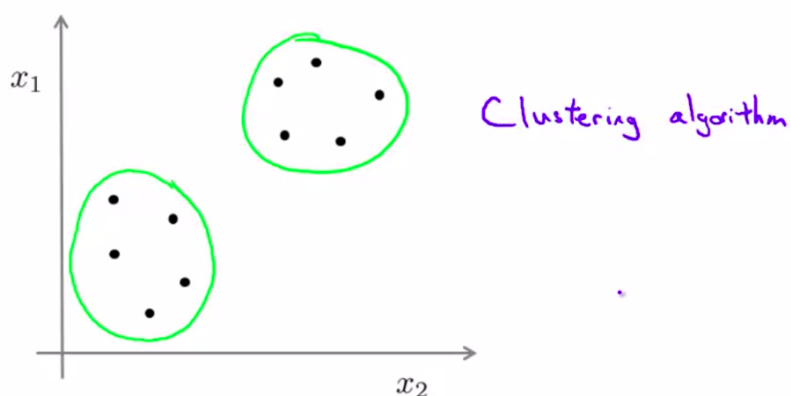


Week 8 - Lecture 1

Unsupervised Learning

In an unsupervised learning problem we have a dataset with unlabeled (unstructured) data, that may look like the following:

Unsupervised learning



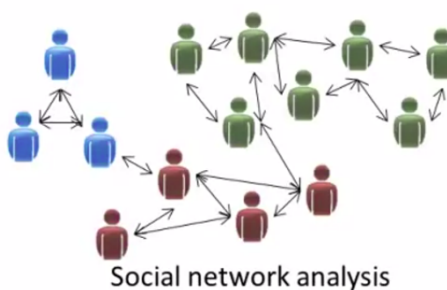
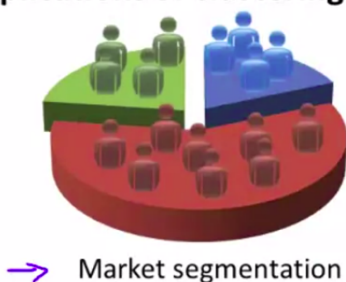
Training set: $\{\underline{x^{(1)}}, \underline{x^{(2)}}, \underline{x^{(3)}}, \dots, \underline{x^{(m)}}\}$ ←

Andrew Ng

and we let the algorithm find some structure on the data for us.

Use cases of clustering:

