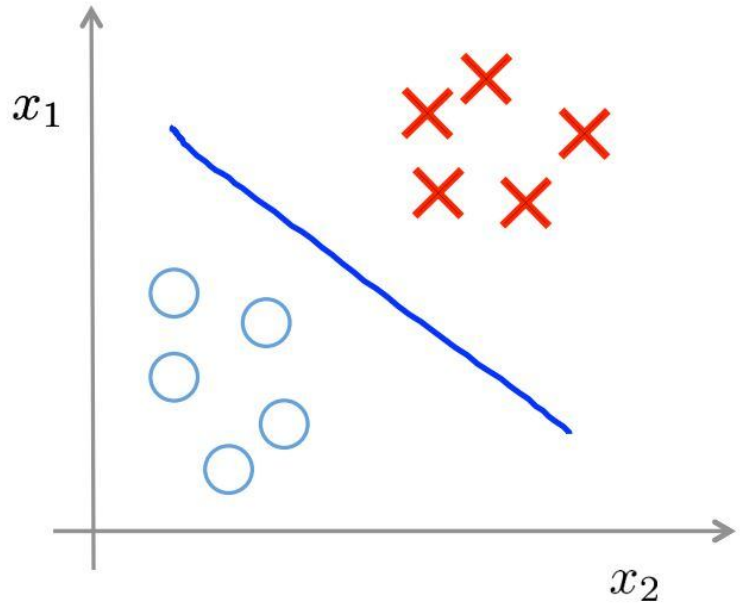
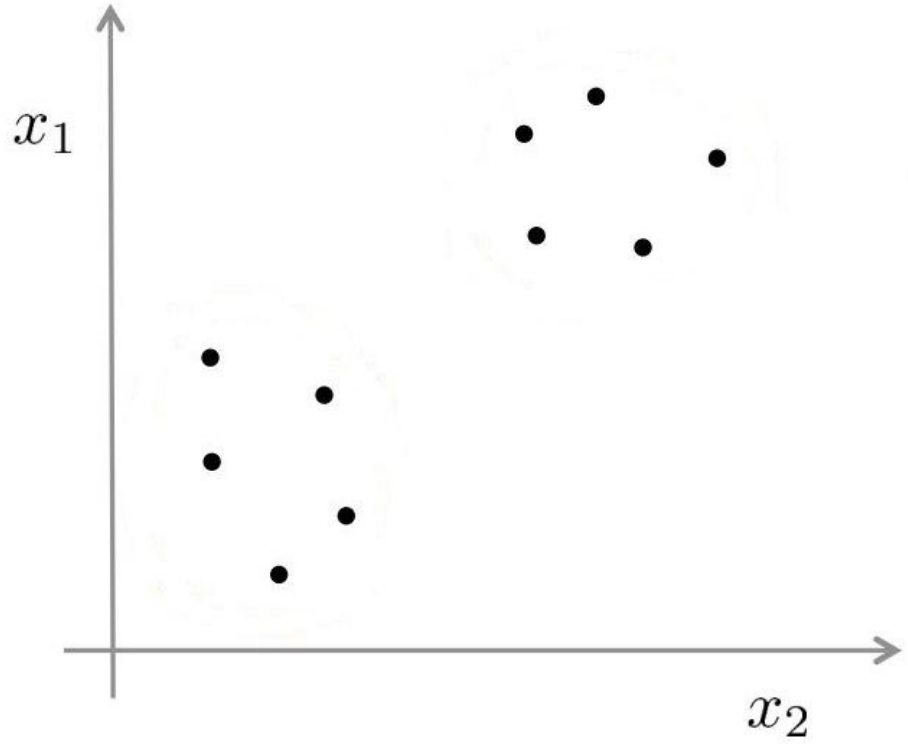
K-means Clustering Algorithm

