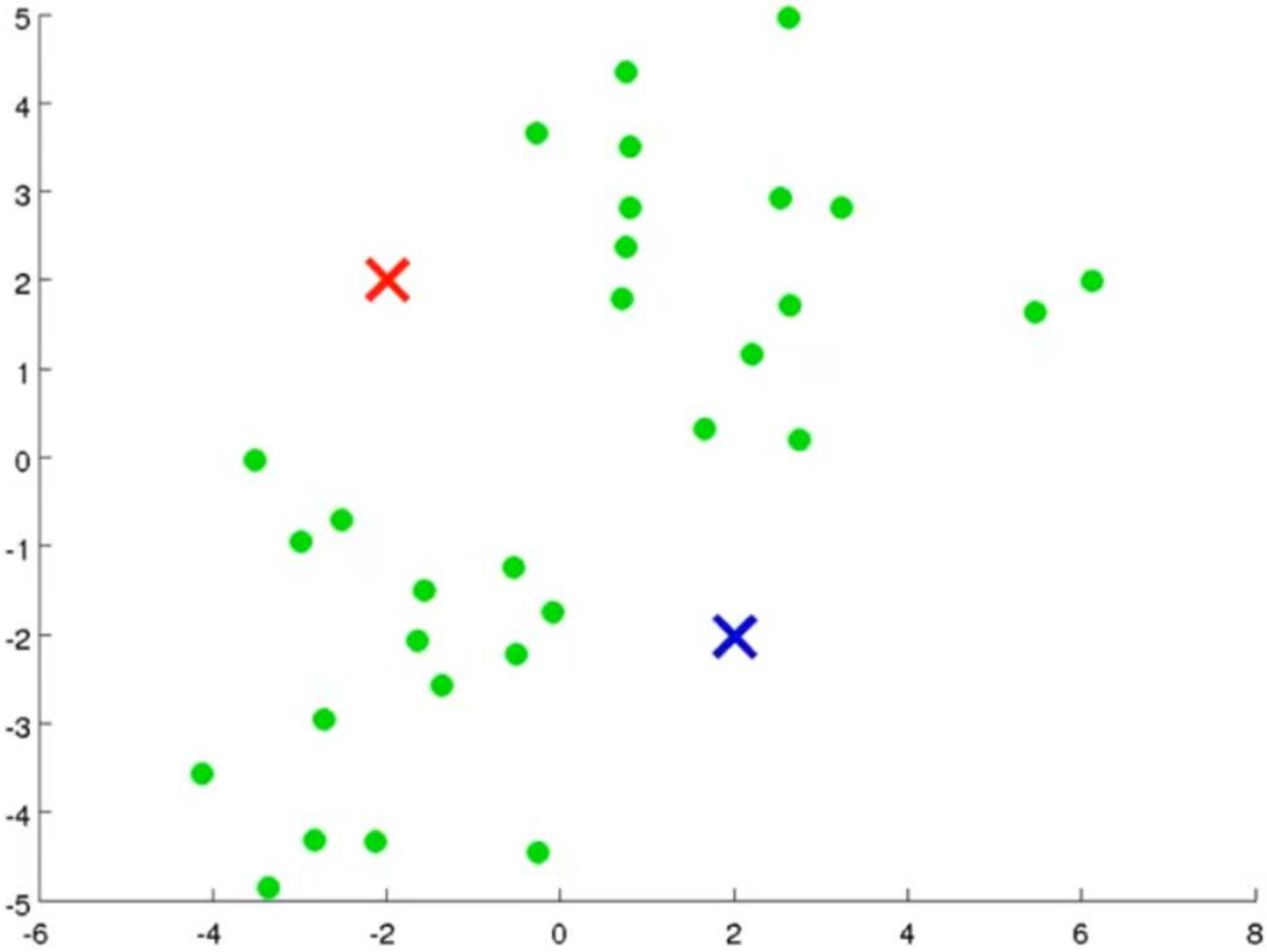
# K-Means Algorithm

*Clustering*

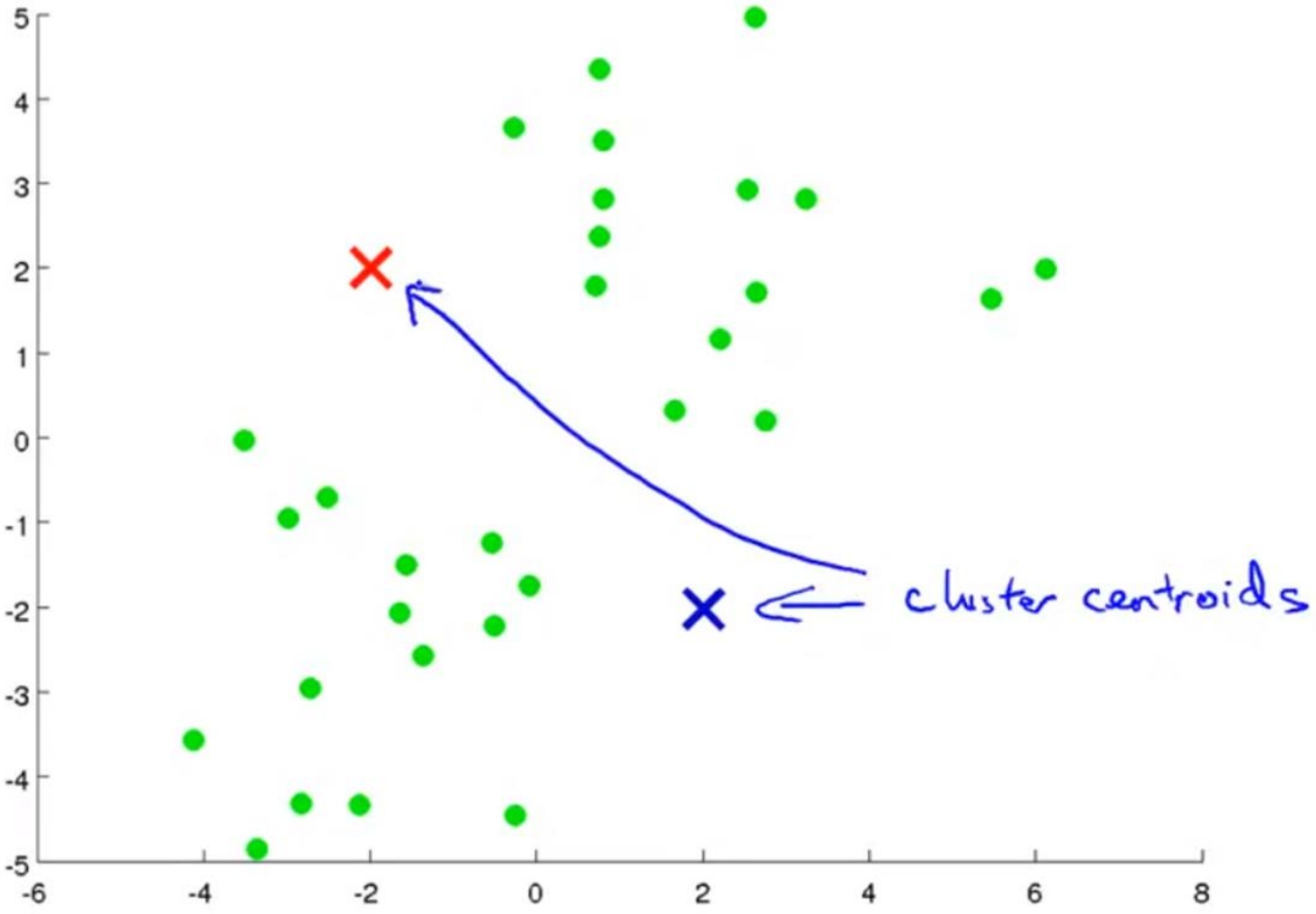
Unsupervised Learning



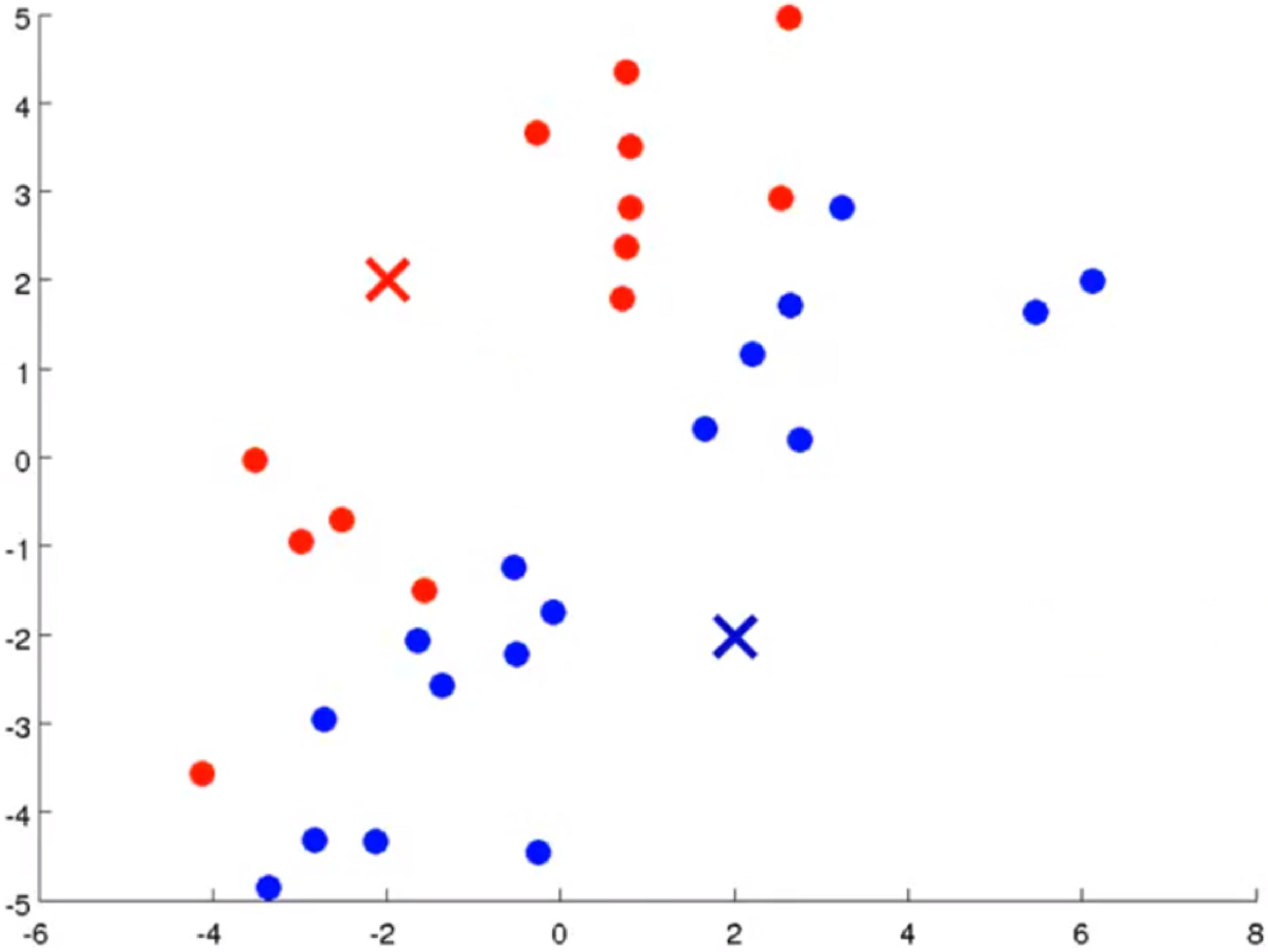

