Week 1

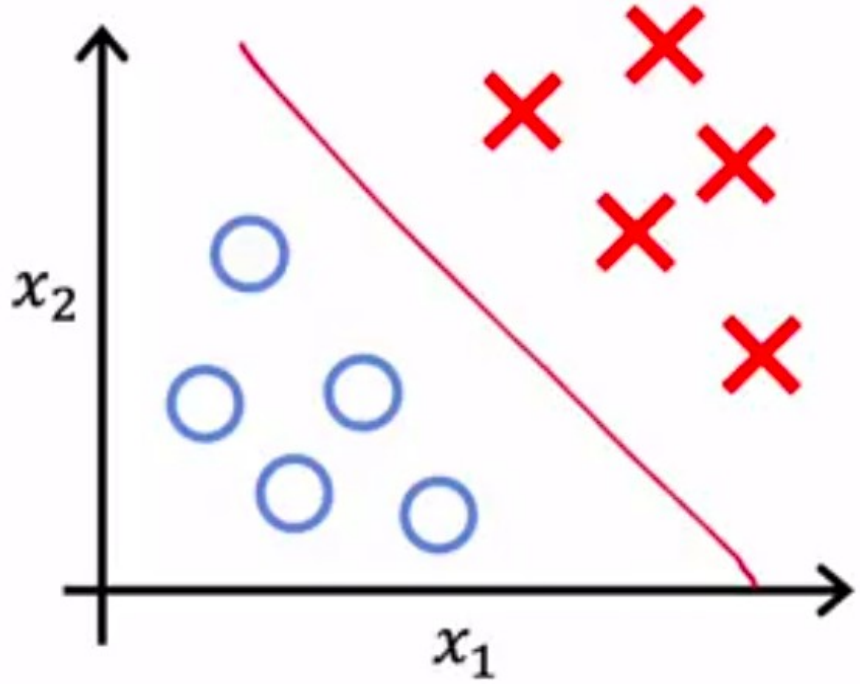
DeepLearning.AI

Stanford
ONLINE

Clustering

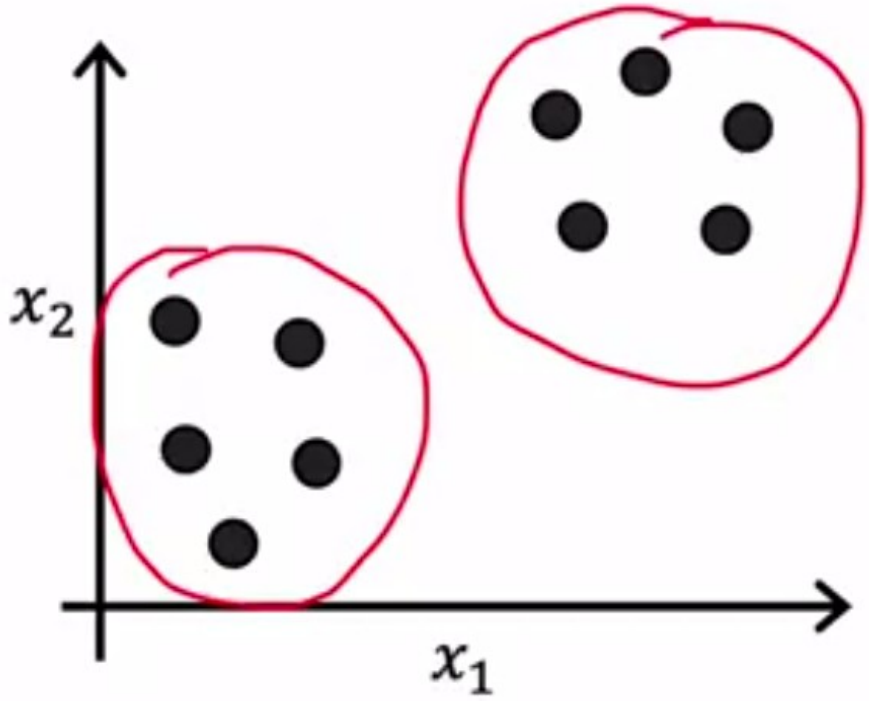
What is clustering?

Supervised learning



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

Unsupervised learning



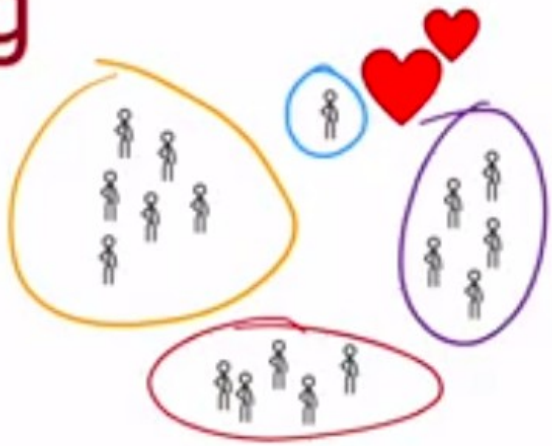
Clustering

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

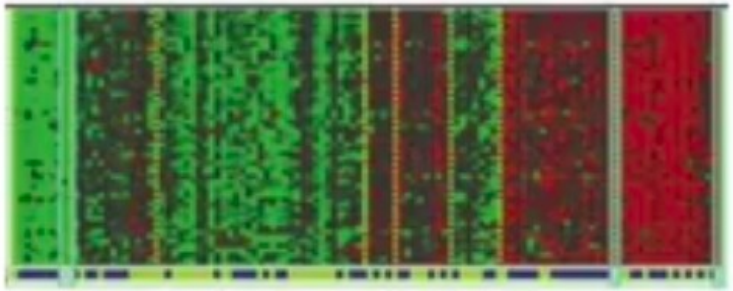
Applications of clustering

Grouping similar news

- Growing skills
- Develop career
- Stay updated with AI, understand how it affects your field of work