Supervised Learning



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

Unsupervised Learning

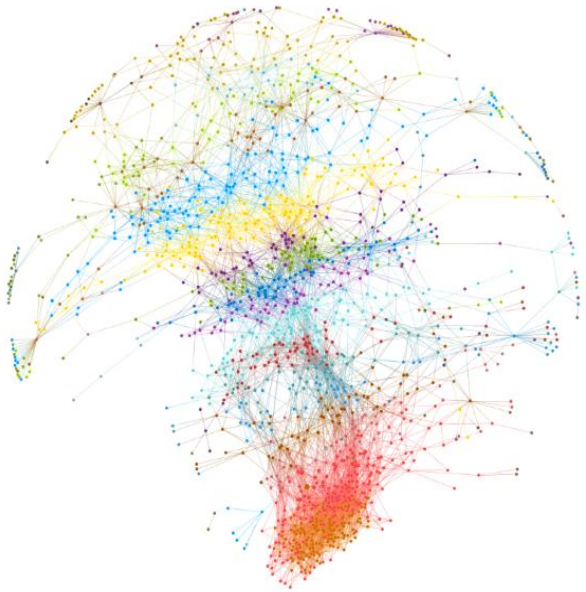


Application on Clustering

- Market Segmentation

Application on Clustering (Cont'd)

- Social Network Analysis



Application on Clustering (Cont'd)