Andrew Ng



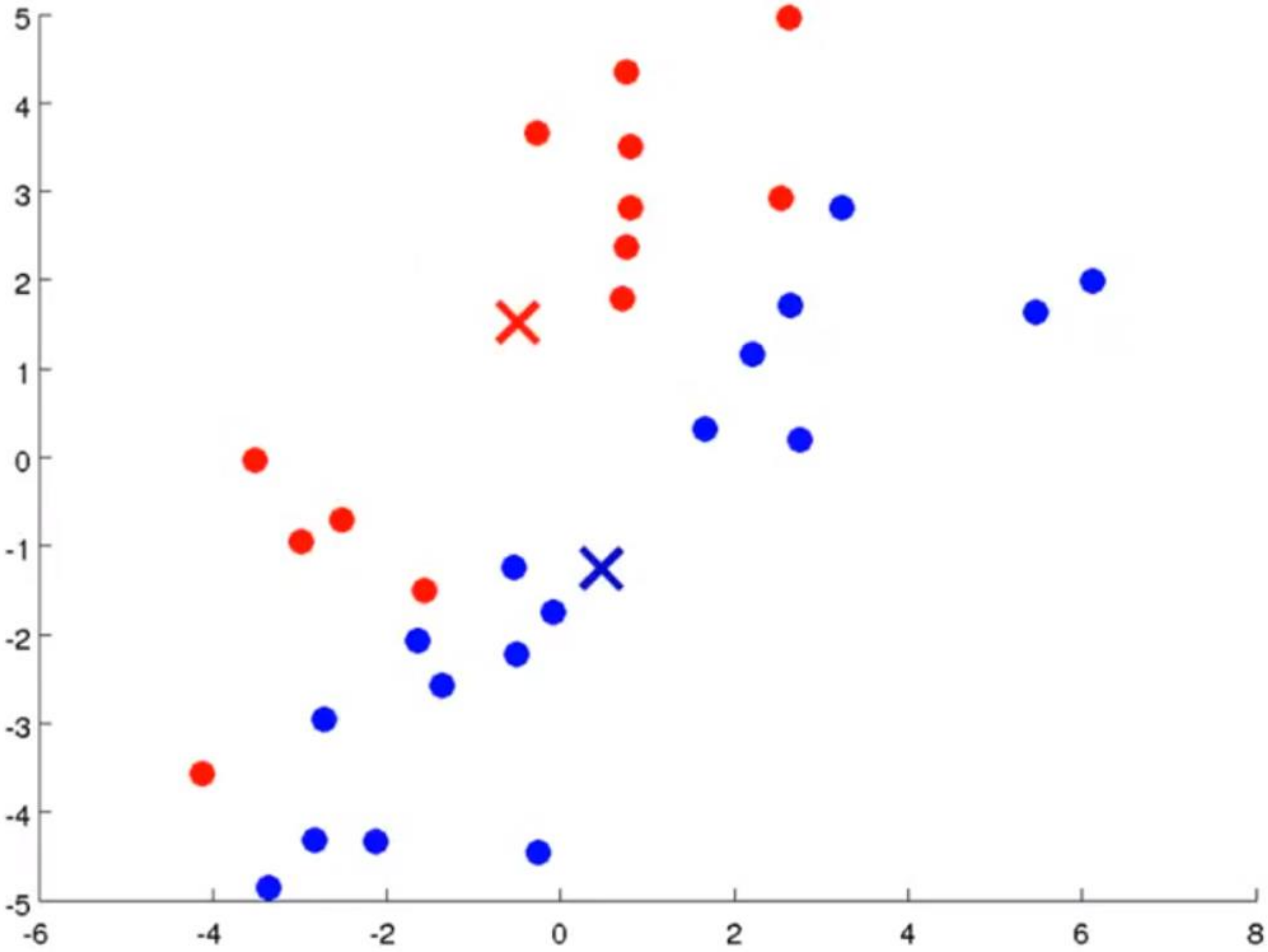

Andrew Ng



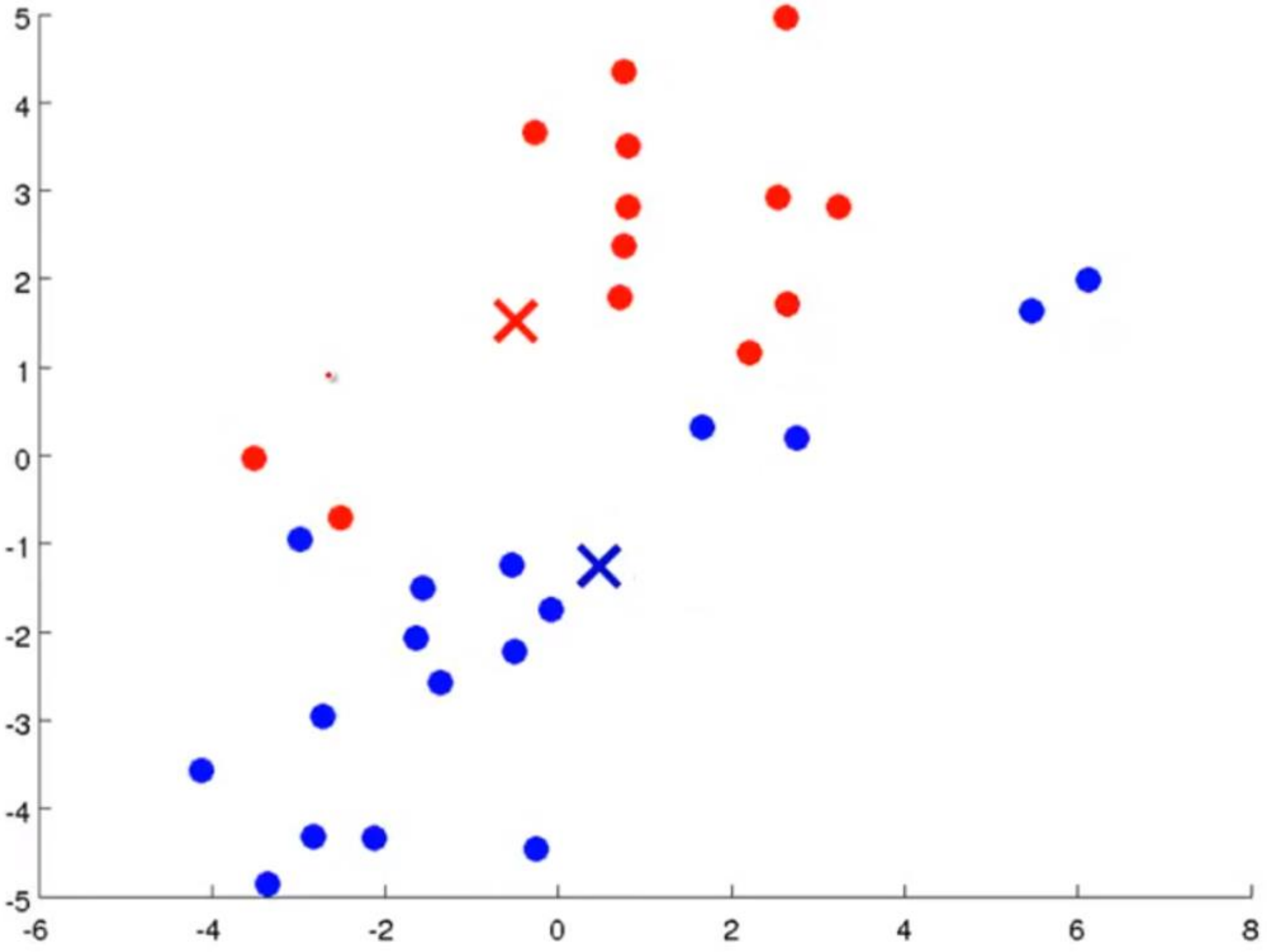
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8
cluster centroids



