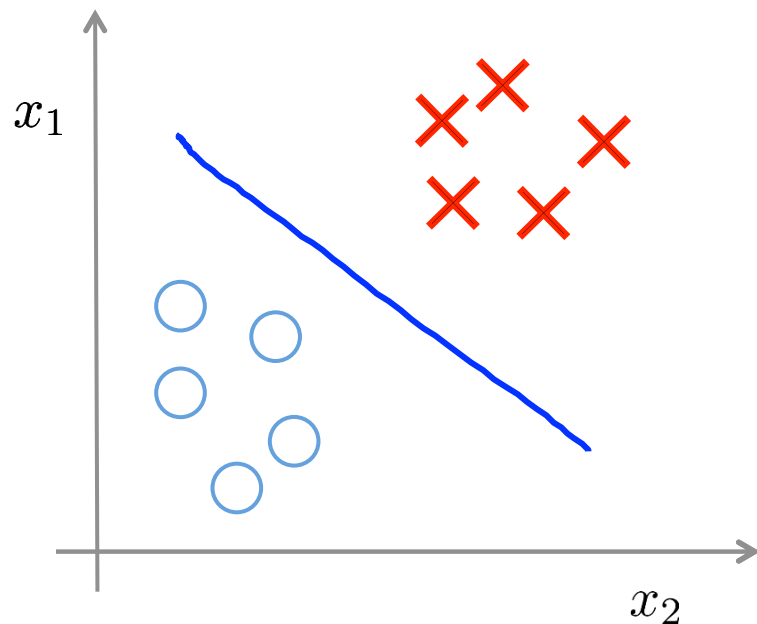


Machine Learning

Clustering

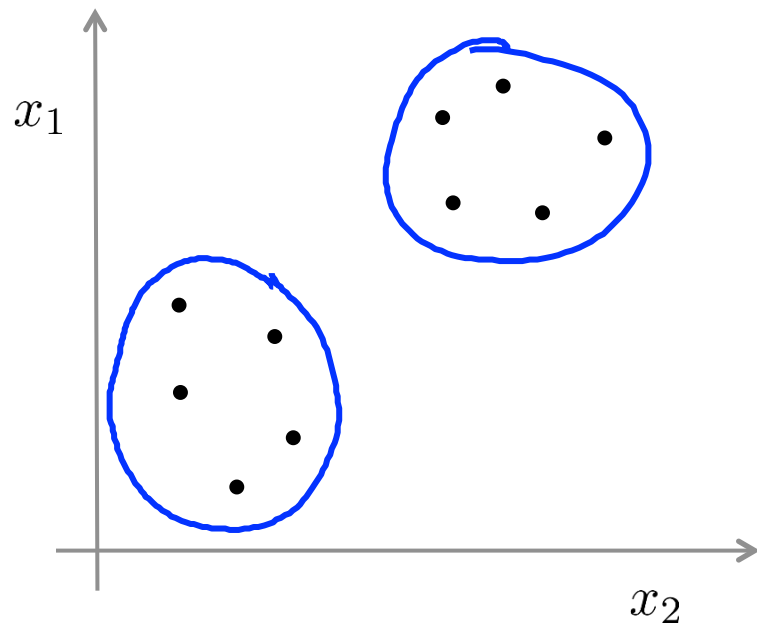
Unsupervised learning
introduction

Supervised learning



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

Unsupervised learning



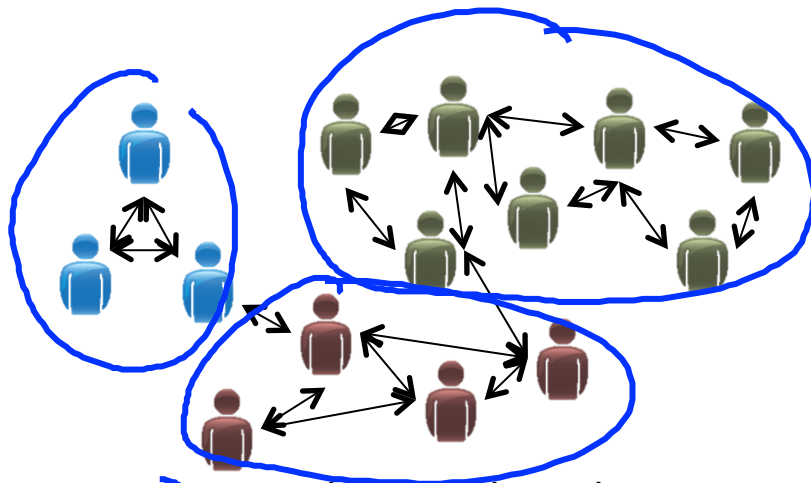
Clustering algorithm

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

Applications of clustering



→ Market segmentation



→ Social network analysis



→ Organize computing clusters

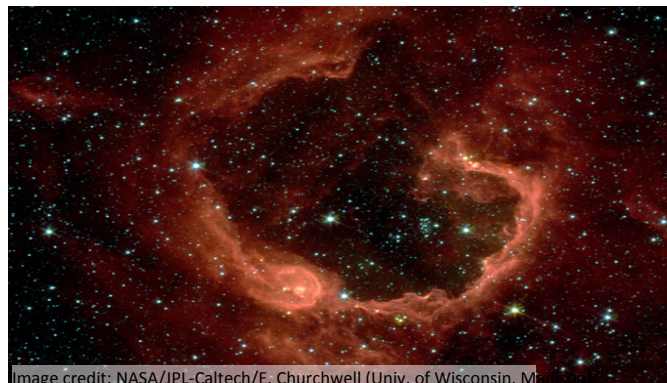
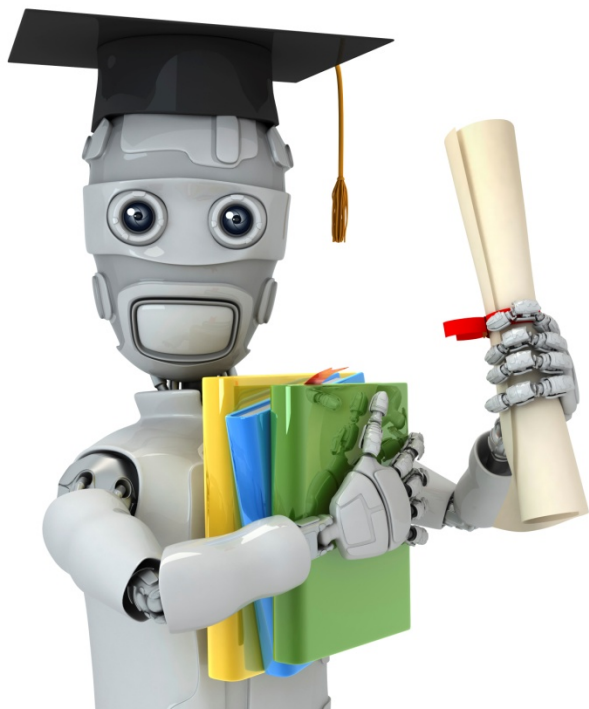