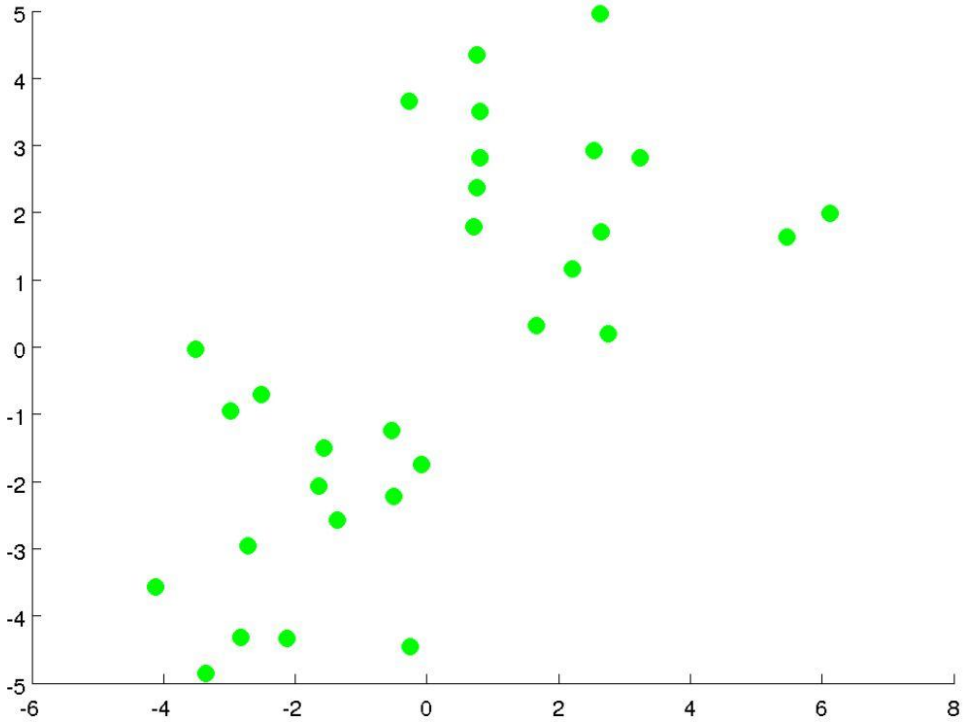
- Astronomical Data Analysis

K-means Algorithm

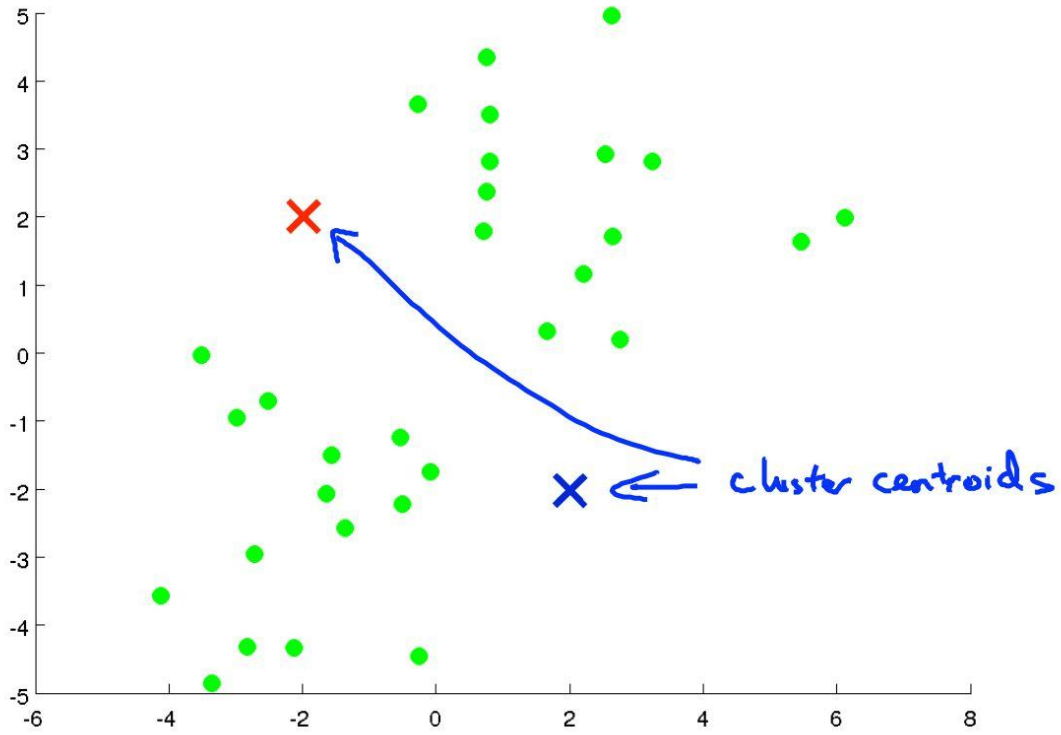
- What does K represents in this clustering algorithm?

Answer: The number of cluster we want to create.

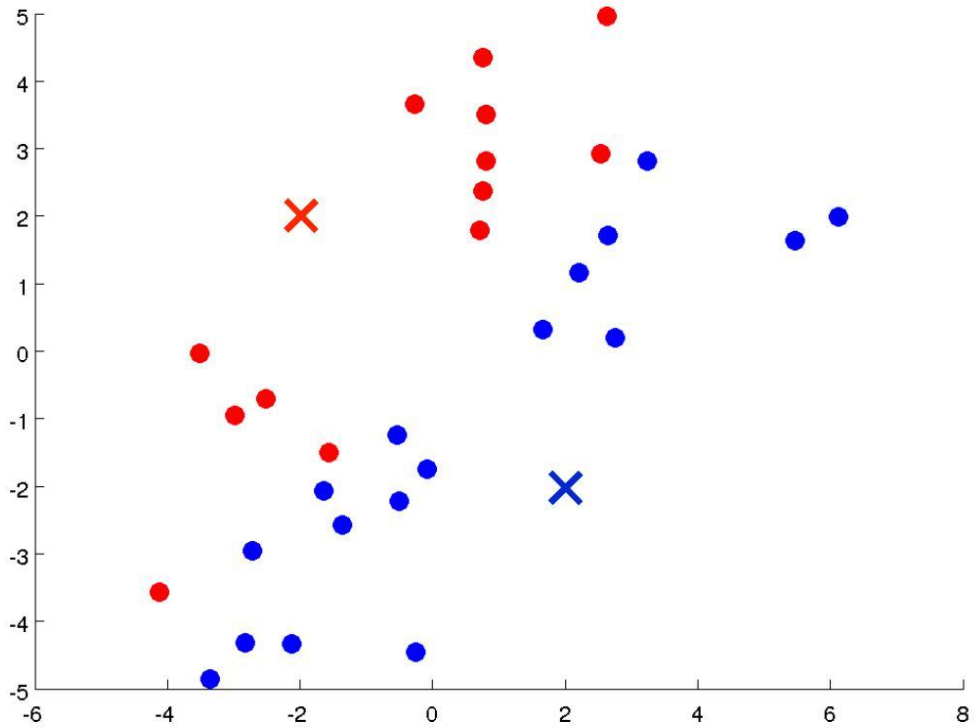
K-means Algorithm (Cont'd)