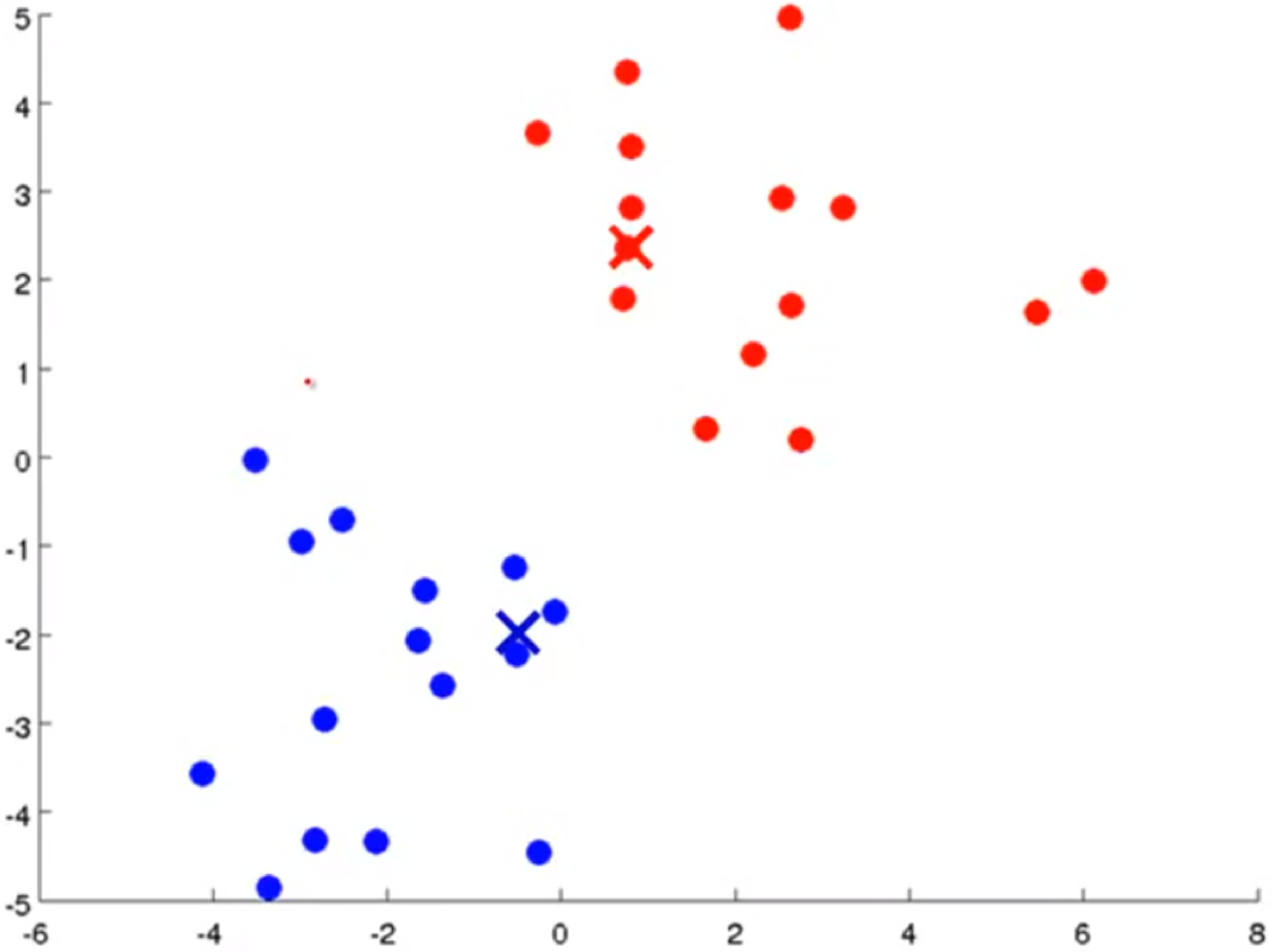

Andrew Ng



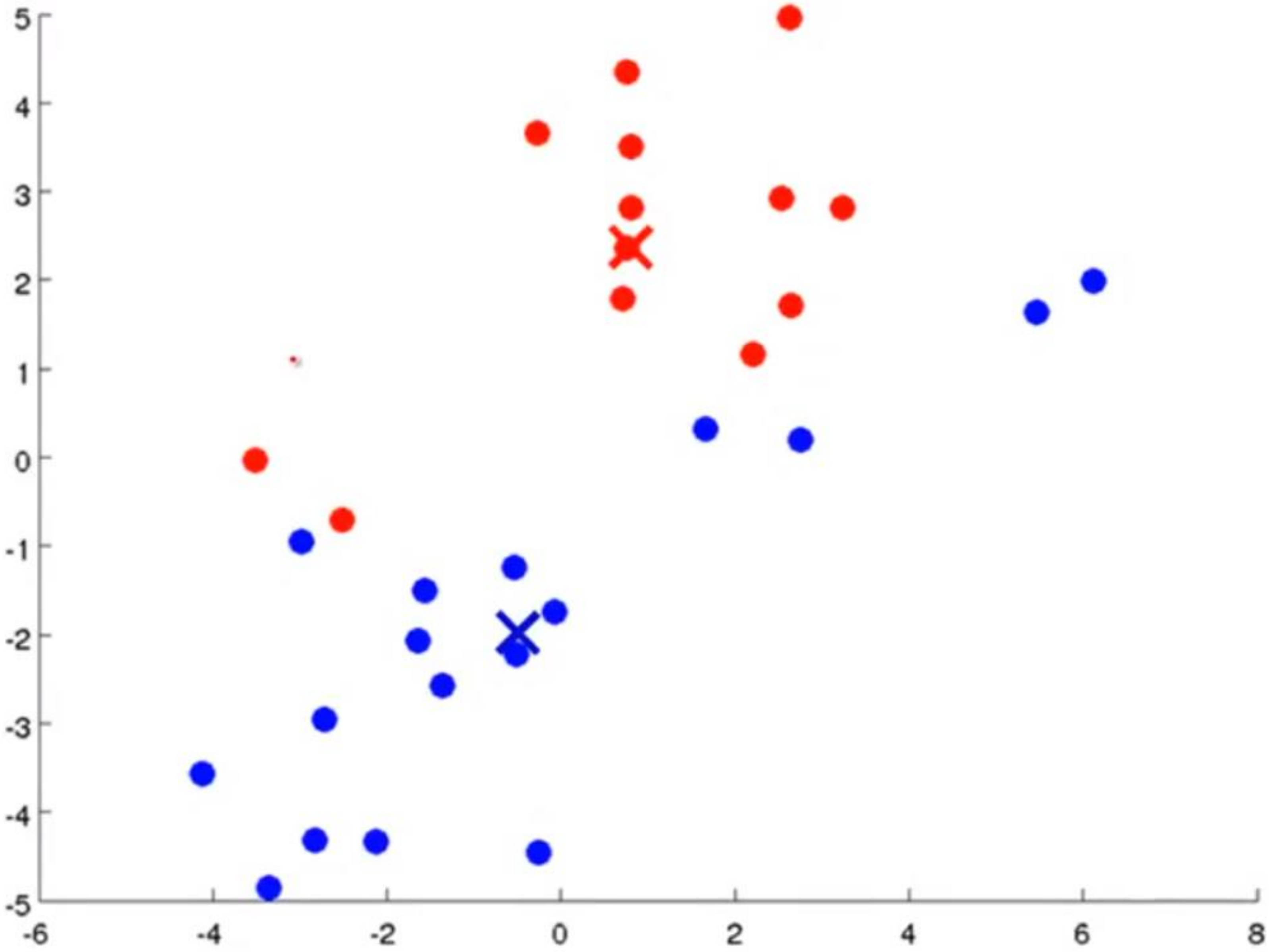

Andrew Ng



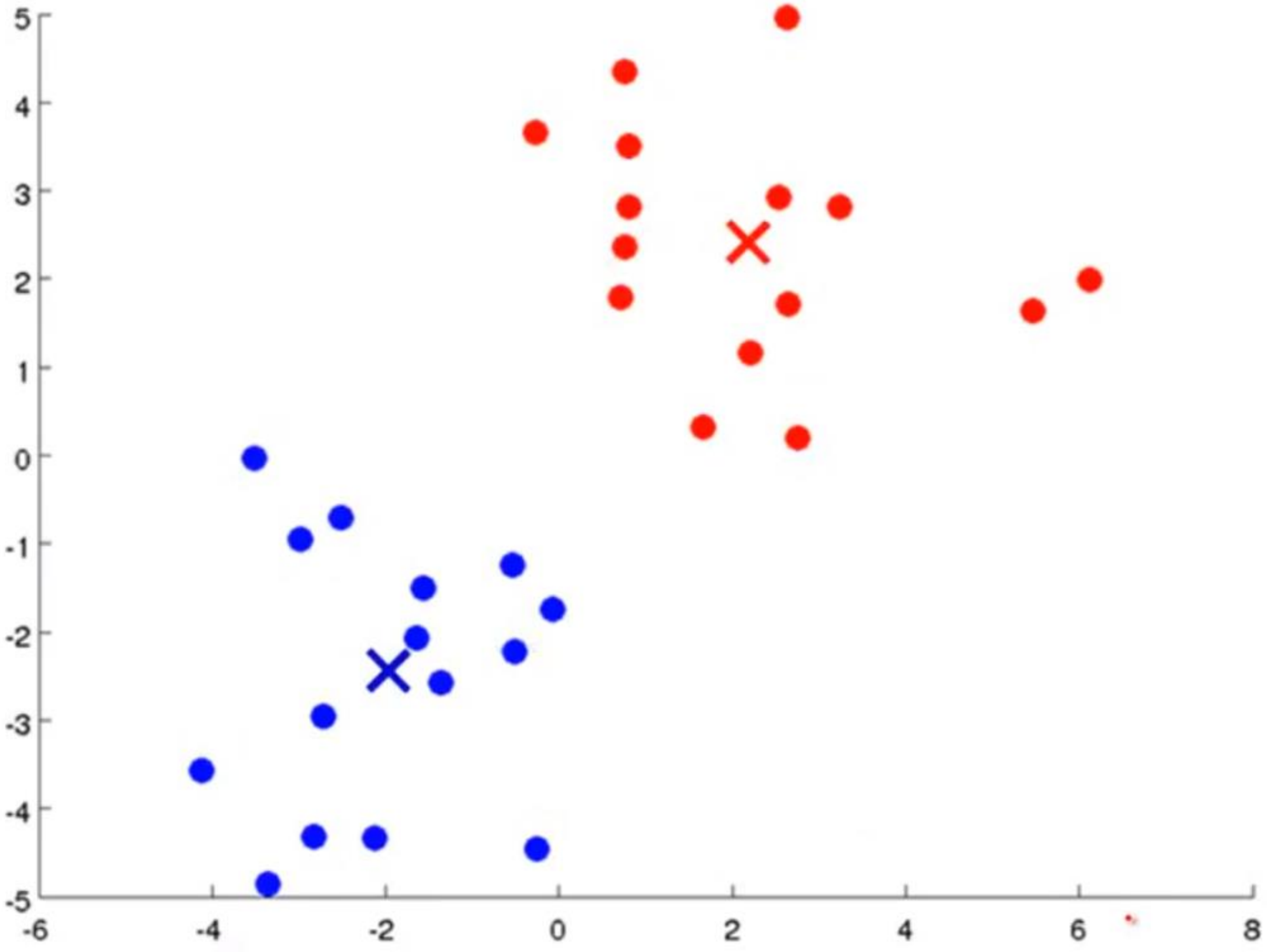

Andrew Ng




Andrew Ng



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 $x^{(i)}$

for $k$ = 1 to $K$

$\mu_k$ := average (mean) of points assigned to cluster $k$

}

# K-means algorithm

$\mu_1$ $\times$ $\mu_2$ $\times$

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

Repeat {

Cluster assignment step

for $i$ = 1 to $m$

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

$\min_k \|x^{(i)} - \mu_k\|$ $\rightarrow c^{(i)}$

for $k$ = 1 to $K$

$\mu_k$ := average (mean) of points assigned to cluster $k$