Image credit: NASA/JPL-Caltech/E. Churchwell (Univ. of Wisconsin, M

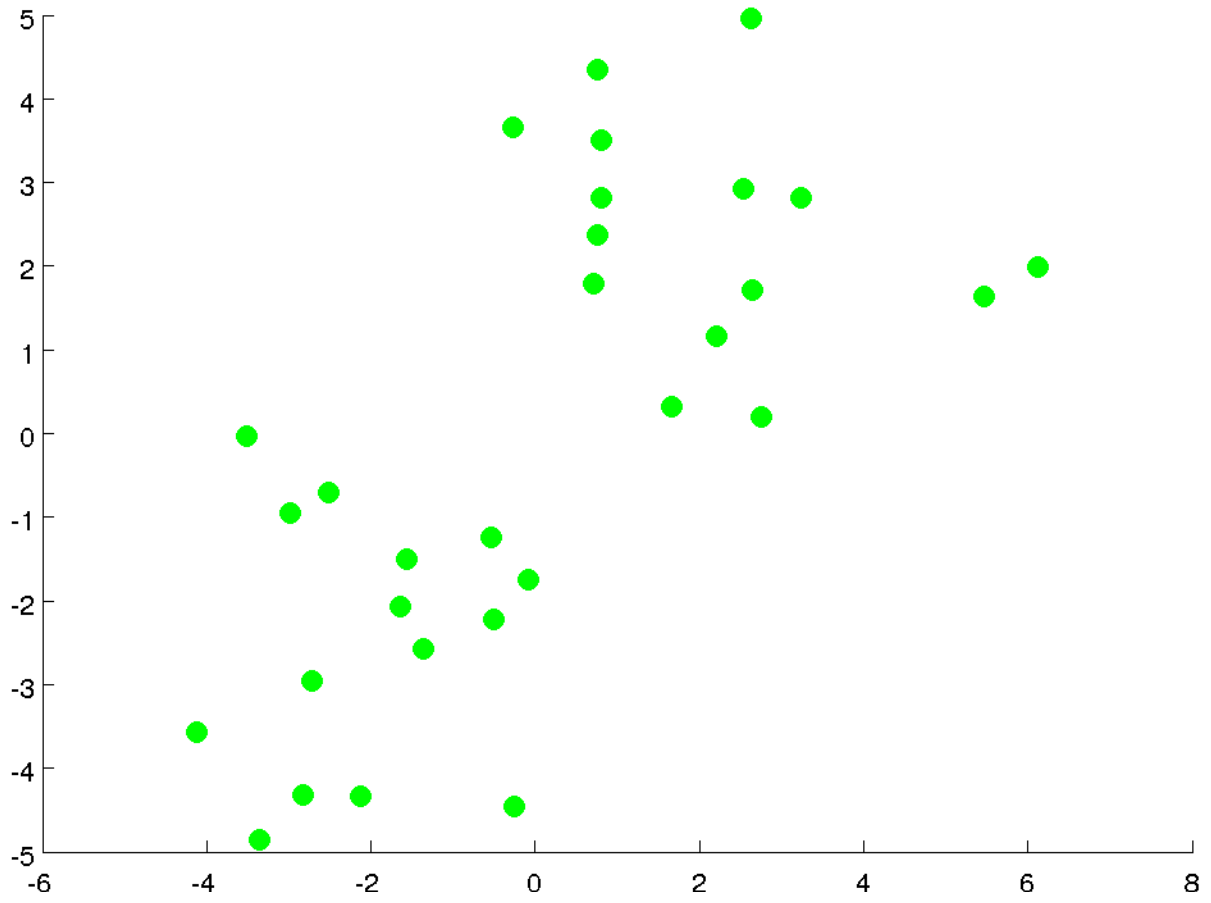
→ Astronomical data analysis

