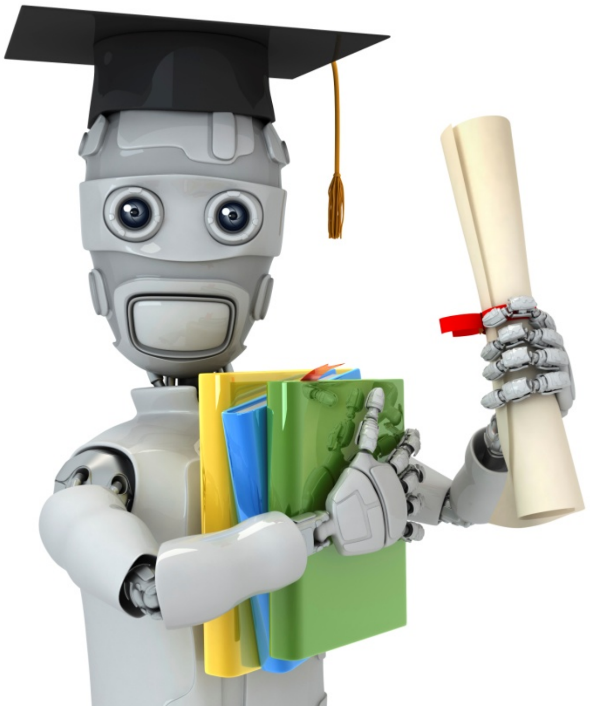


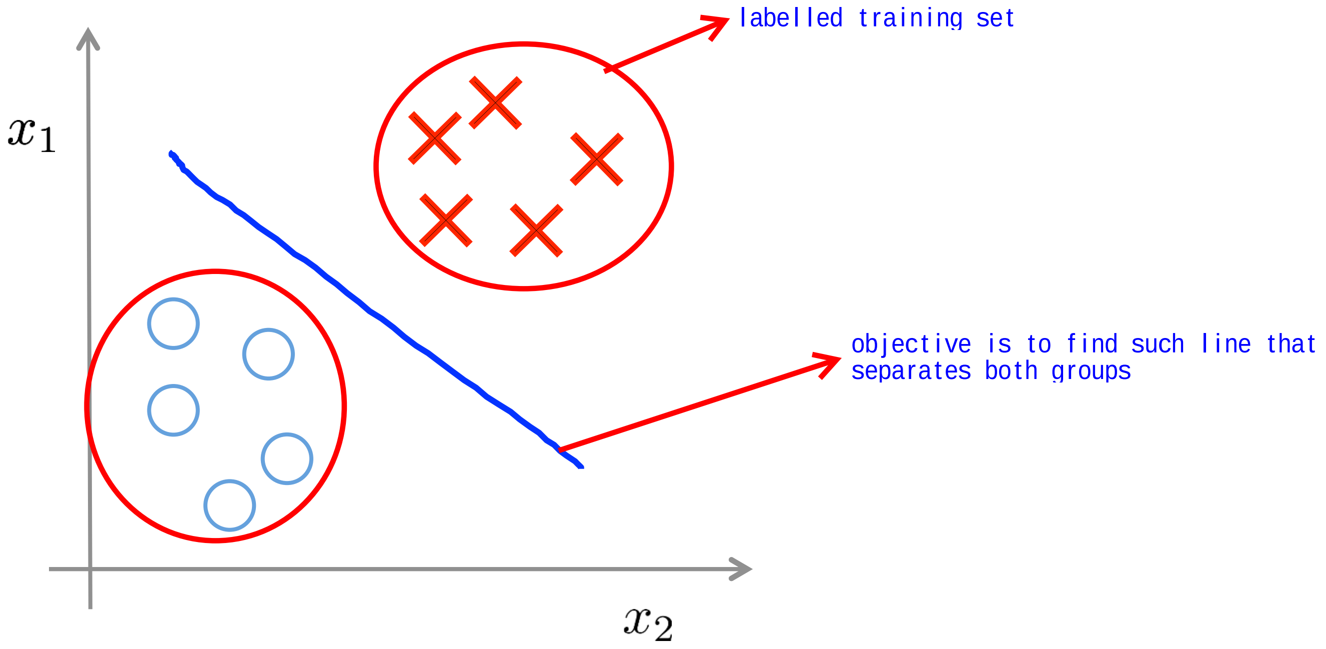
Machine Learning

Clustering

Unsupervised learning introduction

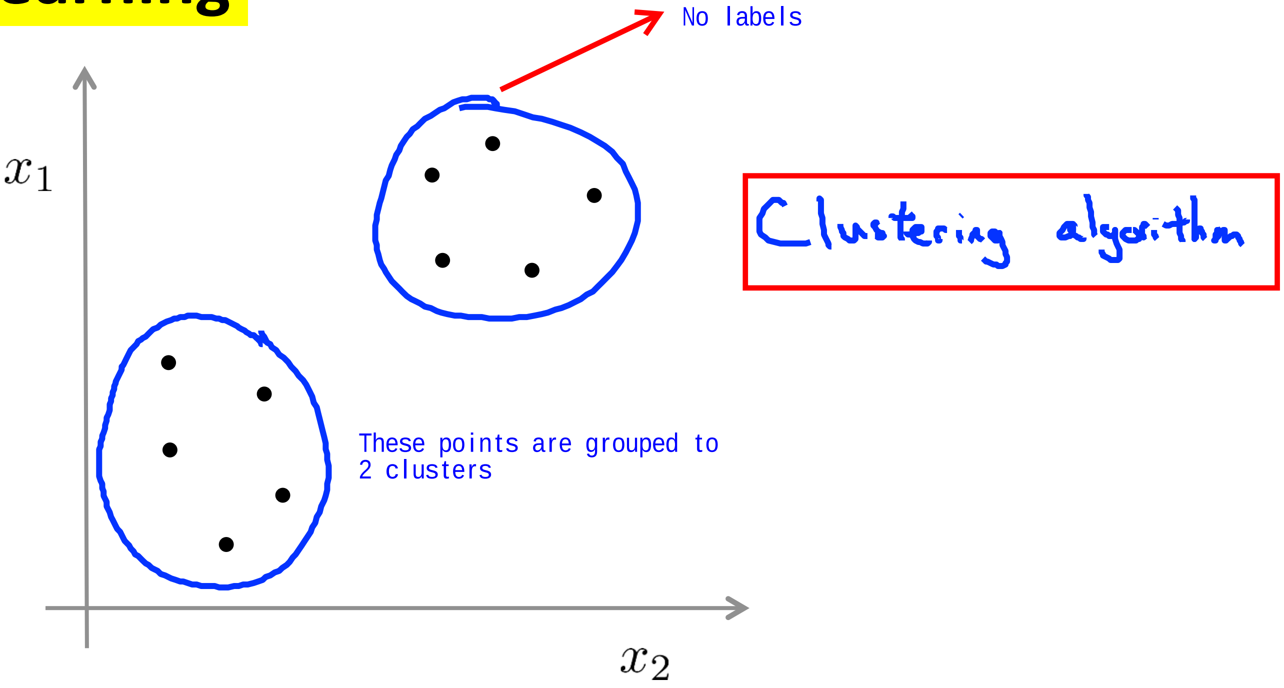
Learn From unlabelled data!