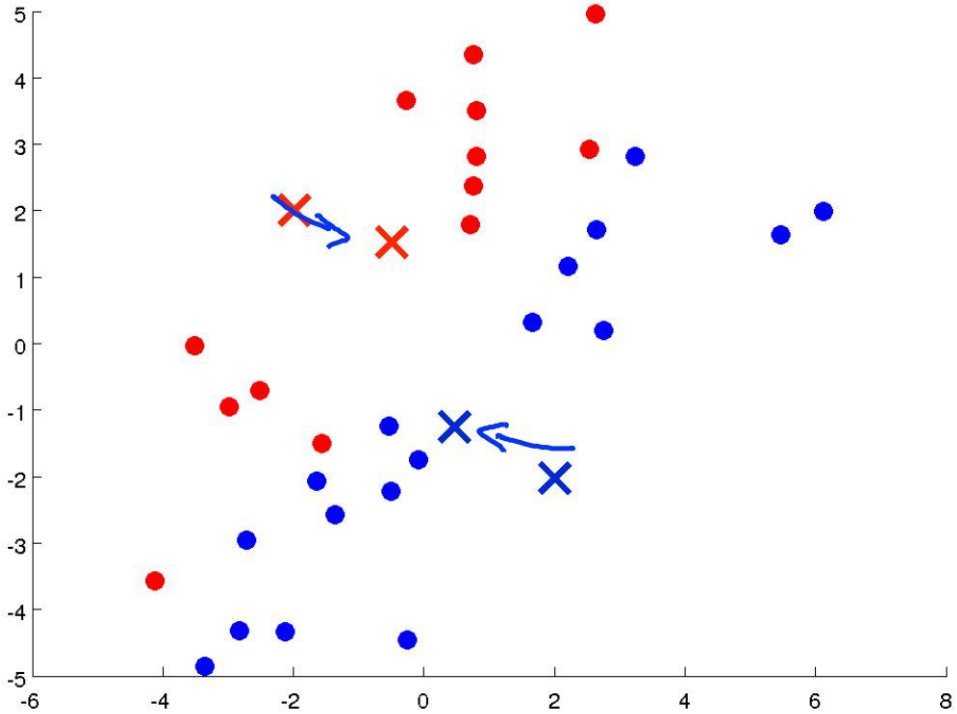
K-means Algorithm (Cont'd)



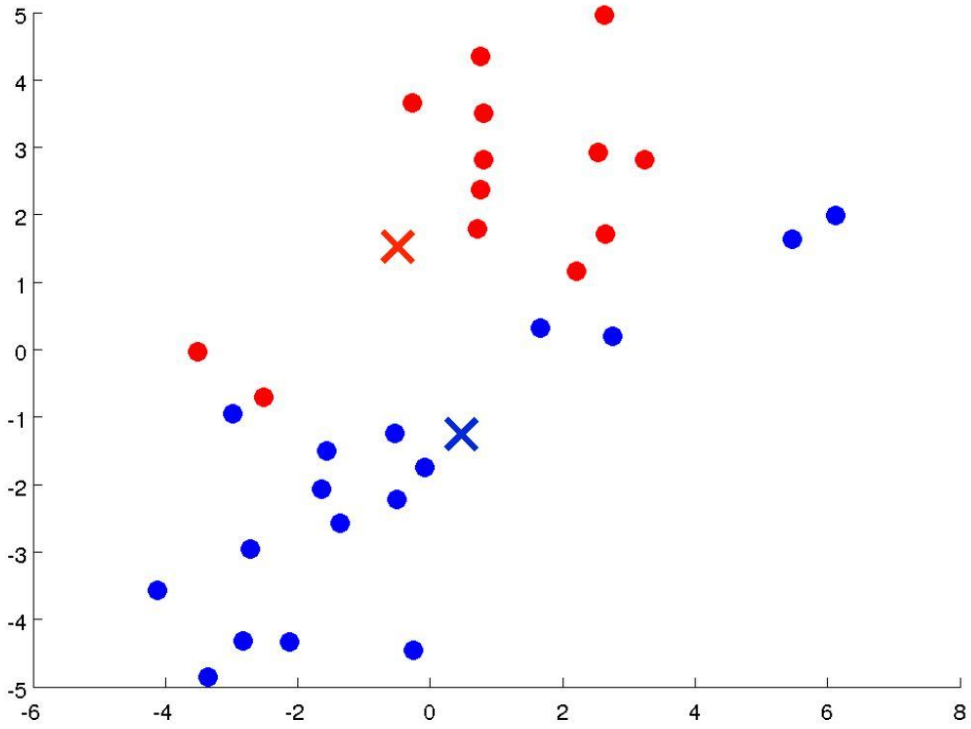
K-means Algorithm (Cont'd)