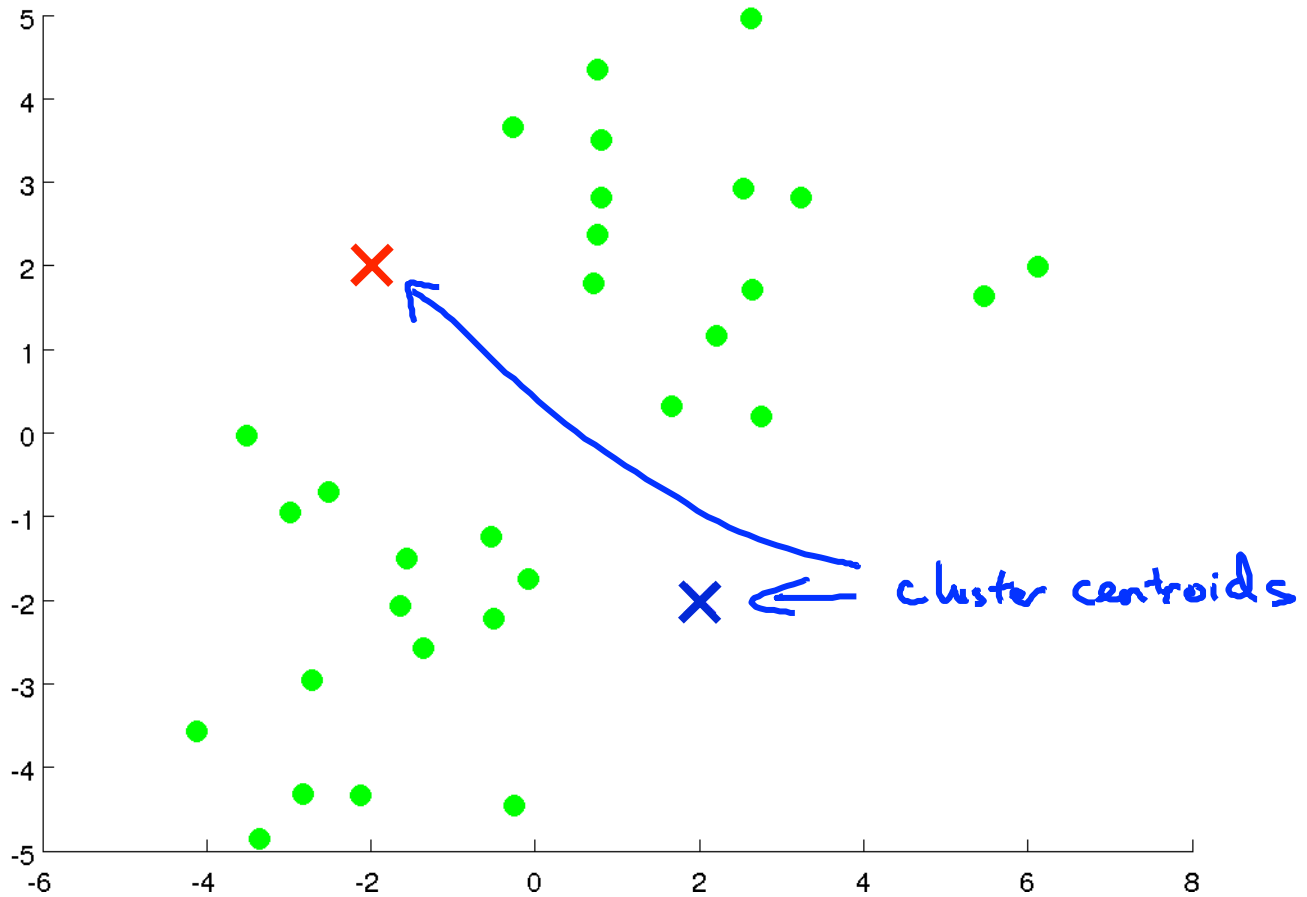


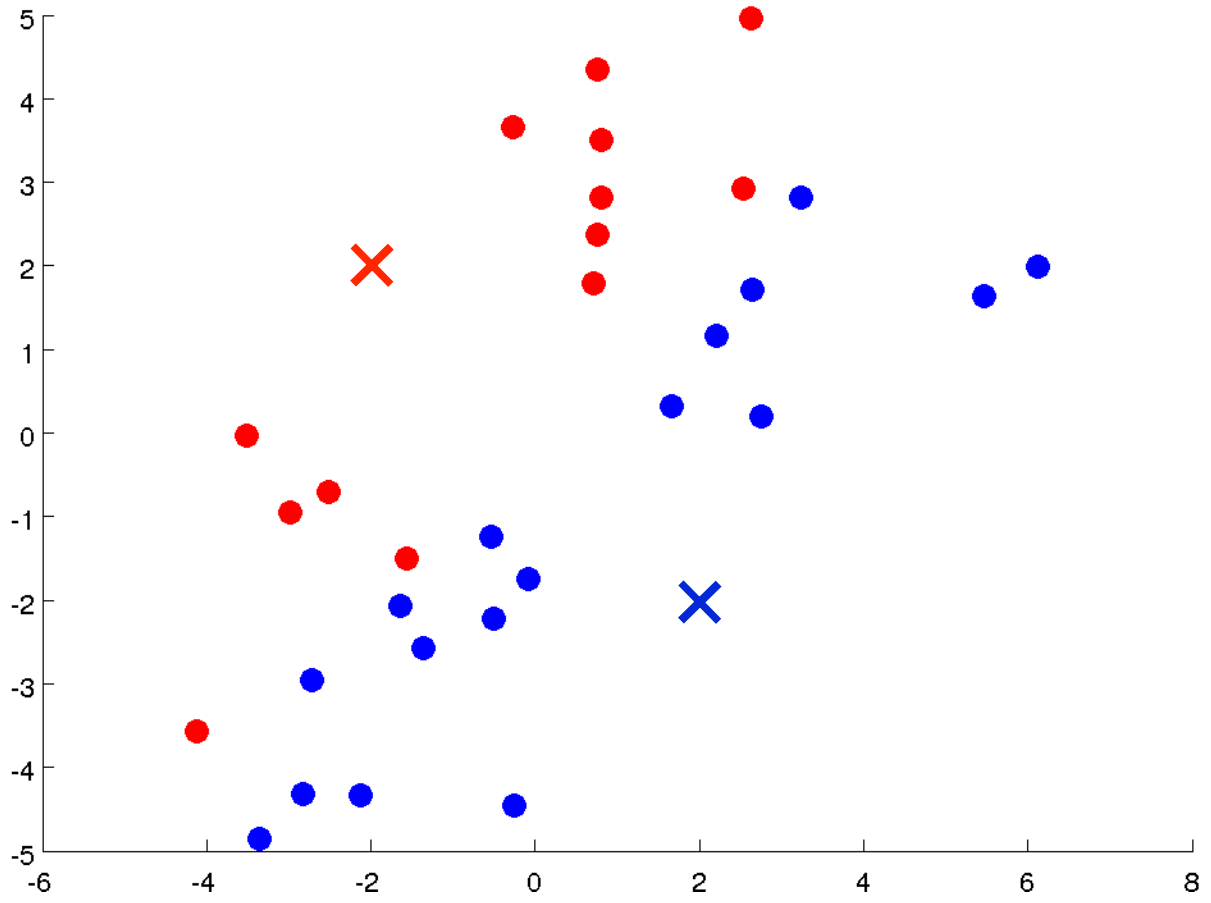
Machine Learning

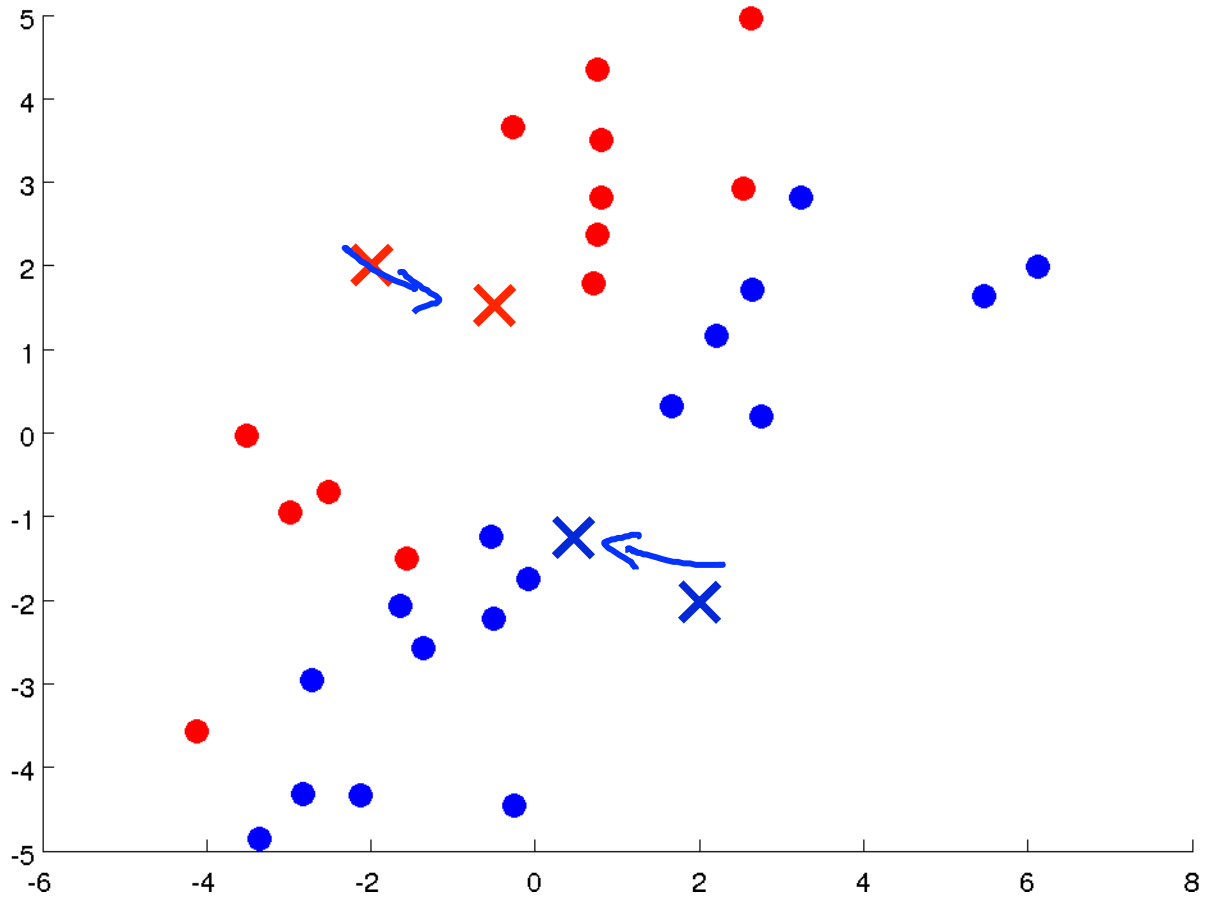