Supervised learning



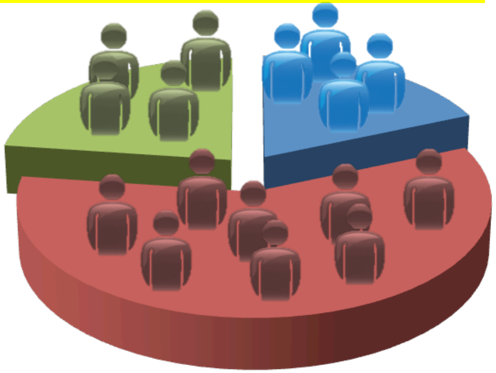
Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \ldots, (x^{(m)}, y^{(m)})\}$

Unsupervised learning

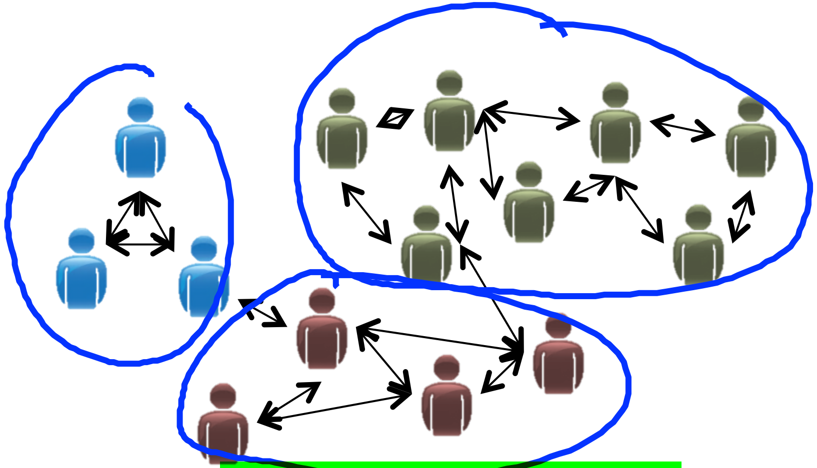


Training set: $\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(m)}\}$

Applications of clustering



Market segmentation



Social network analysis

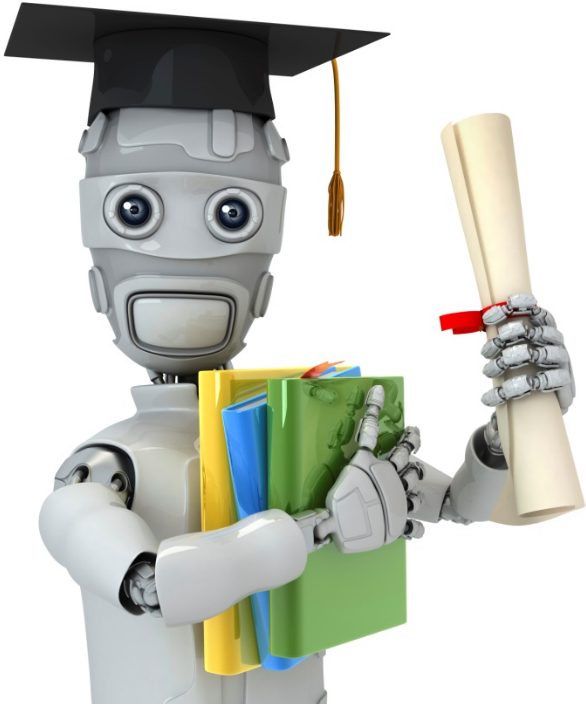
In data center, some computers tend to work together.

Organize computing clusters

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

Astronomical data analysis

Andrew Ng



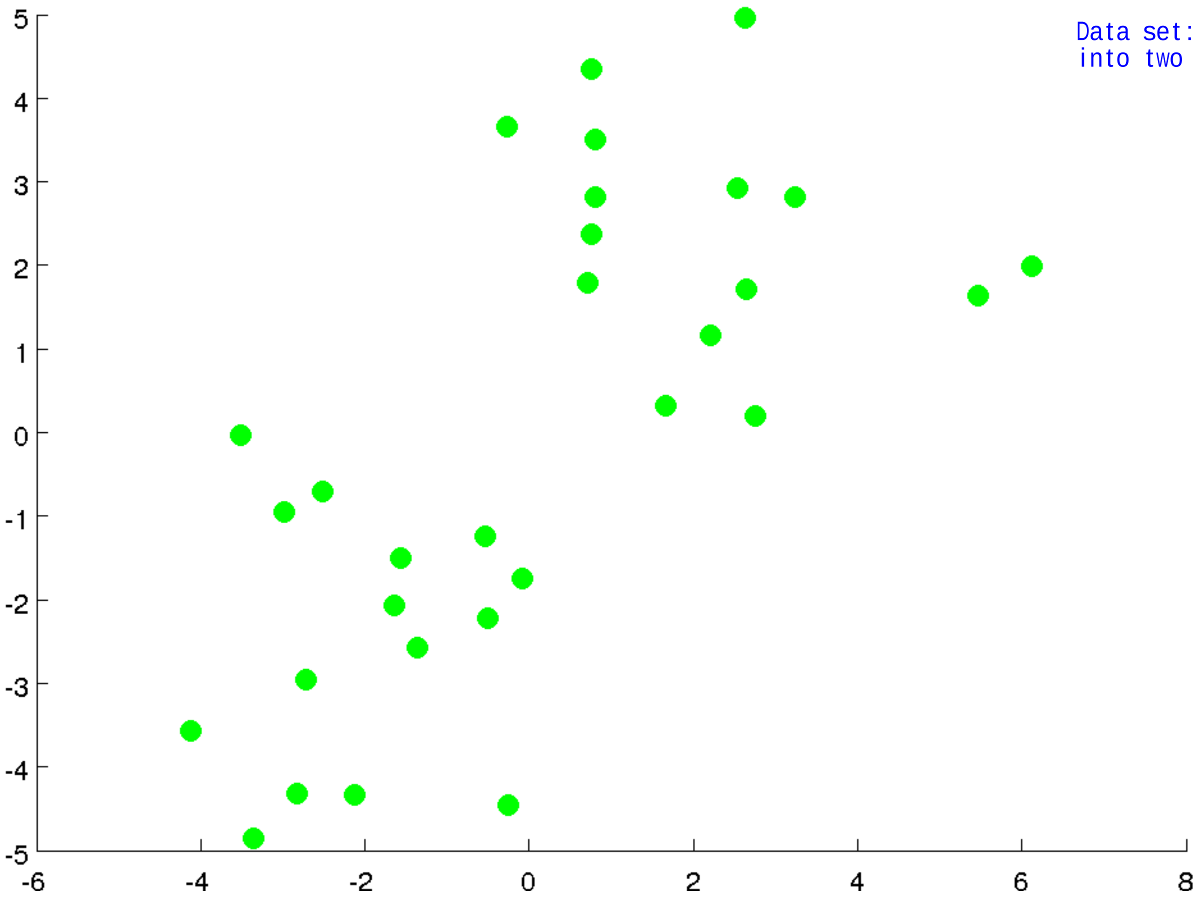
Machine Learning

Clustering

K-means algorithm

iterative algorithm

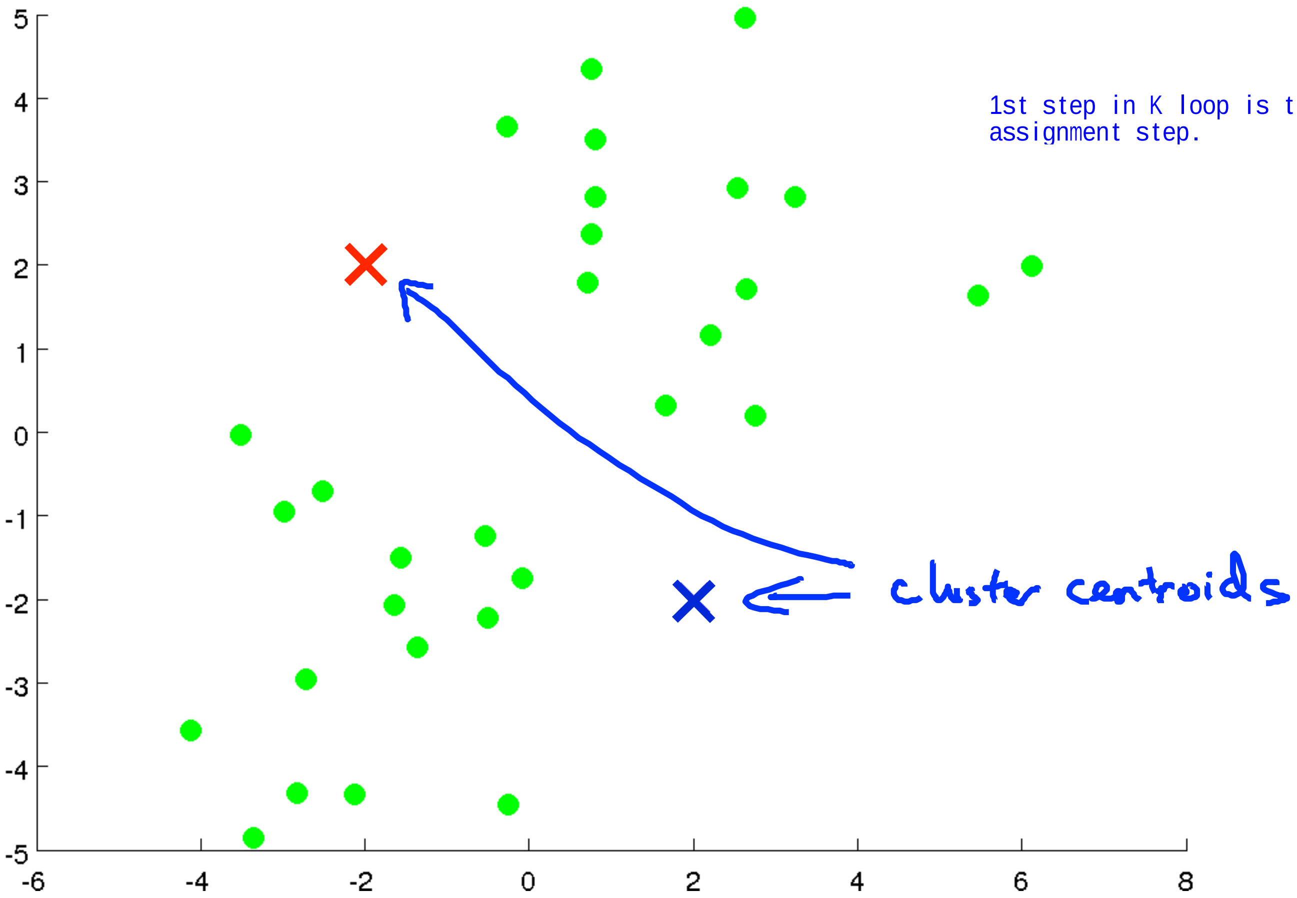
By far, the most popular and most commonly used algorithm