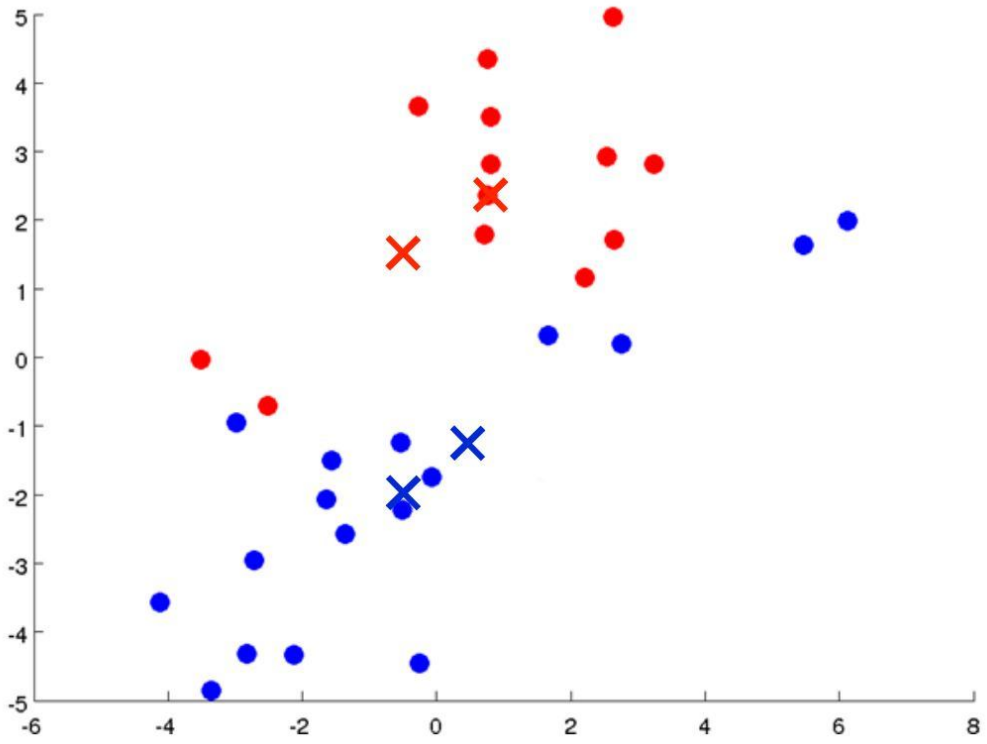
K-means Algorithm (Cont'd)