Applications of clustering

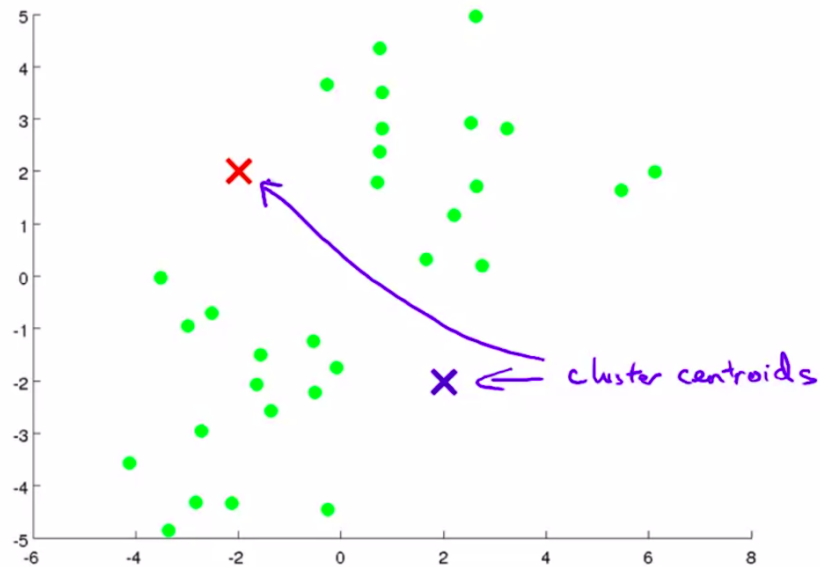


Andrew Ng

K-Means Algorithm

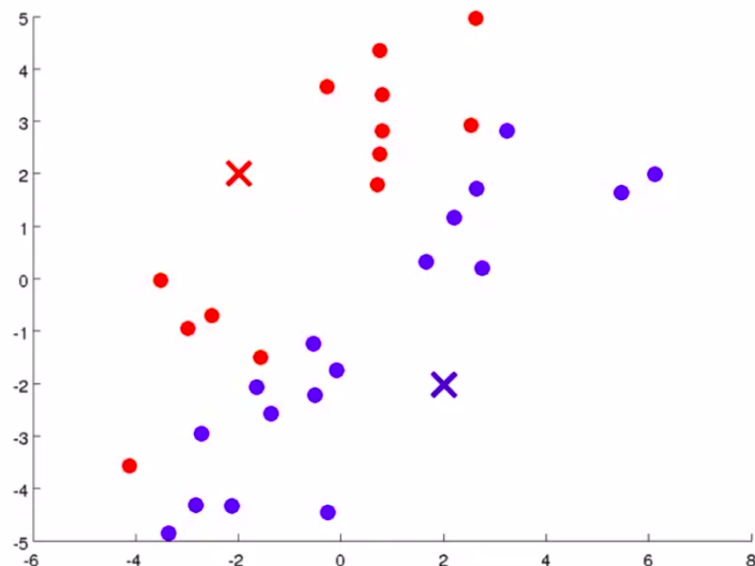
In the clustering problem, we are given a dataset and we would like to find some structure in it, i.e., to divide it into some coherent groups. K-Means is one clustering algorithm.

- **Step 1:** Initialization. The algorithm randomly chooses k centroids.



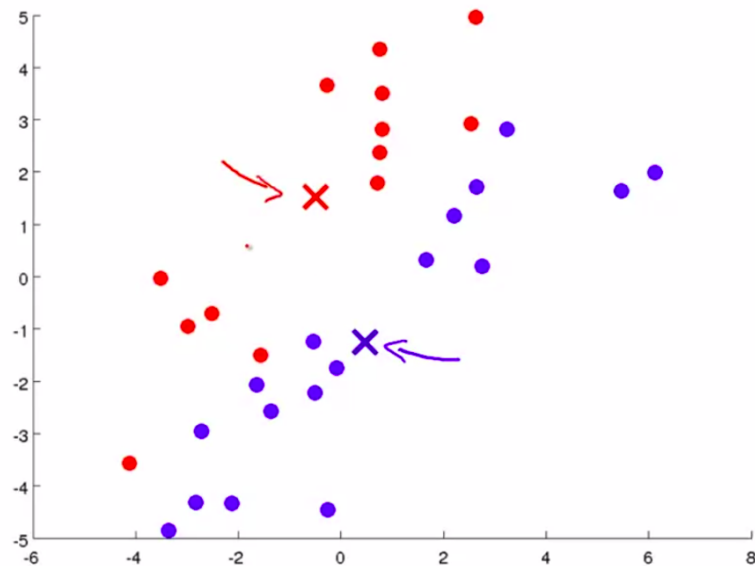
Andrew Ng

- **Step 2:** (Cluster assignment step) The algorithm assigns a label to each data point depending on which centroid is closer.



Andrew Ng

- **Step 3:** (Move centroids step) The algorithm moves the centroids and then repeats itself from step 2.



Andrew Ng

Note: the algorithm repeats steps 2 and 3 until it converges.

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)

Andrew Ng

K-means algorithm

μ_1 μ_2
x x

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

Repeat {

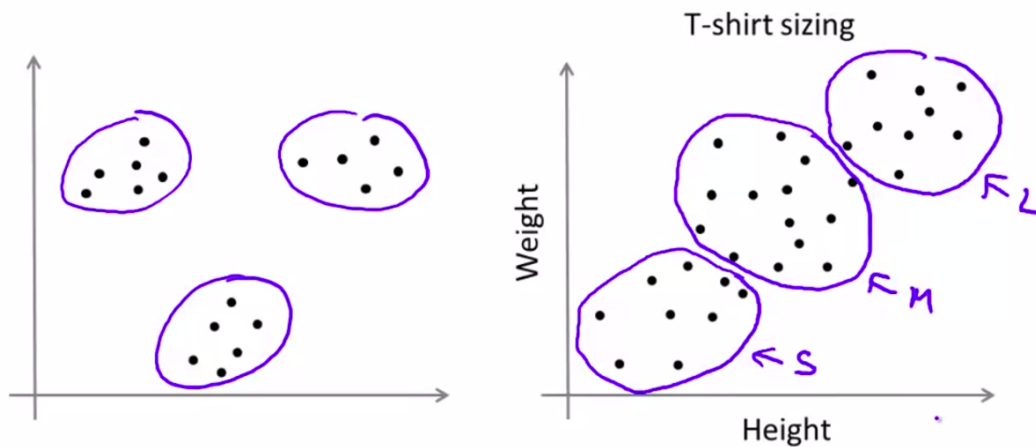
Cluster assignment step for $i = 1$ to m
 $c^{(i)} :=$ index (from 1 to K) of cluster centroid closest to $x^{(i)}$
 $\min_k \|x^{(i)} - \mu_k\|^2 \rightarrow c^{(i)}$
Move centroid for $k = 1$ to K
 $\rightarrow \mu_k :=$ average (mean) of points assigned to cluster k
 $x^{(1)}, x^{(5)}, x^{(6)}, x^{(10)} \rightarrow c^{(1)}=2, c^{(5)}=2, c^{(6)}=2, c^{(10)}=2$
 $\mu_2 = \frac{1}{4} [x^{(1)} + x^{(5)} + x^{(6)} + x^{(10)}] \in \mathbb{R}^n$