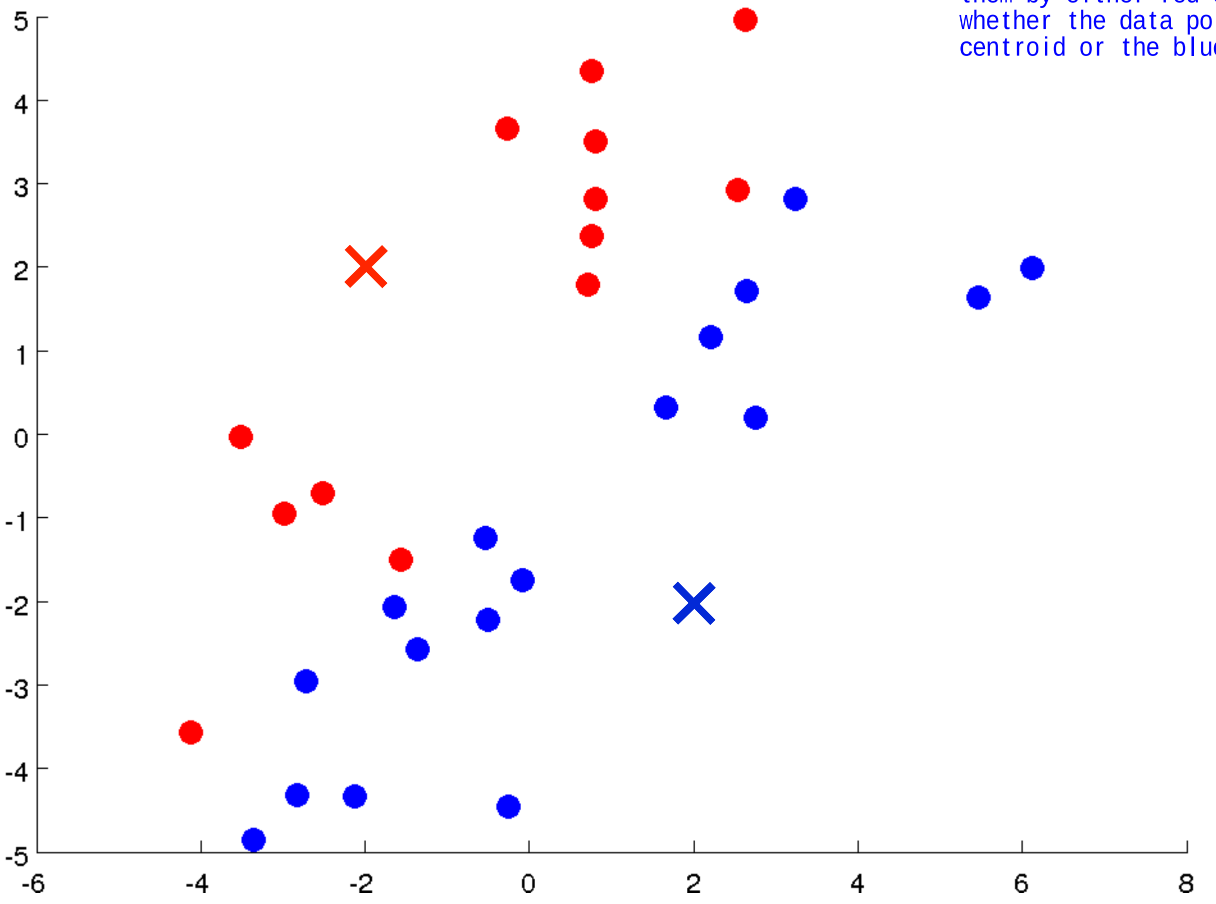
Data set: the goal is to group them into two clusters

1st step: randomly initialize two points called cluster centroids.

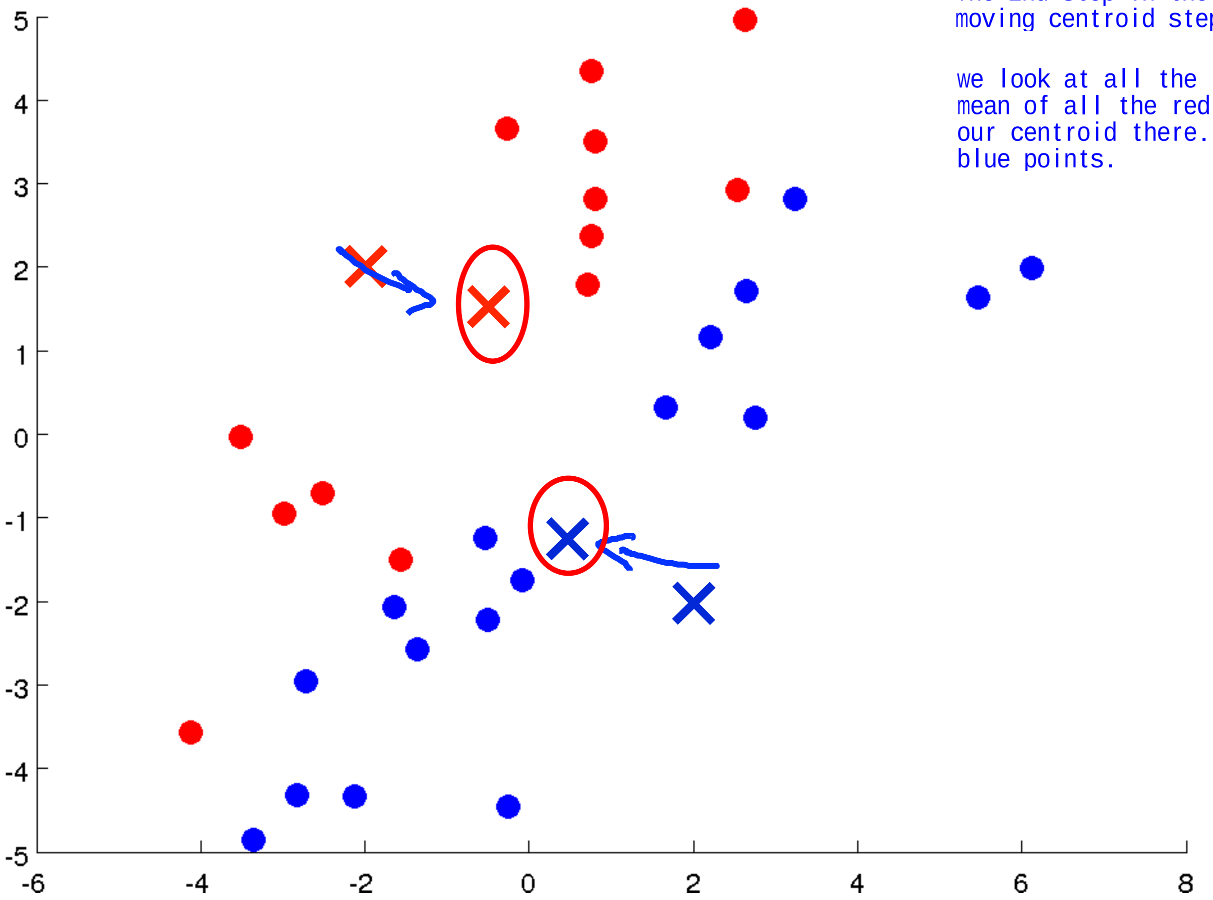
1st step in K loop is the cluster assignment step.



go through all the data points and label them by either red or blue depending on whether the data point is closer to the red centroid or the blue centroid.