Market segmentation



DNA analysis

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

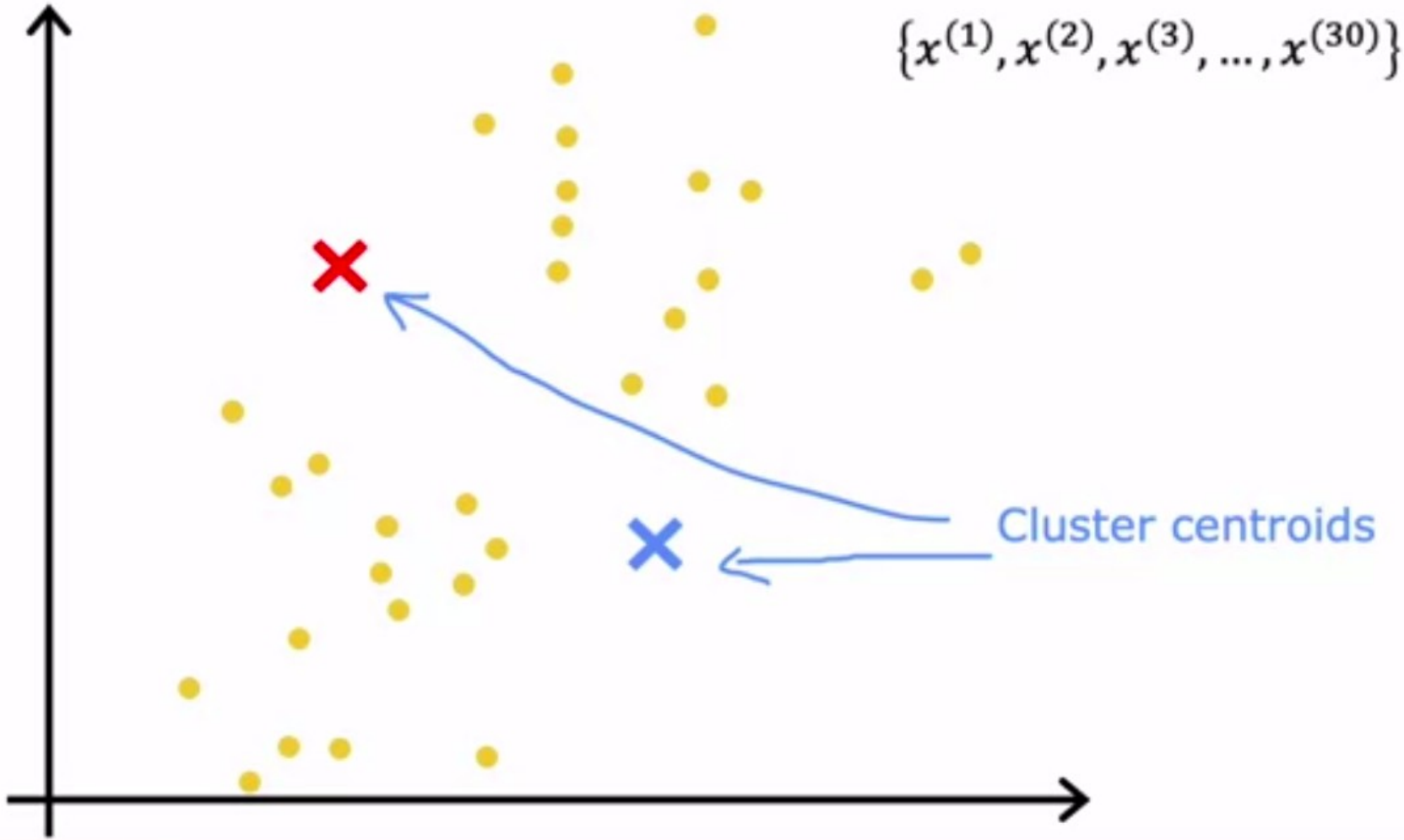
Astronomical data analysis

DeepLearning.AI

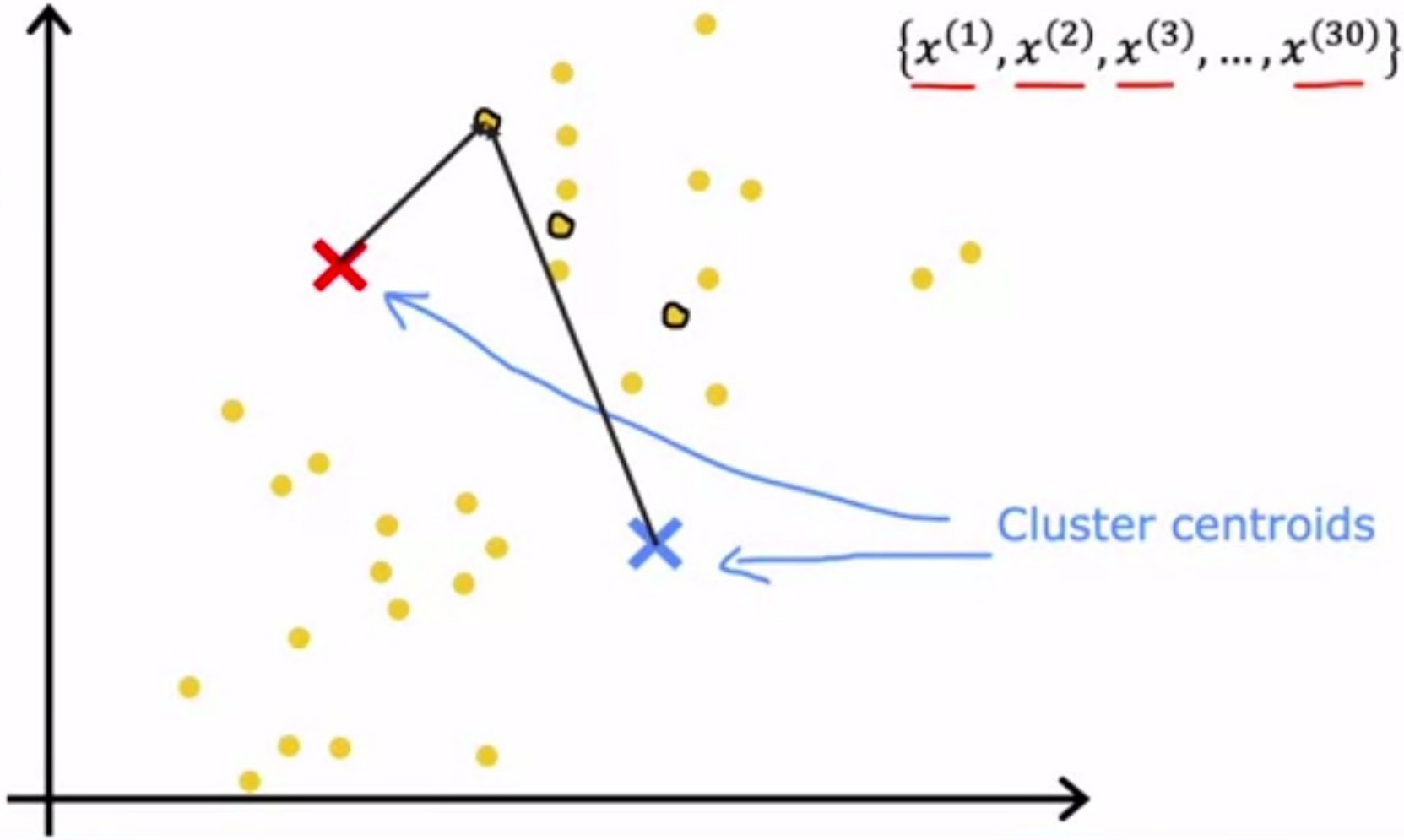
Stanford
ONLINE

Clustering

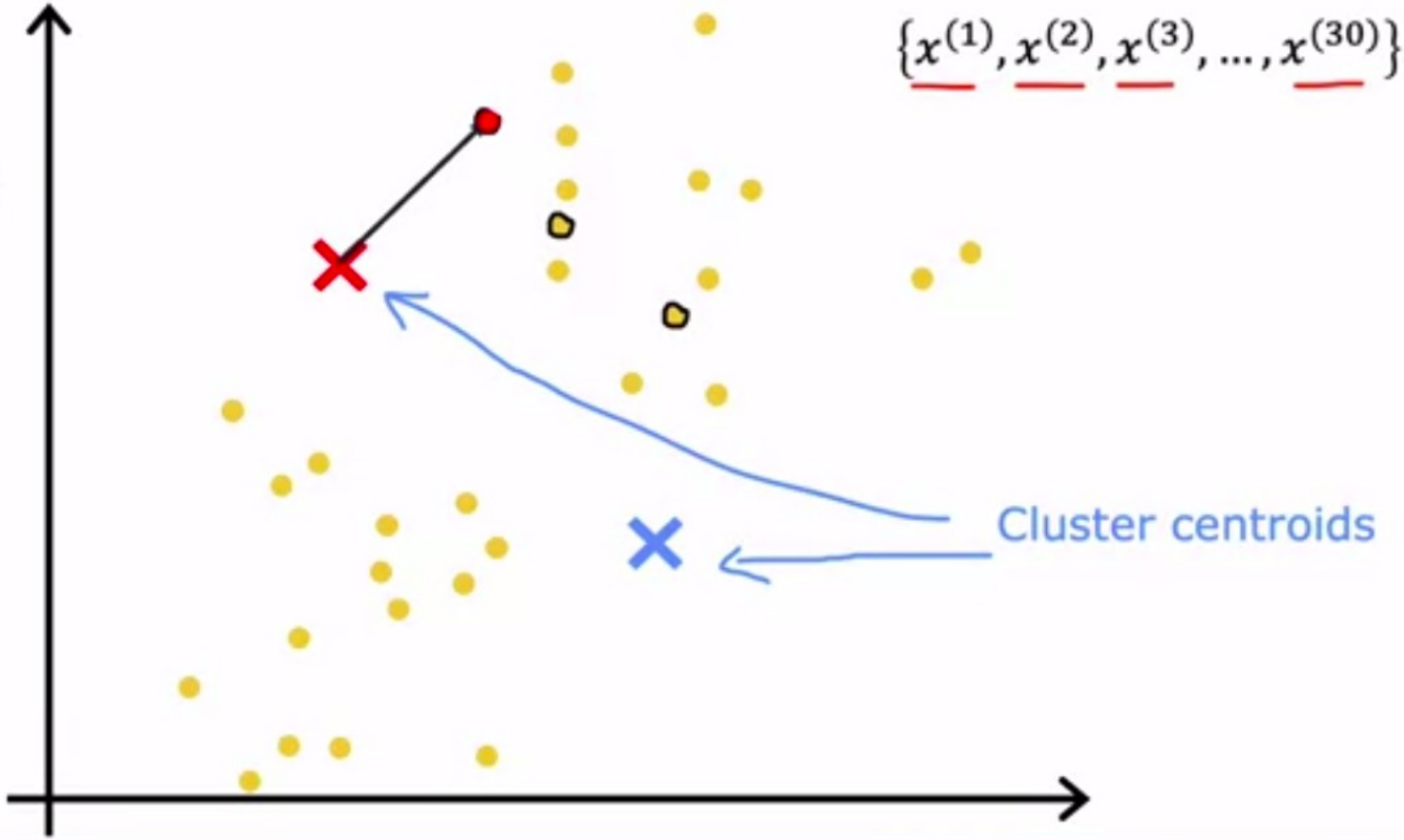
K-means intuition


$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$
Cluster centroids

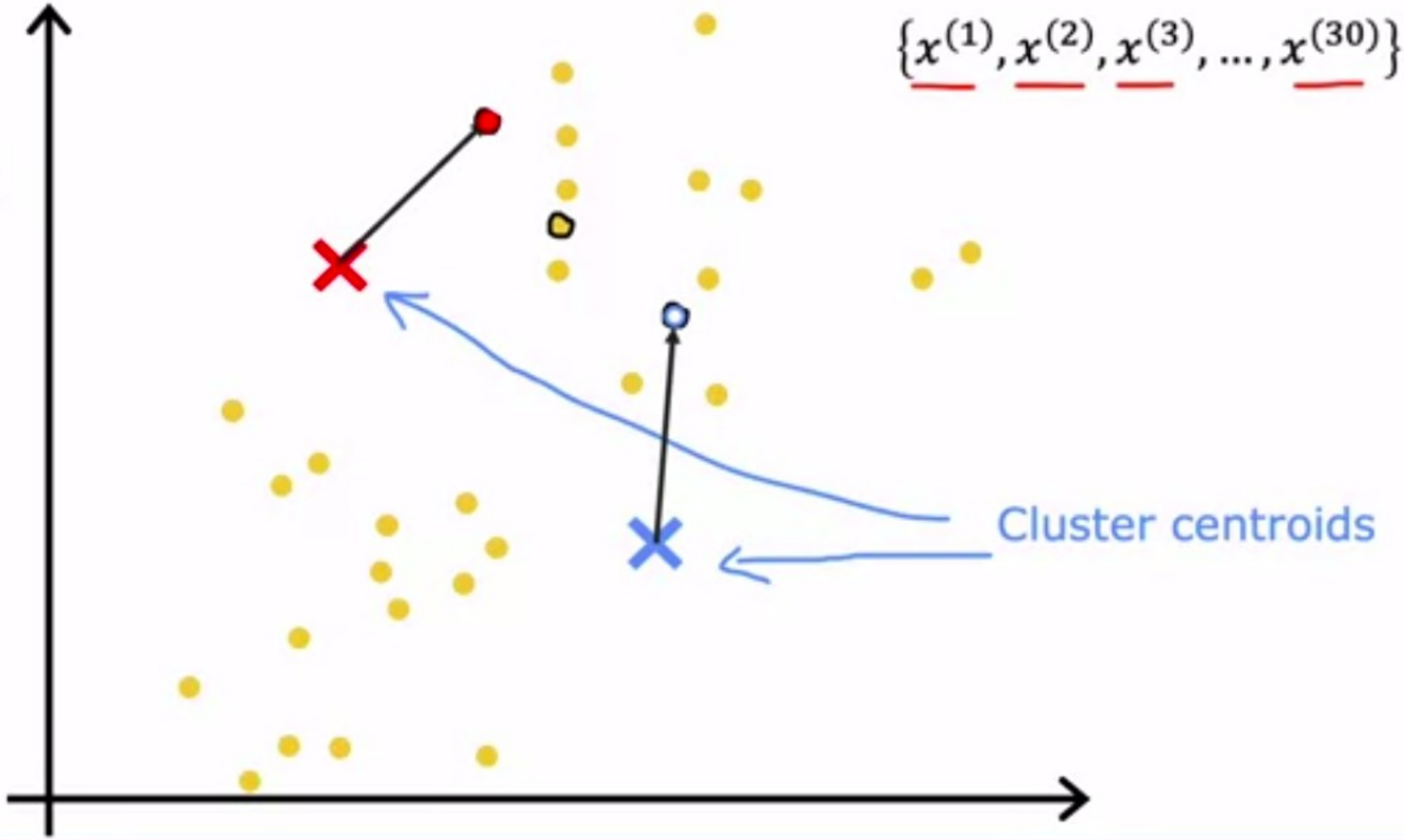
Step 1:
Assign each point to its closest centroid



