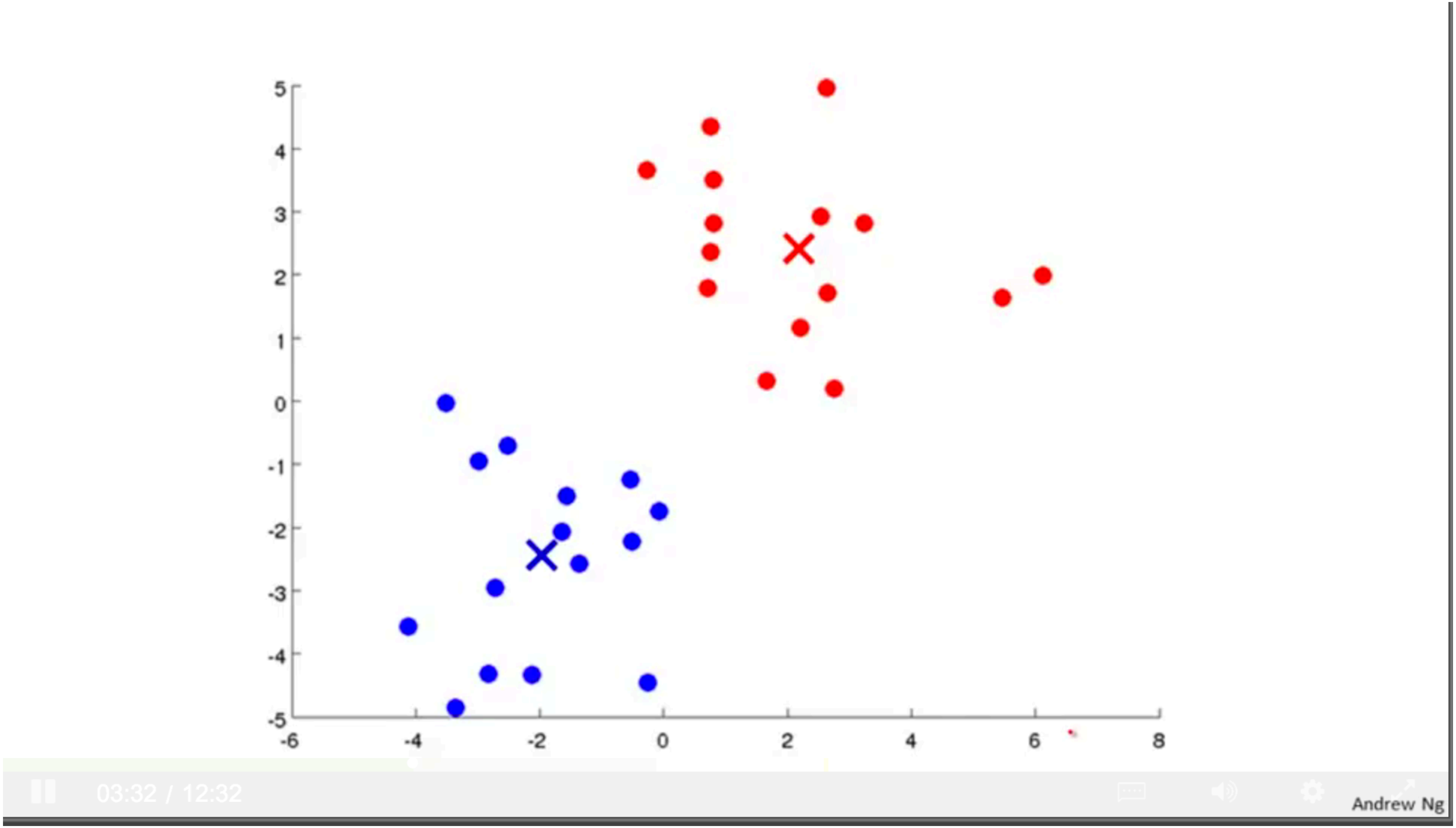


5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8
03:32 / 12:32
Andrew Ng

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

μ_1 $\times$ μ_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 $x^{(i)}$

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

for k = 1 to K

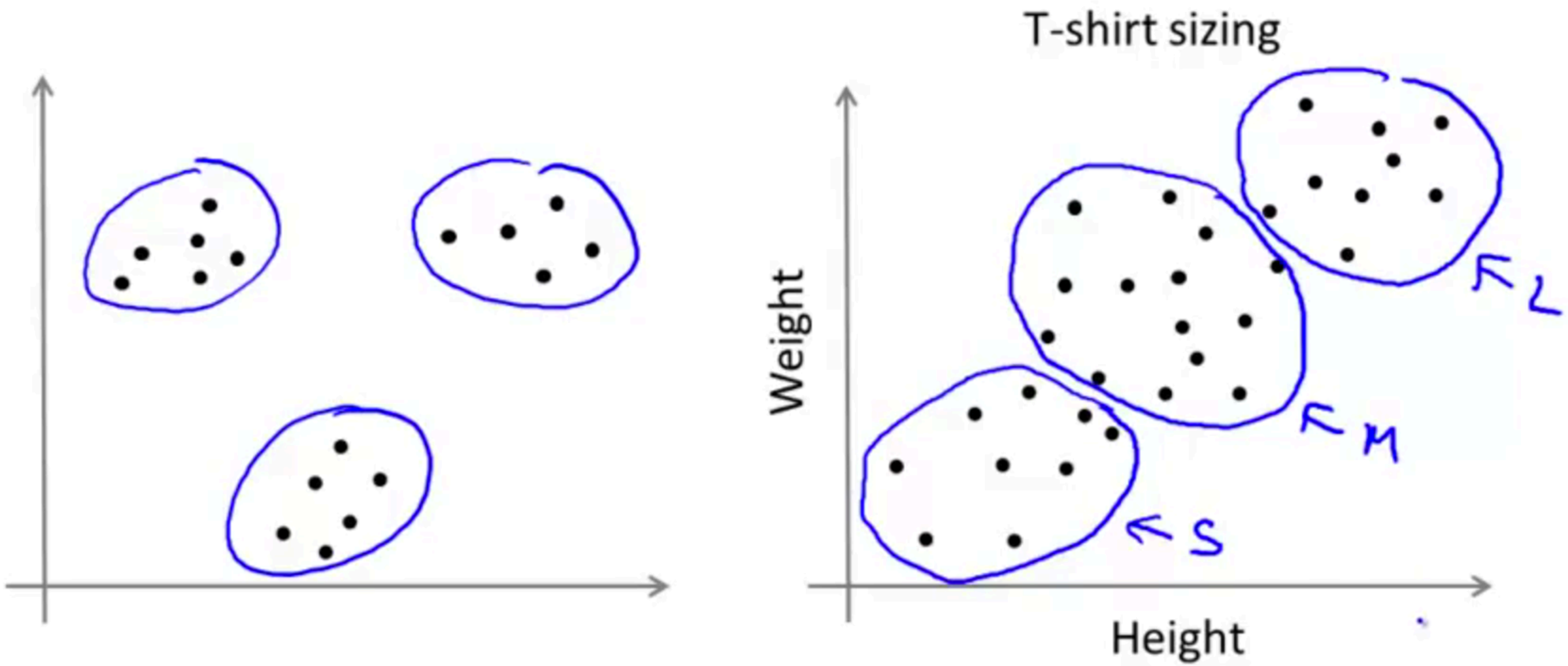
μ_k := average (mean) of points assigned to cluster k

}

Andrew Ng

K-means for non-separated clusters

S, M, L



K-means optimization objective

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

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

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

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

Optimization objective:

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

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

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 $\boxed{c^{(1)}, c^{(2)}, \ldots, c^{(m)}}$ ←

