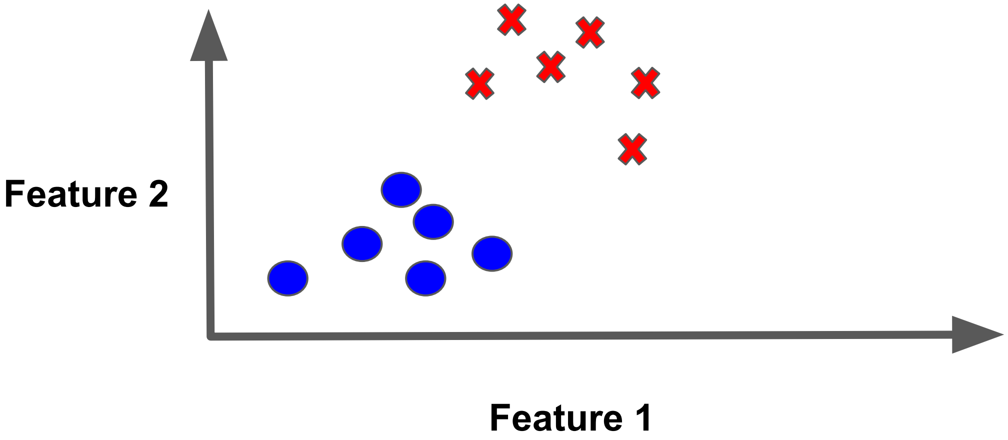
Clustering

Wajahat Hussain

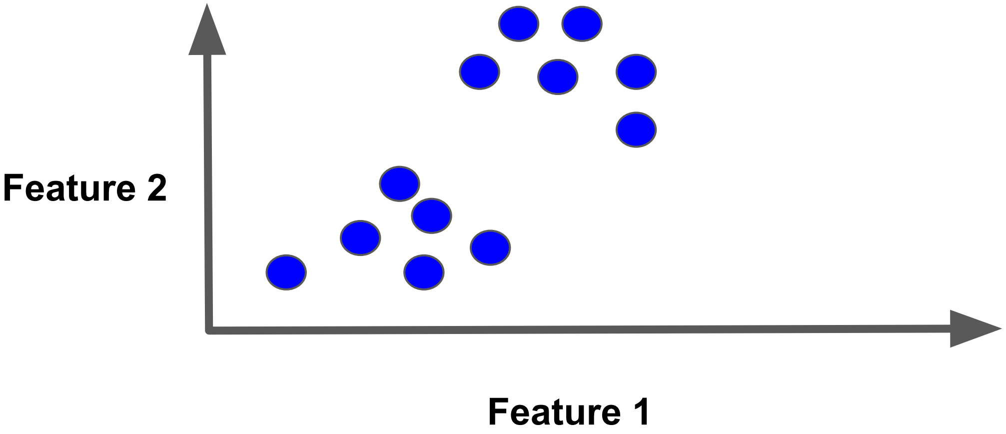
Divide this data into two clusters

X	Y
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
3	2
4	3
5	4
6	5
7	6
8	7
9	8
10	9

Supervised Learning



Unsupervised Learning



Inbox - wajahat.hussain@se | Recent - Google Drive | Copy of ML_Clustering_AI - | Login | Qalam | Geogebra - 3D Graphing –

geogebra.org/m/DRfuvhxR

GeoGebra

ASSIGN

Geogebra - 3D Graphing

Author: Scott R. Franklin, Ph.D.

ABC

G = (1, 32, 500)

H = (-33.24, 93.29, 105

I = (50, 19, 102)

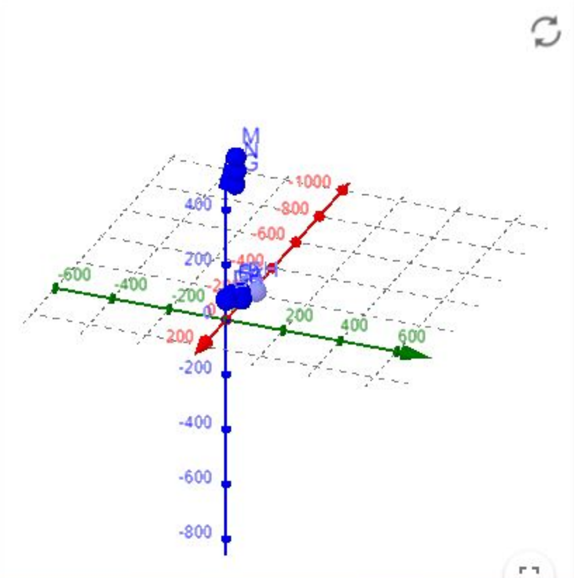
J = (1, 57, 98)

K = (1, 57, 91)

L = (1, 55, 85)

M = (1, 35, 600)

N = (1, 36, 550)



Easy Assigning

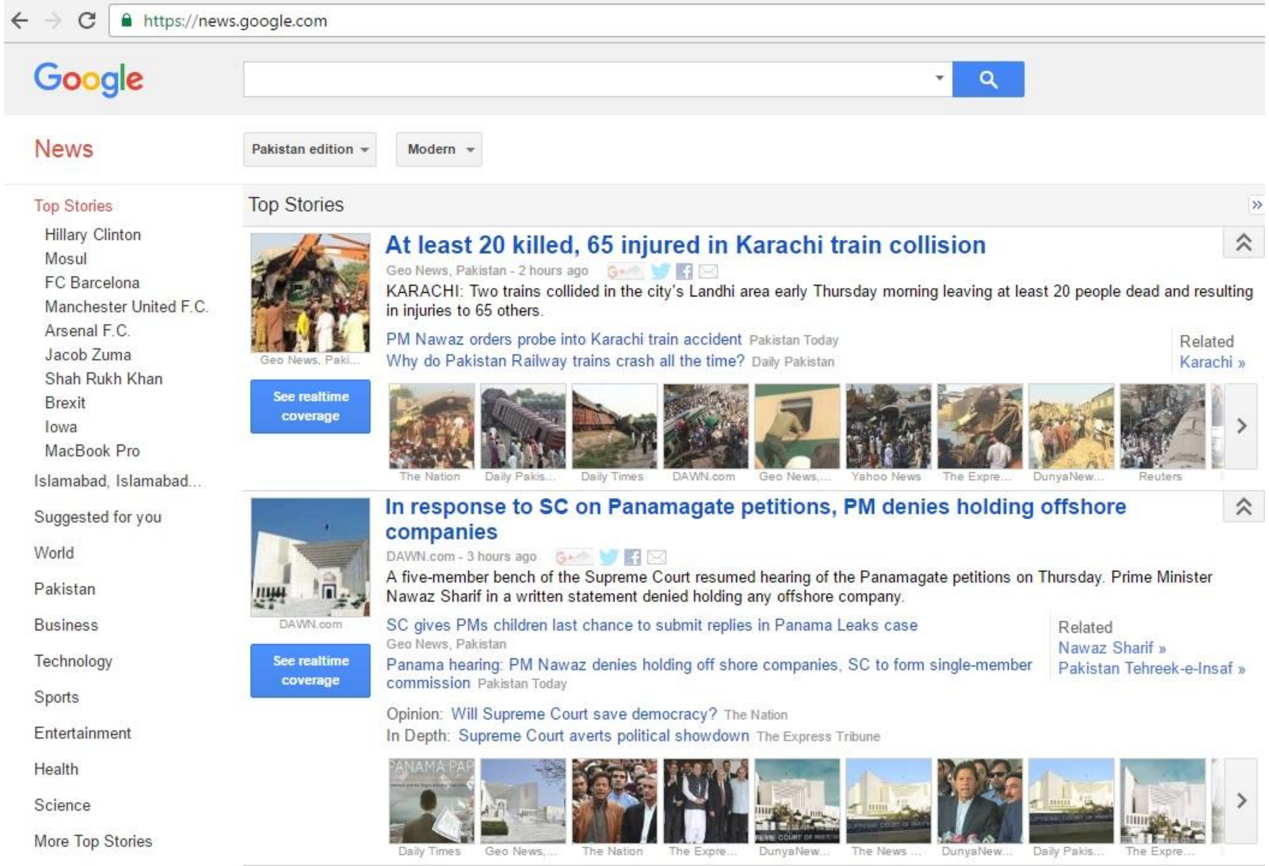
Assign to your Google classes, create a GeoGebra lesson and much more.