Andrew Ng

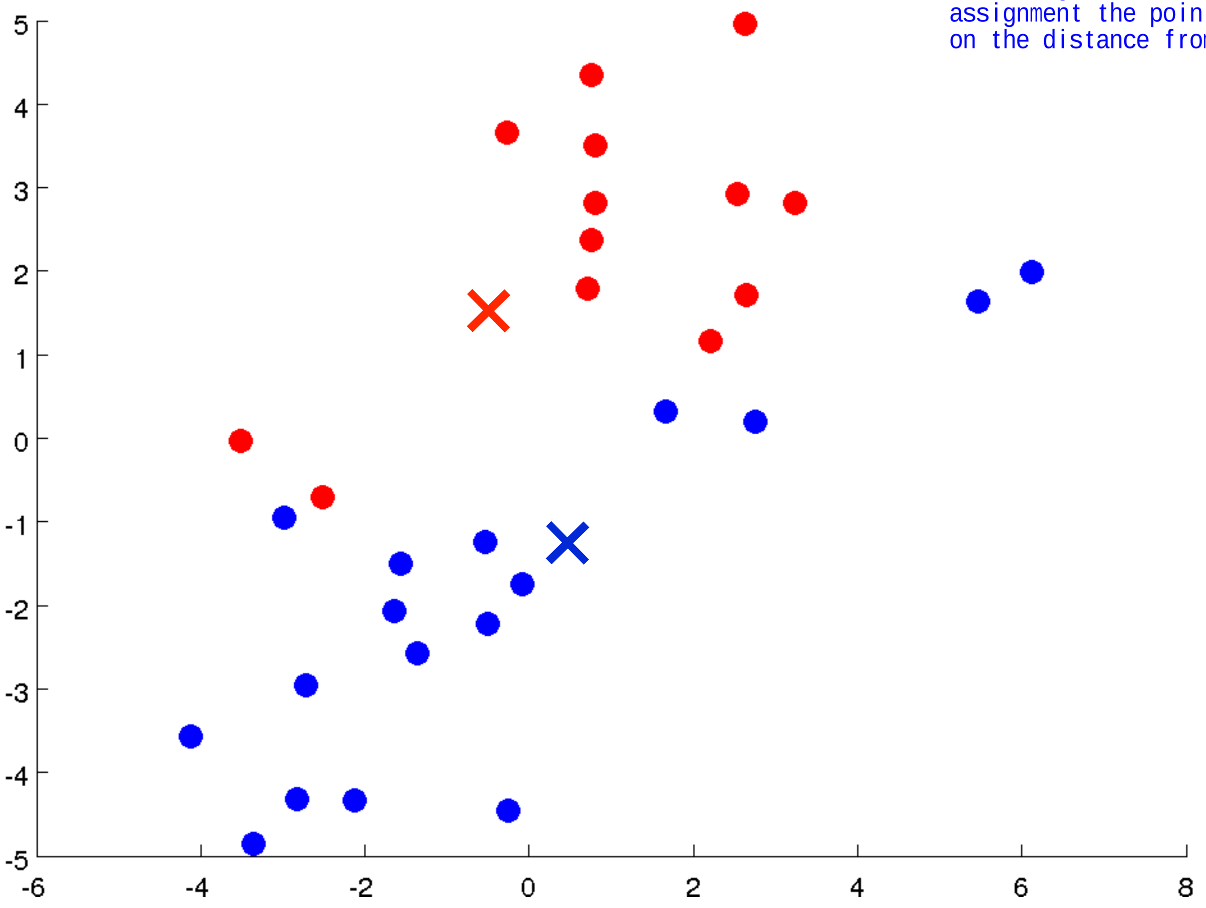


The 2nd step in the K means loop is the moving centroid step.

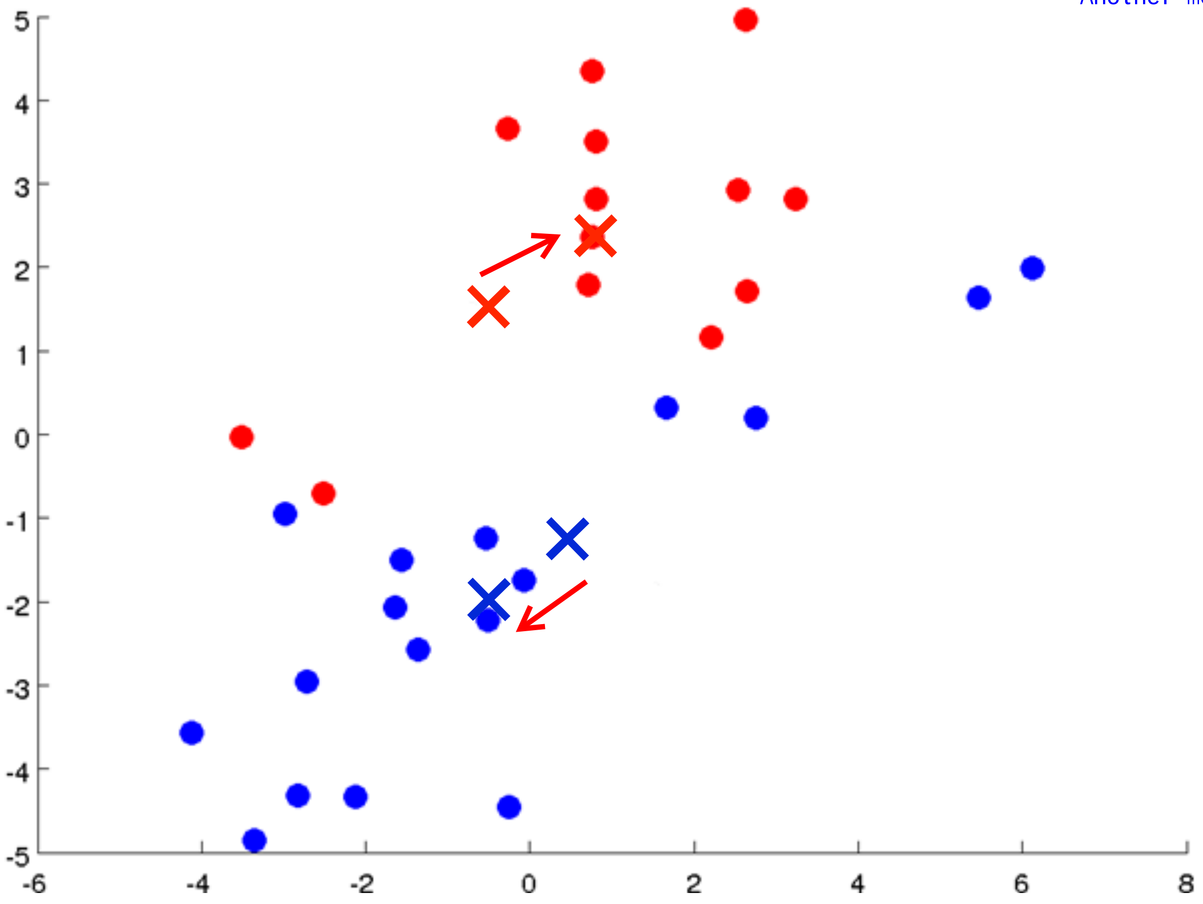
we look at all the red points and compute the mean of all the red points and we then move our centroid there. We perform the same for blue points.