(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

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

}

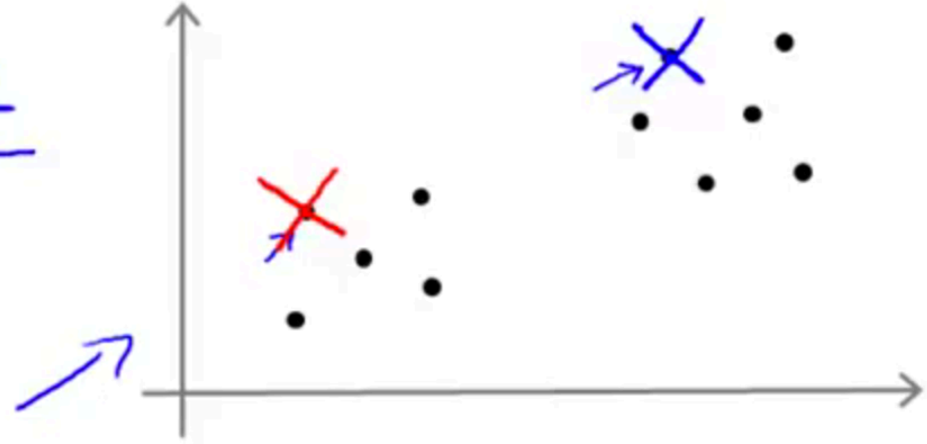
minimize $J(\ldots)$ wrt $\boxed{\mu_1, \ldots, \mu_K}$

Andrew Ng

Random initialization

$K=2$

Should have $K < m$



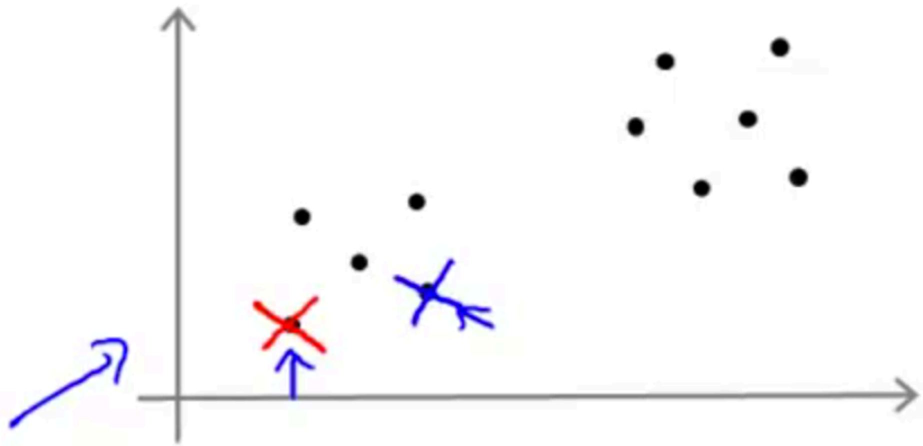
Randomly pick K training examples.