DISMISS

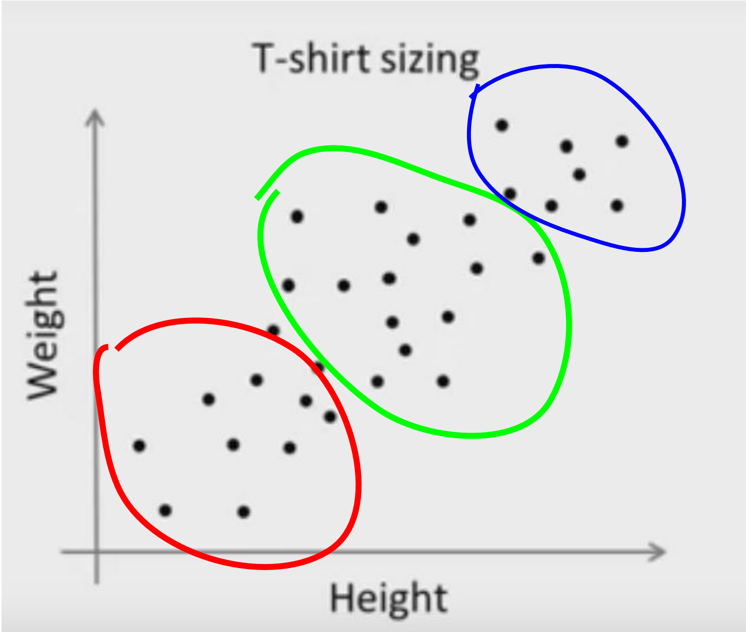
Divide this data into two clusters

X	Y
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
3	2
4	3
5	4
6	5
7	6
8	7
9	8
10	9