Clustering

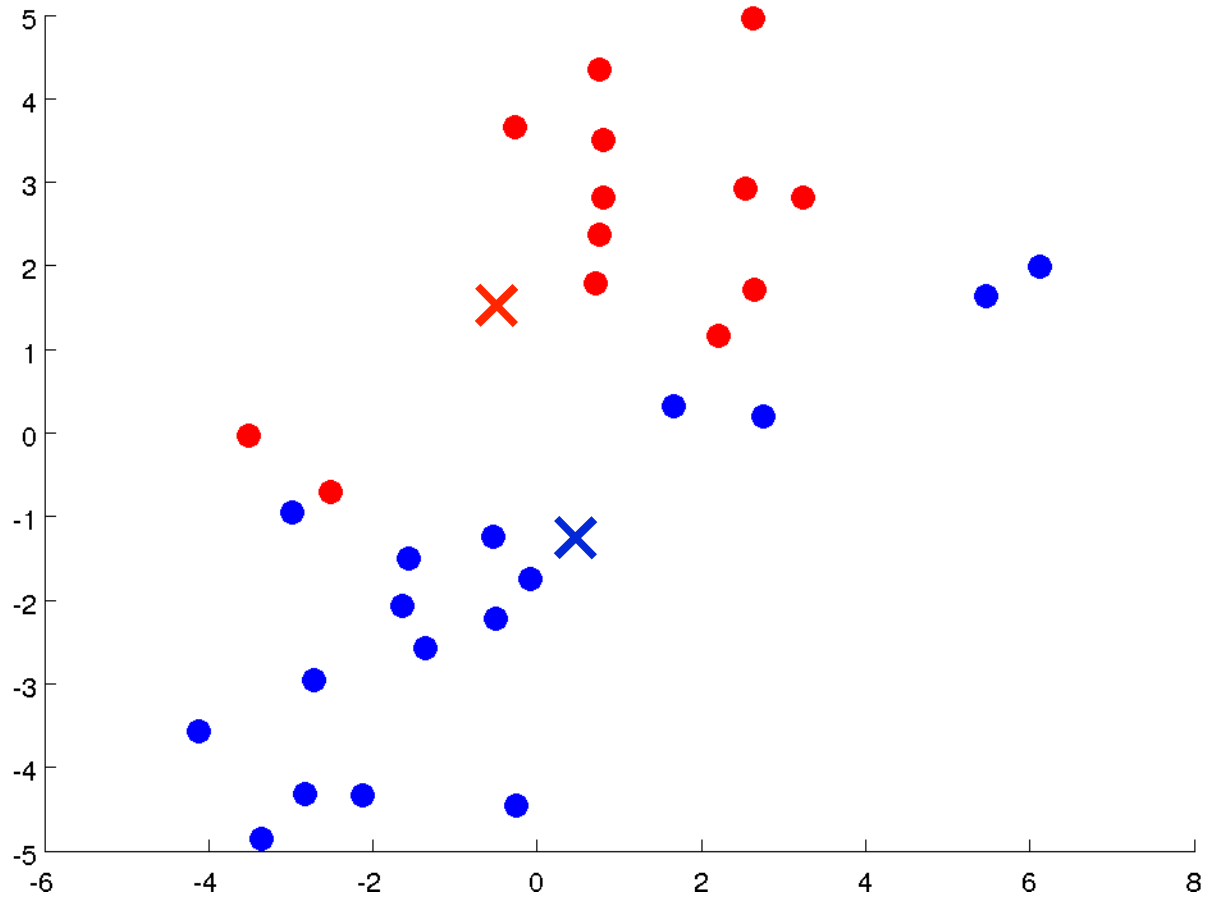
K-means
algorithm

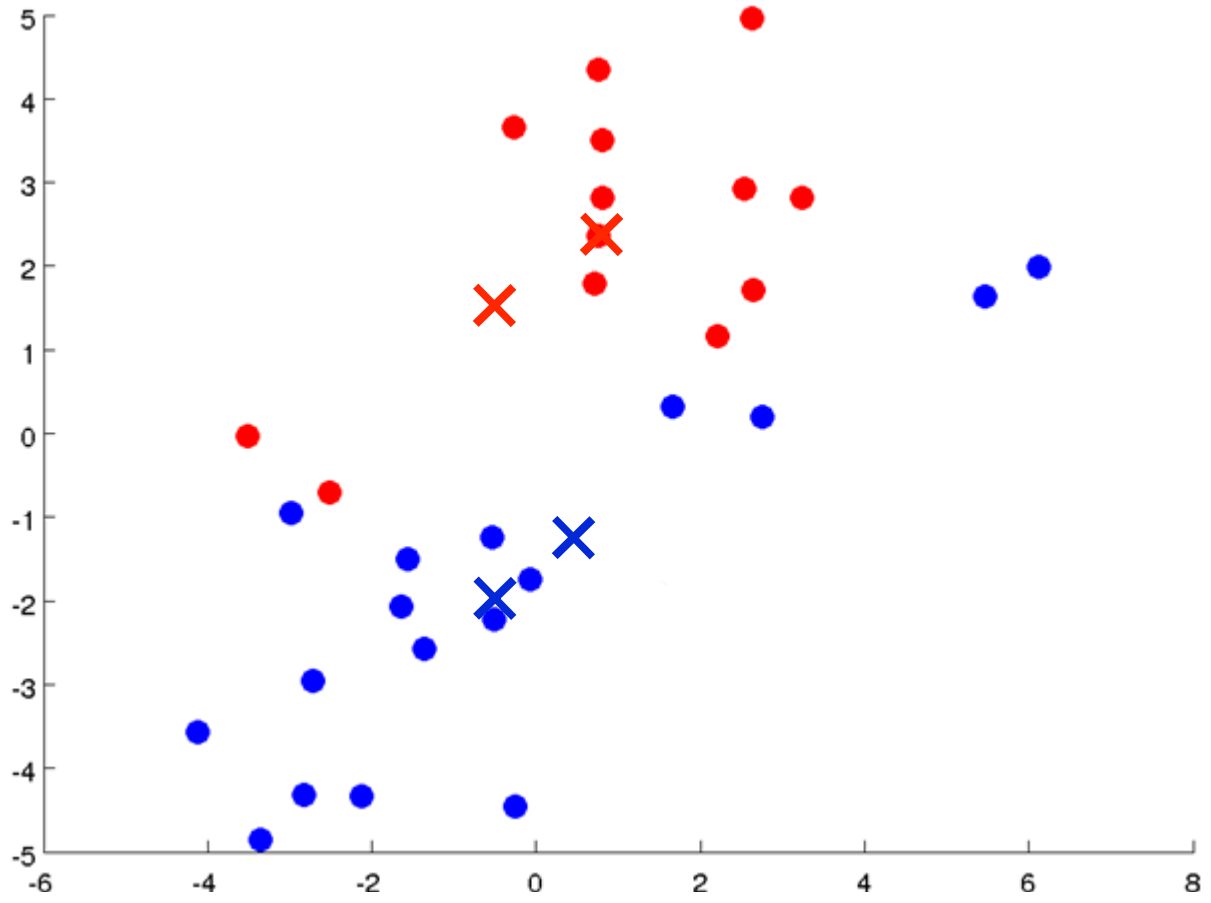


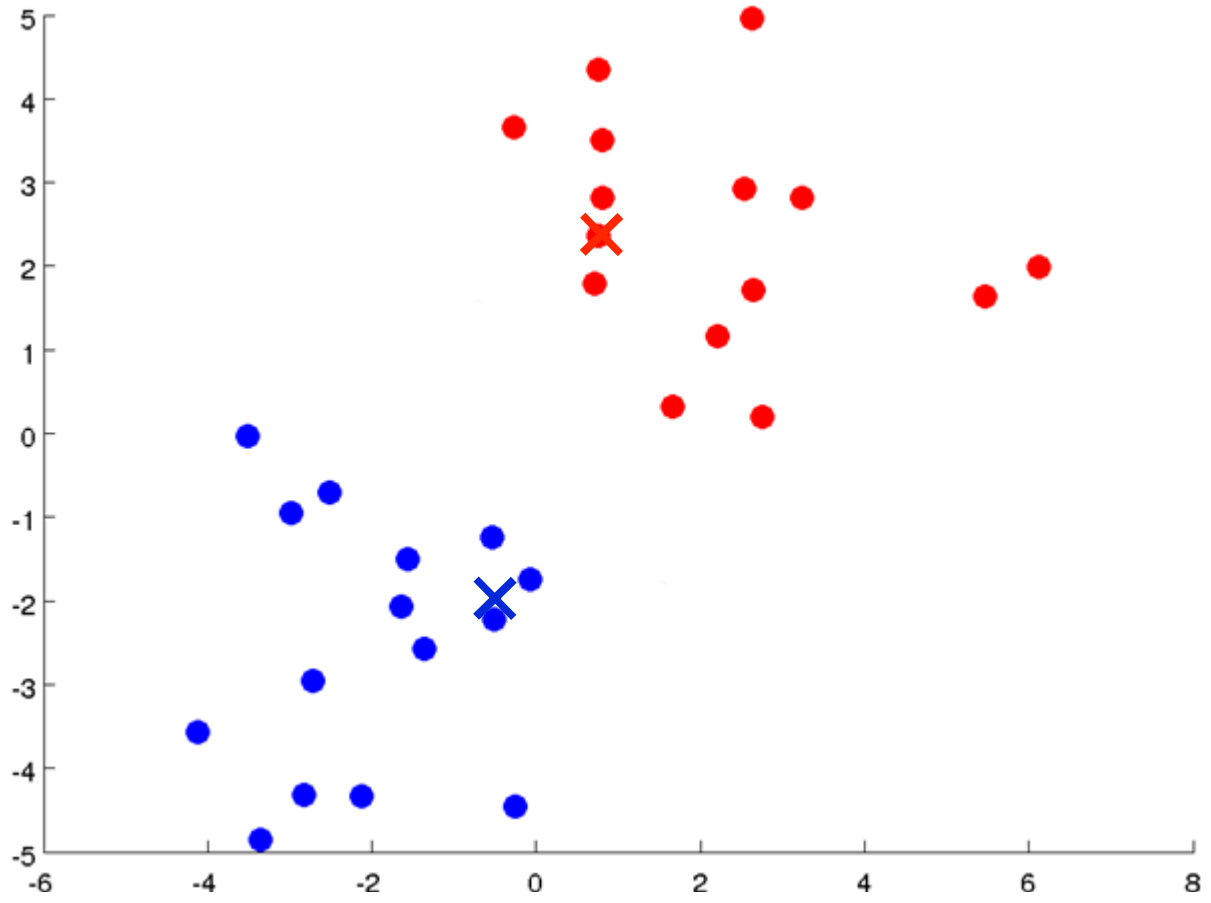