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

$\mu_1 = x^{(i)}$

$\mu_2 = x^{(j)}$

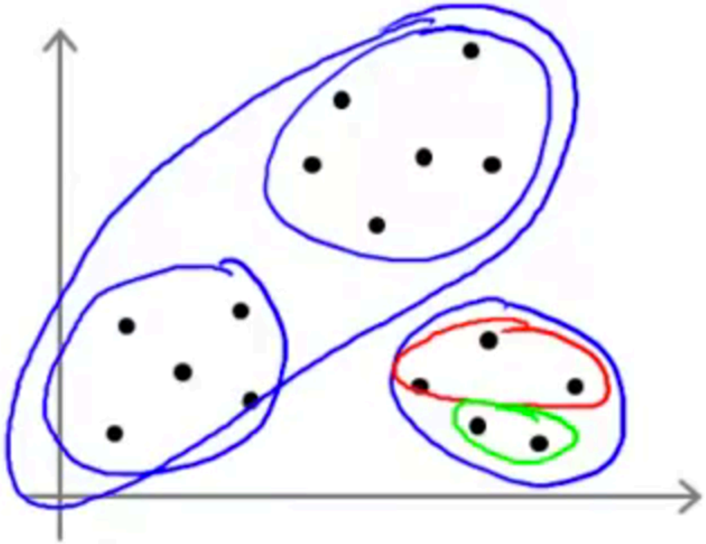
$\vdots$



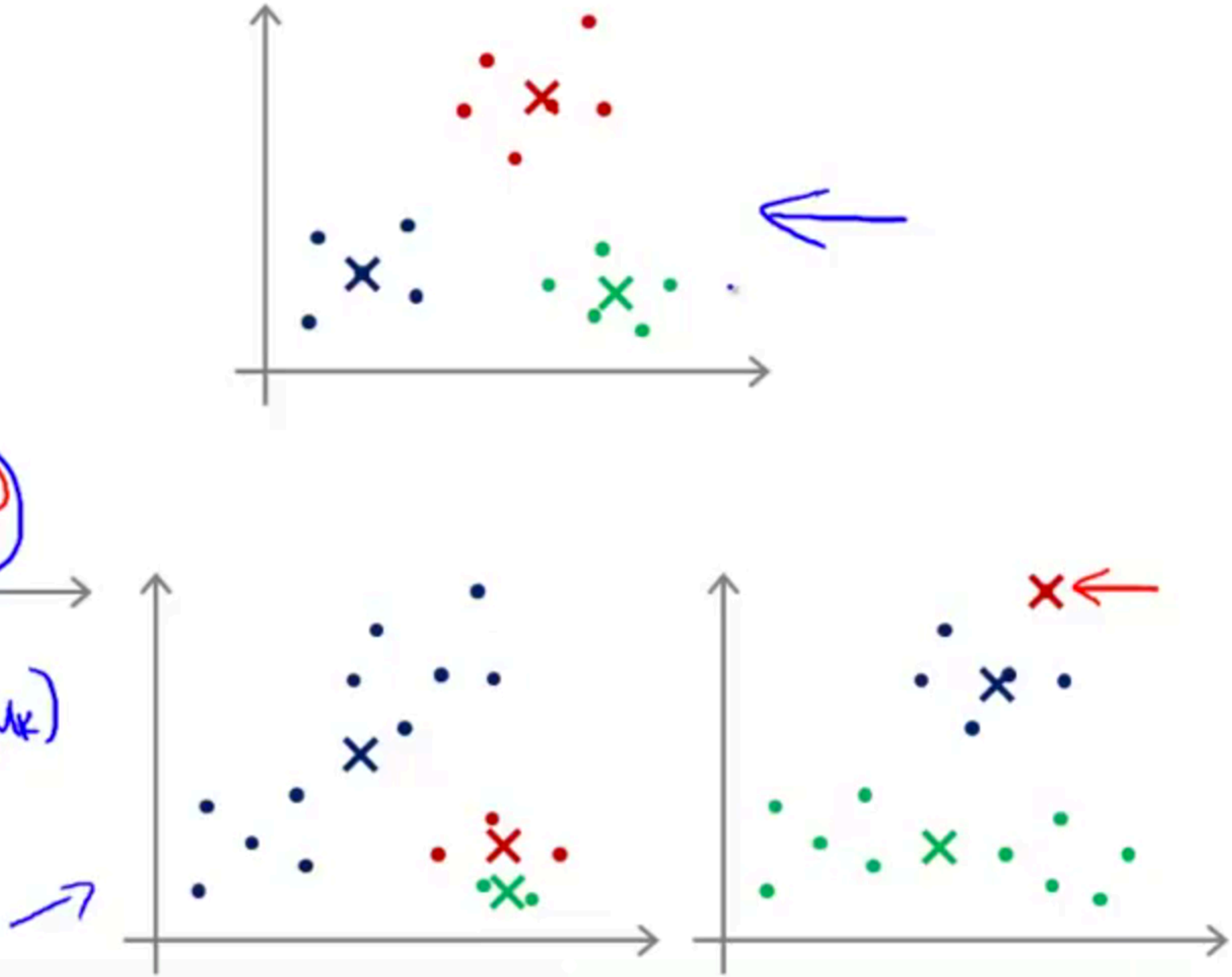
3:18 / 7:49

Andrew Ng

Local optima