Unsupervised Learning

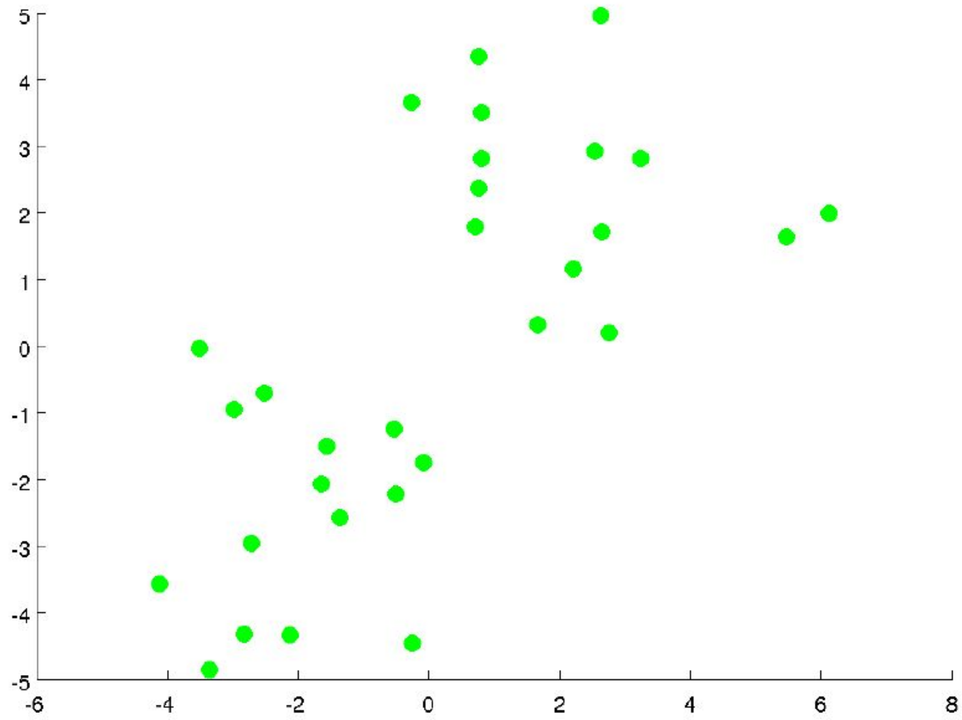


Unsupervised Learning

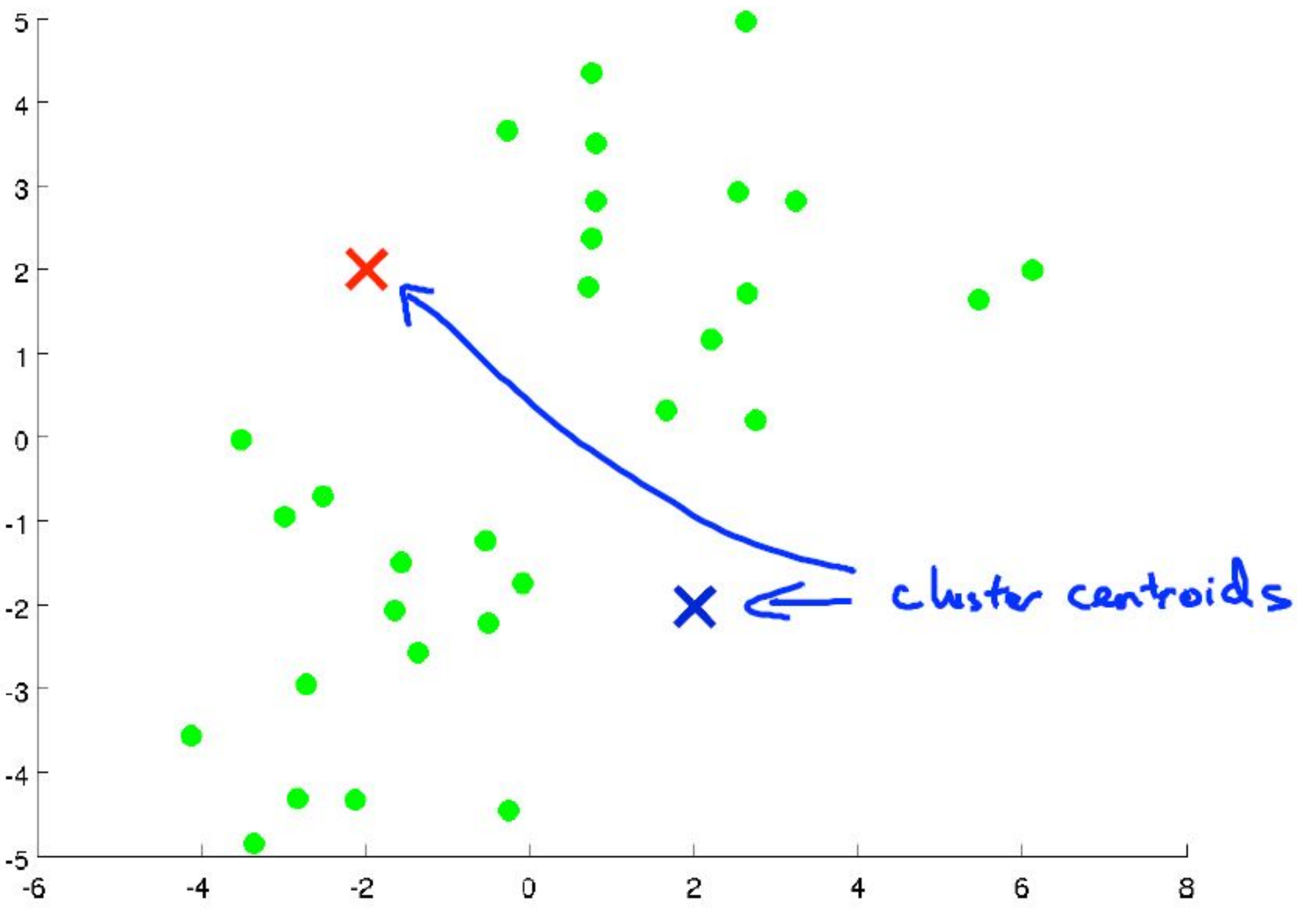


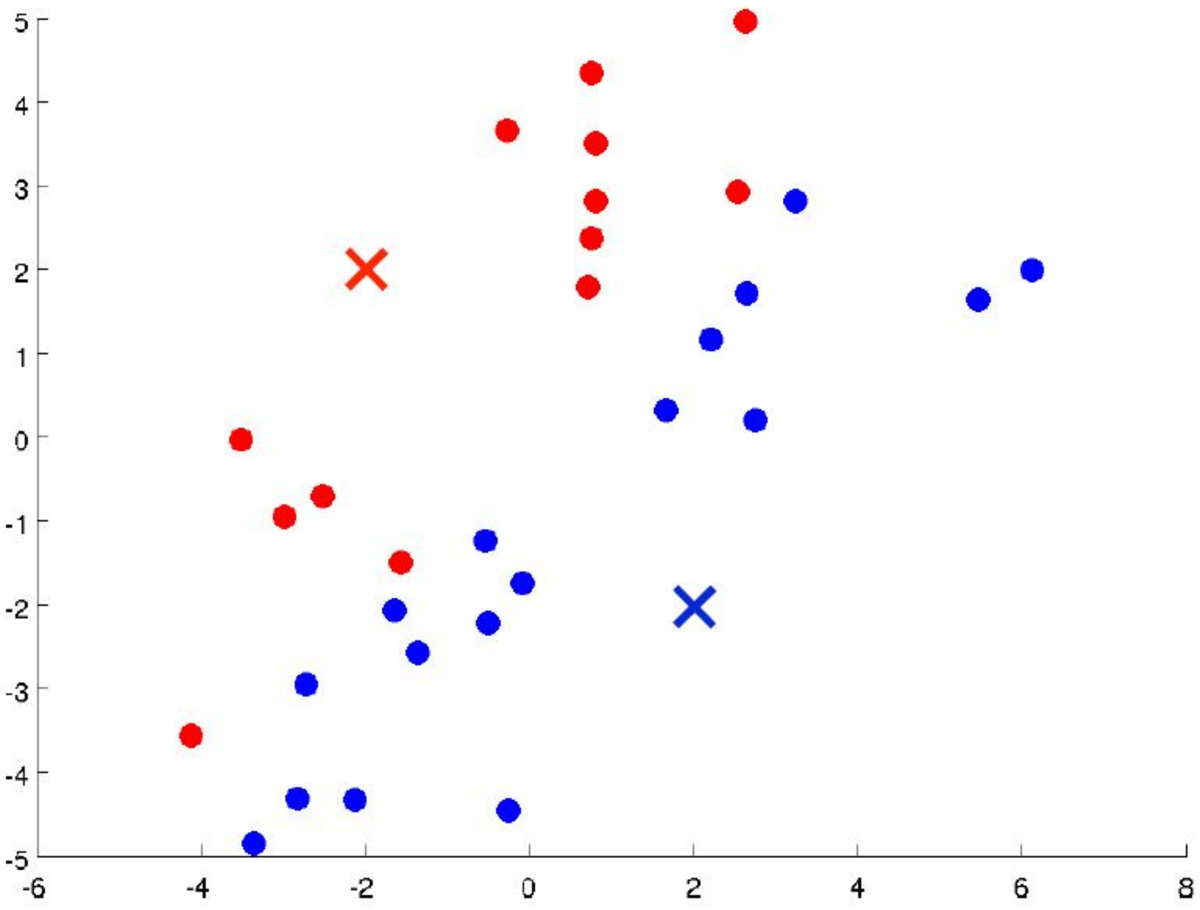
You want to design a shirt. How many sizes should there be?