}

# K-means algorithm

$\mu_1$ × $\mu_2$ ×

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

Repeat {

Cluster assignment step

for $i$ = 1 to $m$

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

$\min_k \|x^{(i)} - \mu_k\|^2 \rightarrow c^{(i)}$

Move centroid

for $k$ = 1 to $K$

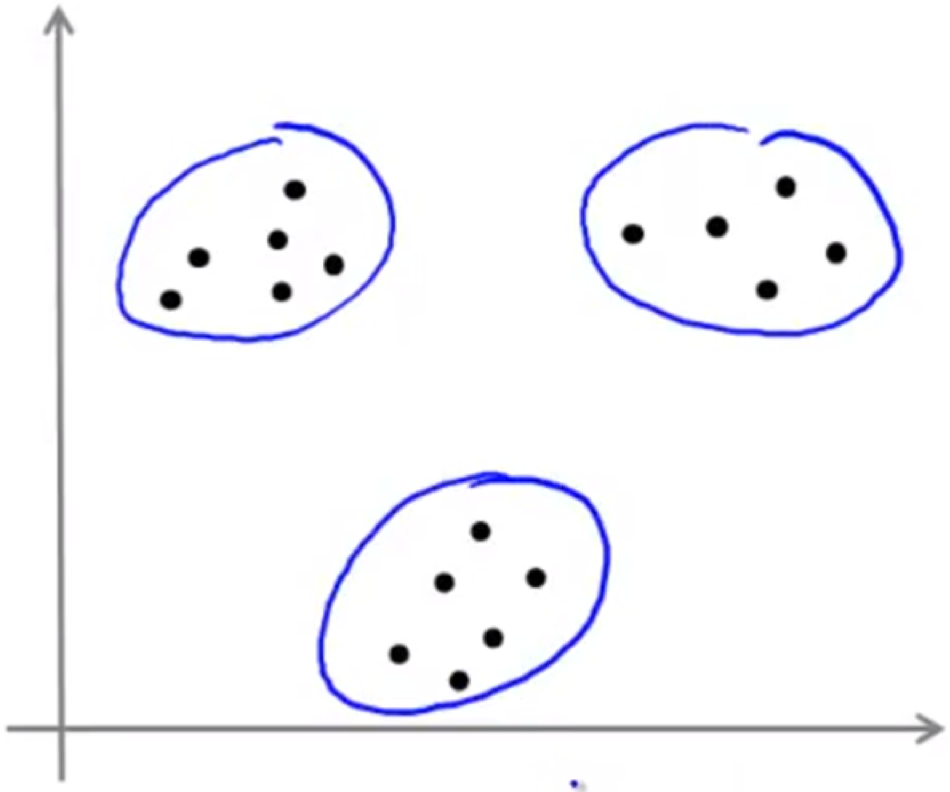
$\mu_k$ := average (mean) of points assigned to cluster $k$

$x^{(1)}, x^{(5)}, x^{(6)}, x^{(10)}$ $\rightarrow c^{(1)}=2,\ c^{(5)}=2,\ c^{(6)}=2,\ c^{(10)}=2$

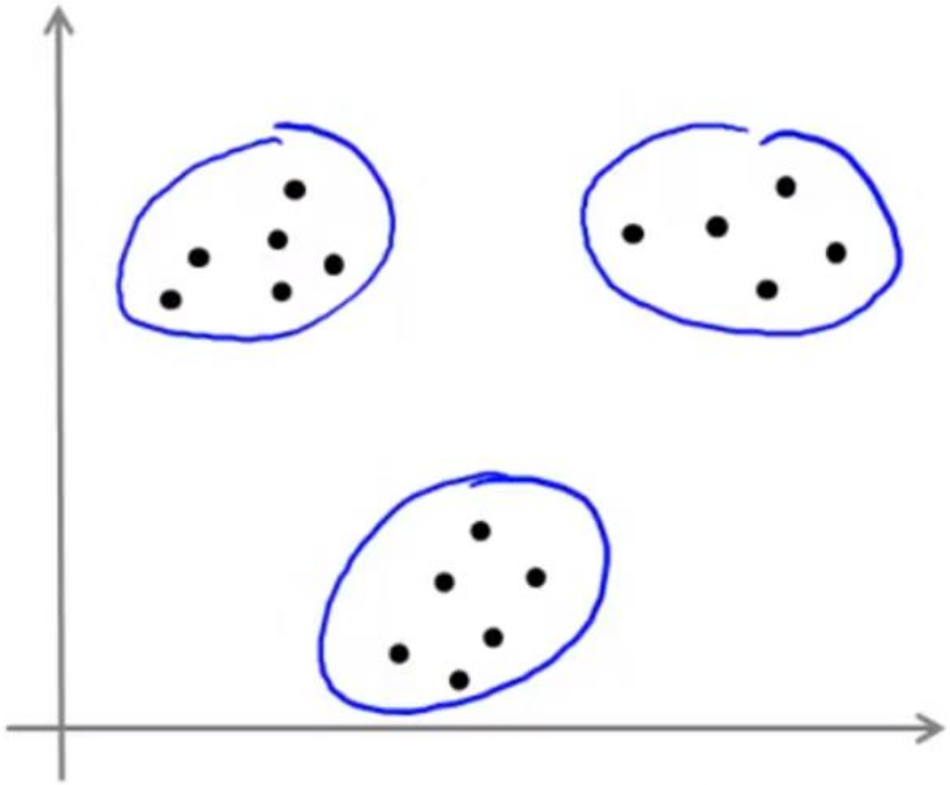
$$\mu_2 = \frac{1}{4}\left[x^{(1)} + x^{(5)} + x^{(6)} + x^{(10)}\right] \in \mathbb{R}^n$$