Andrew Ng

Example:

K-means for non-separated clusters

S, M, L



Andrew Ng

Optimization Objective

K-means optimization objective

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

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

Optimization objective:

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

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

Distortion



Andrew Ng

Random Initialization

Random initialization

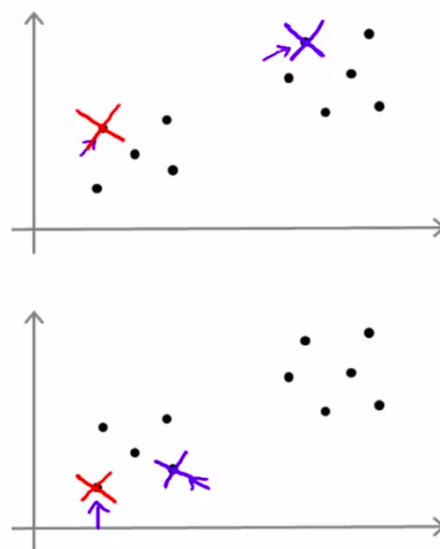
Should have $K < m$

$K=2$

Randomly pick K training examples.

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

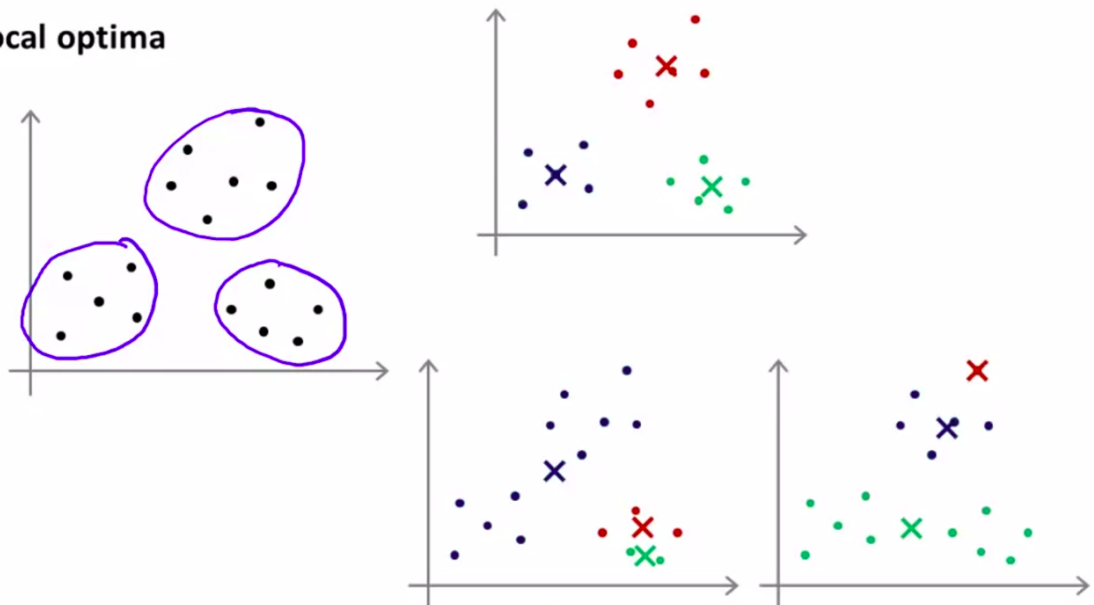
$\mu_1 = x^{(i)}$
 $\mu_2 = x^{(j)}$
⋮



Andrew Ng

Depending on which initialization centroids we choose, K-means might end up in a local optima.

Local optima



Andrew Ng

In the figure above, the upper case might be the global optima and the two on the bottom might be the local optima (suboptimal solutions)

Good practice to avoid this:

Random initialization

For $i = 1$ to 100 { 50 - 1000

- Randomly initialize K-means.
- Run K-means. Get $c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K$.
- Compute cost function (distortion)
- $J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K)$

}

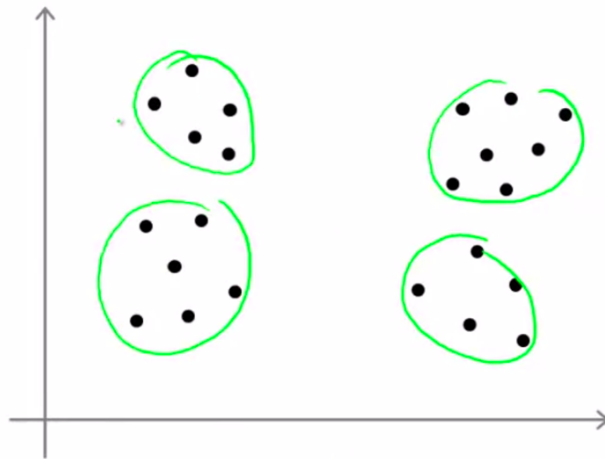
Pick clustering that gave lowest cost $J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K)$