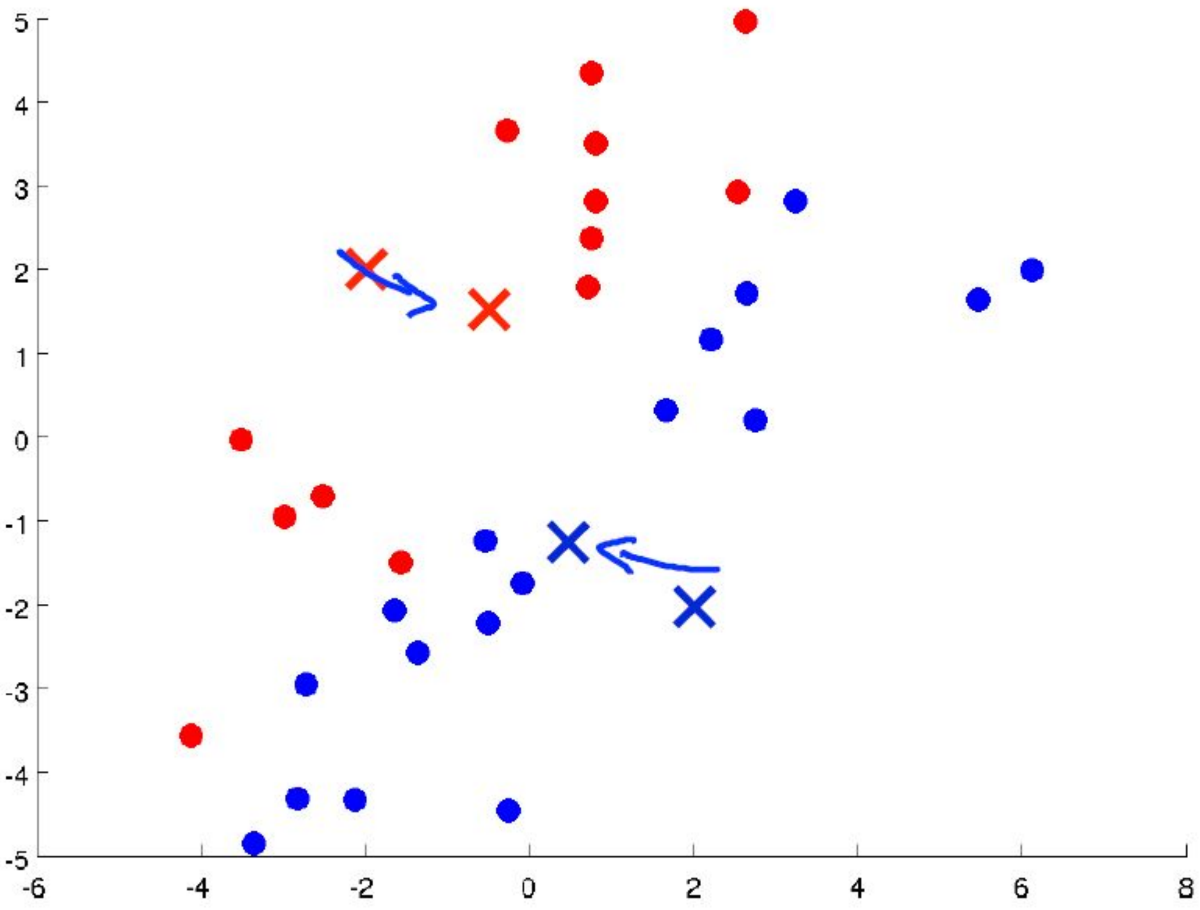
K-means Algorithm

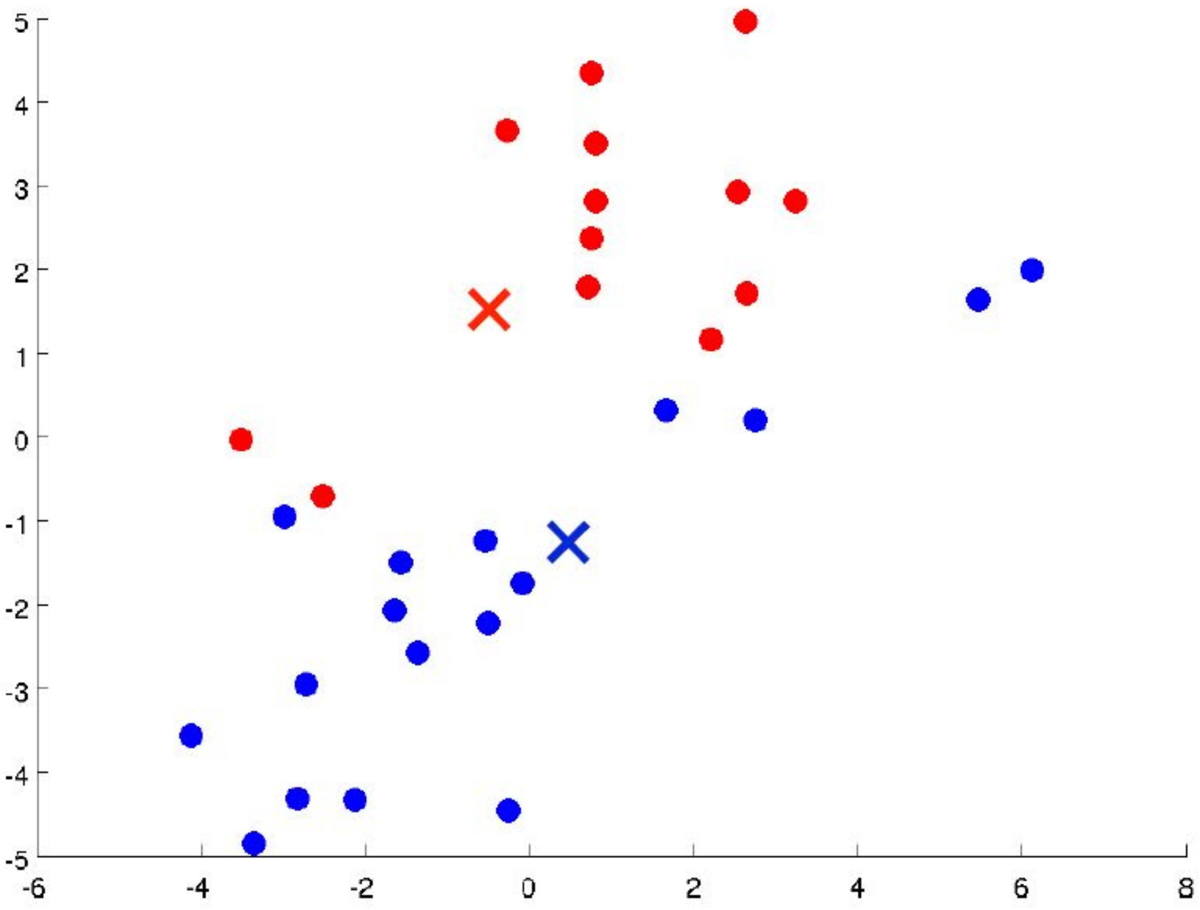


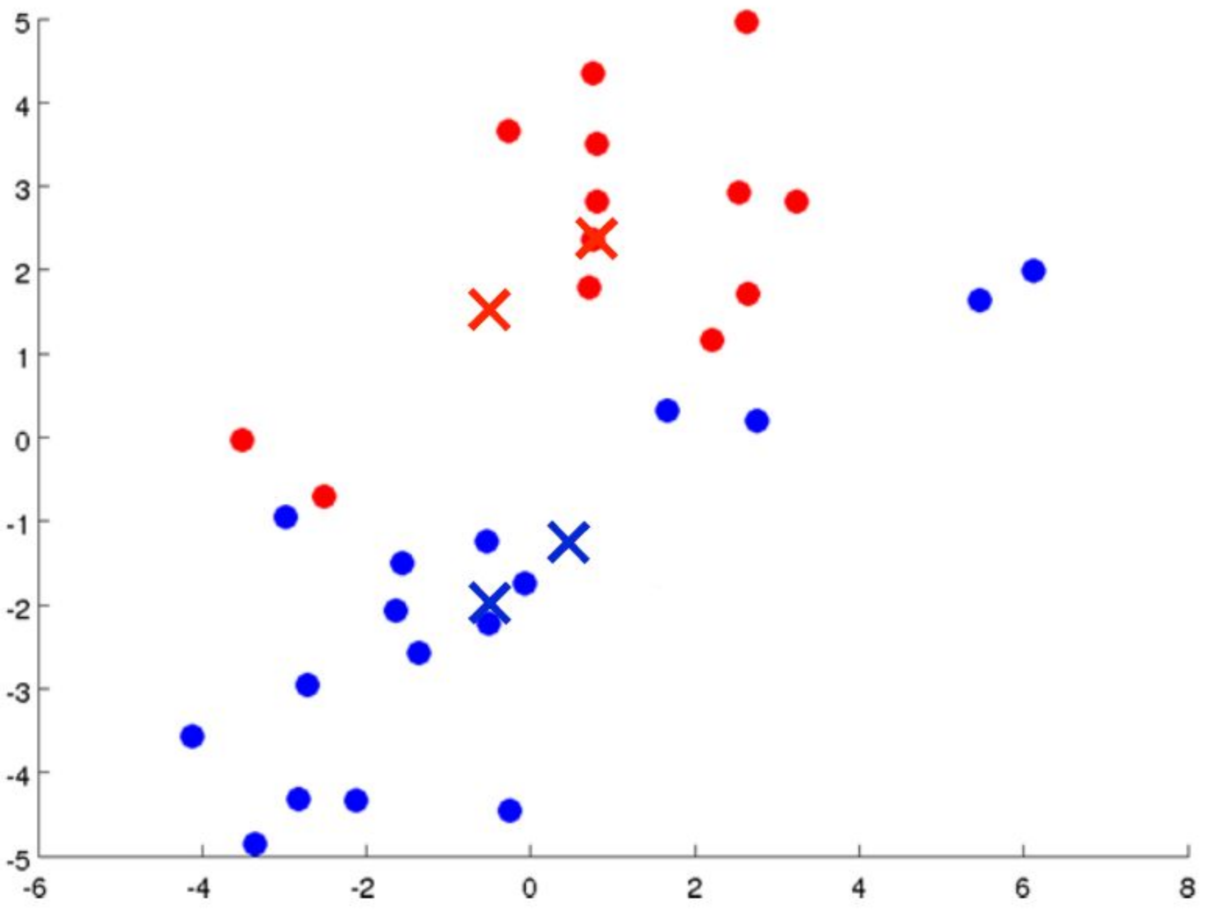
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8