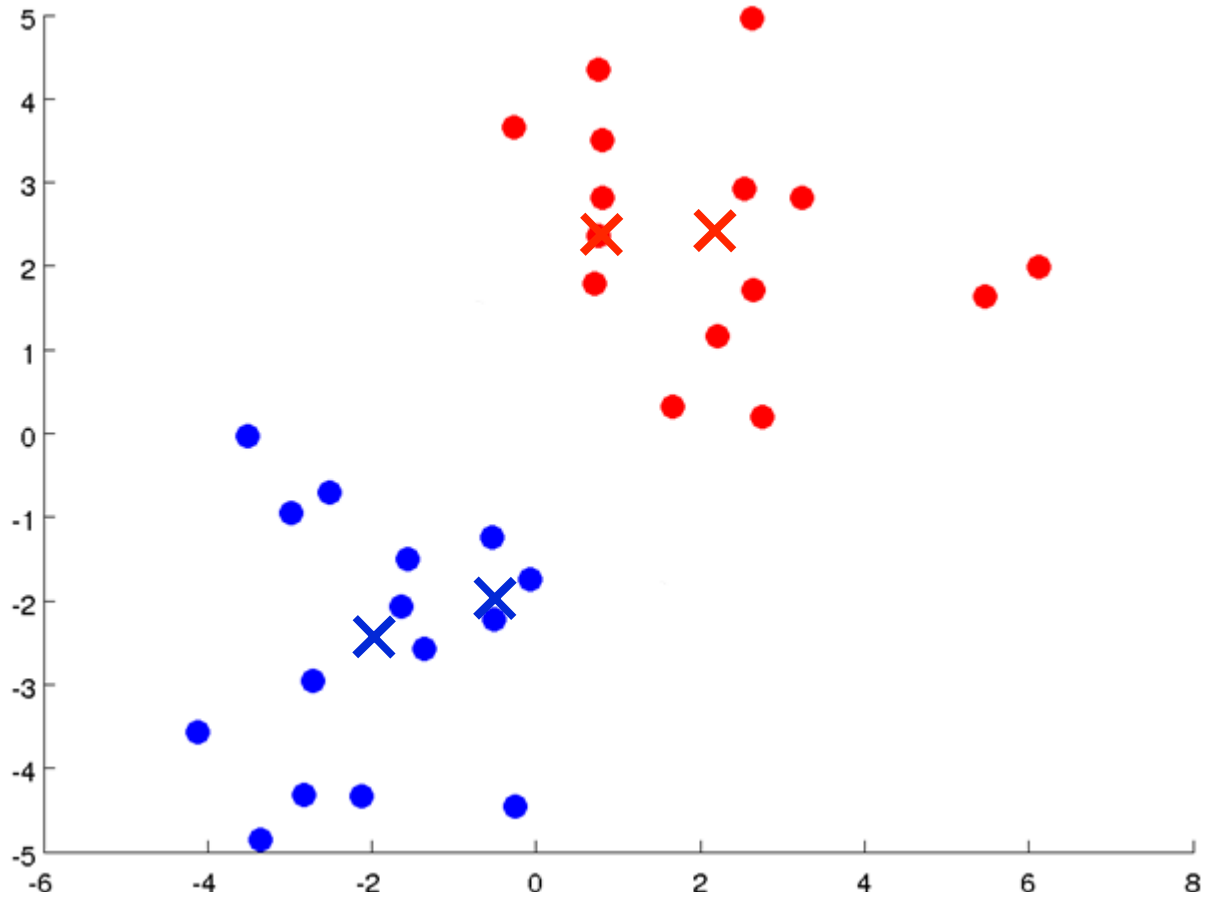


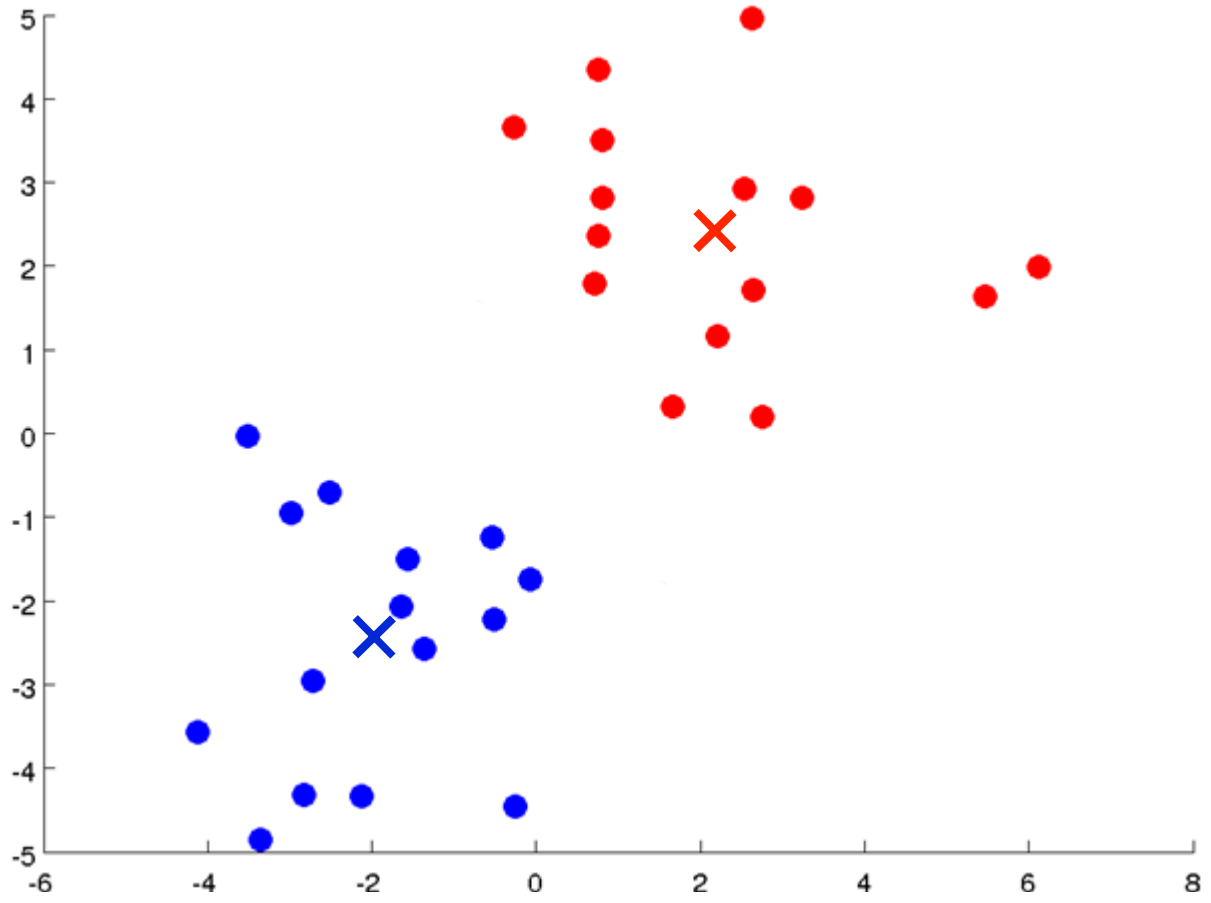














K-means algorithm

Input:

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

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

K-means algorithm

$$\mu_1 \quad \mu_2$$

Randomly initialize K cluster centroids $\underline{\mu}_1, \underline{\mu}_2, \dots, \underline{\mu}_K \in \mathbb{R}^n$

Repeat {

Cluster assignment step

for $i = 1$ to m

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

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

for $k = 1$ to K

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

Move centroid

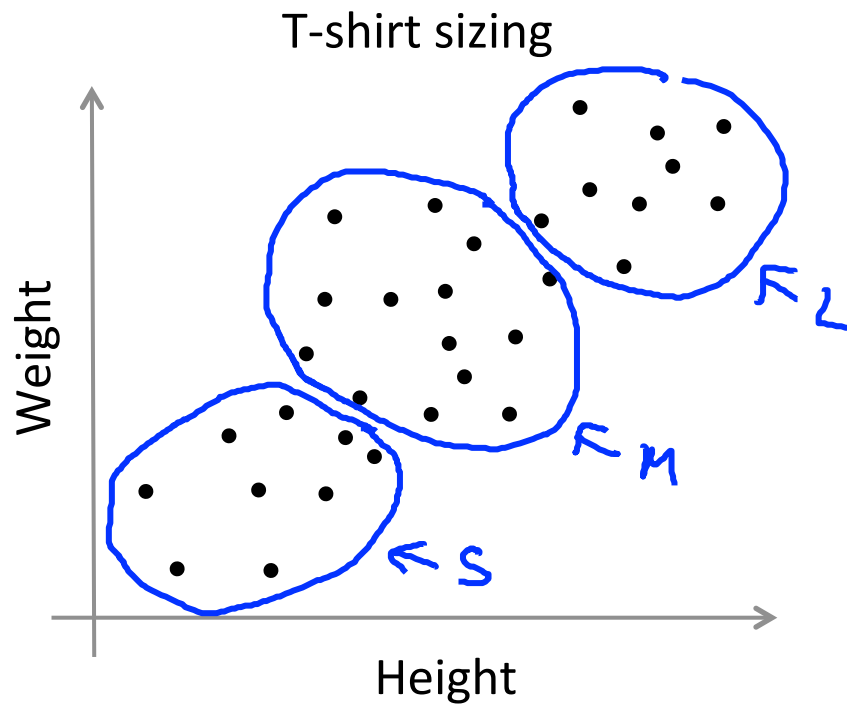
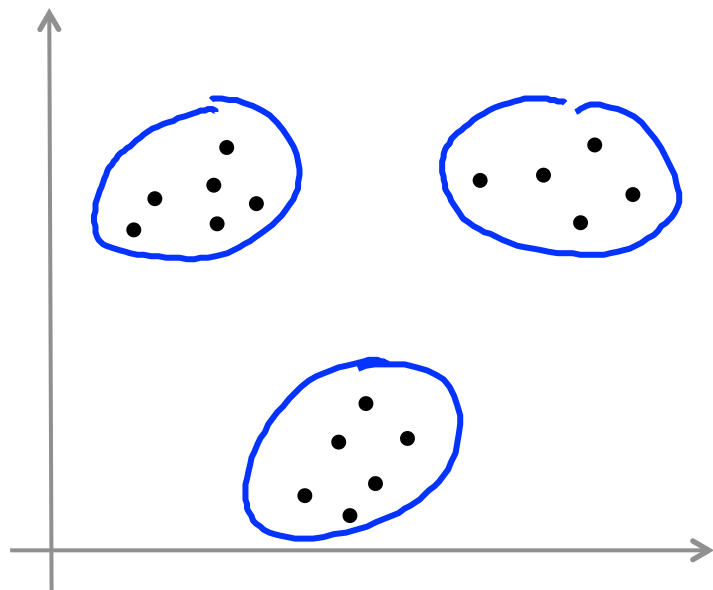
$$x^{(1)}, x^{(5)}, x^{(6)}, x^{(10)}$$

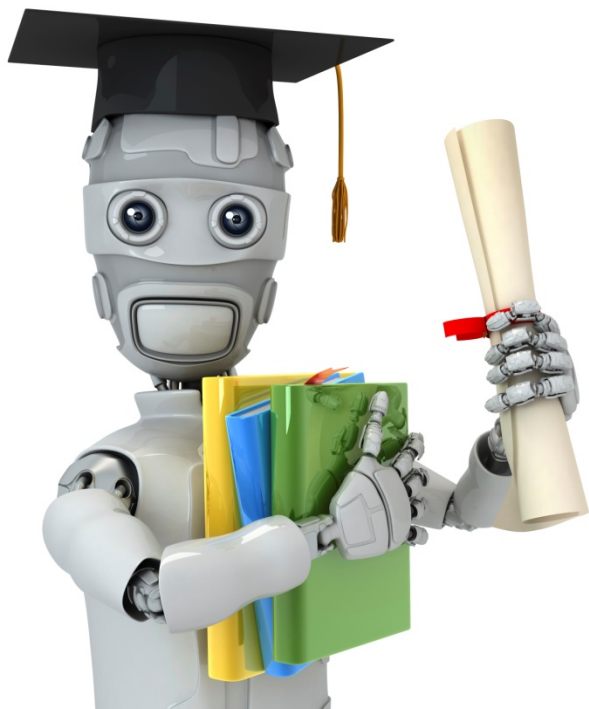
$$\rightarrow c^{(1)}=2, c^{(5)}=2, c^{(6)}=2, c^{(10)}=2$$

$$\mu_2 = \frac{1}{4} \begin{bmatrix} x^{(1)} + x^{(5)} + x^{(6)} + x^{(10)} \\ - \\ - \\ - \end{bmatrix} \in \mathbb{R}^n$$

K-means for non-separated clusters

S, M, L





Machine Learning

Clustering Optimization objective

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

K $k \in \{1, 2, \dots, K\}$
 $x^{(i)} \rightarrow 5$ $c^{(i)} = 5$ $\mu_{c^{(i)}} = \mu_5$

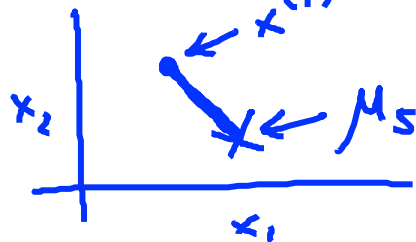
Optimization objective:

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

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

→ μ_1, \dots, μ_K

Distortion



K-means algorithm

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

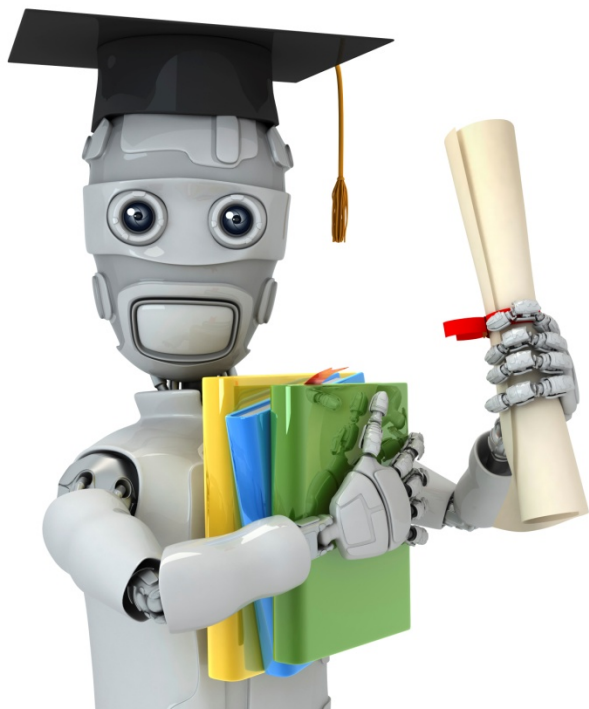
Repeat {

Cluster assignment step
Minimize $J(\dots)$ w.r.t. $c^{(1)}, c^{(2)}, \dots, c^{(m)}$ ←
(holding μ_1, \dots, μ_K fixed)

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 $J(\dots)$ w.r.t. μ_1, \dots, μ_K



Machine Learning

Clustering Random initialization

K-means algorithm

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

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

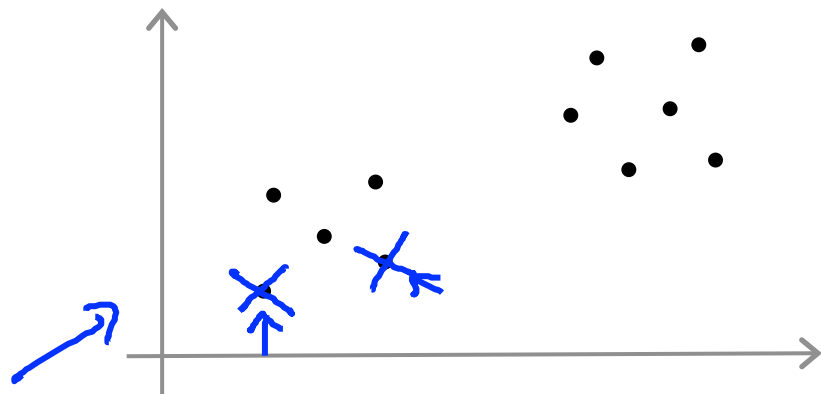
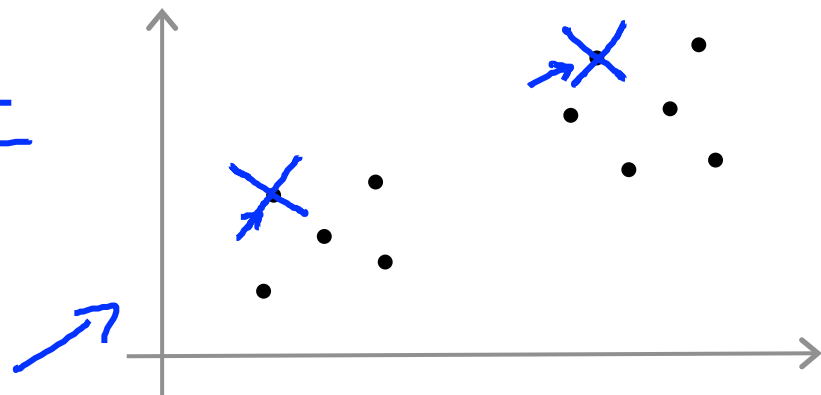
Should have $K < m$

Randomly pick K training examples.

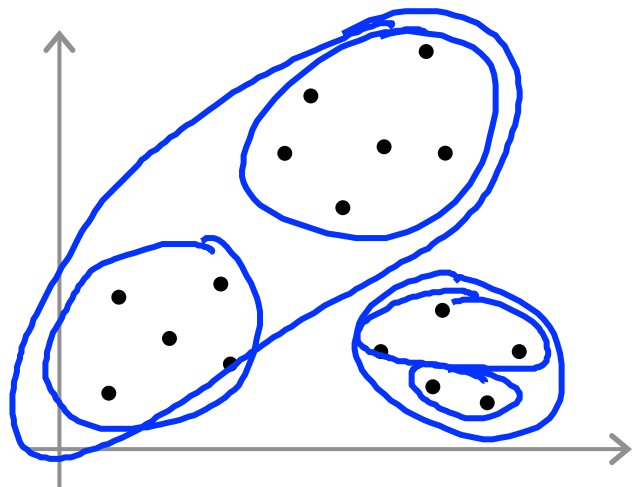
Set μ_1, \dots, μ_K equal to these K examples.

$$\begin{aligned}\mu_1 &= x^{(i)} \\ \mu_2 &= x^{(j)} \\ &\vdots\end{aligned}$$

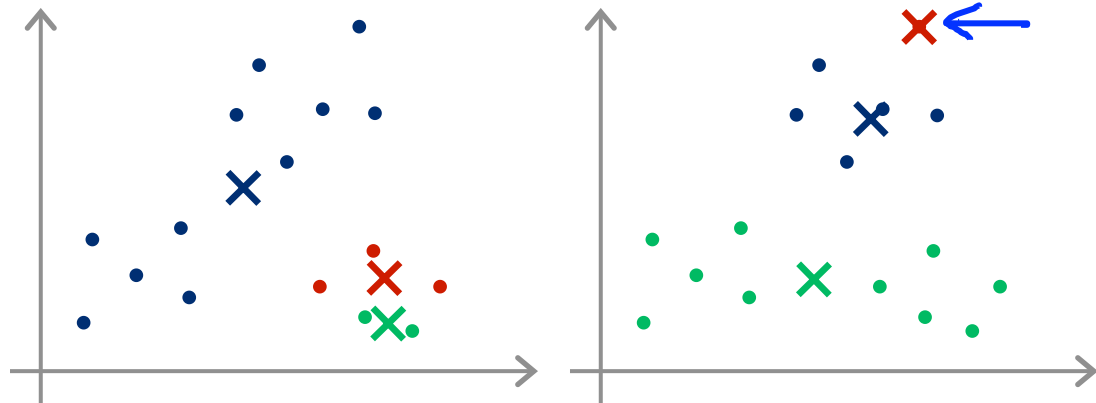
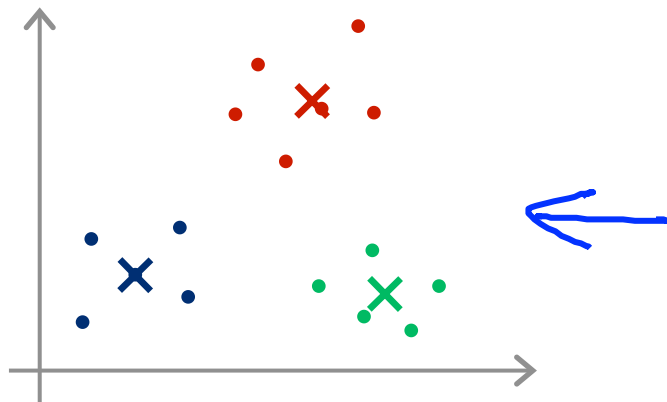
$K=2$



Local optima



$$J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_k)$$



Random initialization

For $i = 1$ to 100 {

Randomly initialize K-means.