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



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

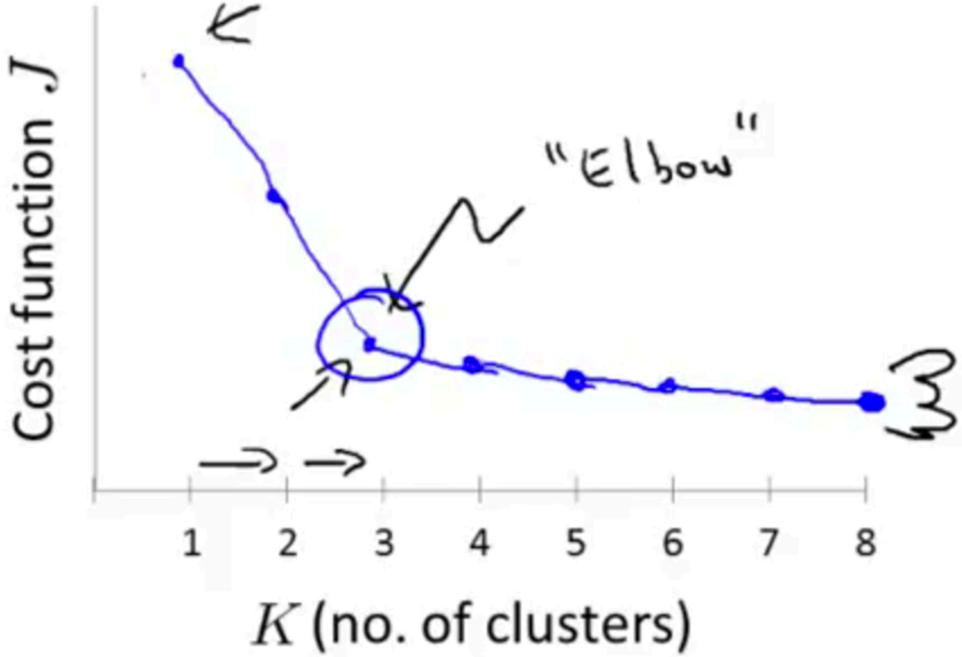
K = 2 - 10

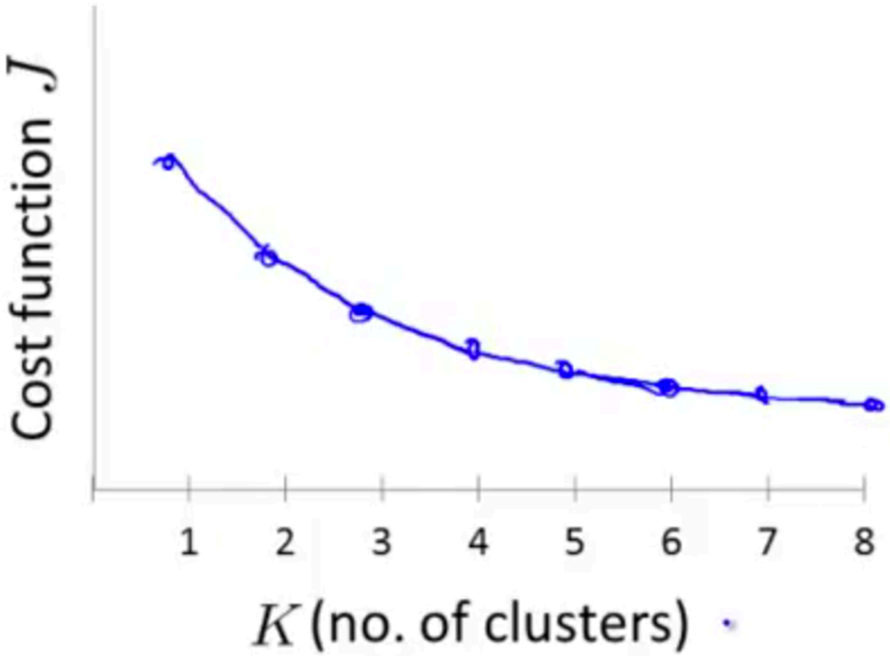
7:12 / 7:49

Andrew Ng

Choosing the value of K

Elbow method:





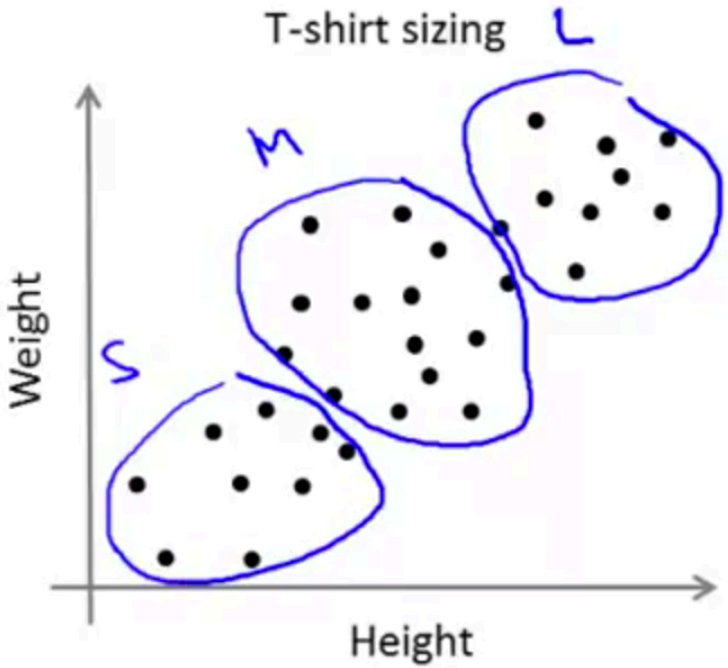
Choosing the value of K

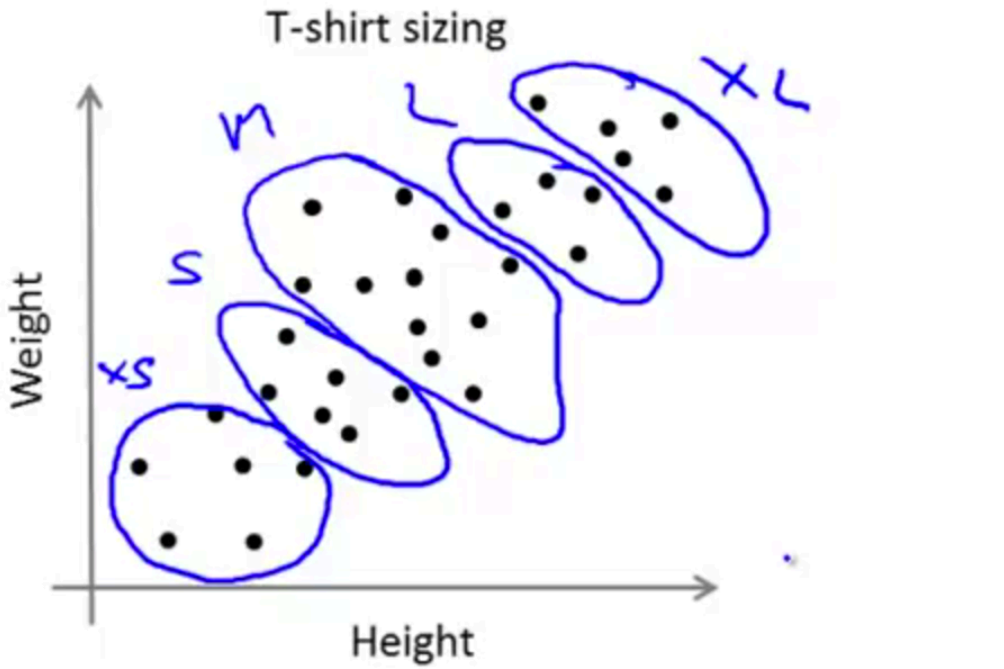
Sometimes, you're running K-means to get clusters to use for some later/downstream purpose. Evaluate K-means based on a metric for how well it performs for that later purpose.

K=3 S, M, L

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

E.g.





Andrew Ng