cluster centroids
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

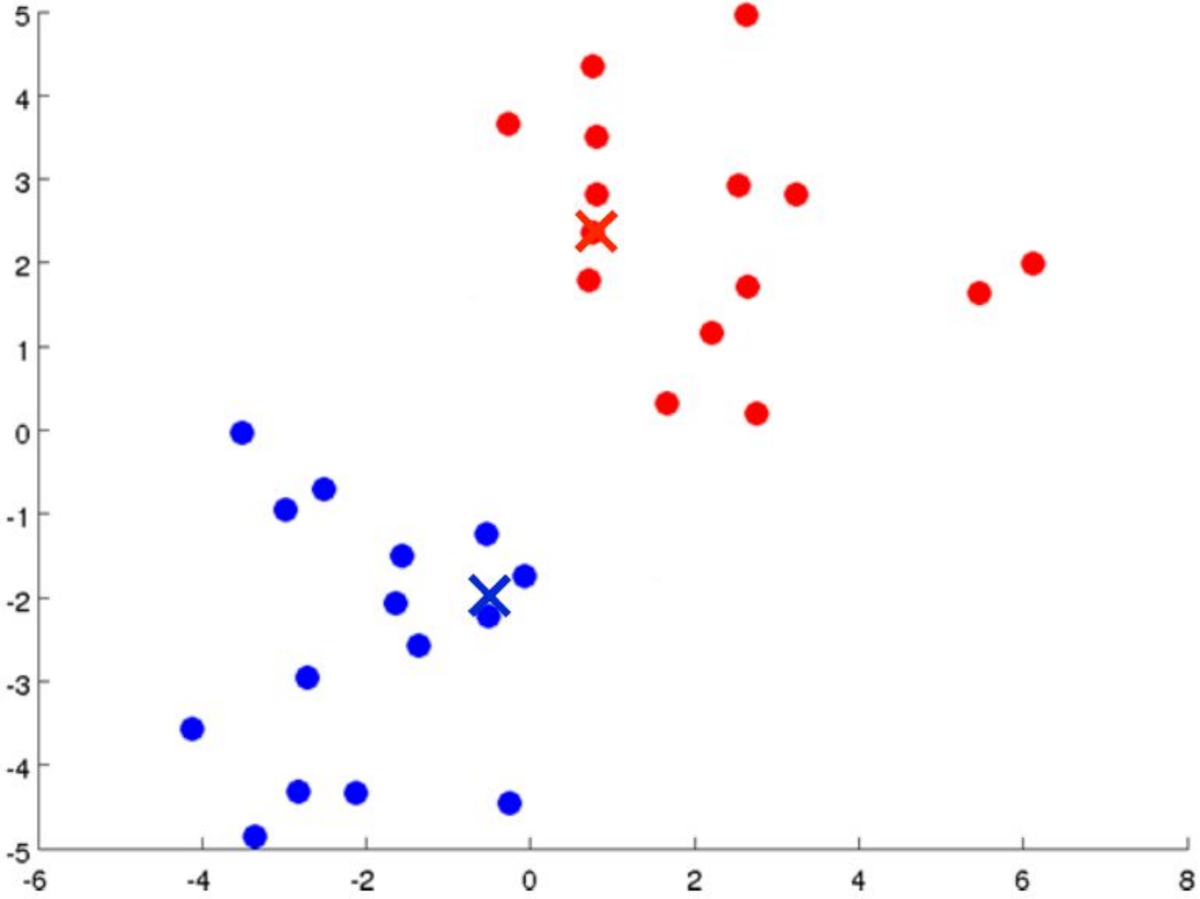

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

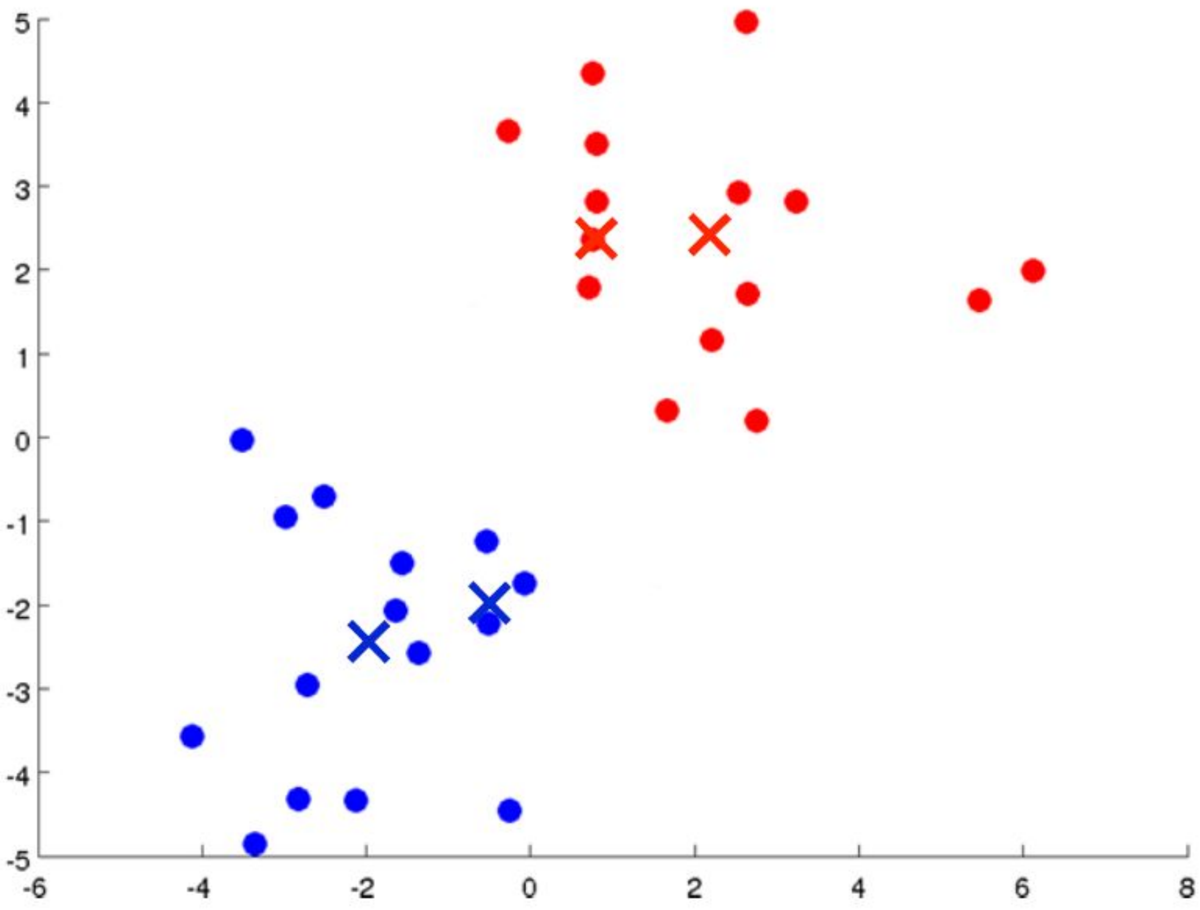

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8


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

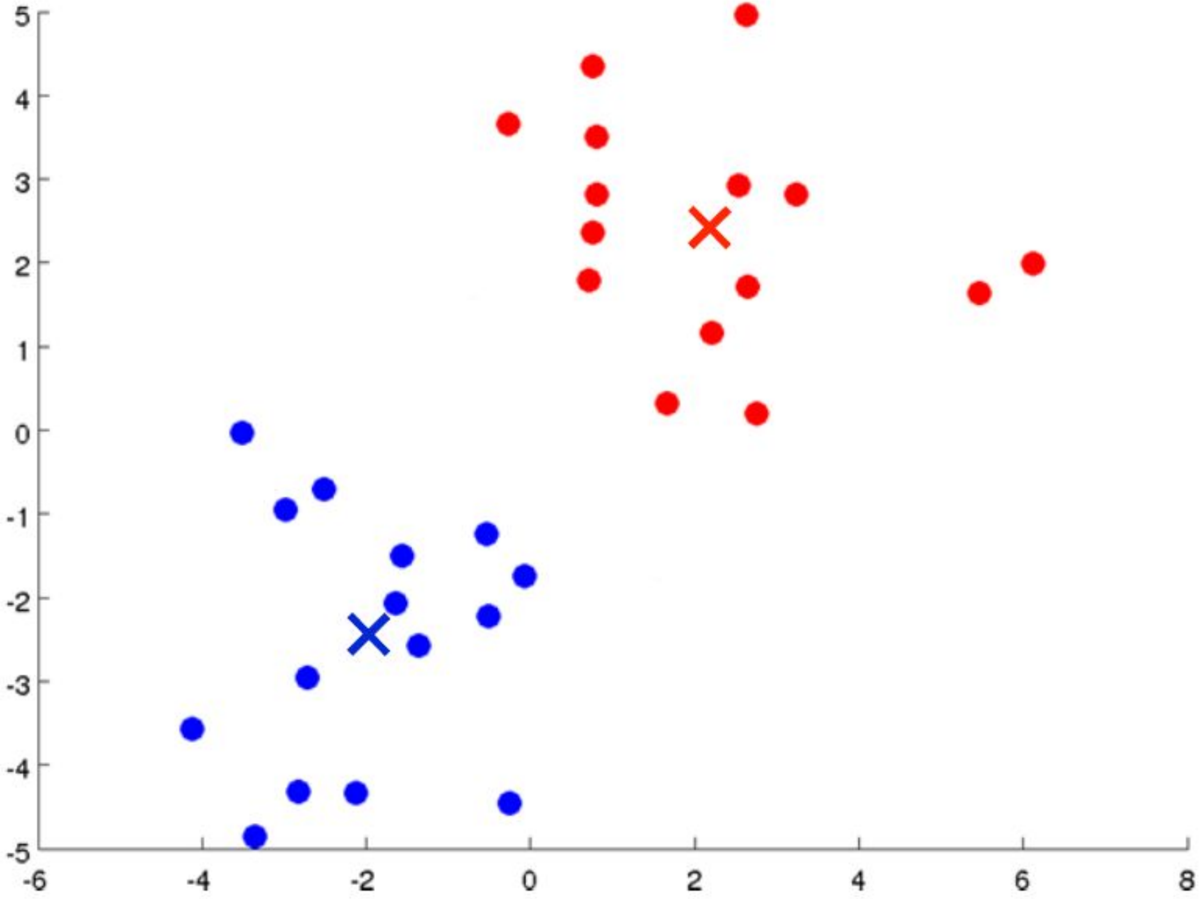


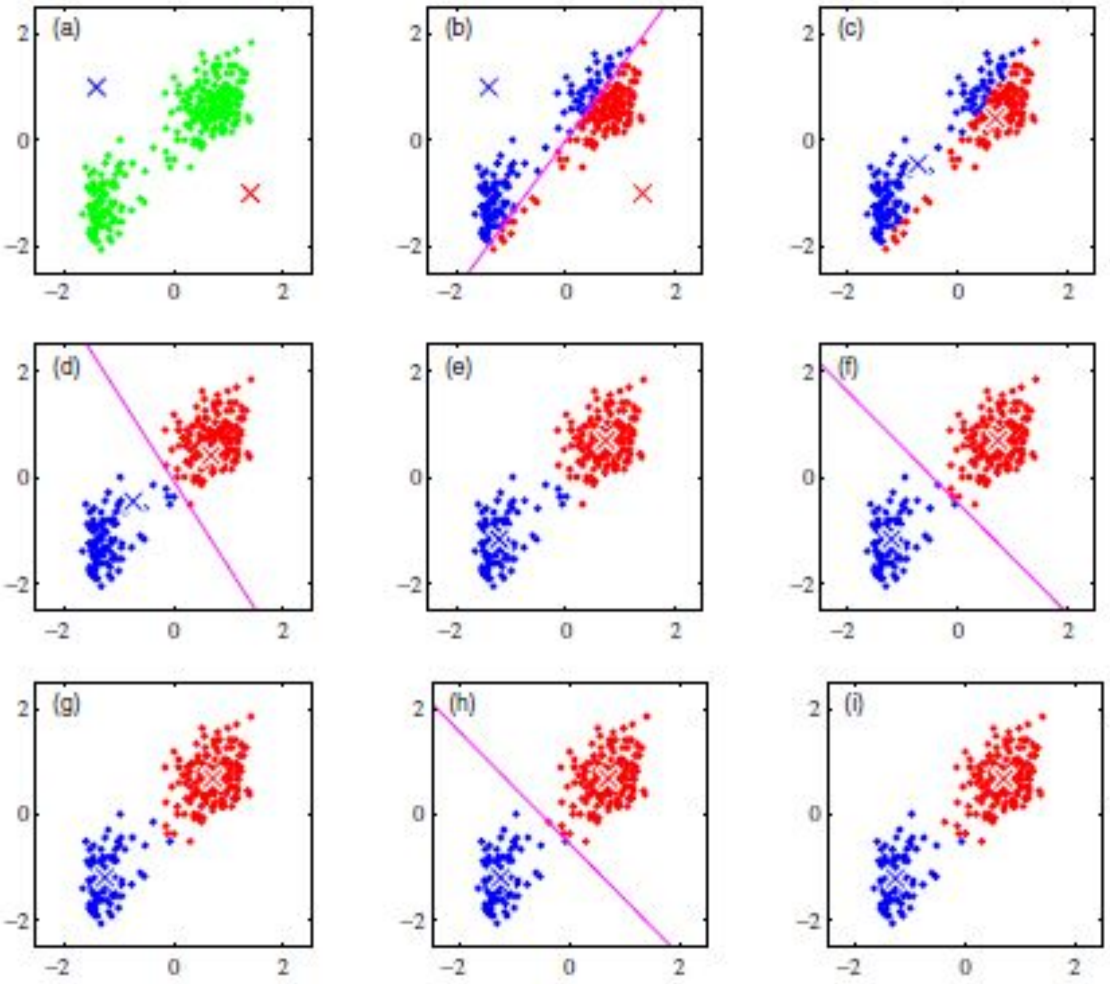
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8


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



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

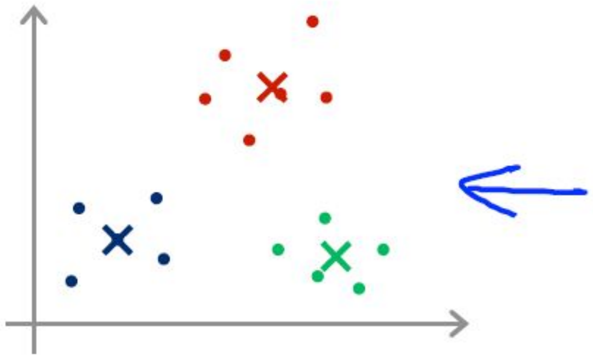

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

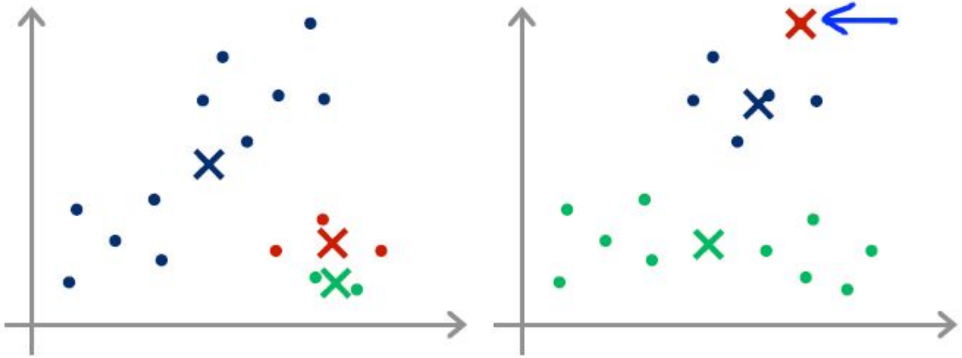


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
2
0
−2
−2
0
2

Is there any failure case?

Local optima





Andrew Ng

K-means algorithm

Input:

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

$x^{(i)} \in \mathbb{R}^n$

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

}

K-means optimization objective

$c^{(i)}$ = 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