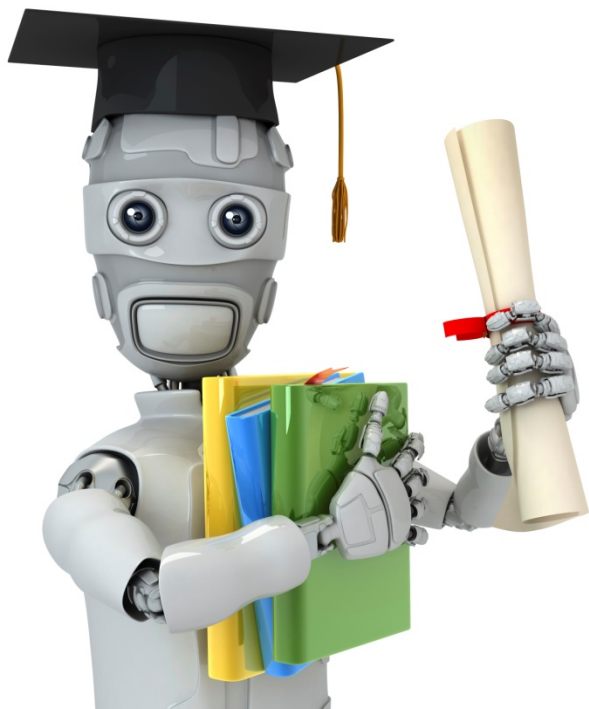
Run K-means. Get $c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K$.

Compute cost function (distortion)

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

}

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

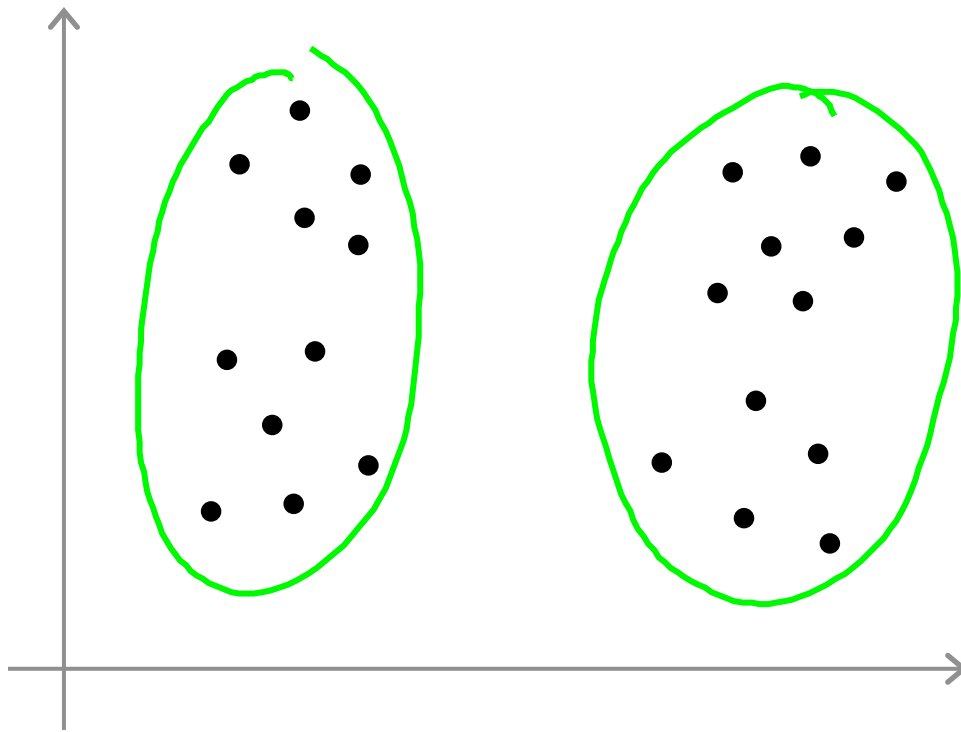


Machine Learning

Clustering

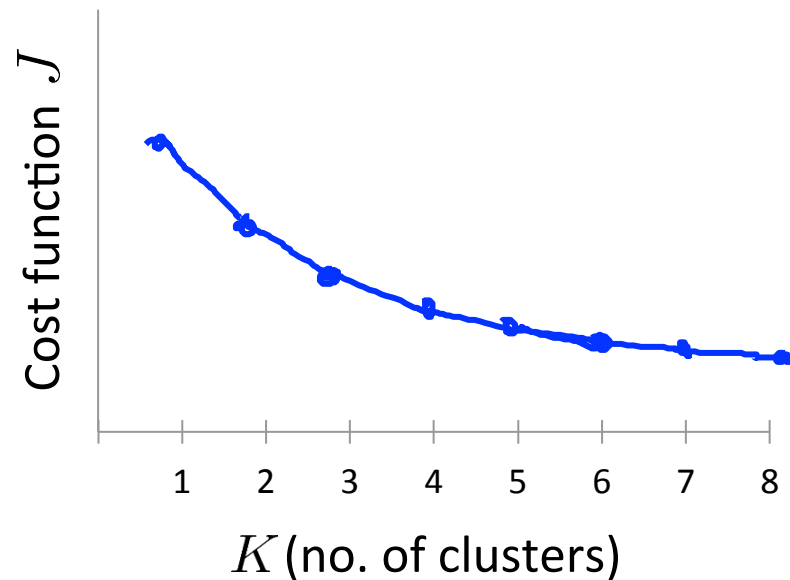
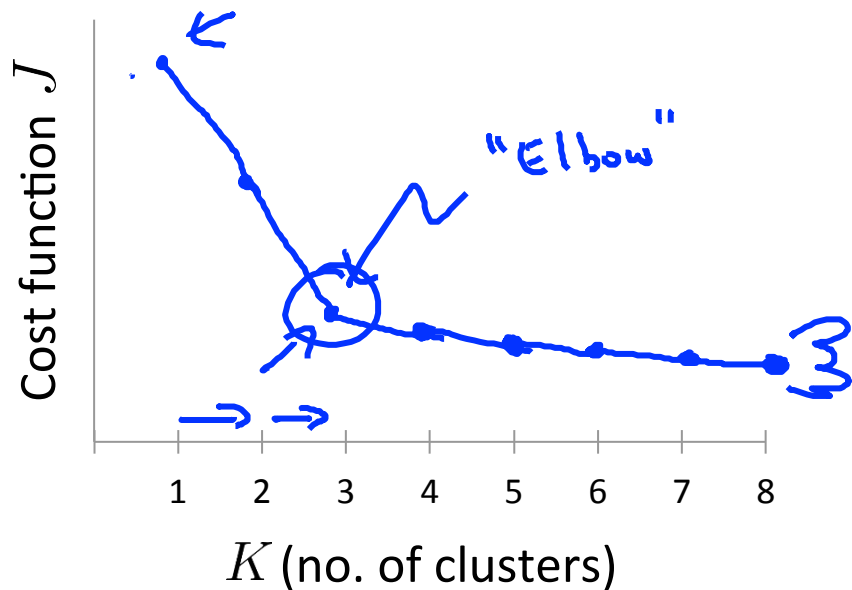
Choosing the
number of clusters

What is the right value of K?



Choosing the value of K

Elbow method:

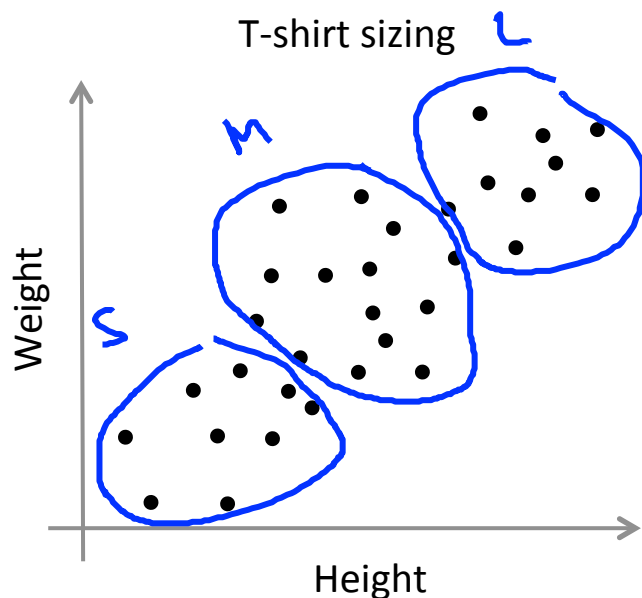


Choosing the value of K

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

E.g.



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