Andrew Ng

Choosing the Number of Clusters

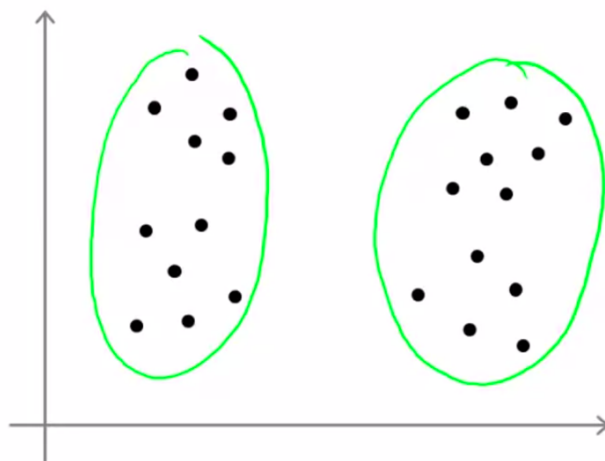
The most common way of choosing the number K of clusters is by doing it manually, looking at visualizations or other kind of intuition. So this is a somewhat intuitive step and, to help develop this intuition, some examples are presented below.

What is the right value of K?



Andrew Ng

What is the right value of K?



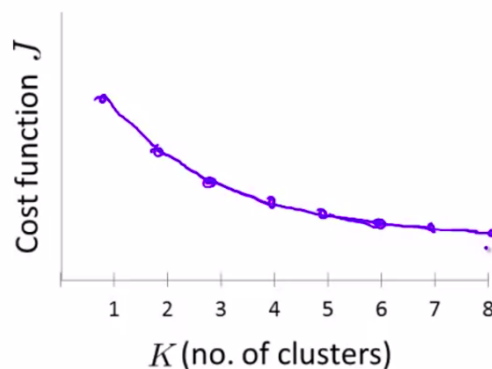
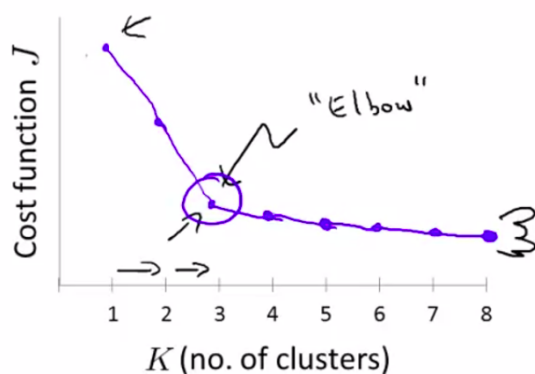
⏮ ⏪ ⏩ ⏭

Andrew Ng

Although, some methods might help with this choice:

Choosing the value of K

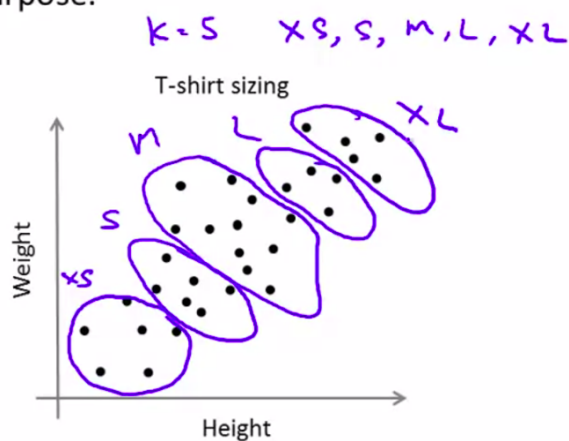
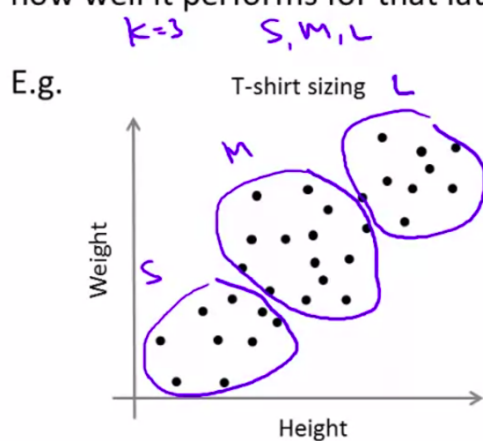
Elbow method:



Andrew Ng

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.



Andrew Ng

References

[1] [Machine Learning - Stanford University \(https://www.coursera.org/learn/machine-learning\)](https://www.coursera.org/learn/machine-learning).