Step 1:
Assign each point to its closest centroid

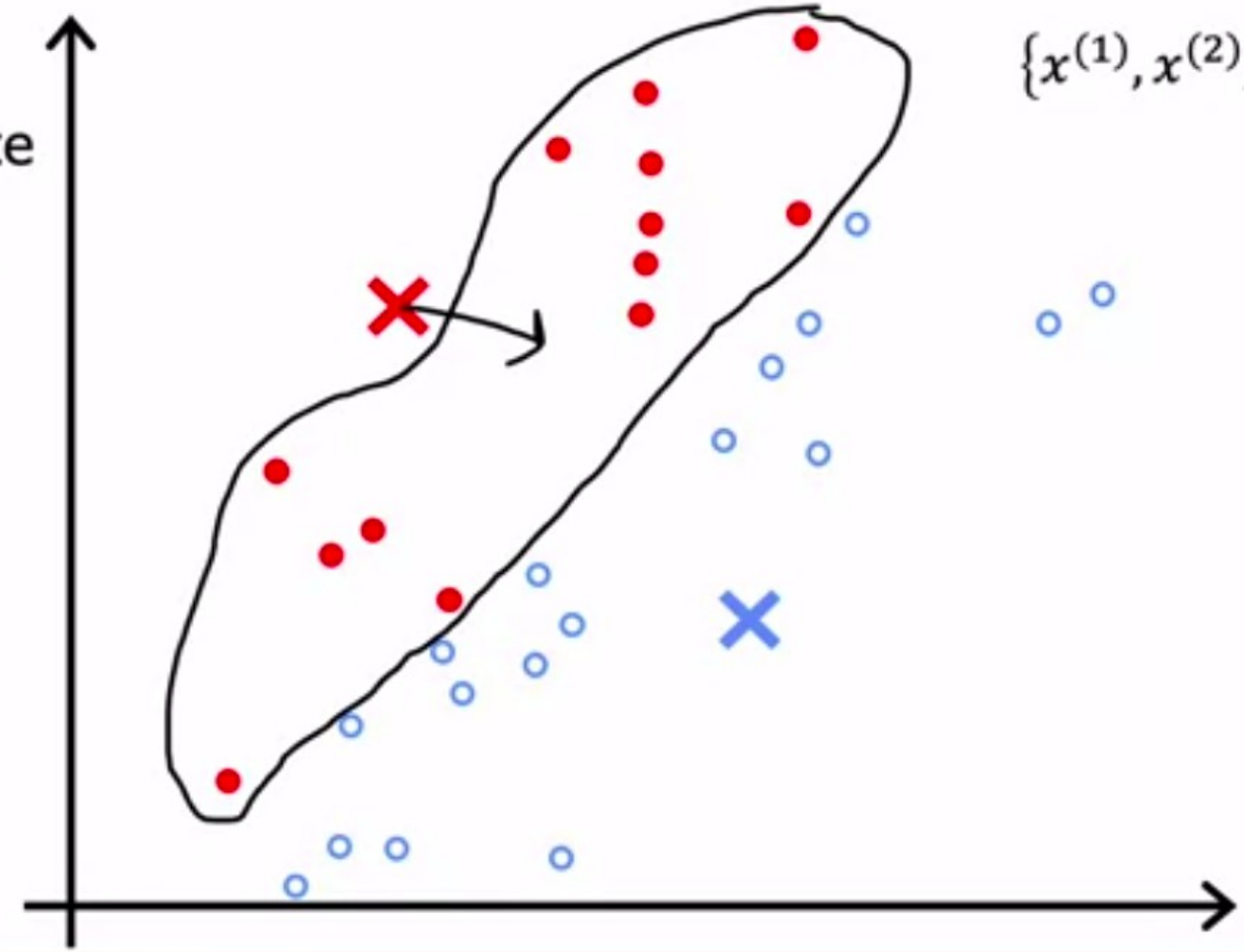


Step 1:
Assign each point to its closest centroid



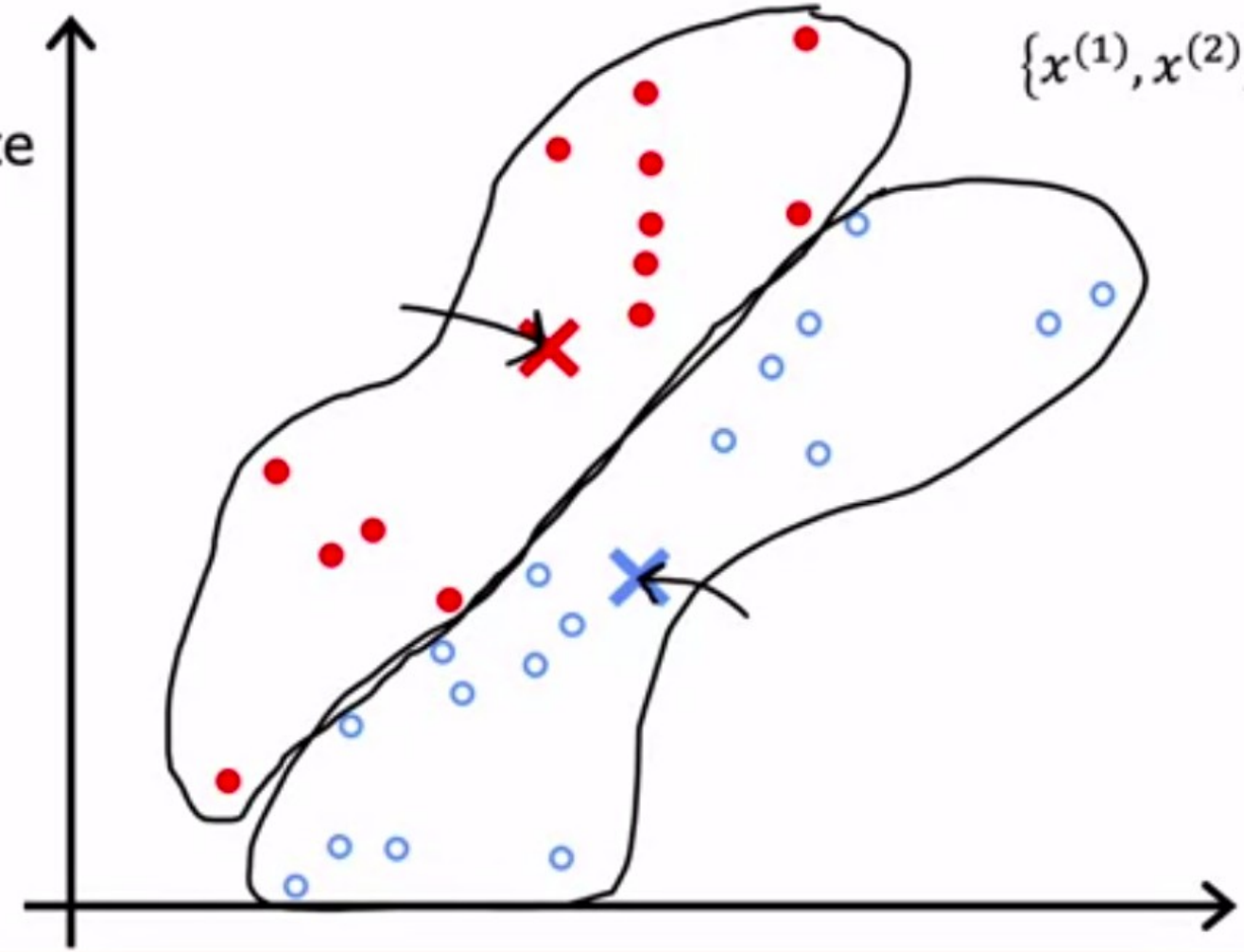
Step 2:
Recompute the centroids

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$

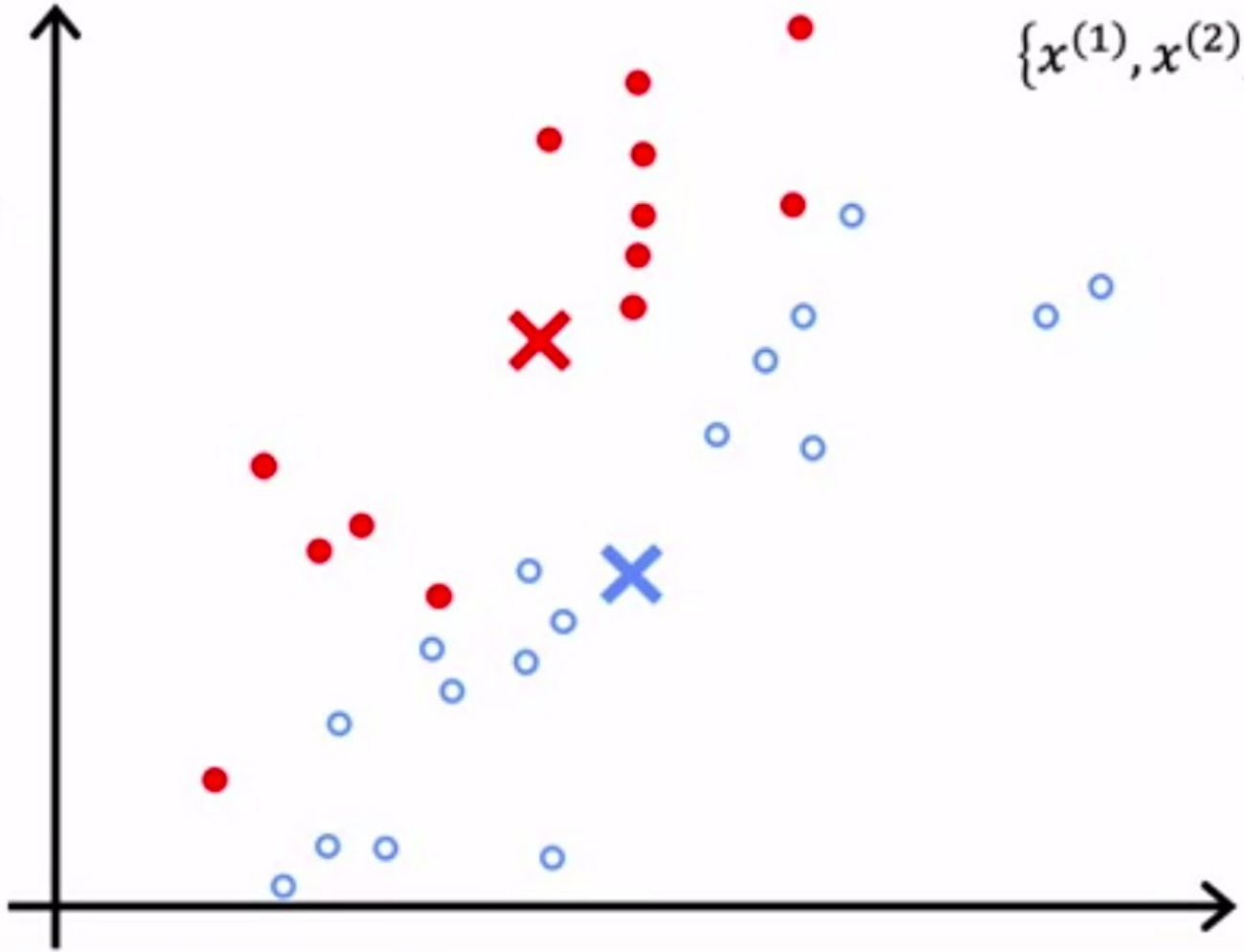


Step 2:
Recompute the centroids

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$