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

Optimization objective:

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

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

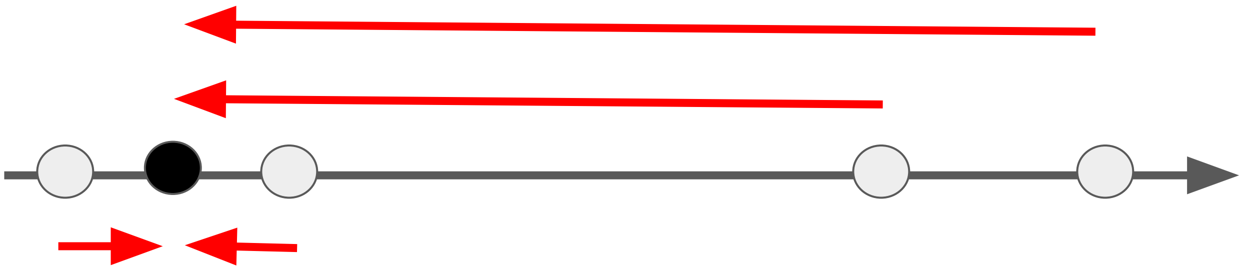
One dimensional example. Why average the points?

- We assume there is only a single cluster. The cluster center is blue/black circle. Which is a better cluster center? Black or blue?
- Red arrows shows the total error.
- Average of data points represents better cluster representation.

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

One dimensional example

- We assume there is only a single cluster. The cluster center is black circle
- Red arrows shows the total error.



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

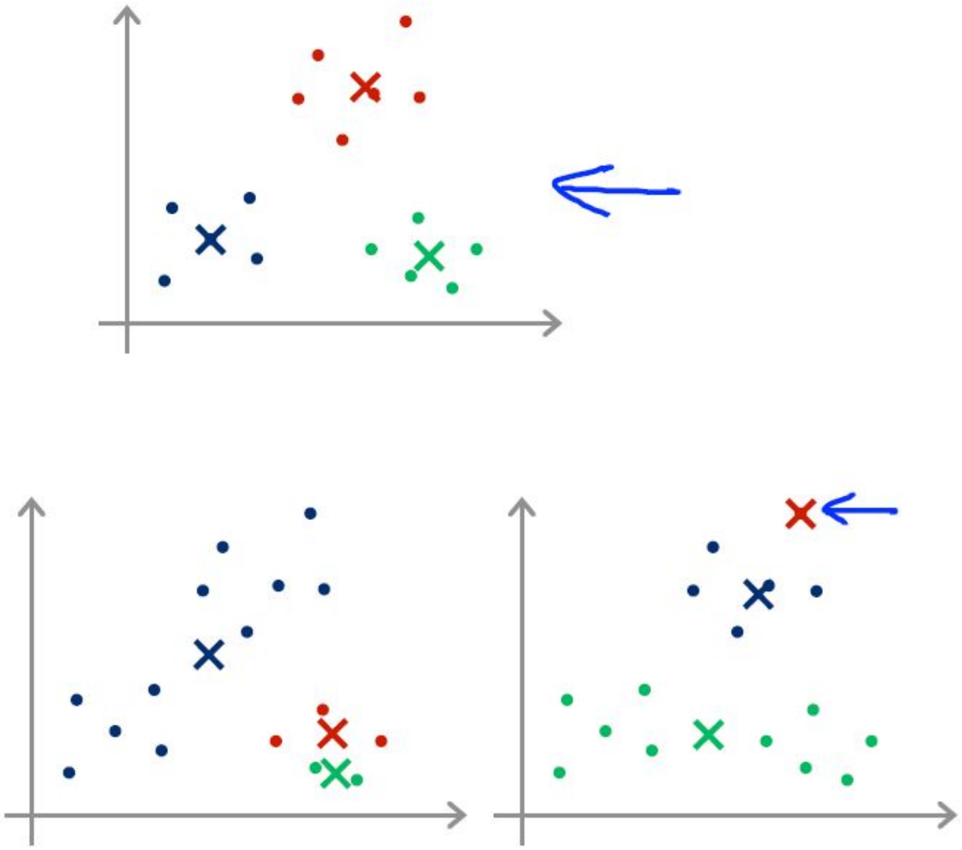
One dimensional example

- Now we assume there are two clusters. The cluster centers are black circles.
- Red arrows shows the total error.
- Increasing clusters reduces the error.

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

Is there any failure case?

Local optima



Andrew Ng

How to avoid the local optima?