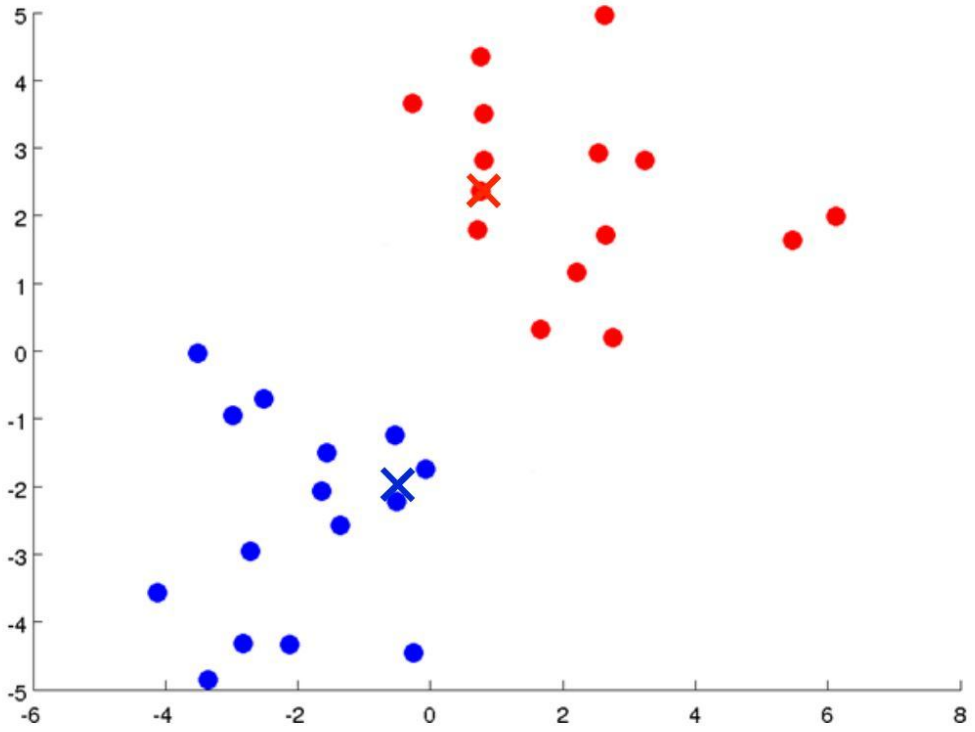
K-means Algorithm (Cont'd)



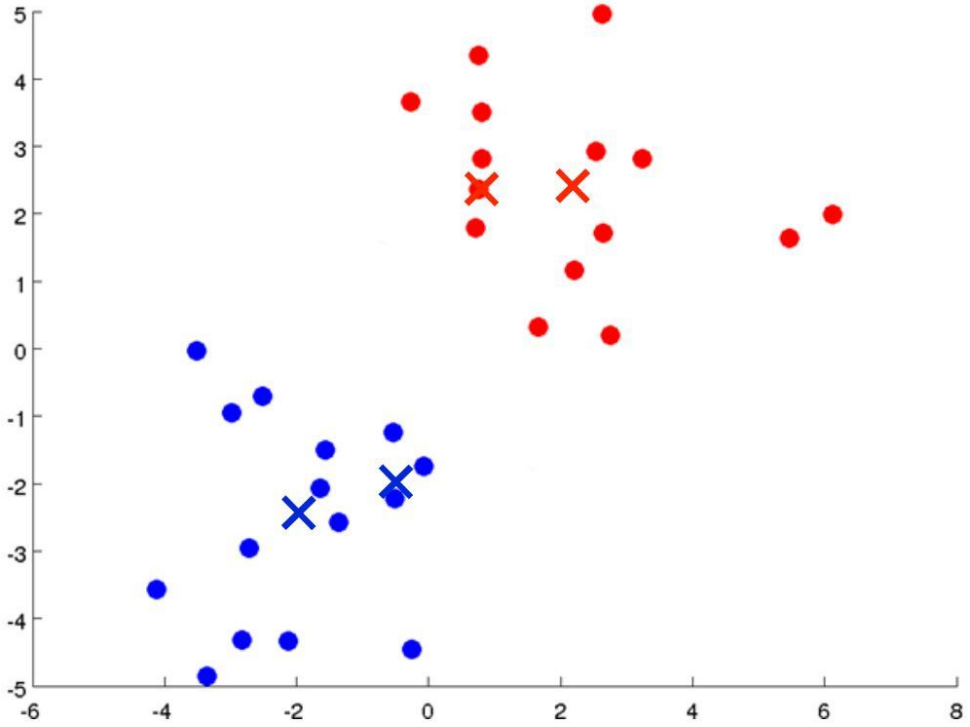
K-means Algorithm (Cont'd)