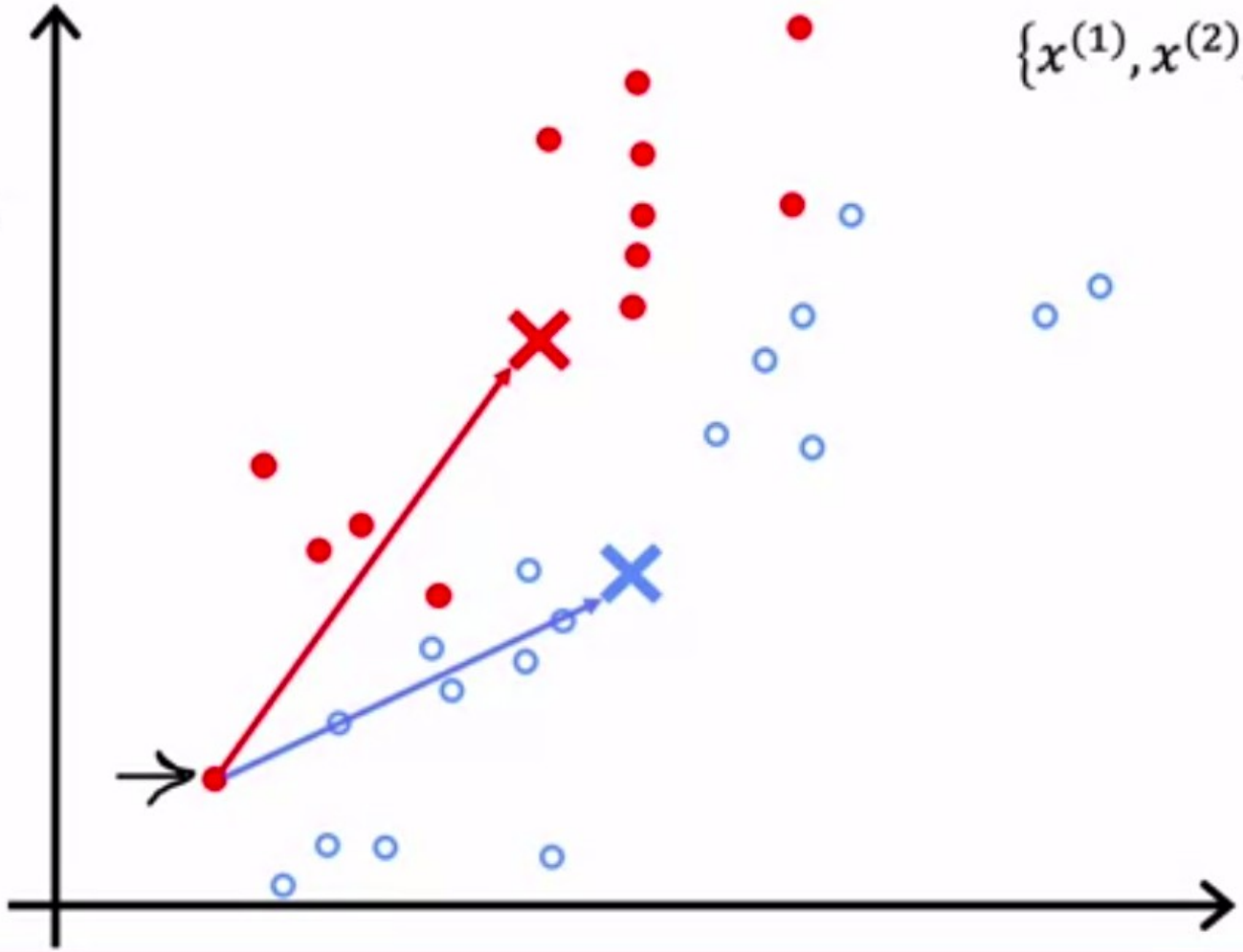
Step 1:
Assign each point to its closest centroid



$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$

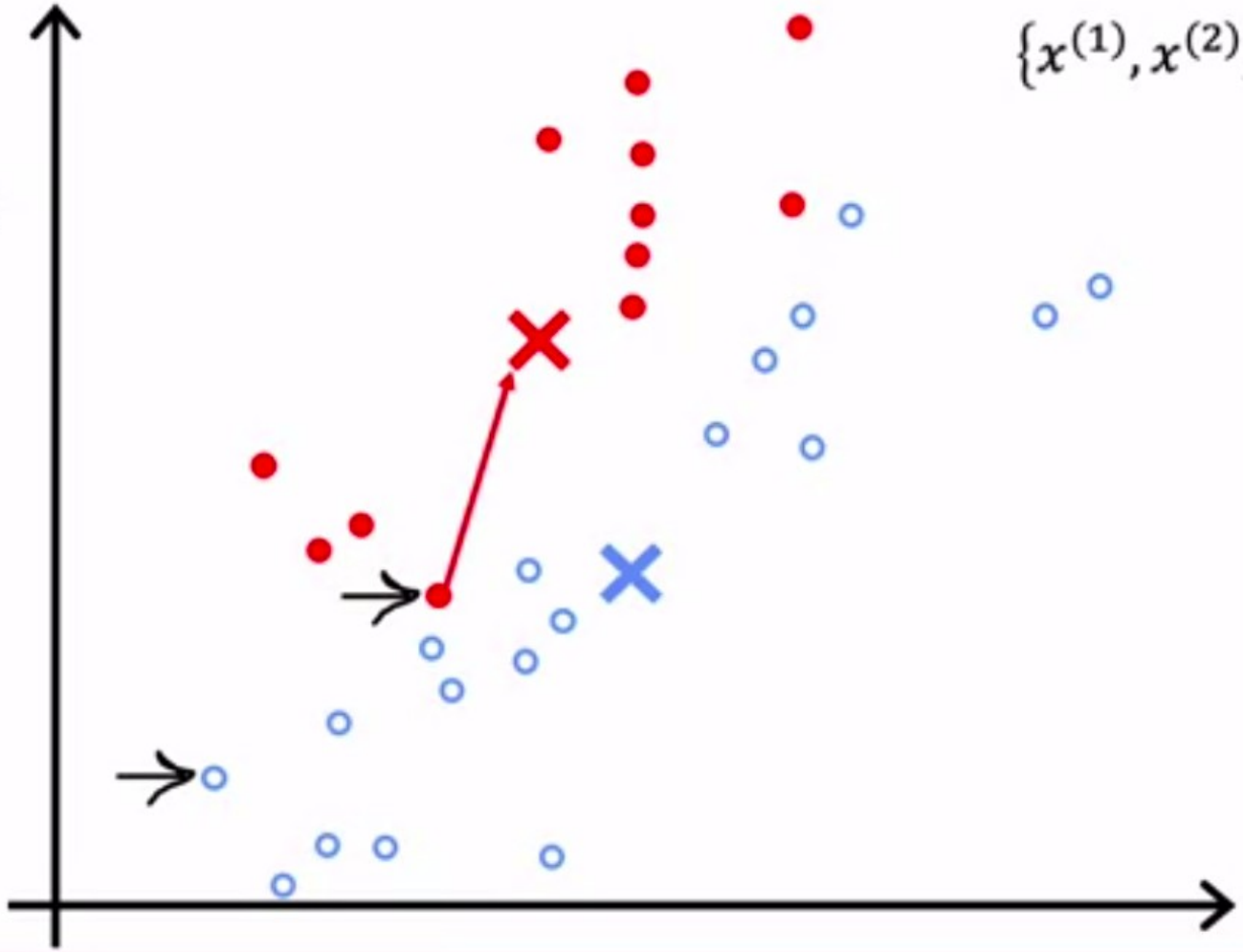
Step 1:
Assign each point to its closest centroid

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$



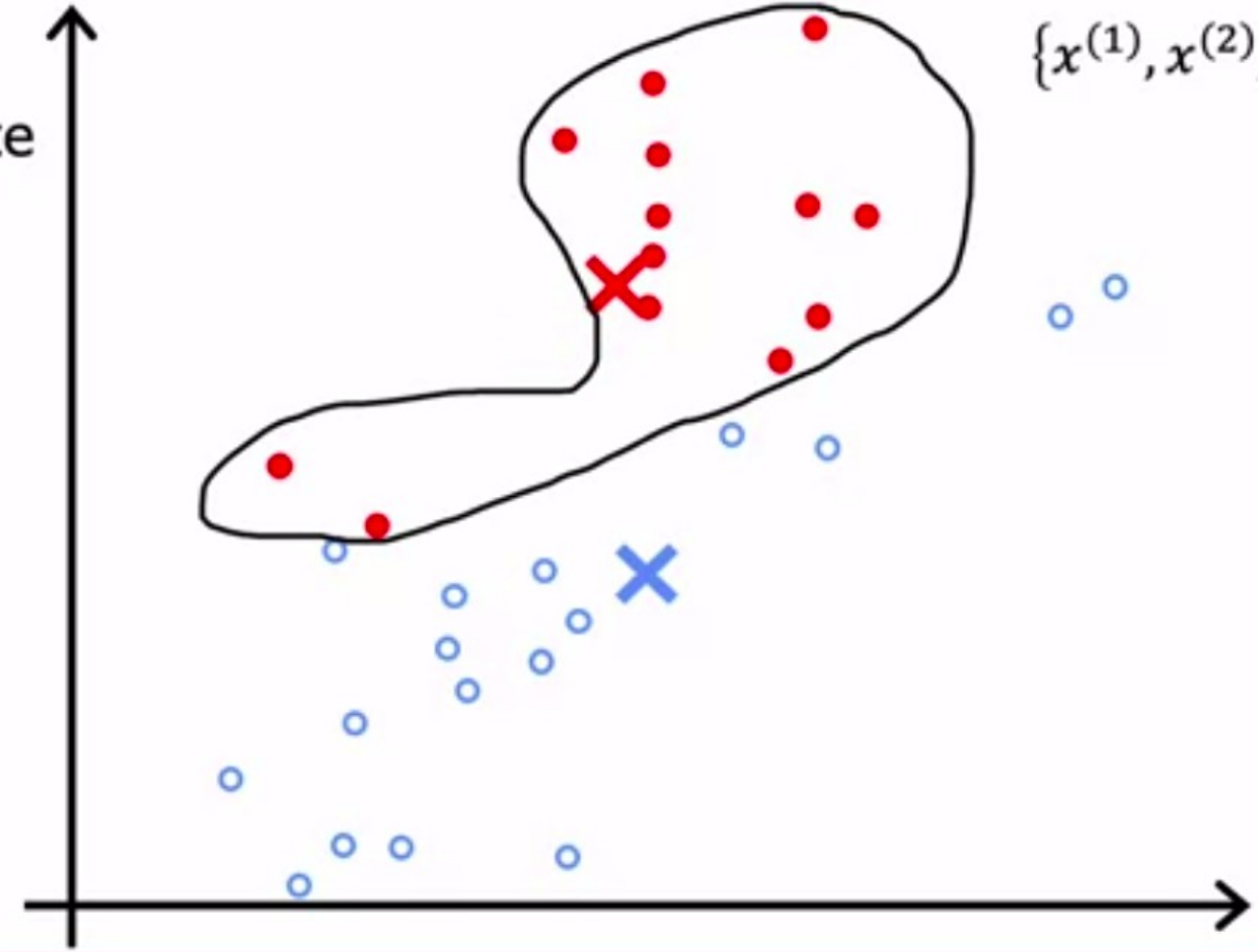
Step 1:
Assign each point to its closest centroid