Random initialization

For i = 1 to 100 {

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

Andrew Ng

One dimensional example

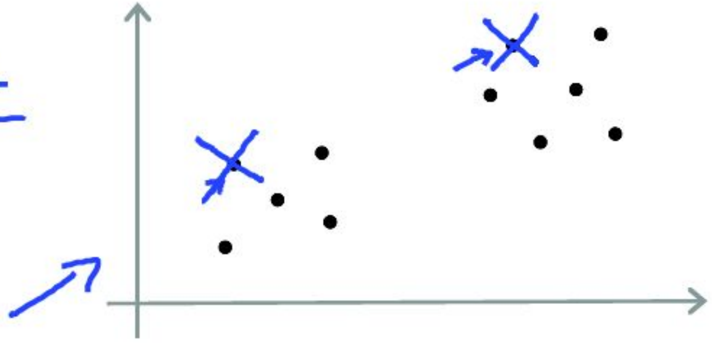
How to randomly initialize the clusters centers?

Random initialization

$K=2$

Should have $K < m$

Randomly pick K training examples.

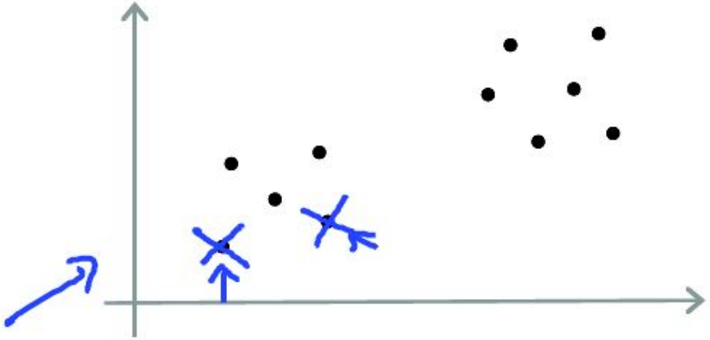


Set $\mu_1, \dots, \mu_K$ equal to these K examples.