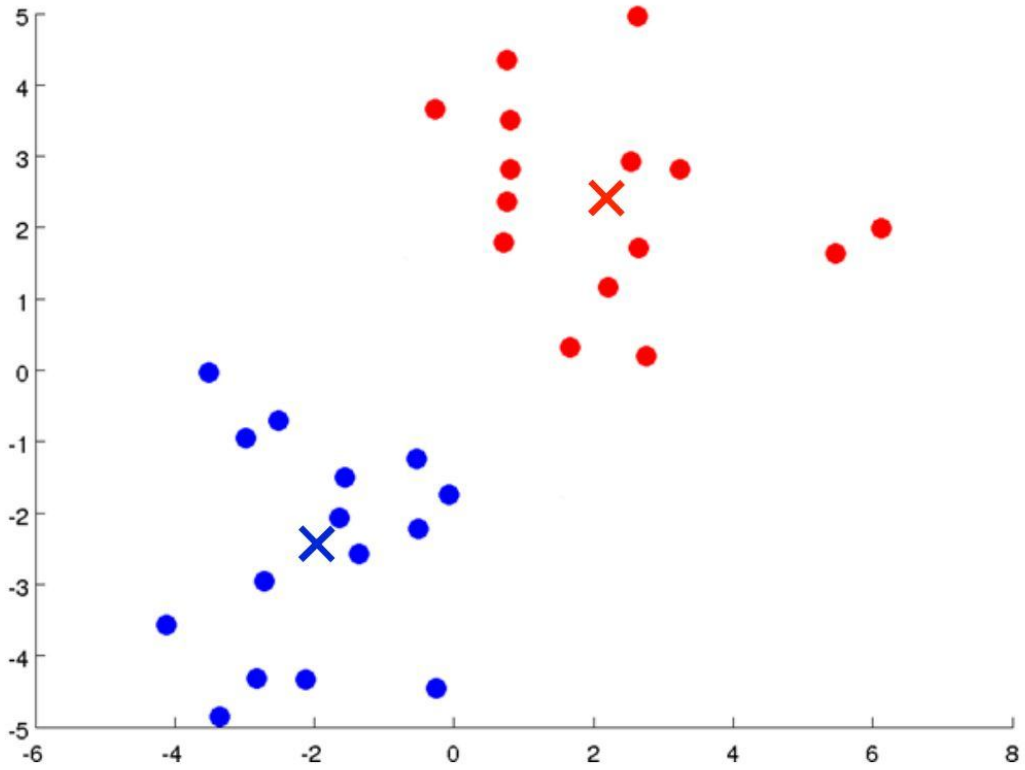
K-means Algorithm (Cont'd)



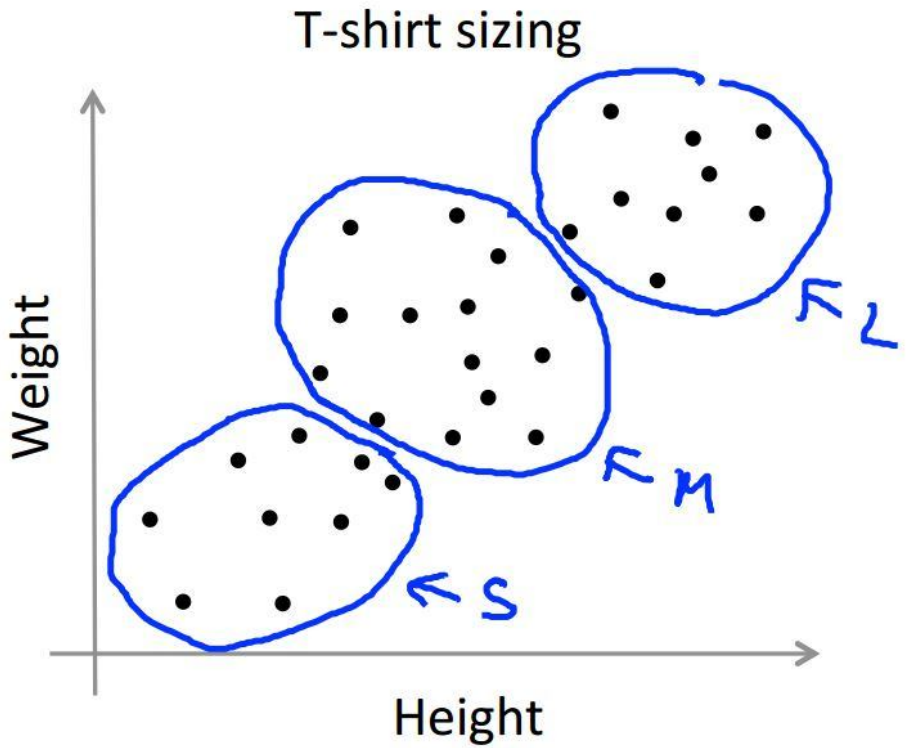
K-means Algorithm (Cont'd)