}

# K-means for non-separated clusters



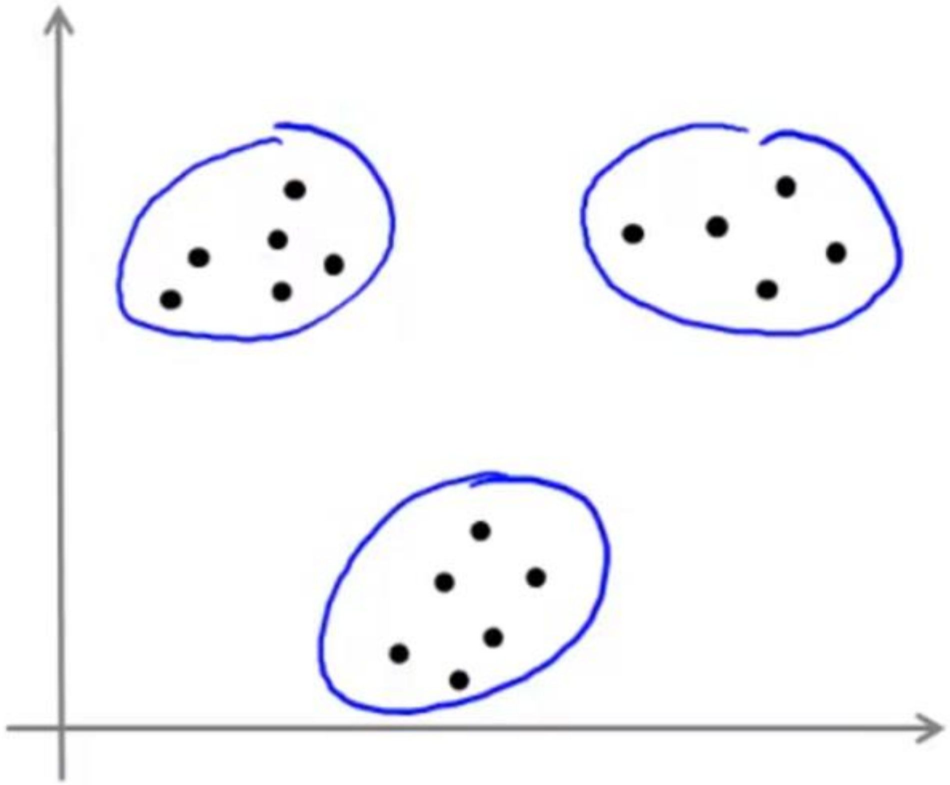
# K-means for non-separated clusters

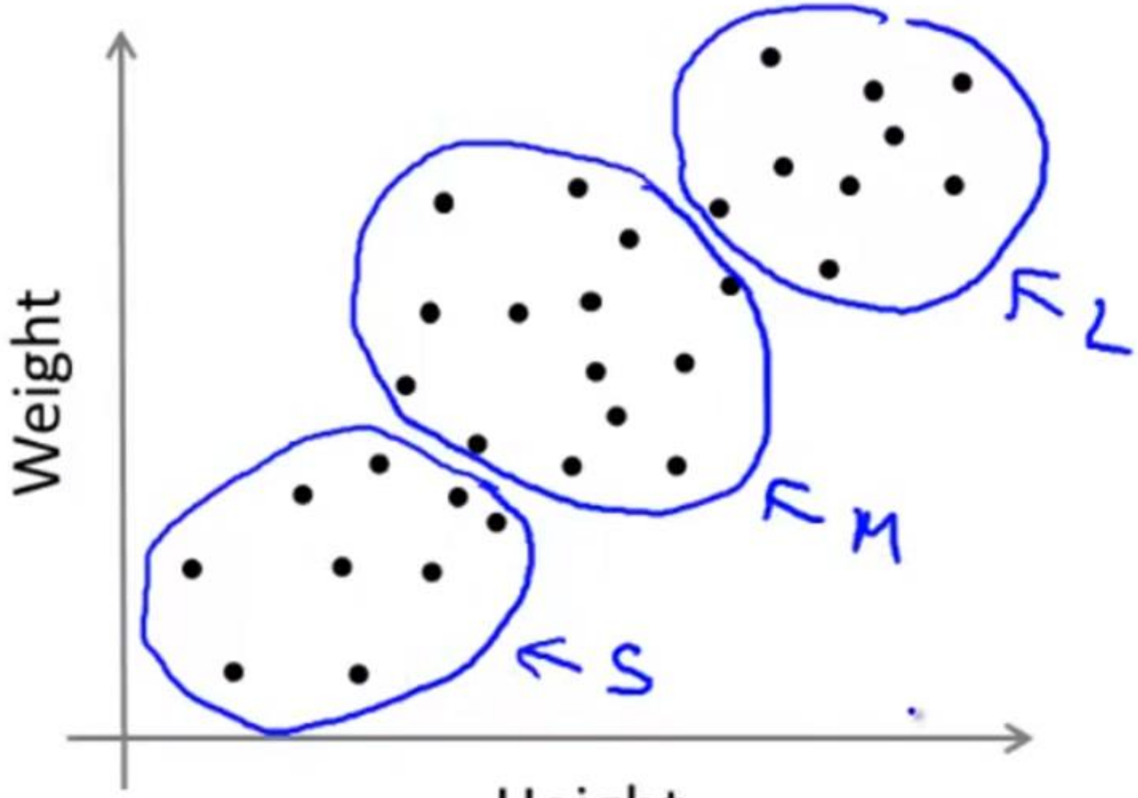


T-shirt sizing

# K-means for non-separated clusters

S, M, L





# Optimization Objective

*Clustering*

Unsupervised Learning

# K-means optimization objective

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

$\mu_k$ = cluster centroid $k$ ($\mu_k \in \mathbb{R}^n$)    $K$    $k \in \{1, 2, \ldots, K\}$

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

$x^{(i)} \to 5$    $c^{(i)} = 5$    $\mu_{c^{(i)}} = \mu_5$

# K-means optimization objective

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

$\mu_k$ = cluster centroid $\underline{k}$ ($\mu_k \in \mathbb{R}^n$)   K   $k \in \{1, 2, \ldots, K\}$

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

$x^{(i)} \rightarrow \underline{5}$   $\underline{c^{(i)} = 5}$   $\underline{\mu_{c^{(i)}}} = \mu_5$

Optimization objective:

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

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

Andrew Ng

# K-means algorithm

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

Cluster assignment step

Minimize $J(\ldots)$ wrt $c^{(1)}, c^{(2)}, \ldots, c^{(m)}$

(holding $\mu_1, \ldots, \mu_K$ fixed)

Repeat {

for $i$ = 1 to $m$

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

for $k$ = 1 to $K$

$\mu_k$ := average (mean) of points assigned to cluster $k$

}

# K-means algorithm

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

Cluster assignment step

Minimize $J(\ldots)$ wrt $c^{(1)}, c^{(2)}, \ldots, c^{(m)}$ ←

(holding $\mu_1, \ldots, \mu_K$ fixed)

Repeat {

for $i$ = 1 to $m$

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

move centroid

for $k$ = 1 to $K$

$\mu_k$ := average (mean) of points assigned to cluster $k$

}

minimize $J(\ldots)$ wrt $\mu_1, \ldots, \mu_K$

Andrew Ng

# Random Initialization

*Clustering*

Unsupervised Learning

# 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 $K$

$\mu_k$ := average (mean) of points assigned to cluster $k$

}

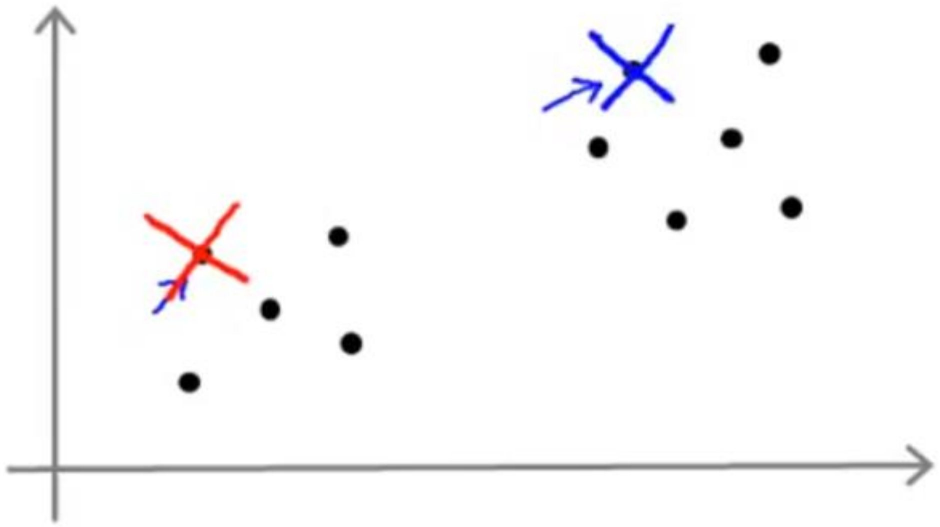
## Random initialization

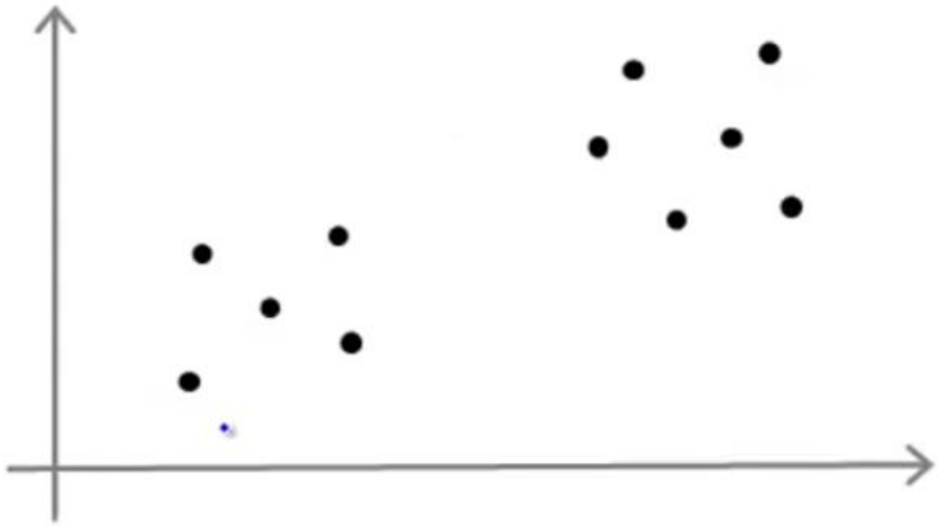
$K=2$

Should have $K < m$

Randomly pick $K$ training examples.

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





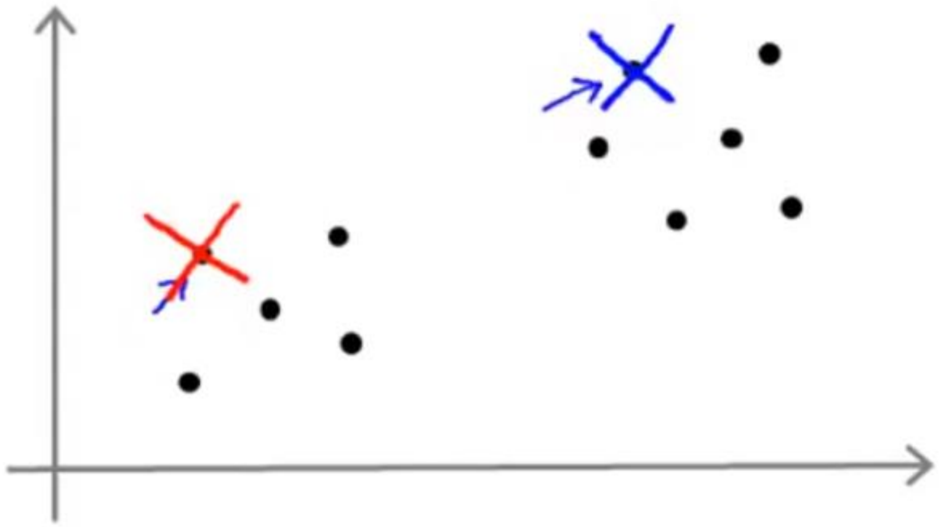
# Random initialization

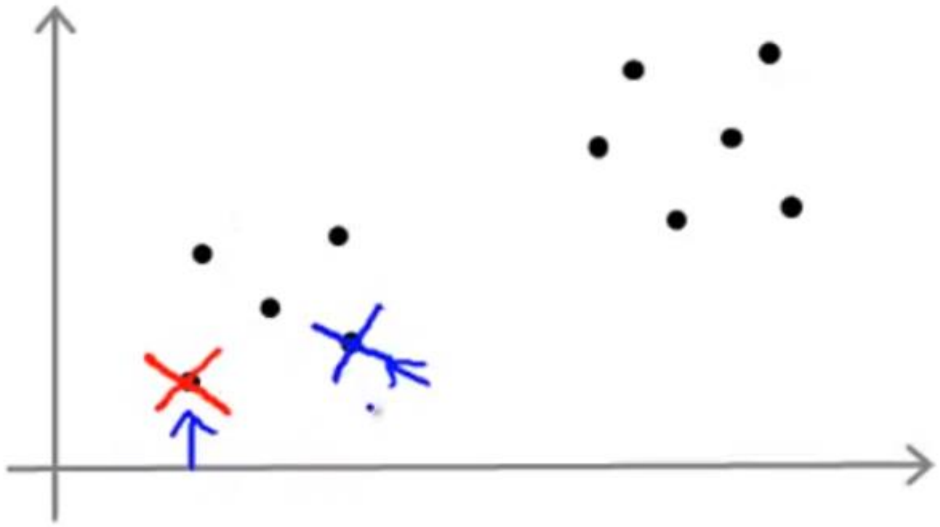
$K=2$

Should have $K < m$

Randomly pick $K$ training examples.