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

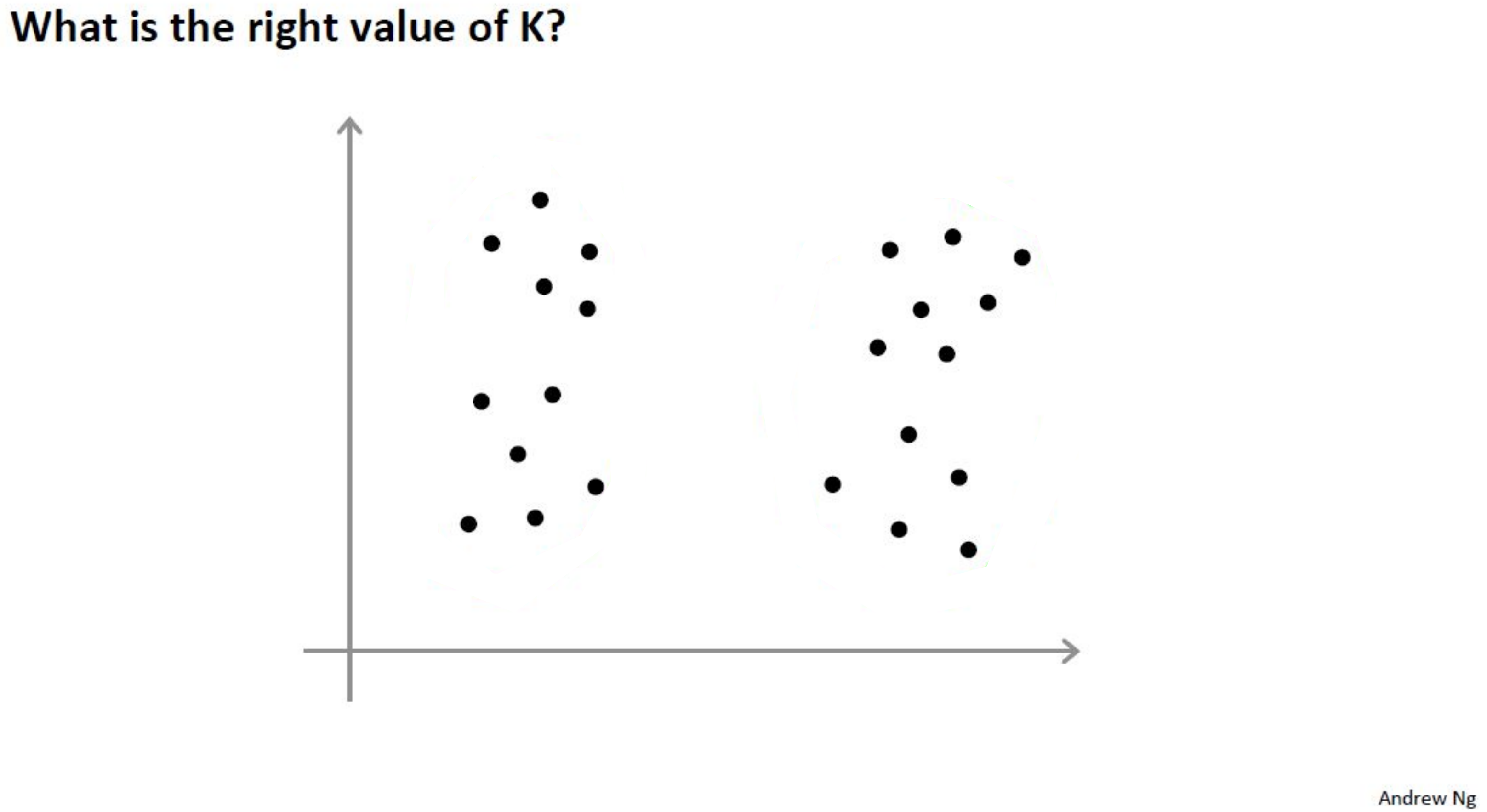
$\mu_2 = x^{(j)}$

$\vdots$

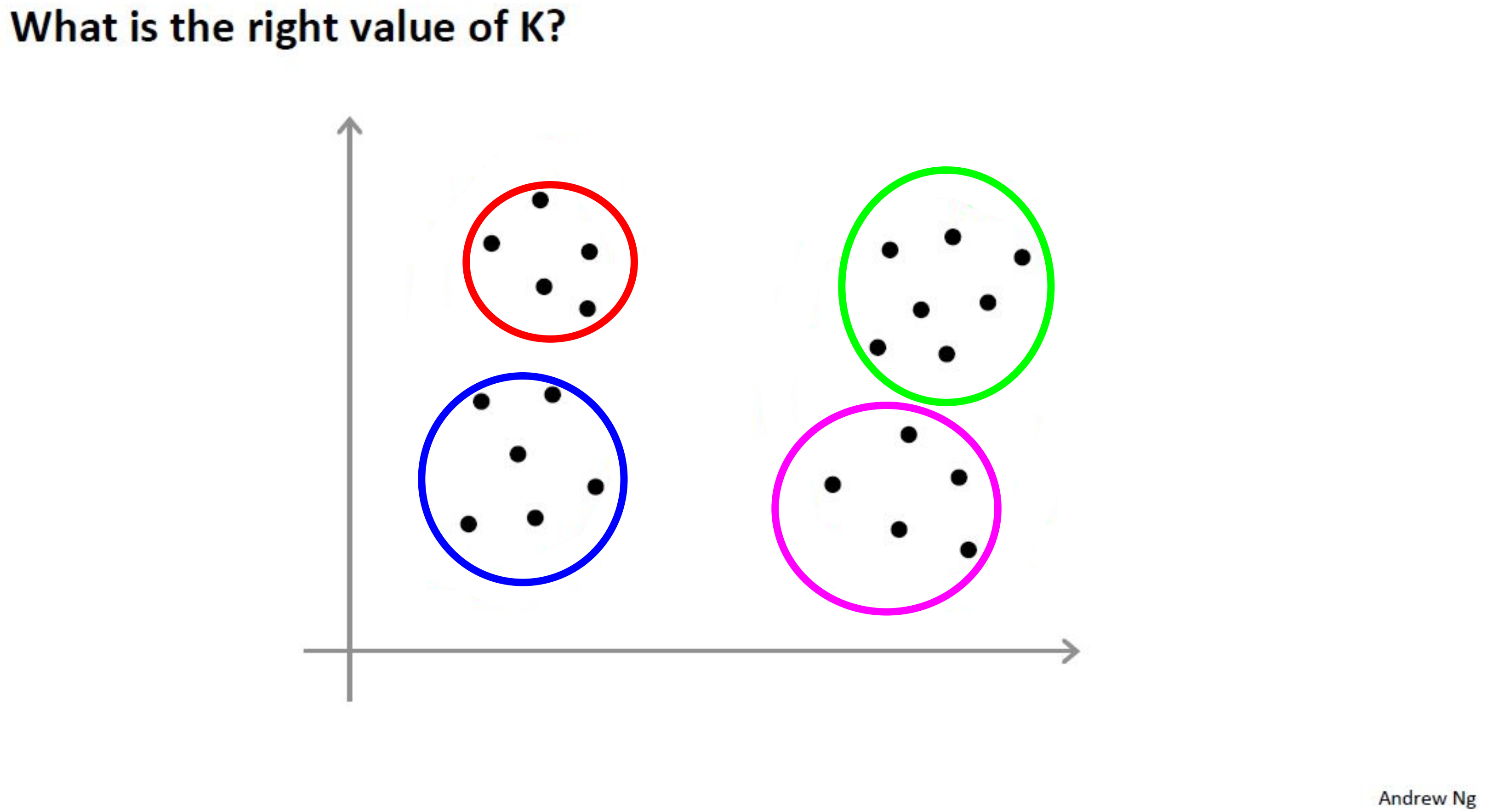


Andrew Ng

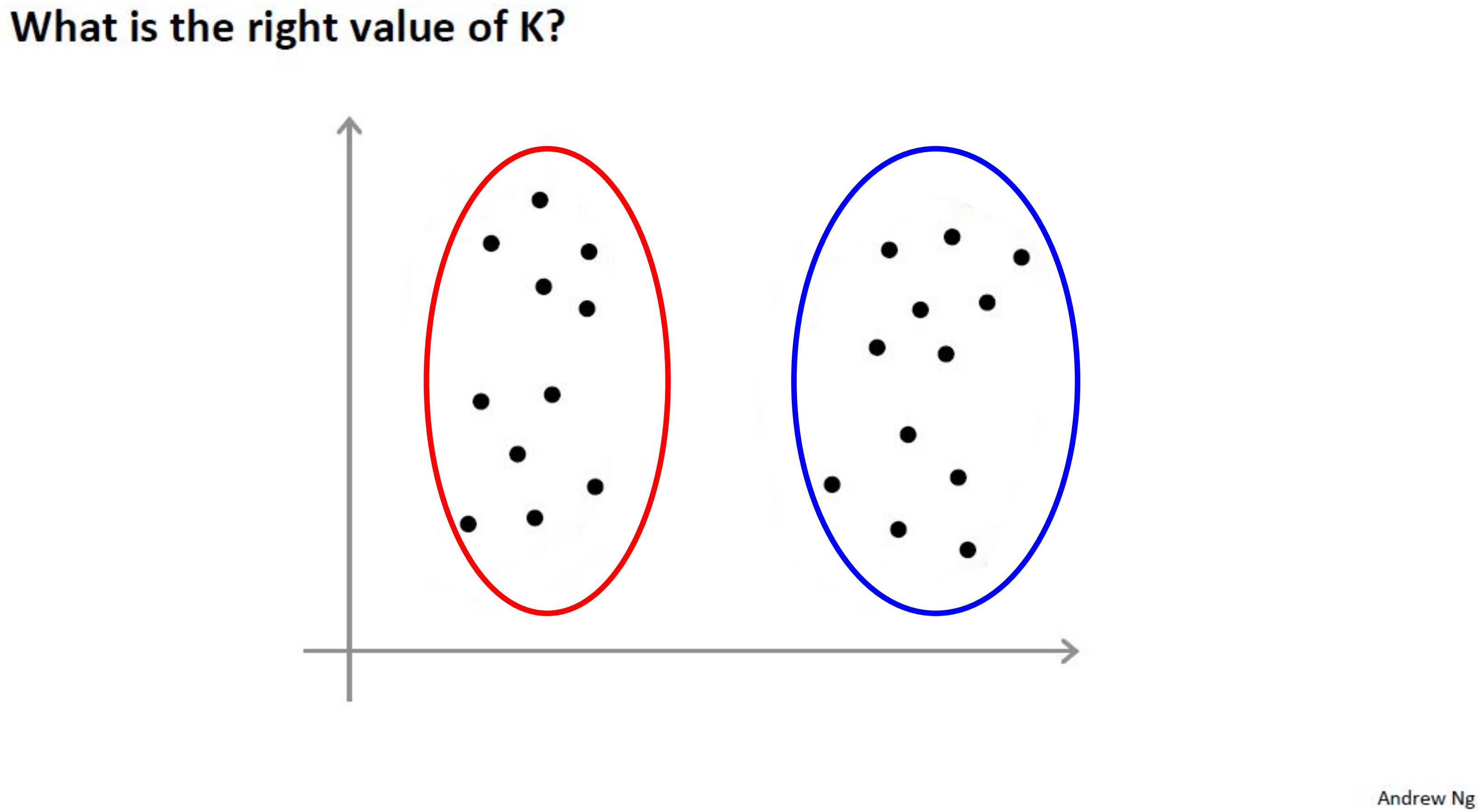
What is the right value of K?



What is the right value of K? 4?

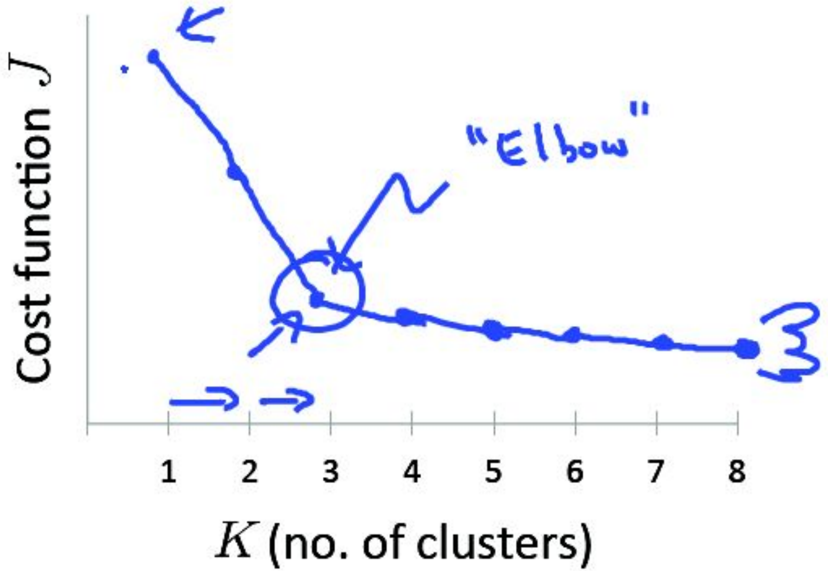


What is the right value of K? 2?



Choosing the value of K

Elbow method:



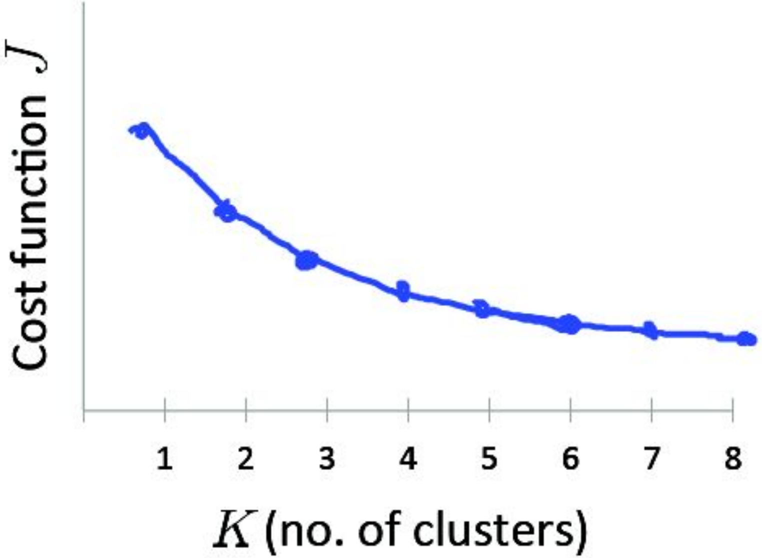


Image Compression Using K-means

Image Compression Using K-means

N pixels

R, G, B Channel

8 bits per value

Total bits = **24 N** bits

K clusters

Total bits =24 K + N $\log_2$ K bits

http://ieeexplore.ieee.org › document

Love it or leave it? A new look at Signal Fidelity Measures

by Z Wang · 2009 · Cited by 2787 — In this article, we have reviewed the reasons why we (collectively) want to **love** or **leave** the venerable (but perhaps hoary) **MSE**.

DOI: 10.1109/MSP.2008.930649