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$



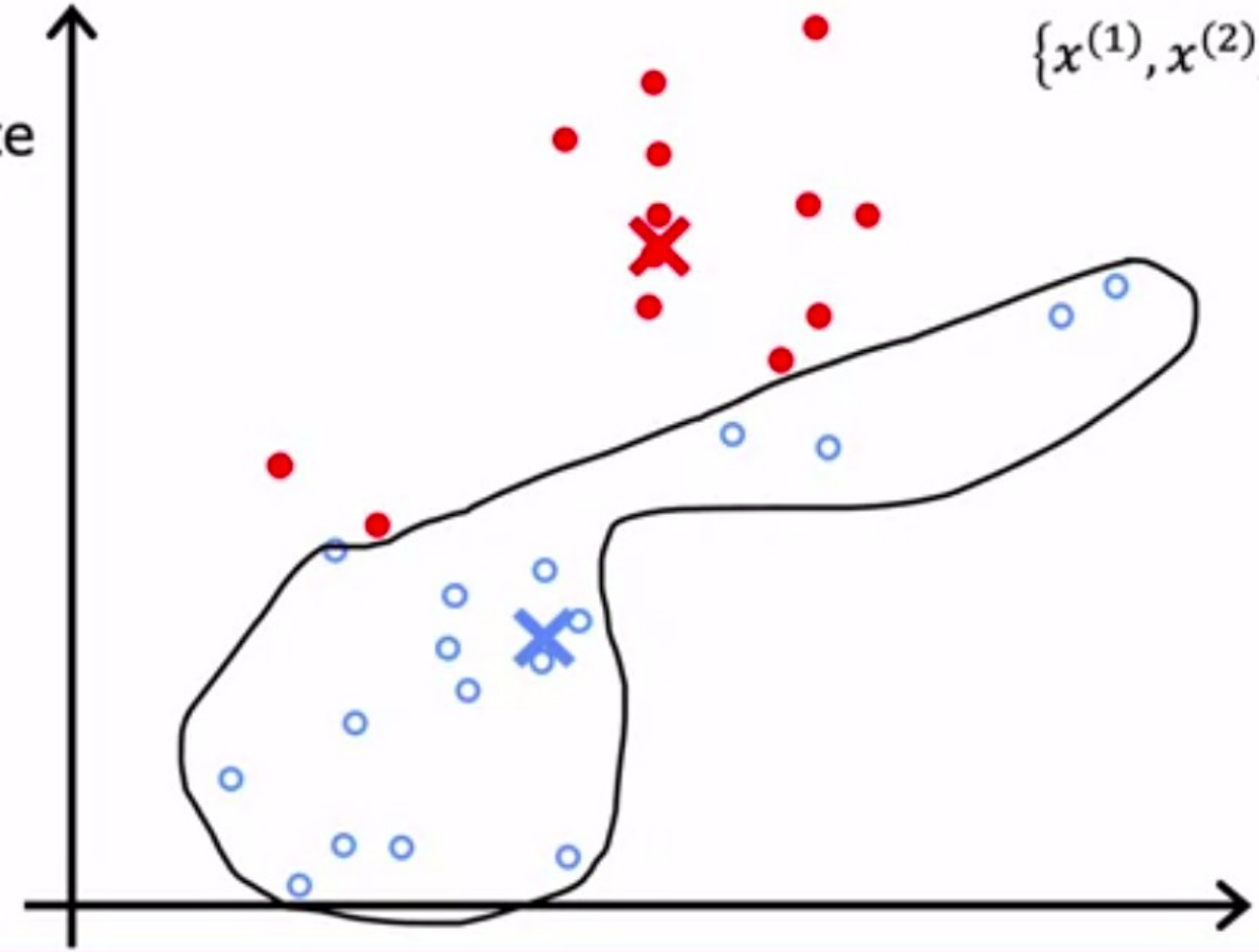
Step 2:
Recompute the centroids

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$



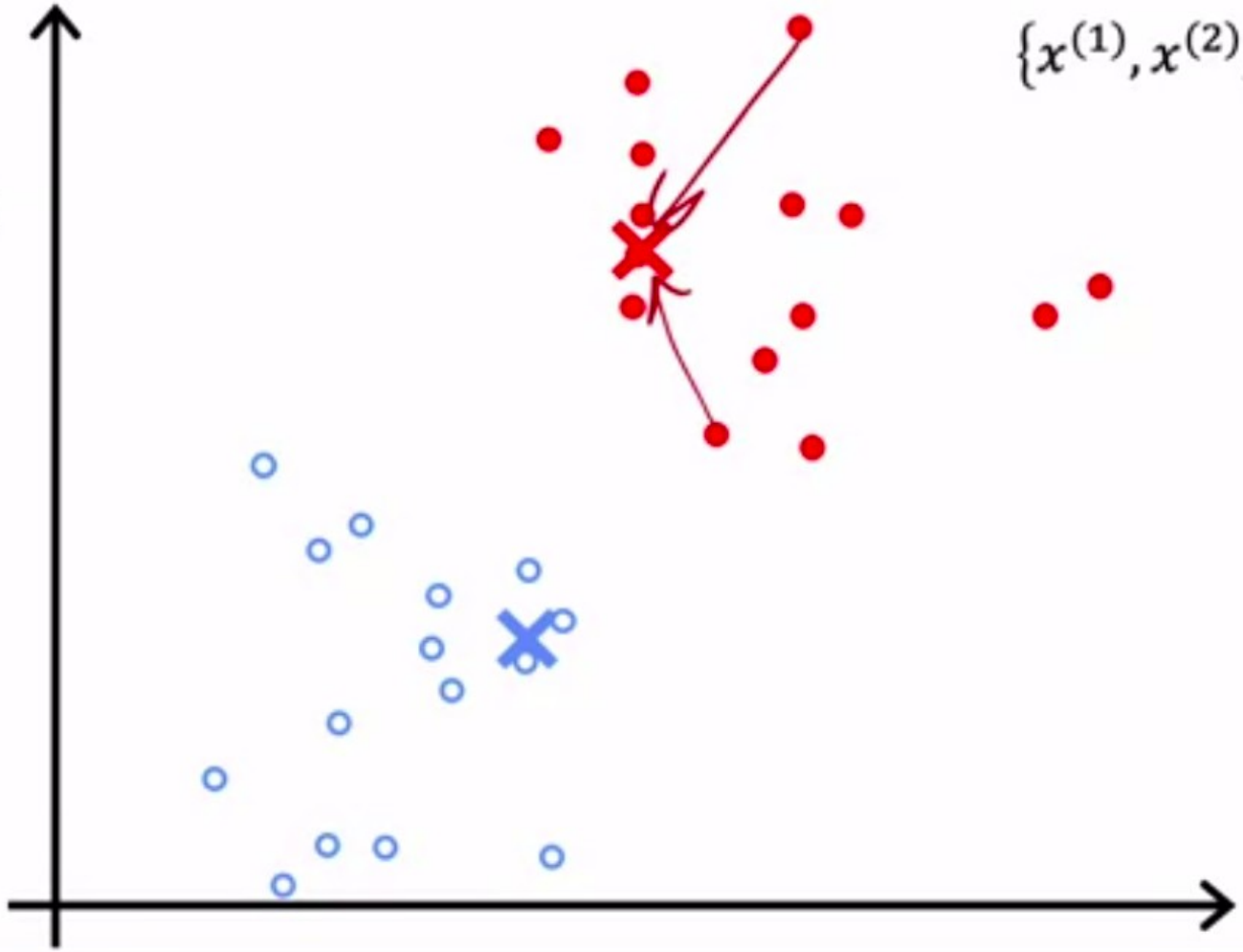
Step 2:
Recompute the centroids

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$



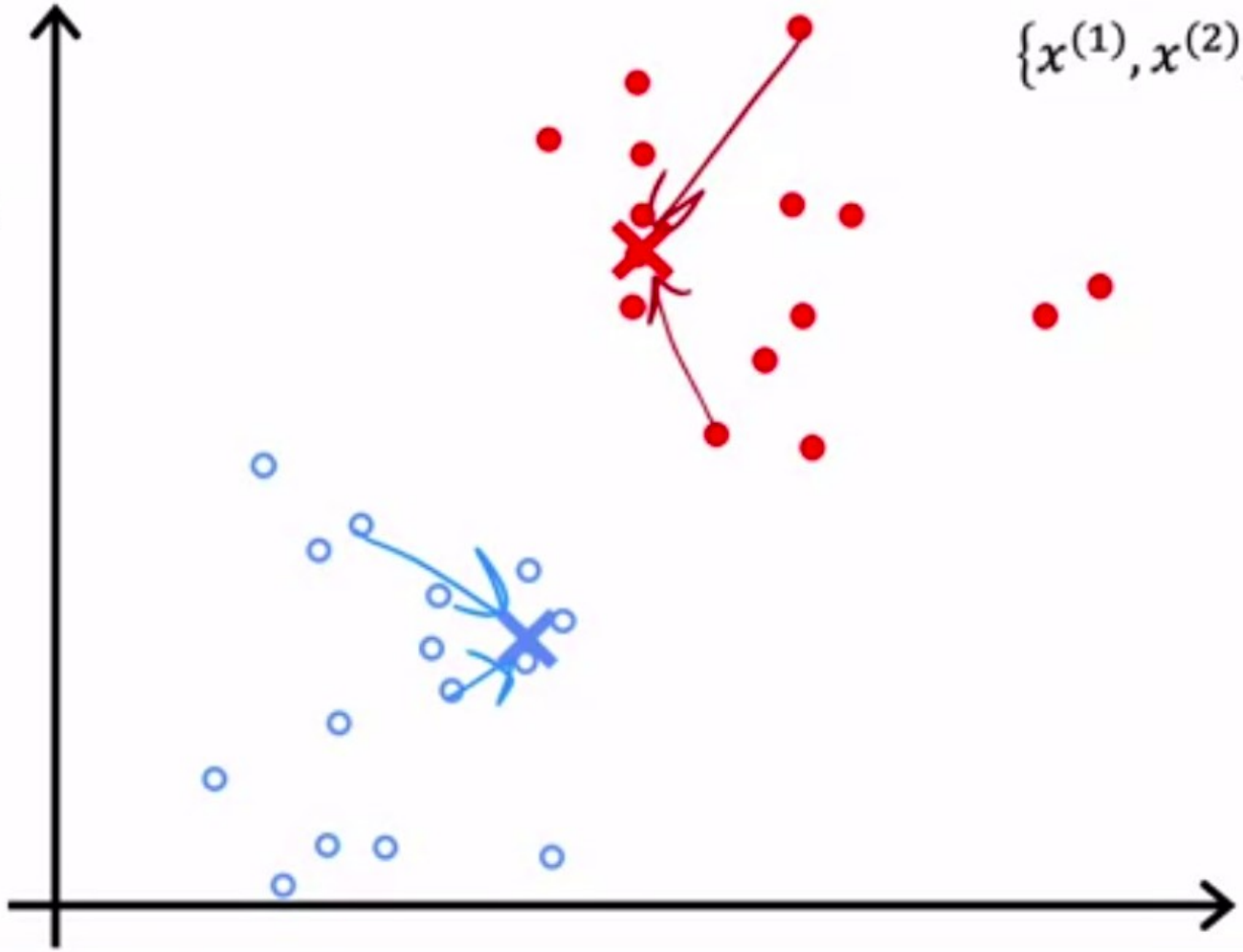
Step 1:
Assign each point to its closest centroid