Andrew Ng

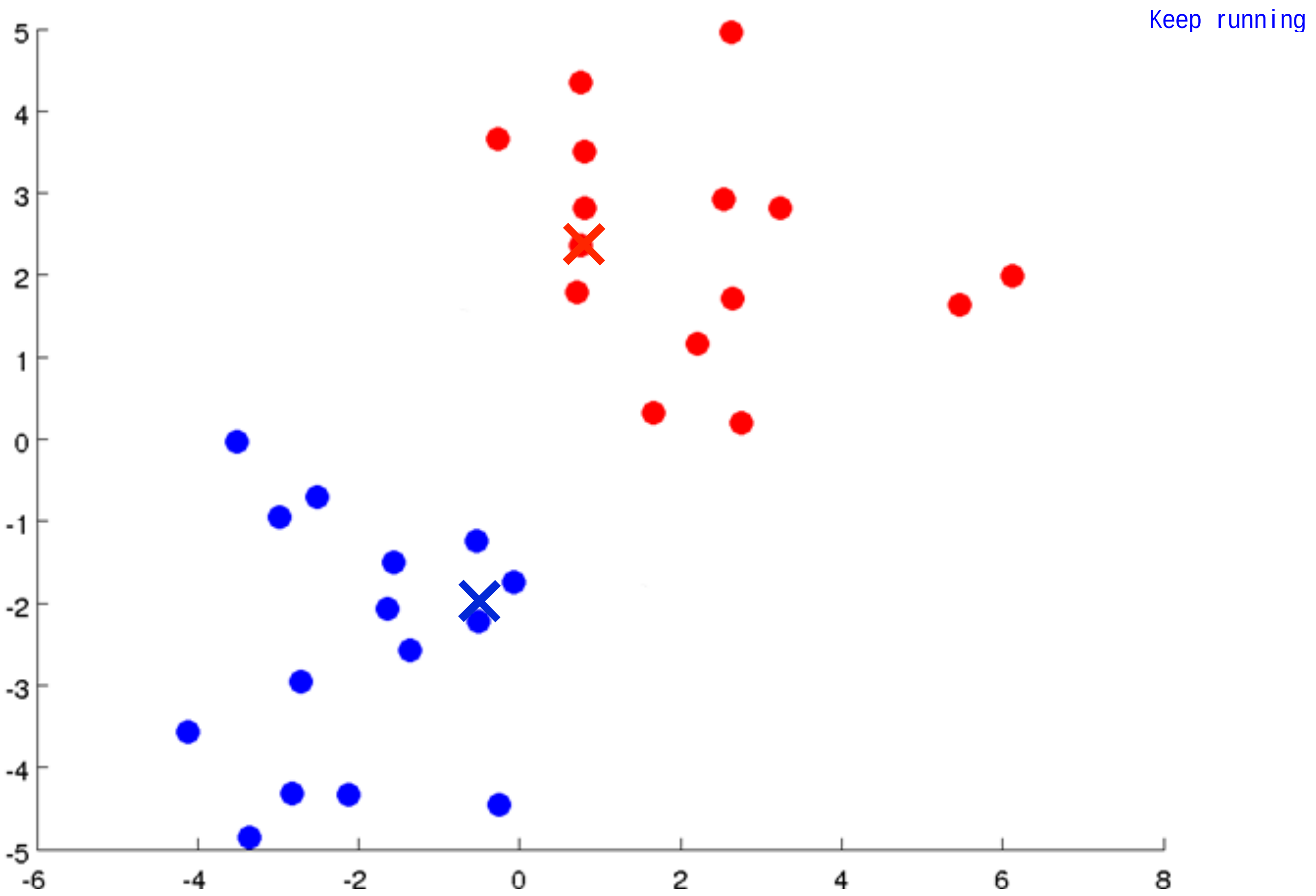
Then we go back to the first step and assignment the points to red or blue based on the distance from the new centroid.

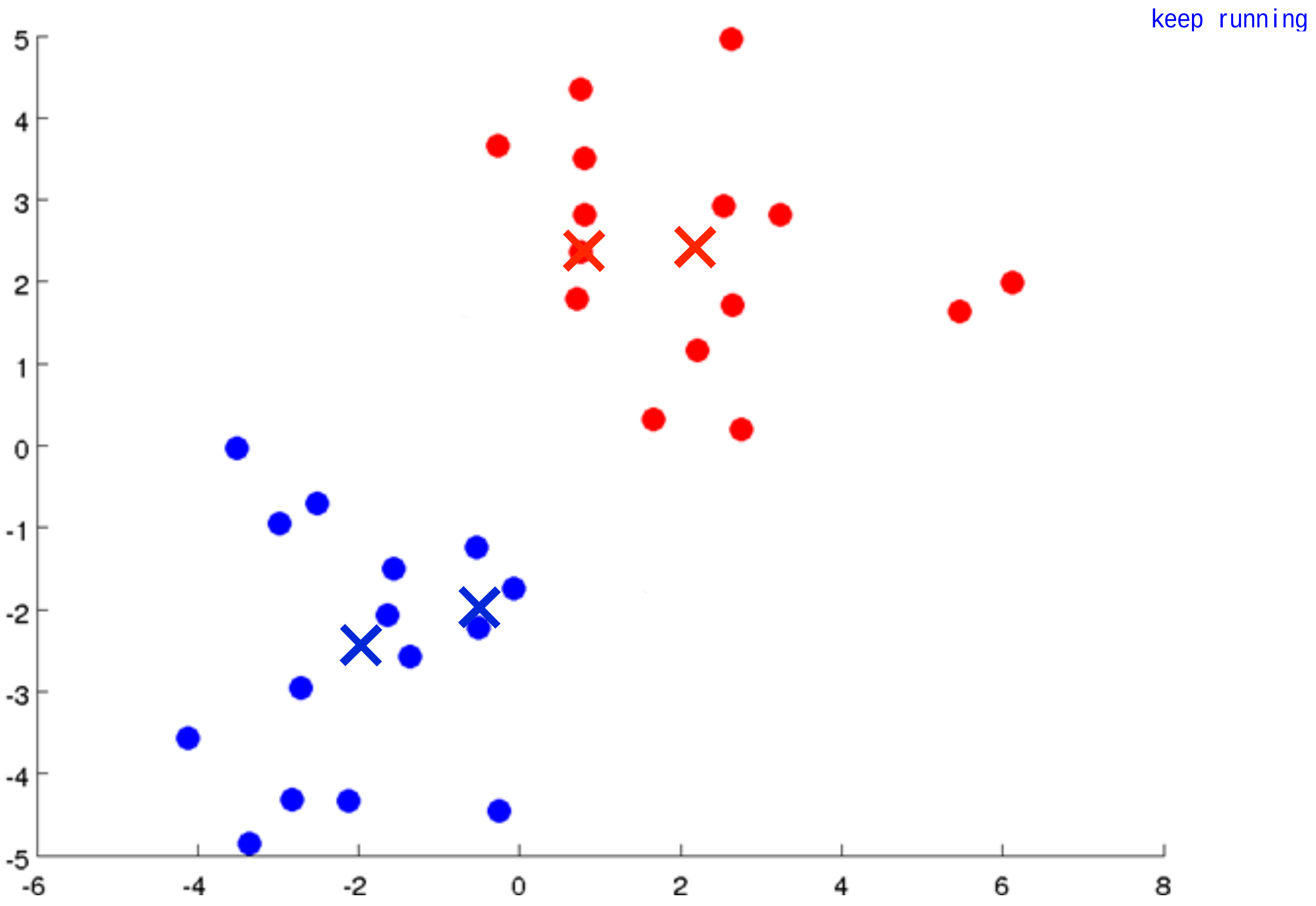


Another move centroid step!