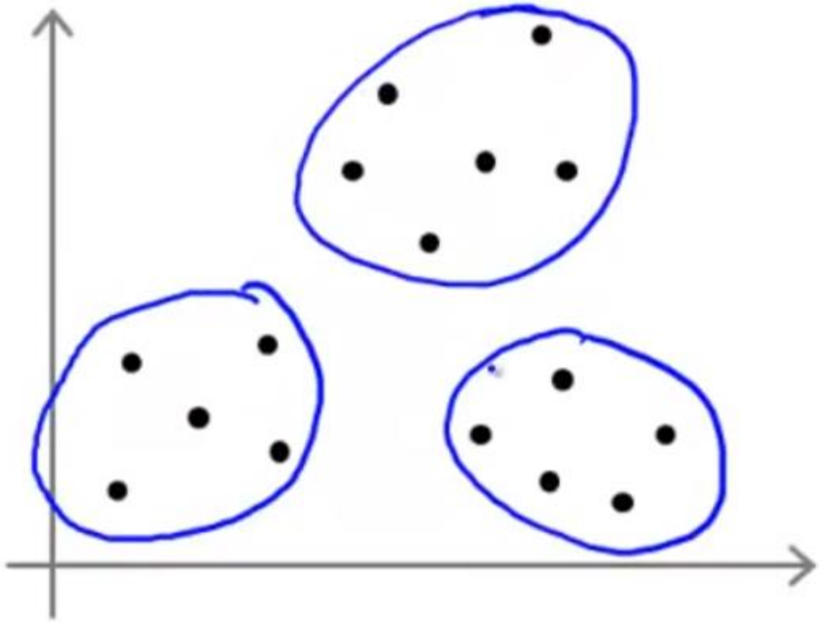
Set $\mu_1, \ldots, \mu_K$ equal to these $K$ examples.

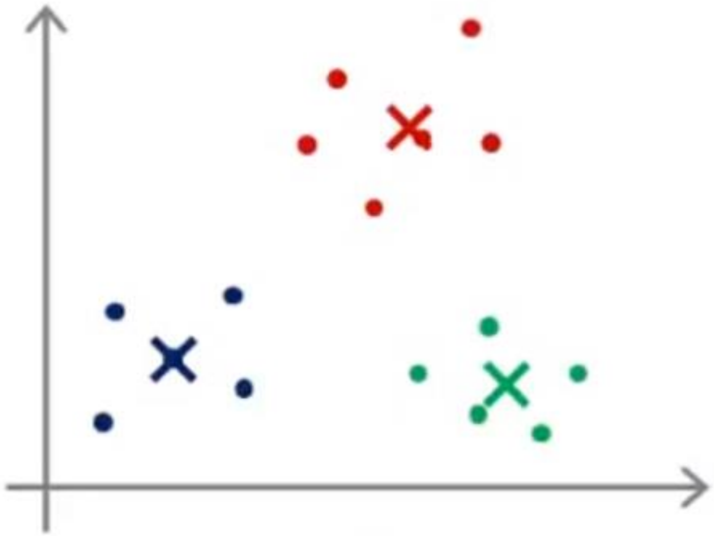




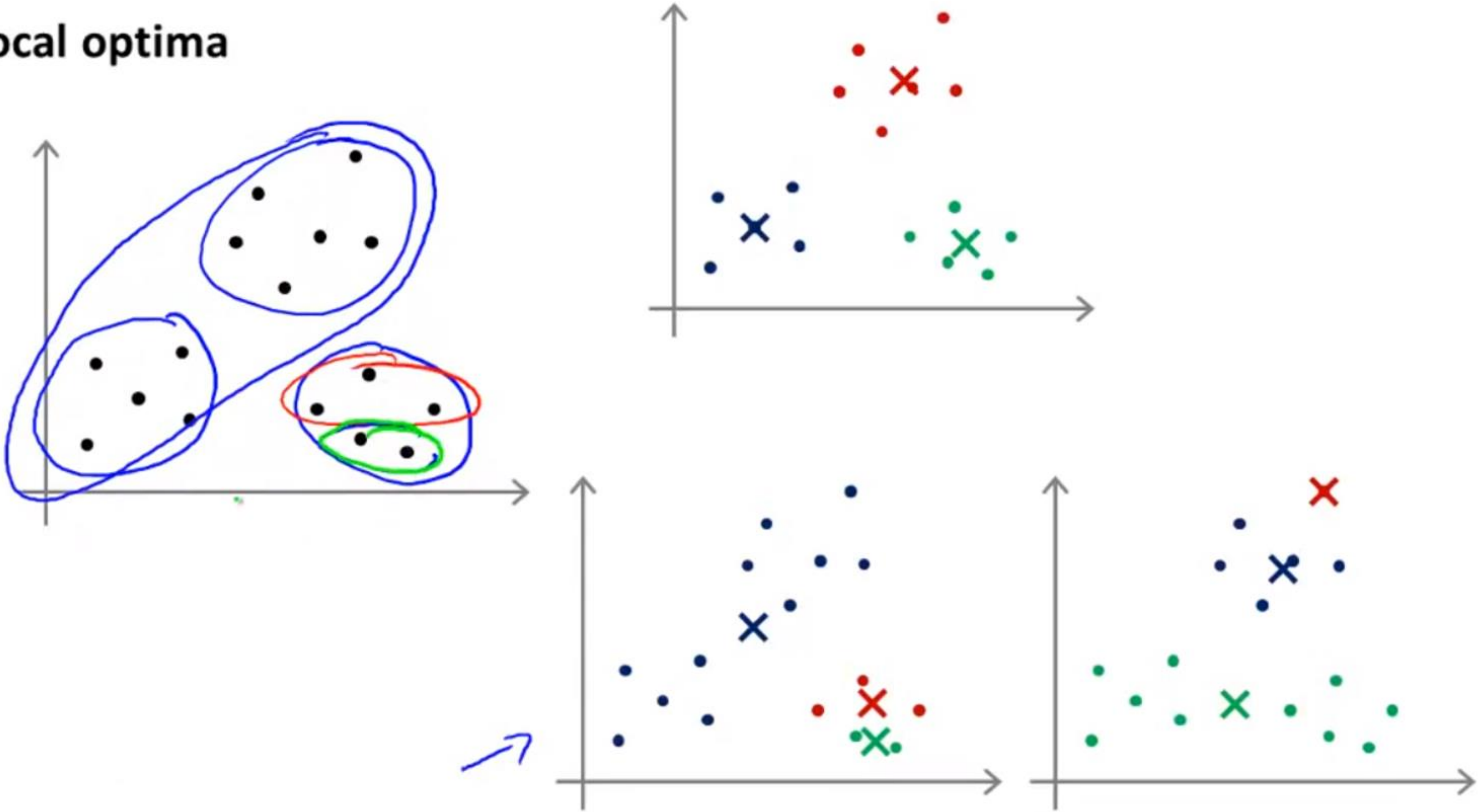
# Local optima

# Local optima





# 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)

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

}

# Random initialization

For i = 1 to 100 {  50 - 1000

Randomly initialize K-means.

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

Compute cost function (distortion)

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

}

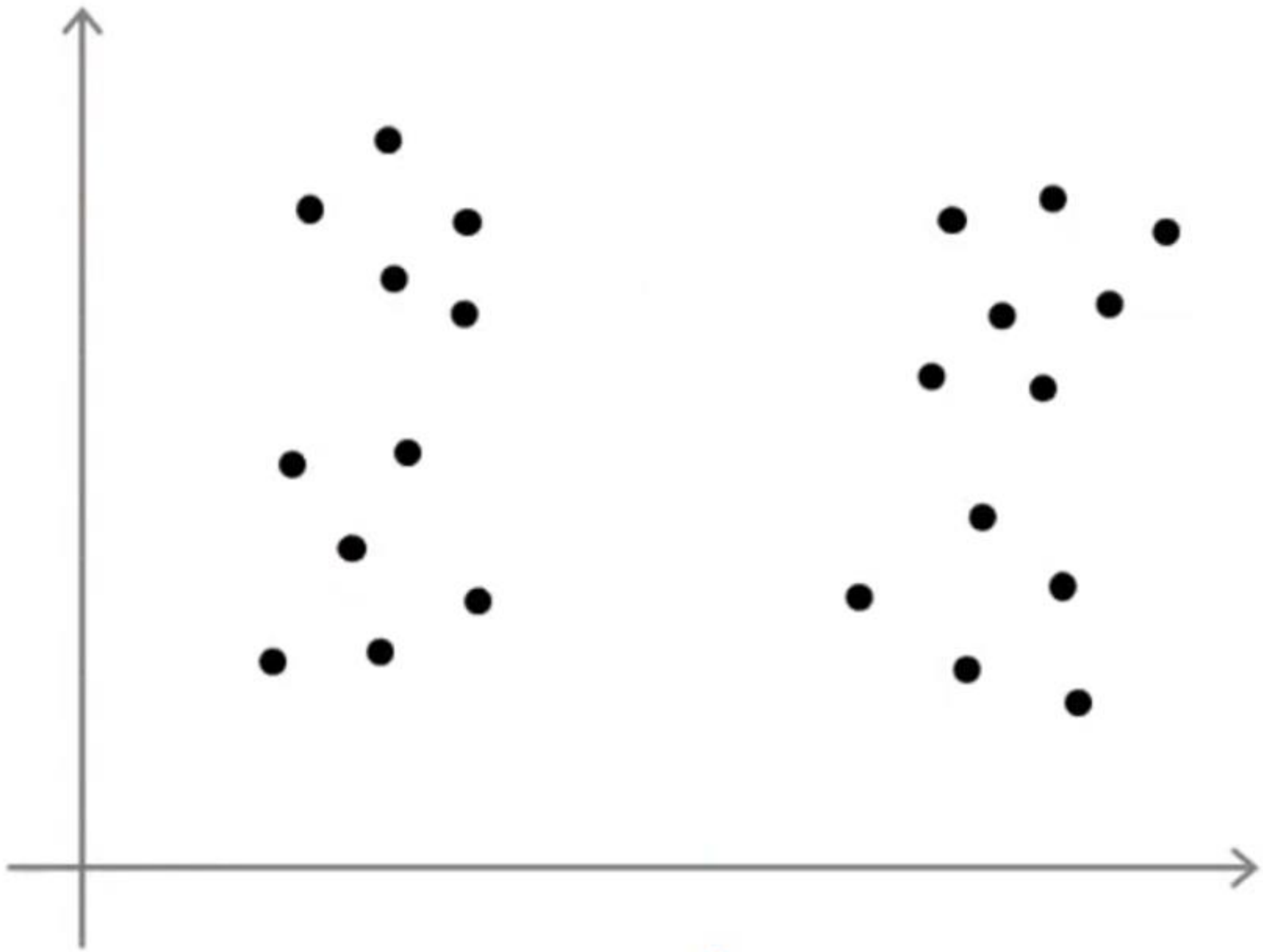
Pick clustering that gave lowest cost $J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K)$