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$



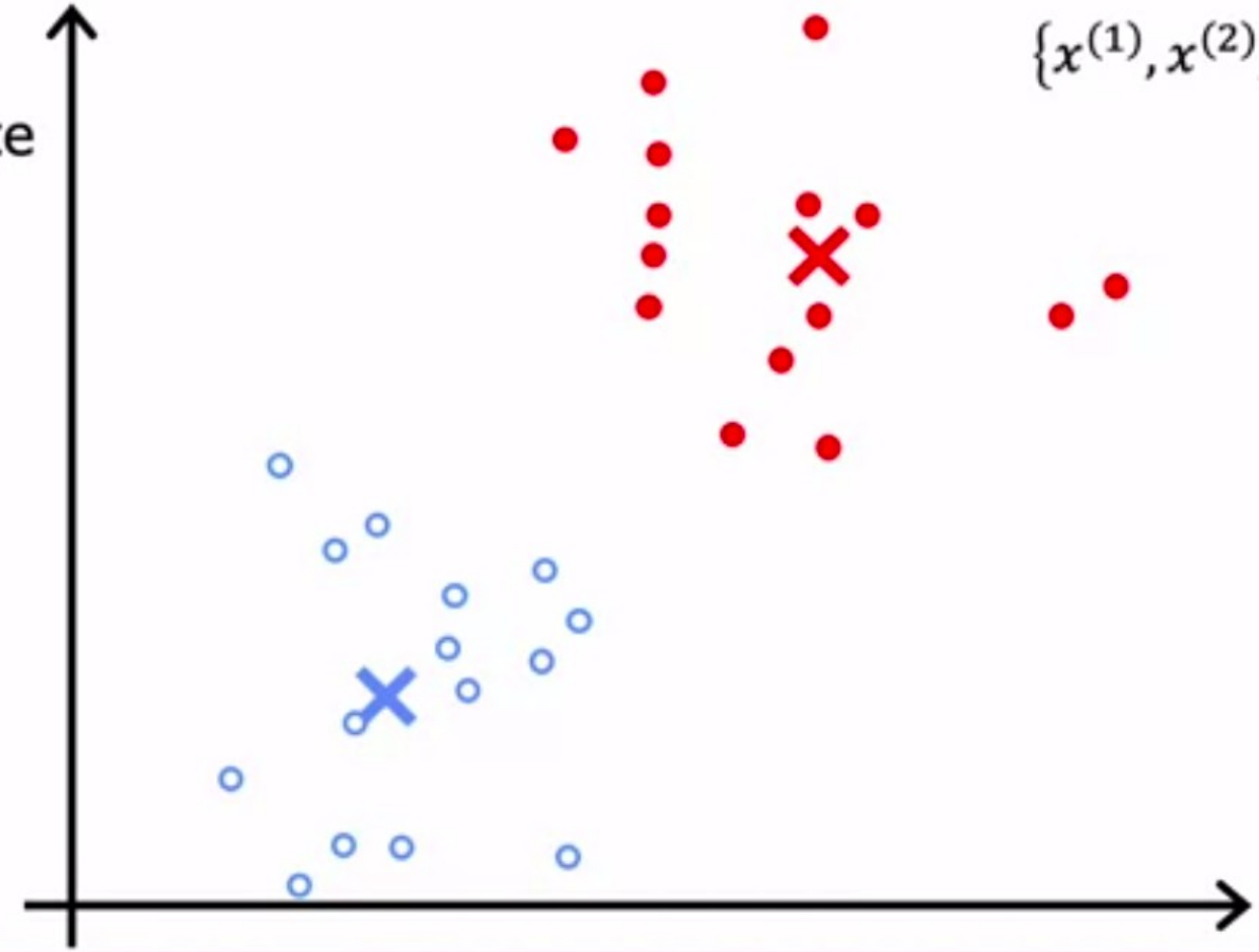
Step 1:
Assign each point to its closest centroid

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$

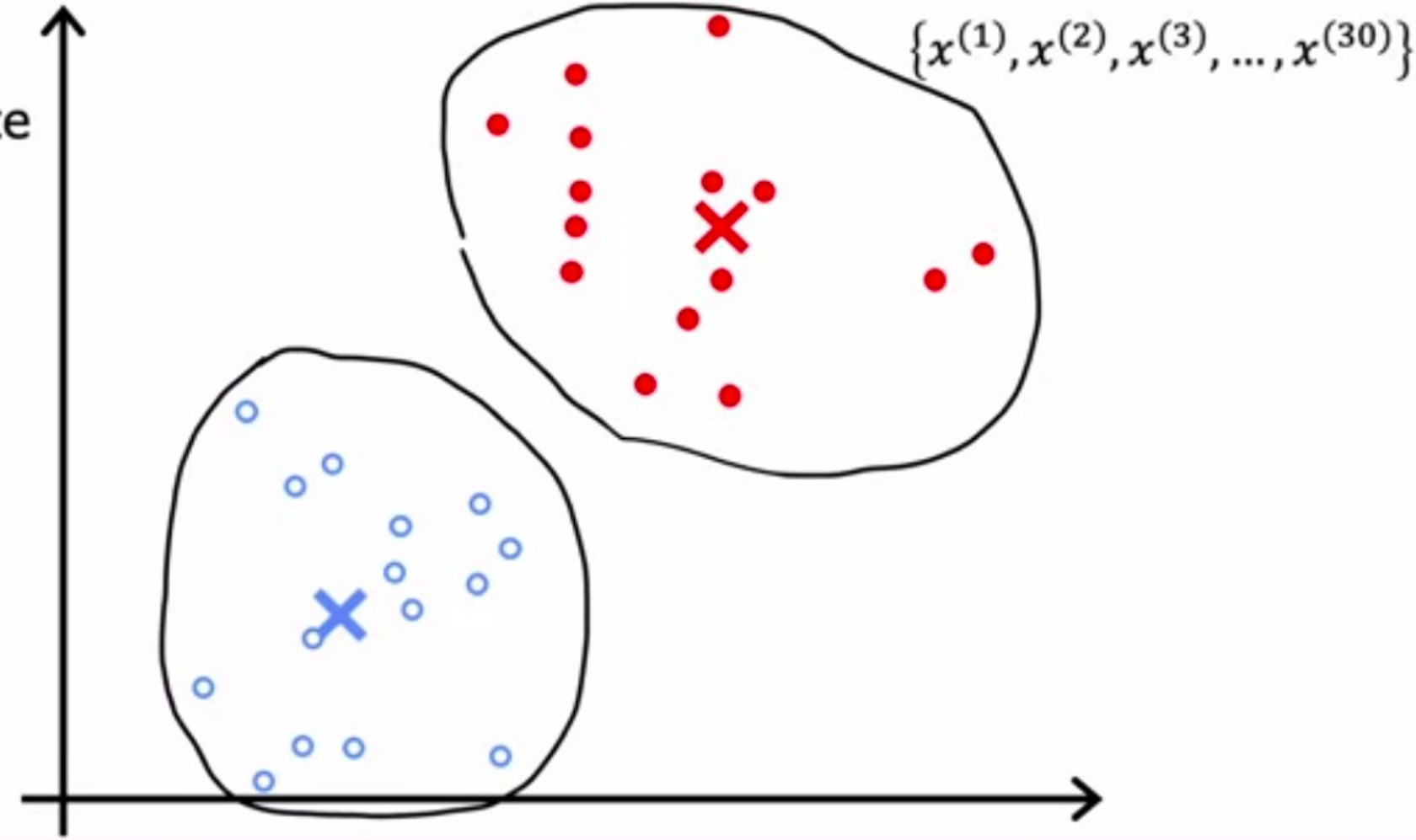


Step 2:
Recompute the centroids

$\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(30)}\}$



Step 2:
Recompute the centroids



DeepLearning.AI

Stanford
ONLINE

Clustering

K-means algorithm

K-means algorithm

μ_1, μ_2 $x^{(1)}, x^{(2)}, \ldots x^{(30)}$

$n = 2$ $K = 2$

Randomly initialize K cluster centroids $\mu_1, \mu_2, \ldots, \mu_K$

Repeat {

 # Assign points to cluster centroids

for i = 1 to m

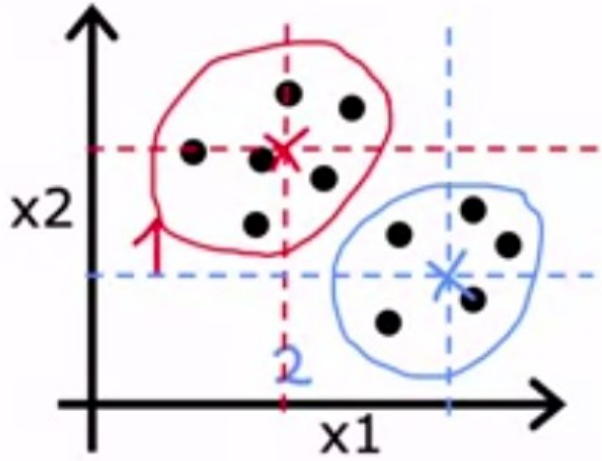
$c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$

 # Move cluster centroids

for k = 1 to K

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

}



$$\mu_1 = \frac{1}{4}\left[x^{(1)} + x^{(5)} + x^{(6)} + x^{(10)}\right]$$

K-means algorithm

μ_1, μ_2 $x^{(1)}, x^{(2)}, \ldots x^{(30)}$

$n = 2$ ~~$k = 2$~~

Randomly initialize K cluster centroids $\mu_1, \mu_2, \ldots, \mu_K$ $k = k - 1$ ✓

Repeat {

 # Assign points to cluster centroids

for i = 1 to m

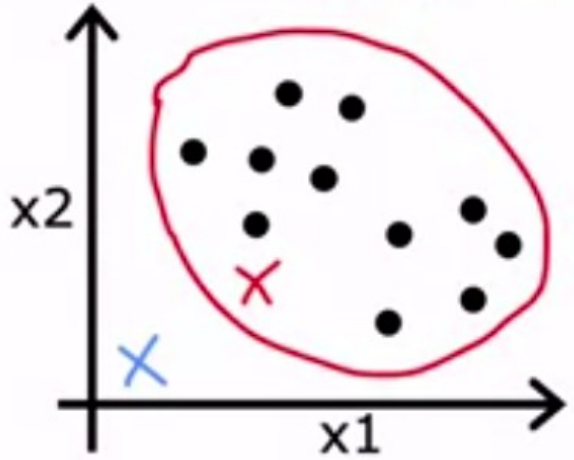
$c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$

 # Move cluster centroids

for k = 1 to K

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

}



K-means algorithm

μ_1, μ_2 $x^{(1)}, x^{(2)}, \ldots x^{(30)}$

$n = 2$ ~~$K = 2$~~

Randomly initialize K cluster centroids $\mu_1, \mu_2, \ldots, \mu_K$

$K = K - 1$

Repeat {

 # Assign points to cluster centroids

for i = 1 to m

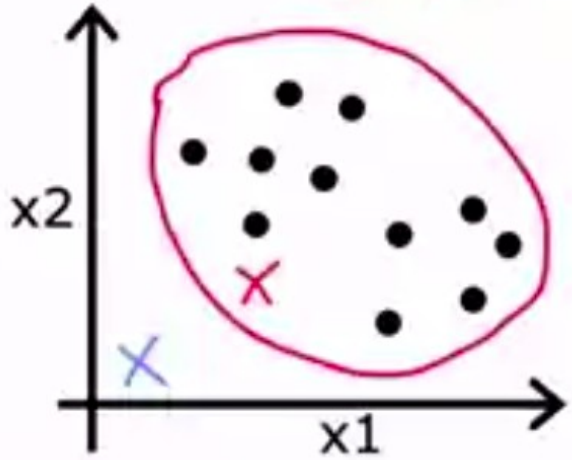
$c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$