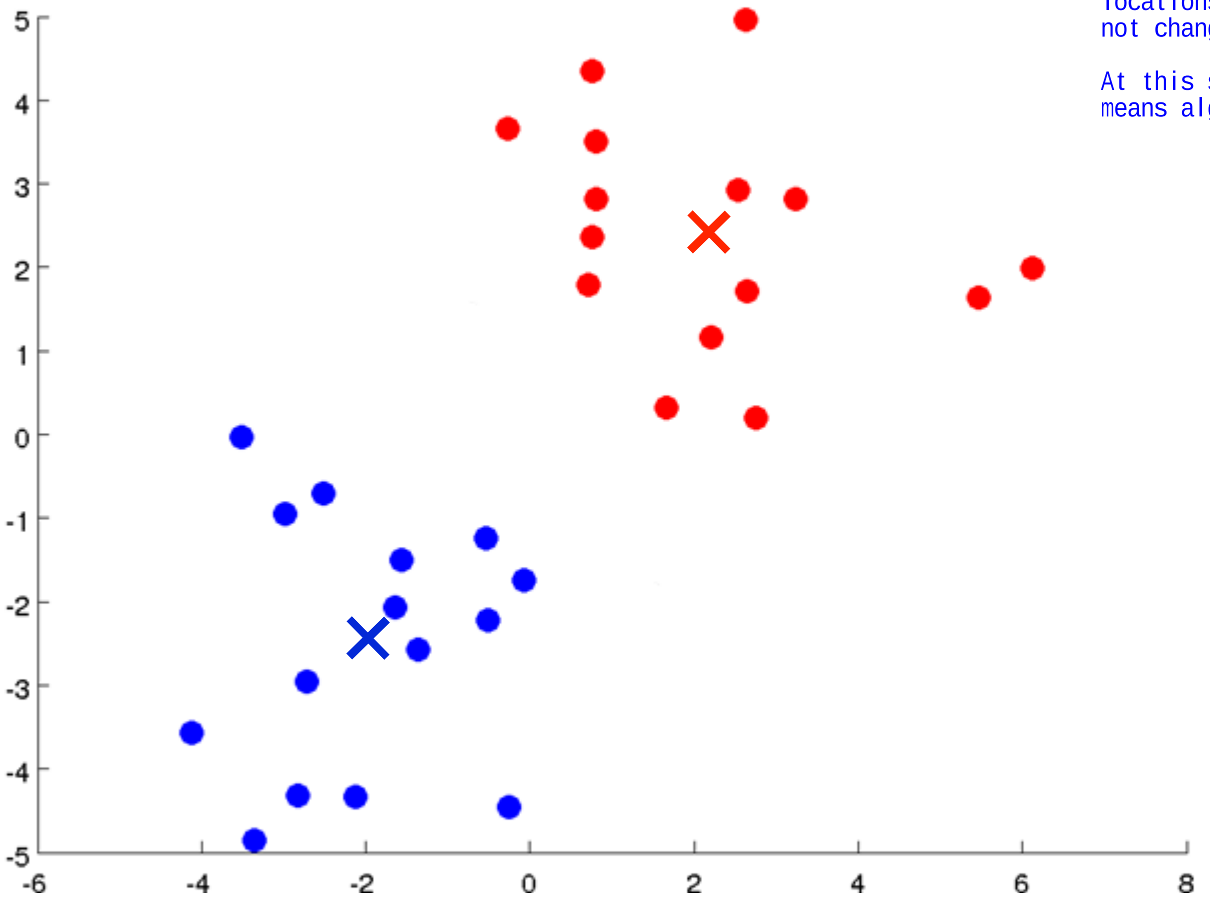


Andrew Ng


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



Andrew Ng



keep running, eventually, the locations of the centroids will not change further.

At this stage, we say that the K means algorithm has converged.

K-means algorithm

formal statement/steps of K means algorithm

For now, we just decide certain number for K.

Input:

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

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

K-means algorithm

cluster assignment step

1st step

μ_1 x μ_2 x

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 $x^{(i)}$

is a number from 1 to K, which indicates which centroid it is closet to x_i

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

Move centroid

for k = 1 to K

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

move centroid step

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

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

average

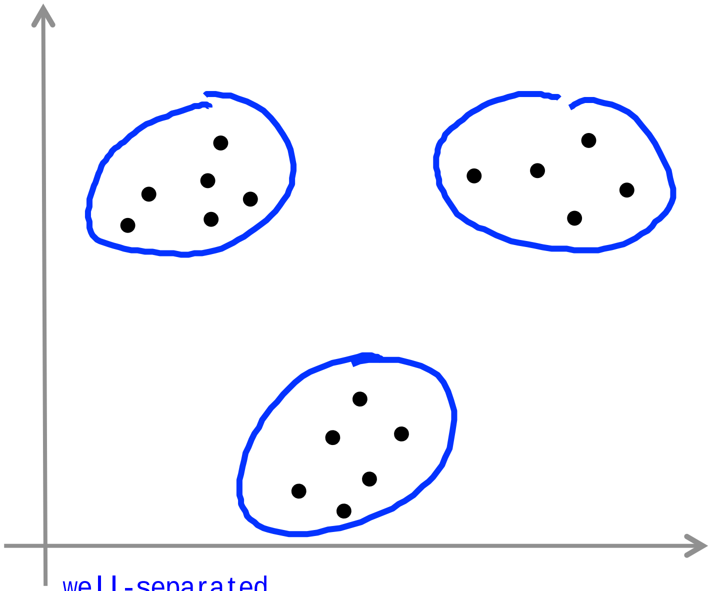
when there is cluster with no points: we can just eliminate that cluster then we have K-1 clusters

}

Andrew Ng

K-means for non-separated clusters

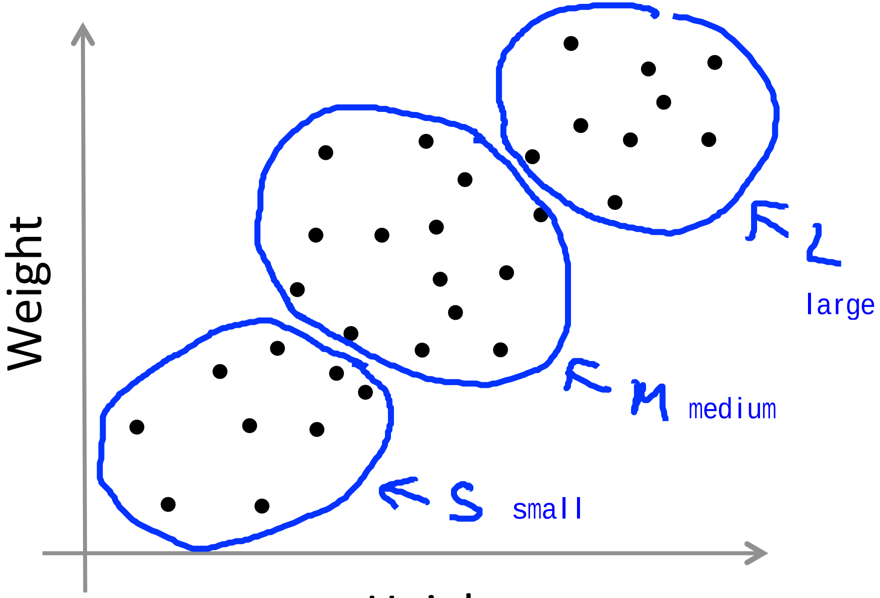
One other common application of K means



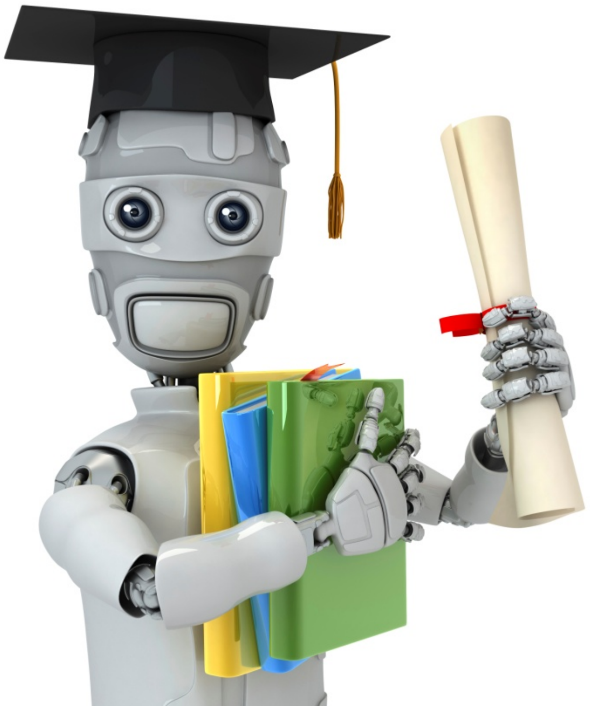
An example of market segregation

S, M, L

T-shirt sizing



not well-separated data set
we can also apply K-means



Machine Learning

Clustering

Optimization objective

K means algorithm also has an optimization objective!

K-means optimization objective

we keep tracking these two parameters

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

μ_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

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

x i is assigned to cluster 5

corresponding centroid location

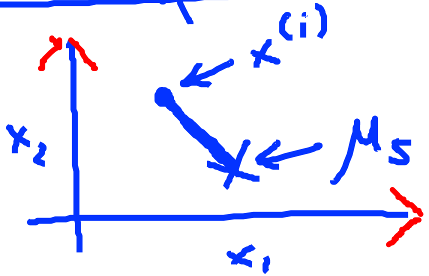
Optimization objective:

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

objective!