# Choosing K

*Clustering*

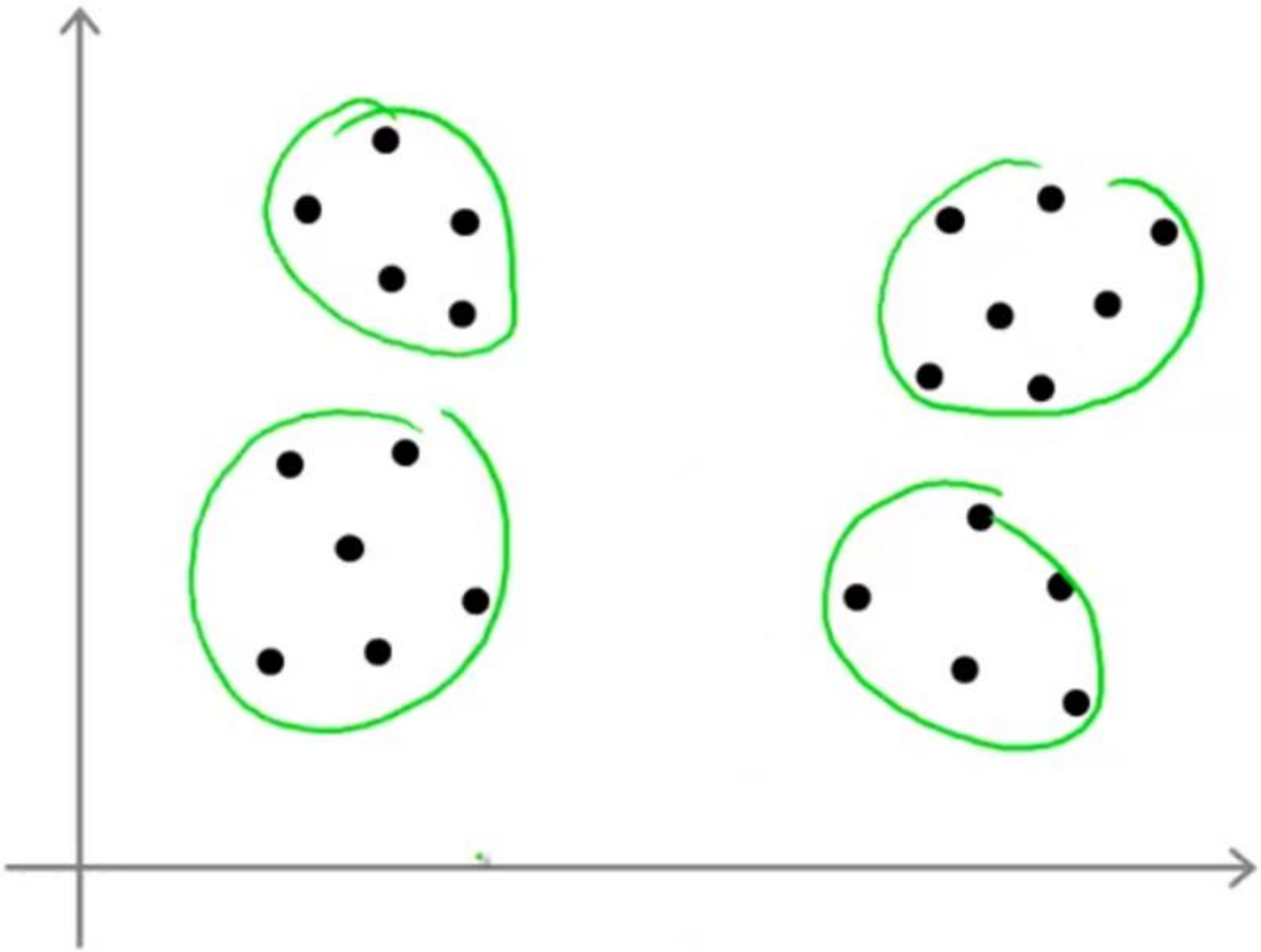
Unsupervised Learning

# What is the right value of K?

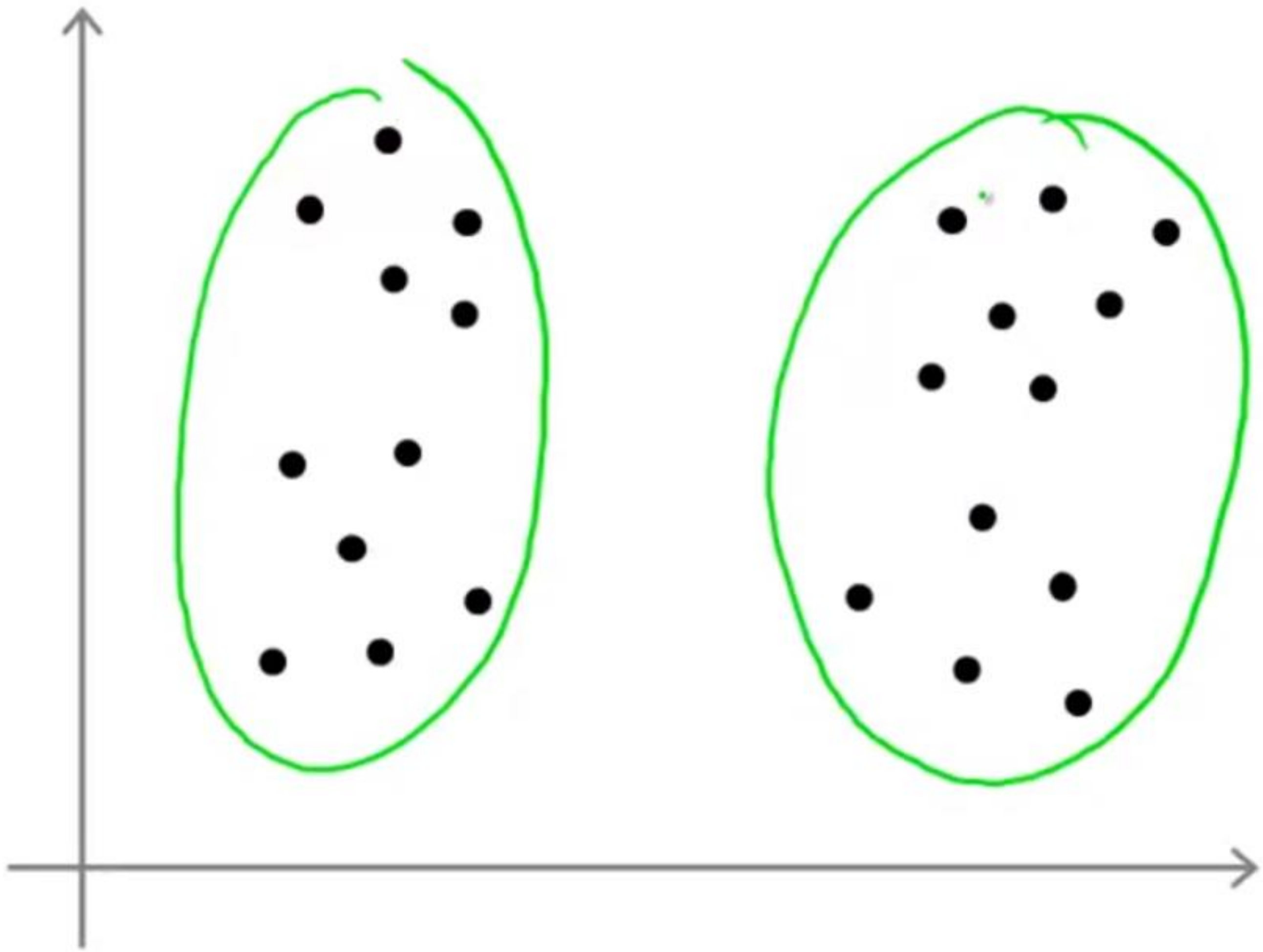


Andrew Ng

# What is the right value of K?

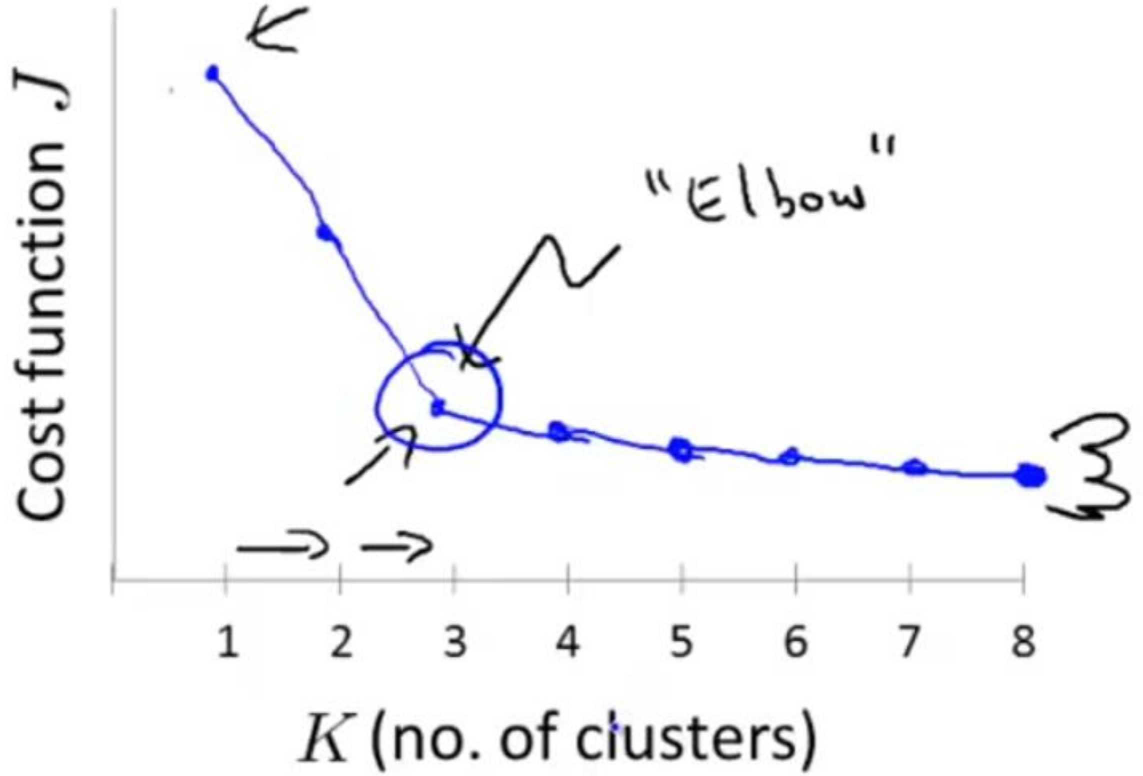


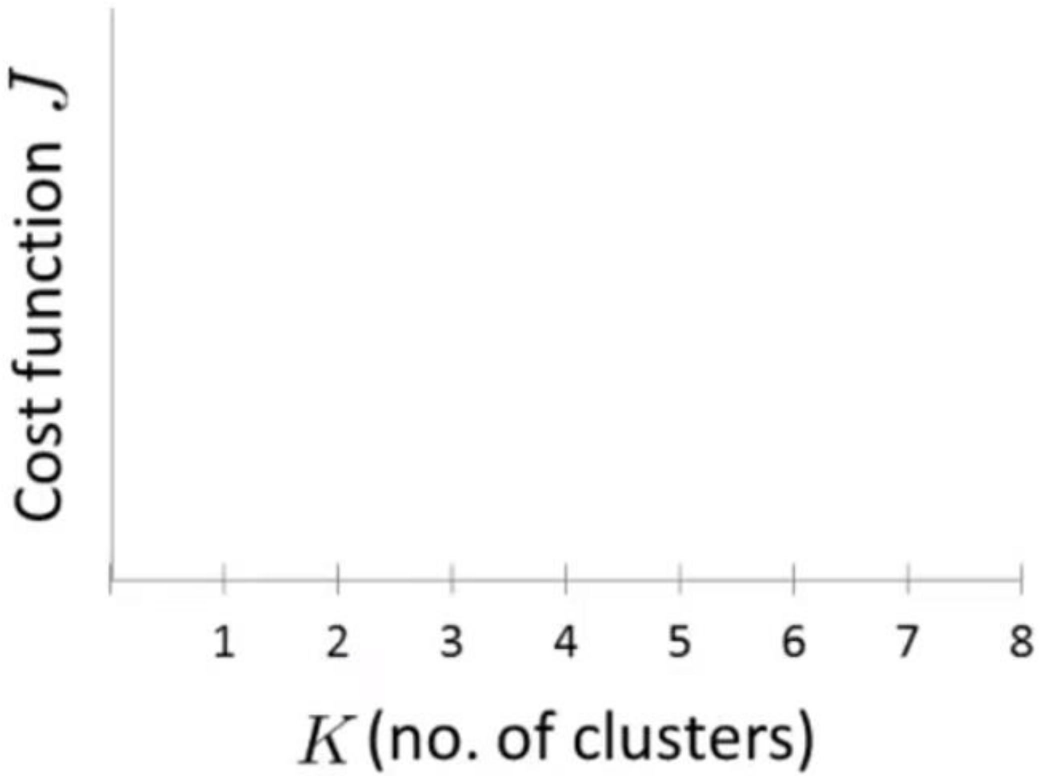
# What is the right value of K?



# Choosing the value of K

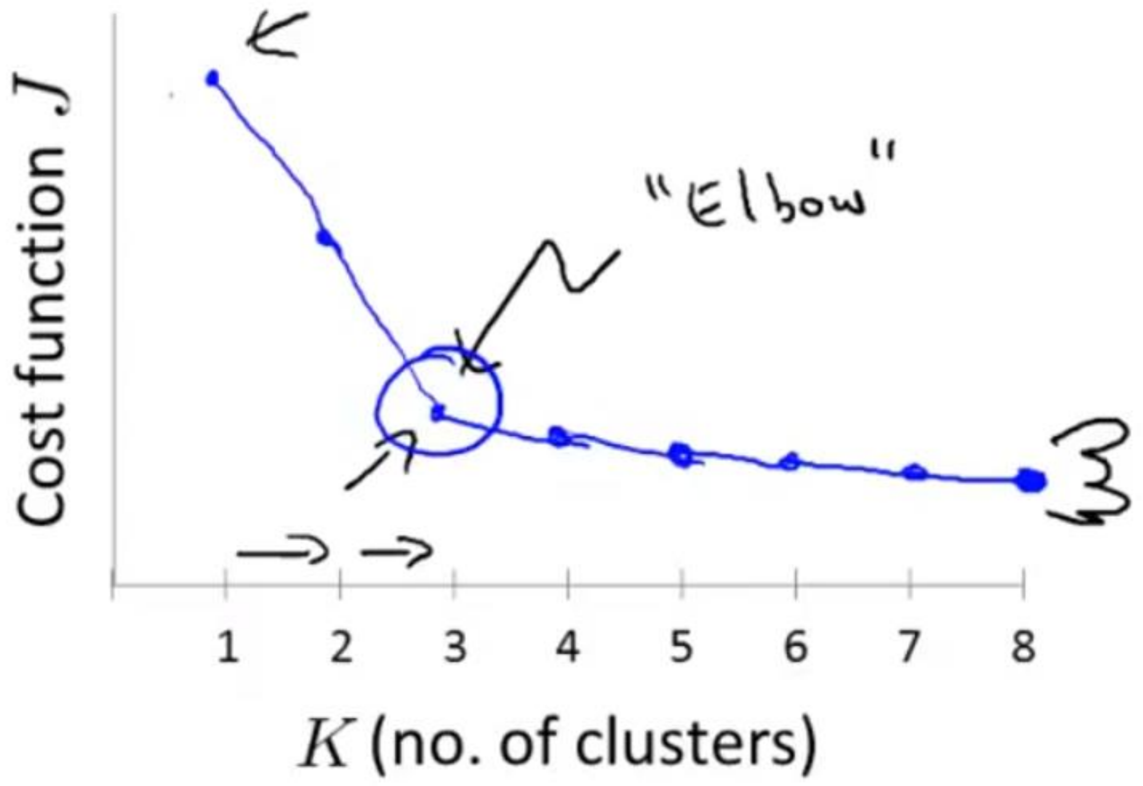