 # Move cluster centroids

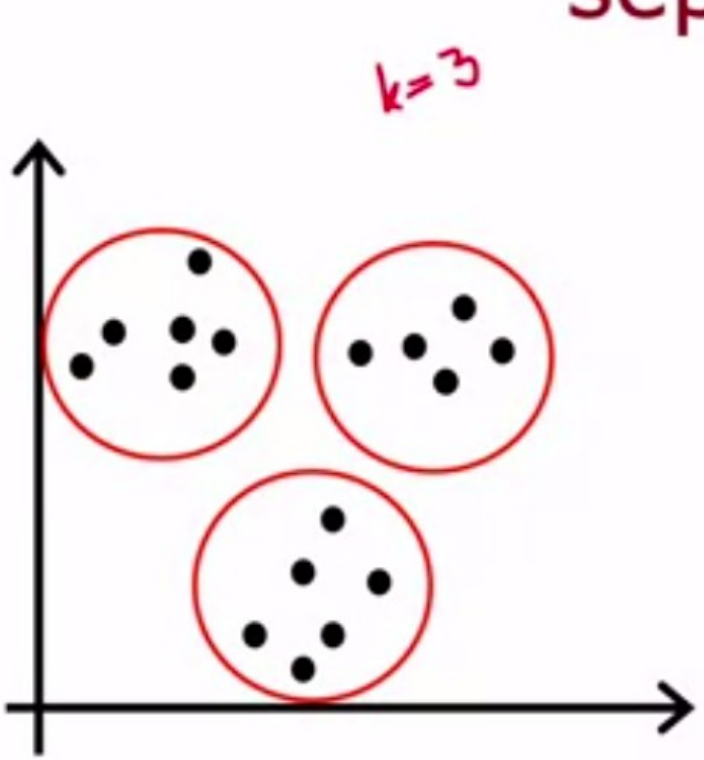
for k = 1 to K

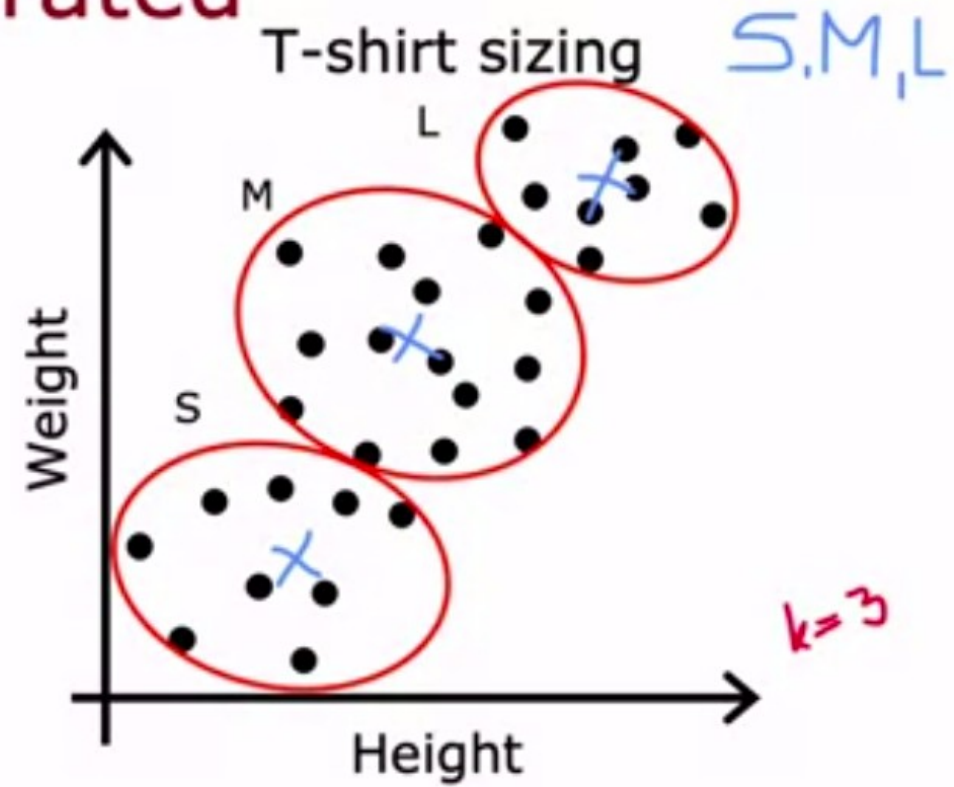
μ_k := average (mean) of points assigned to cluster k

}



K-means for clusters that are not well separated





DeepLearning.AI

Stanford
ONLINE

Clustering

Optimization objective

K-means optimization objective

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

μ_k = cluster centroid k

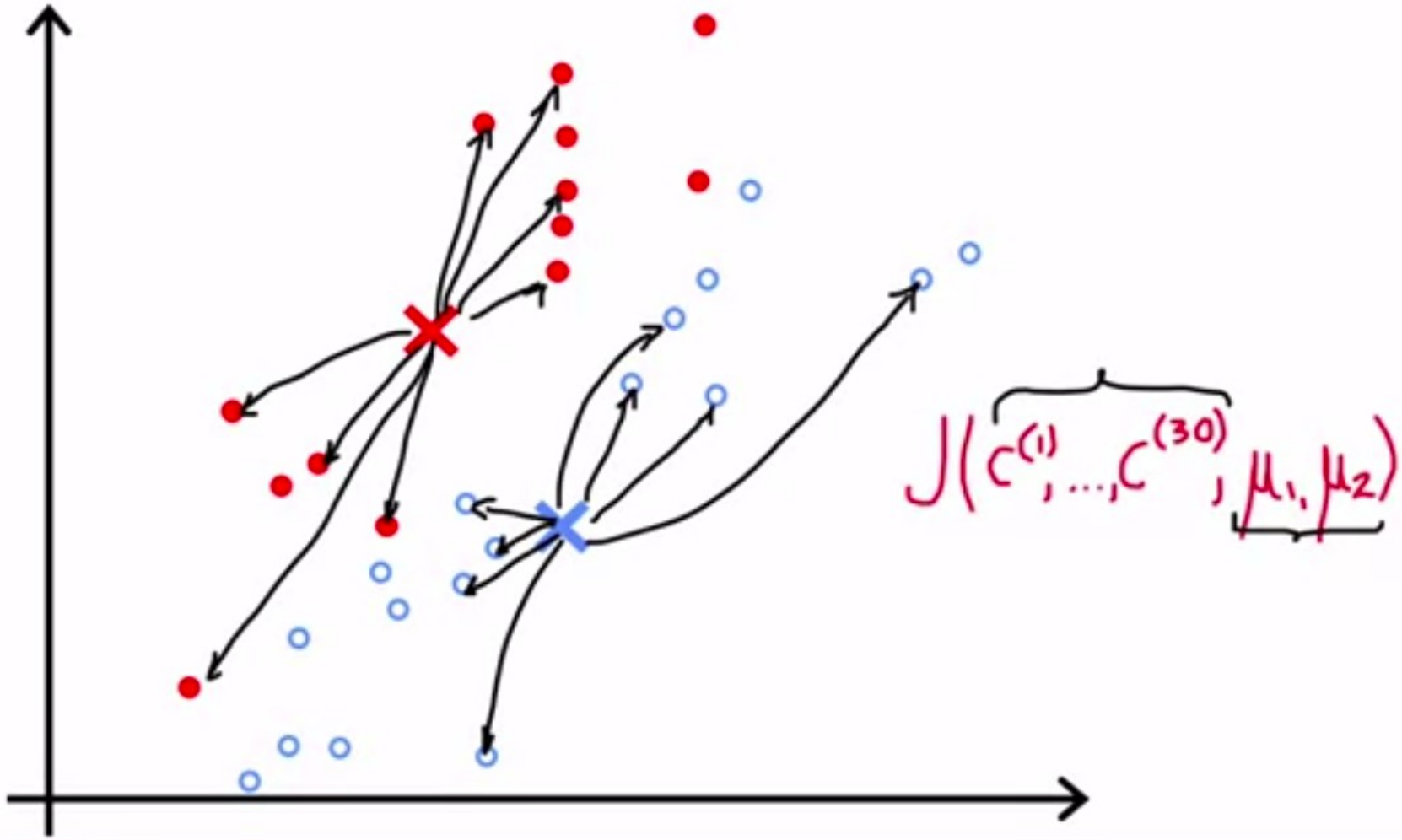
$\mu_{c^{(i)}}$ = cluster centroid of cluster to which example $x^{(i)}$ has been assigned

$x^{(10)}$ $c^{(10)}$ $\mu_c^{(10)}$

Cost function

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

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


$J(c^{(1)}, \ldots, c^{(30)}, \mu_1, \mu_2)$

K-means optimization objective

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

μ_k = cluster centroid k

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

$x^{(10)}$ $c^{(10)}$ $\mu_c^{(10)}$

Cost function

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

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

Distortion

Cost function for K-means

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

Repeat {

 # Assign points to cluster centroids

for i = 1 to m

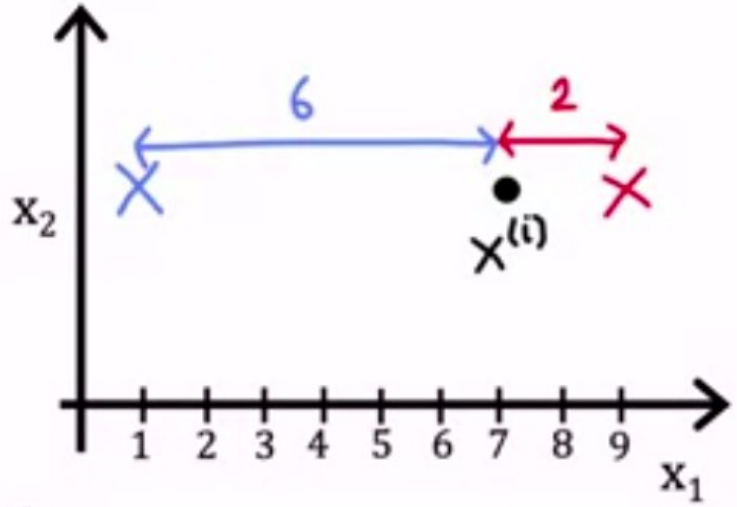
$c^{(i)}$:= index of cluster centroid closest to $x^{(i)}$