K-means Algorithm (Cont'd)



A real life application of K-means Clustering



K-means Clustering Algorithm in general

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 $x^{(i)}$

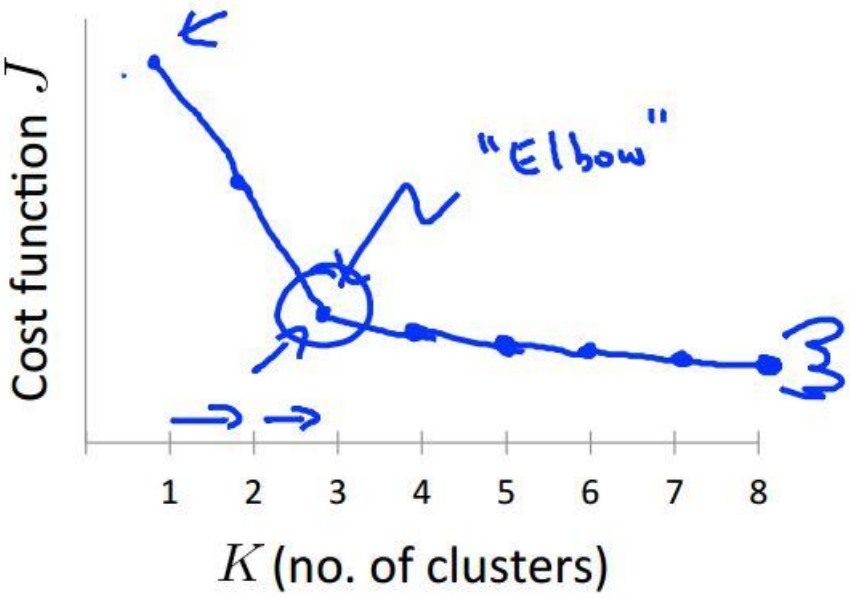
for k = 1 to K

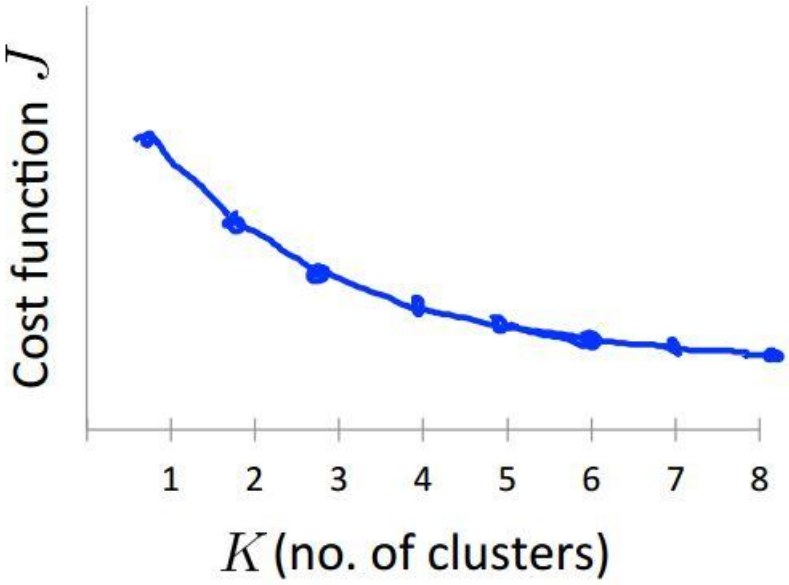
μ_k := average (mean) of points assigned to cluster k

}

How to choose the value of K?

Elbow method:





Thanks for letting me finish the class real quick!

Or did you?

Any Question/Suggestion?