Elbow method:

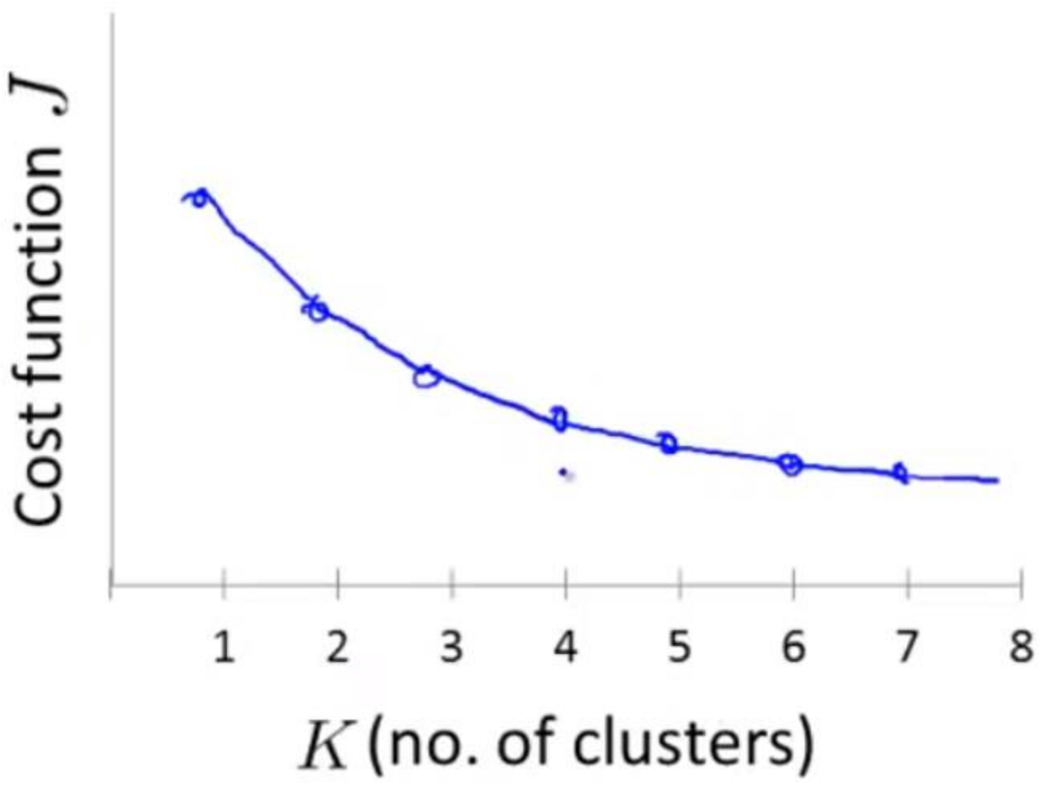


## Choosing the value of K

Elbow method:



## Choosing the value of K

Sometimes, you're running K-means to get clusters to use for some later/downstream purpose. Evaluate K-means based on a metric for how well it performs for that later purpose.

K=3 S, M, L K=5 XS, S, M, L, XL

E.g.