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

Distortion



Andrew Ng

K-means algorithm

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

Cluster assignment step

Minimize $J(\dots)$ wrt $\boxed{c^{(1)}, c^{(2)}, \dots, c^{(m)}}$ obvious!

(holding $\mu_1, \dots, \mu_K$ 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

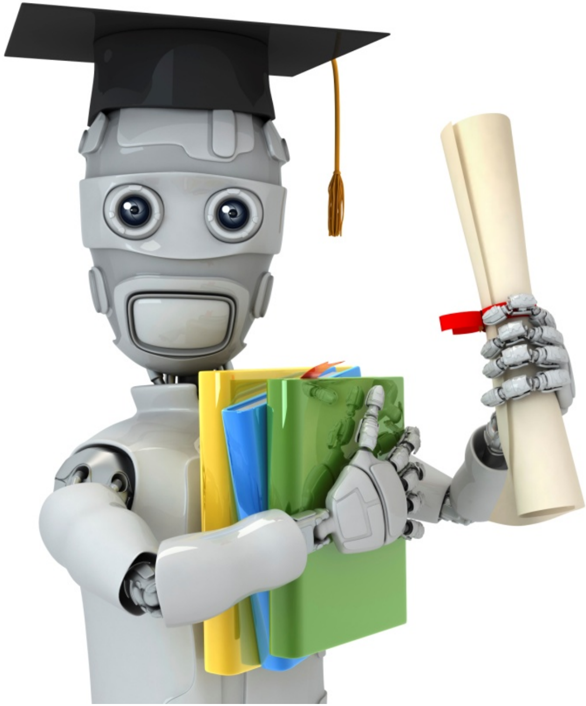
}

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

holding c_i

with respect to

Andrew Ng



Machine Learning

Clustering

Random initialization

To make K-means avoid local optima

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

}

Random initialization

Should have $K < m$

$K=2$

Randomly pick K training examples.

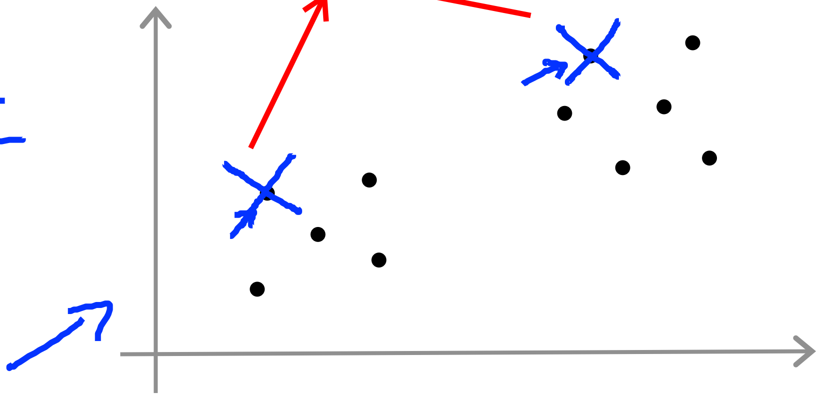
Set $\mu_1, \ldots, \mu_K$ equal to these K examples.

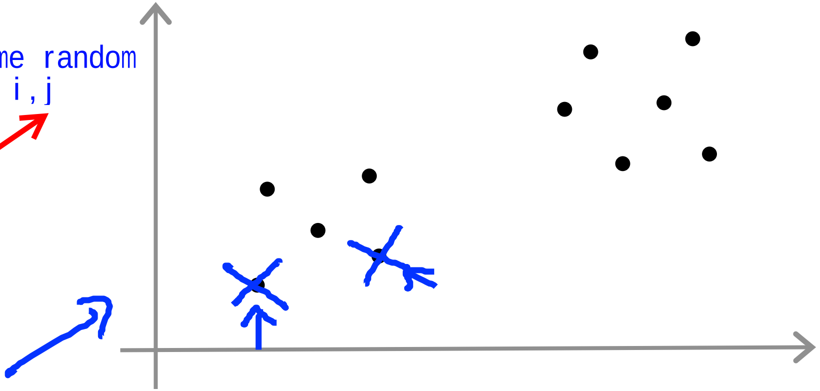
$\mu_1 = x^{(i)}$
$\mu_2 = x^{(j)}$
$\vdots$

for some random values i,j

K-means can end up with different solution depending on your initial conditions

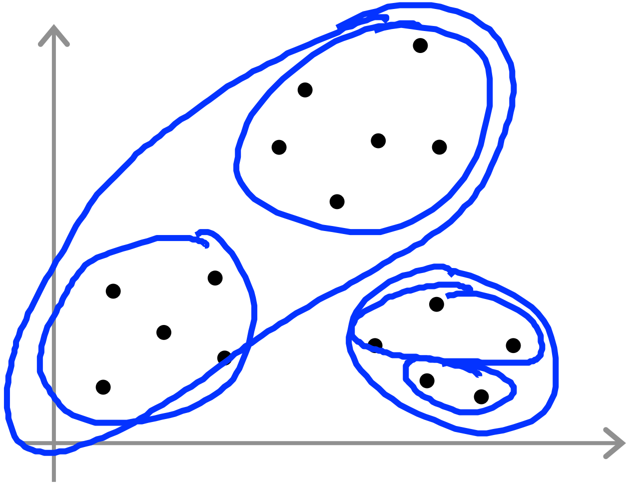
randomly pick two initial centroid



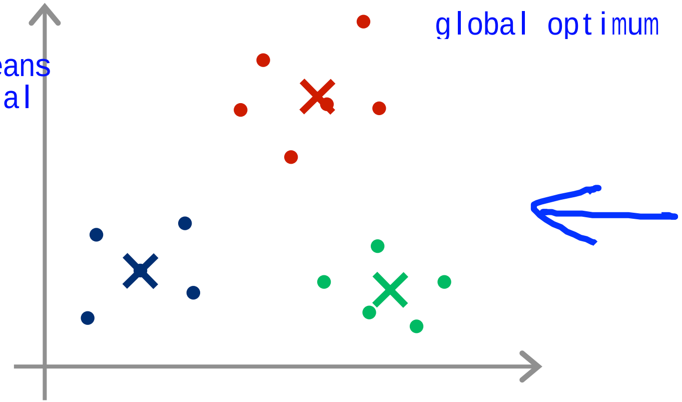


Local optima

Can end up at local optima
To avoid this, we can run K-means multiple times with diff initial conditions

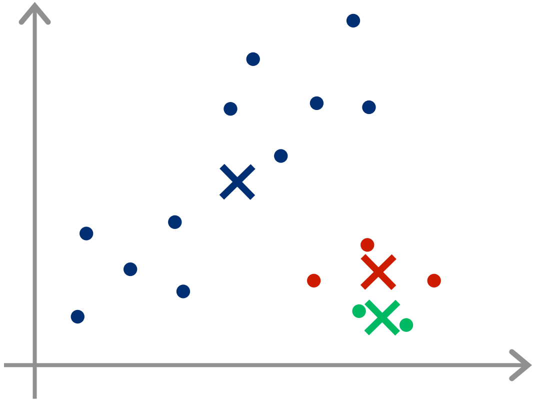


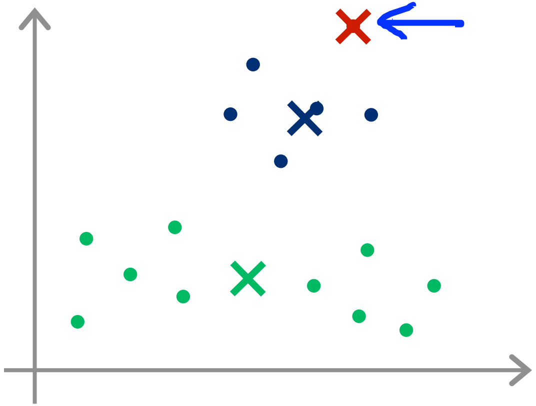
$$J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K)$$



global optimum

different local optima





Random initialization

For i = 1 to 100 { Run K-means 100 times

Randomly initialize K-means.

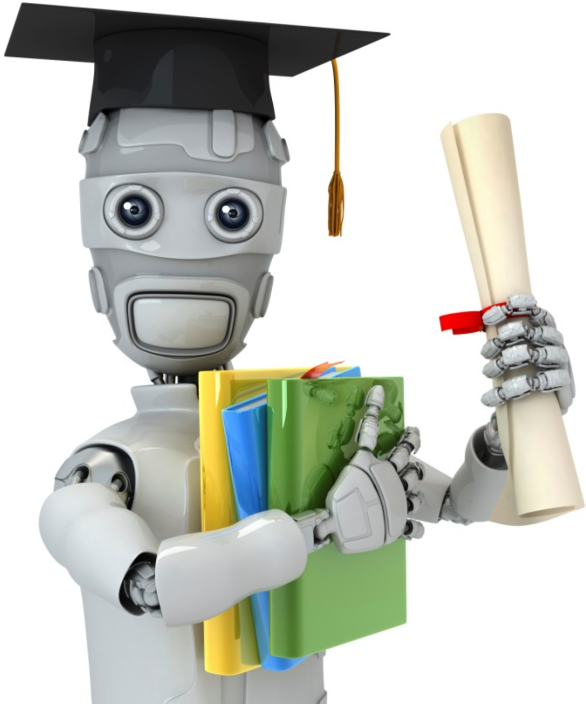
Run K-means. Get $c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K$.

Compute cost function (distortion)

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

}

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



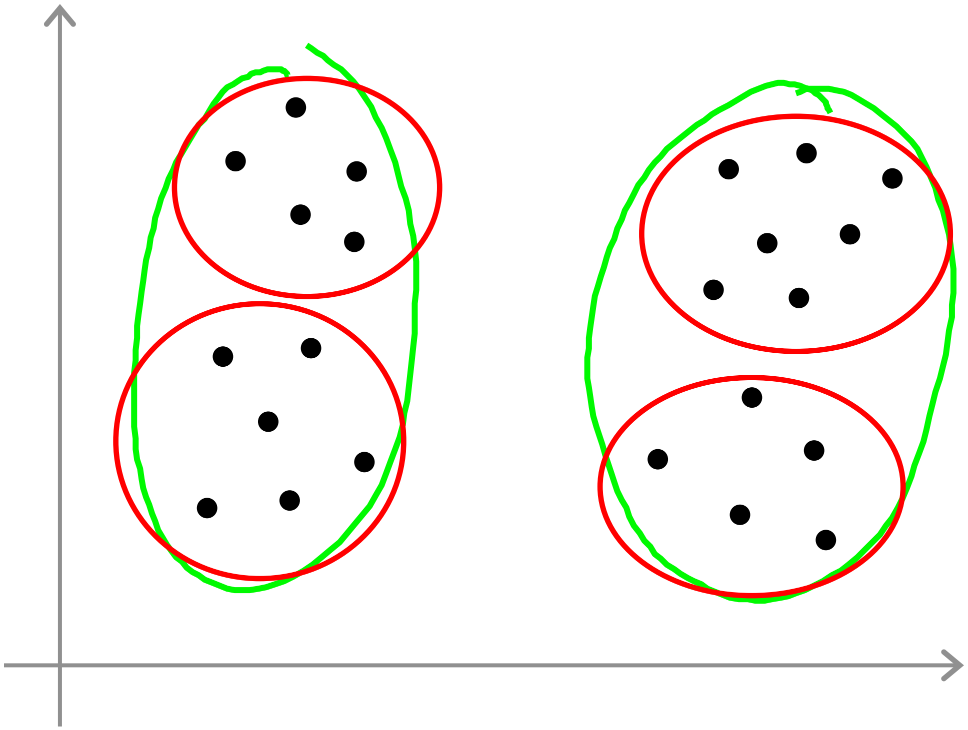
Machine Learning

Clustering

Choosing the number of clusters

What is the right value of K?

So far, the most common way is to choose k by hand!

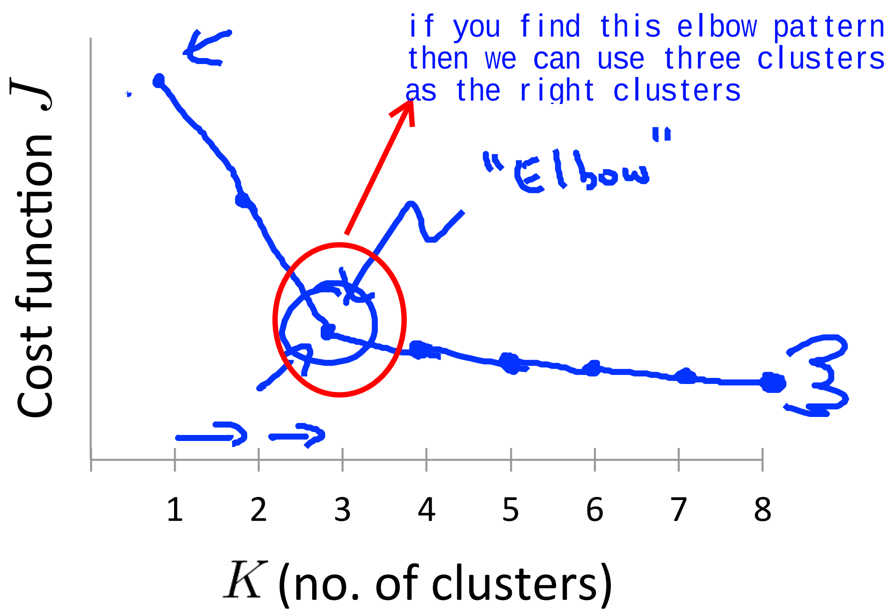


Andrew Ng

Choosing the value of K

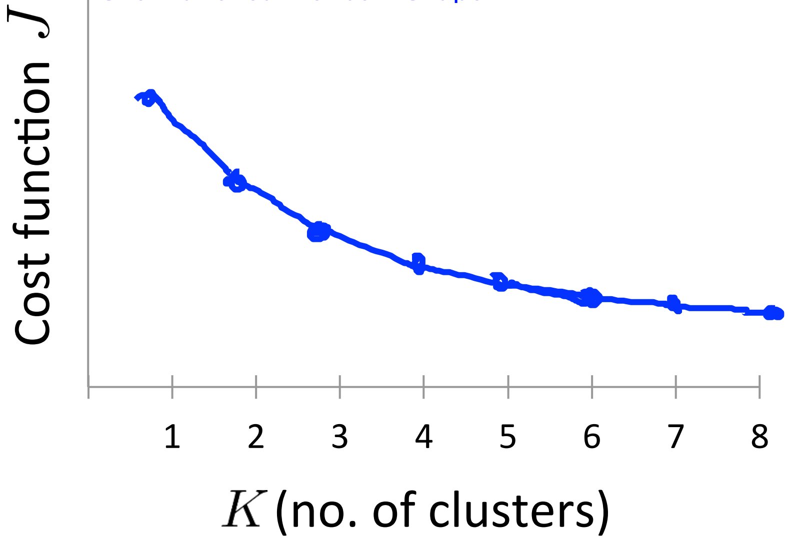
Elbow method:

No high expectation for any particular problem



run k-means with diff clusters

Sometimes, you have this pattern, which does not show a clear elbow shape.



Choosing the value of K

Another way to choose number of clusters

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.

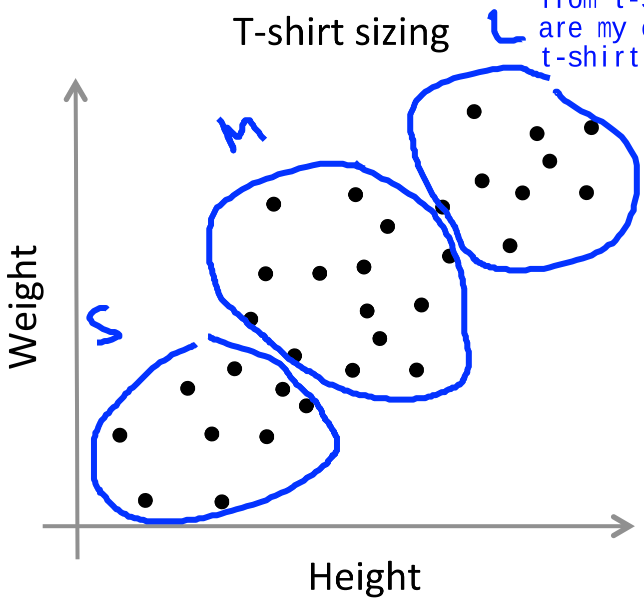
K=3 S,M,L

Do we want to choose 3 sizes or 5 sizes? We can decide this from t-shirt sale business.e.g. are my customer happier with 5 t-shirt sizes?

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

E.g.

T-shirt sizing



T-shirt sizing