 # Move cluster centroids

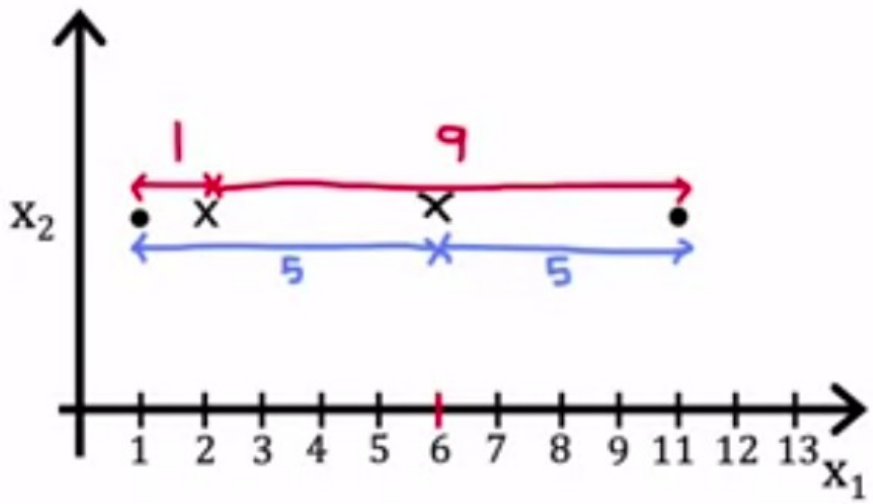
for k = 1 to K

μ_k := average of points in cluster k

}



Moving the centroid



$\frac{1}{2}(1^2 + 9^2) = \frac{1}{2}(1+81) = 41$

$\frac{1}{2}(1 + 11) = 6$

$\frac{1}{2}(5^2 + 5^2) = 25$

DeepLearning.AI

Stanford
ONLINE

Clustering

Initializing K-means

K-means algorithm

Step 0: Randomly initialize K cluster centroids $\mu_1, \mu_1, \ldots, \mu_k$

Repeat {

Step 1: Assign points to cluster centroids

Step 2: Move cluster centroids

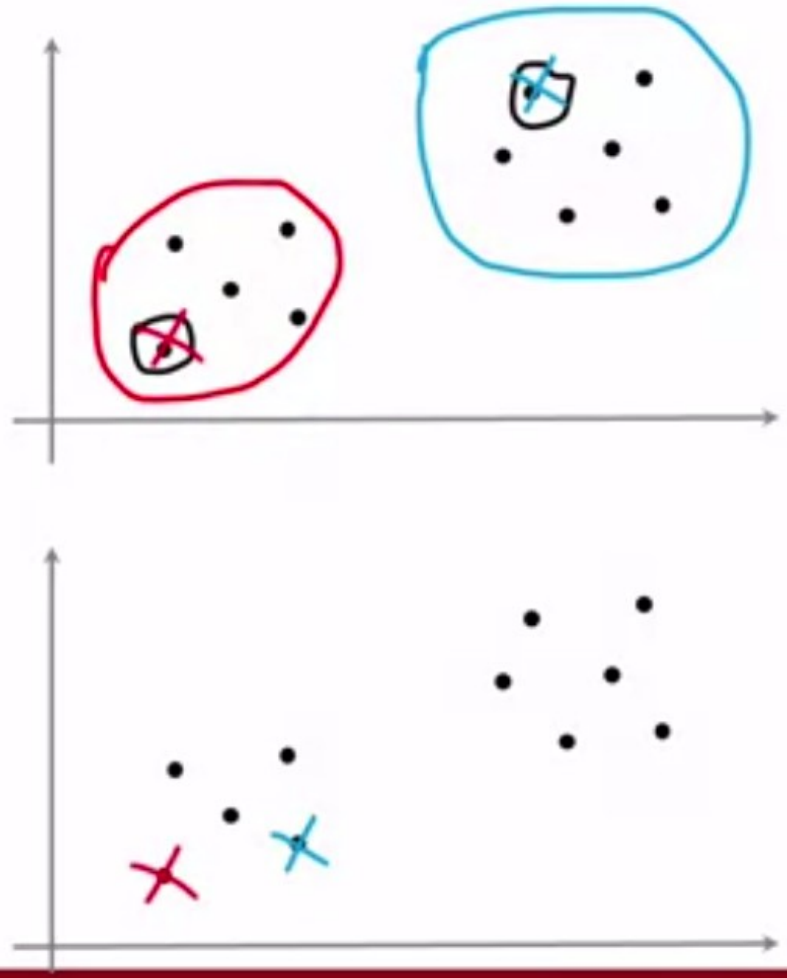
}

Random initialization

Choose $K < m$

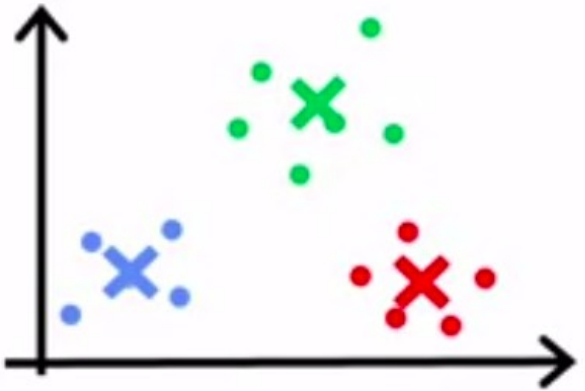
Randomly pick K training examples.

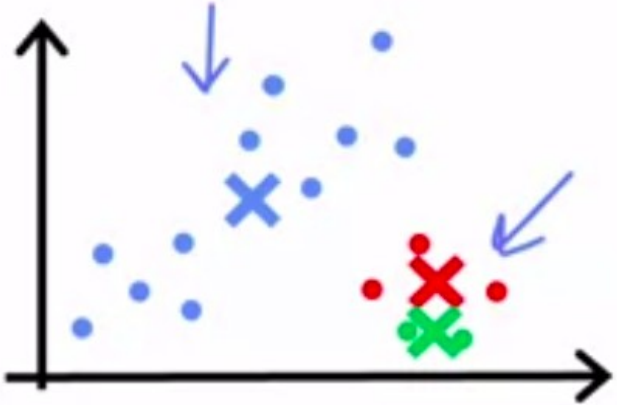
Set $\mu_1, \mu_1, \ldots, \mu_k$ equal to these K examples.

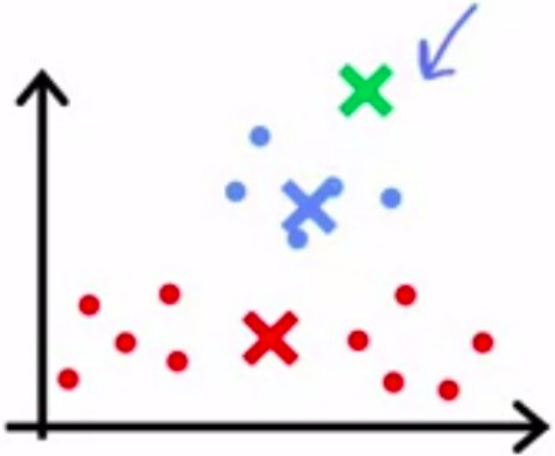


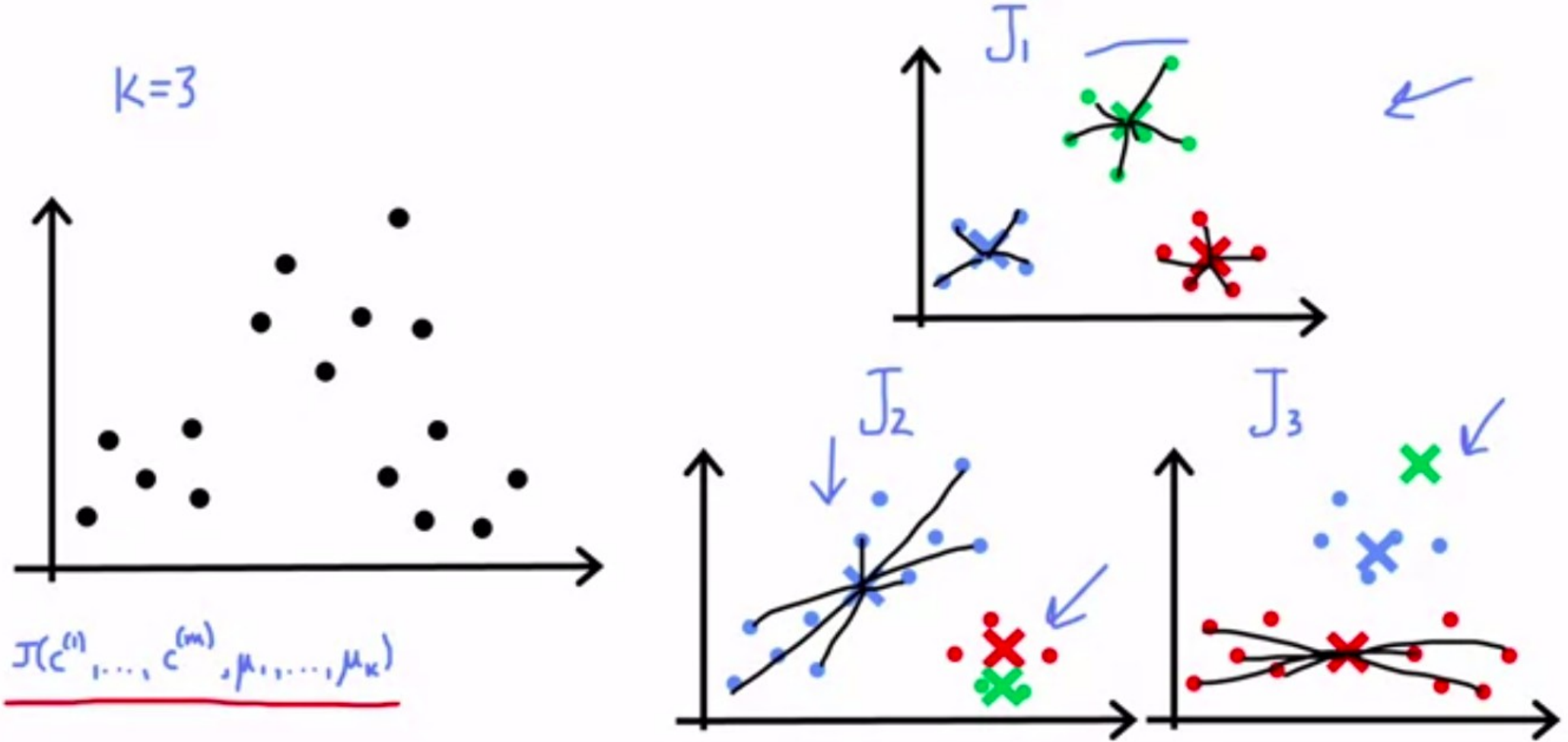
$K=3$

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

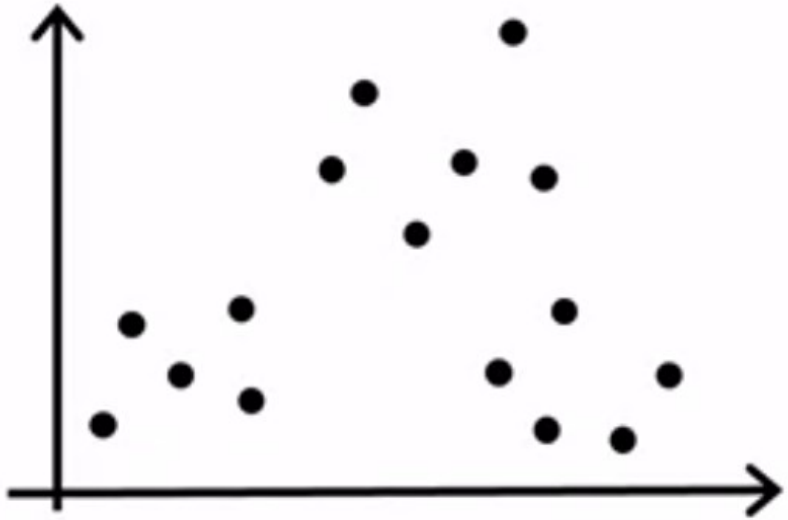




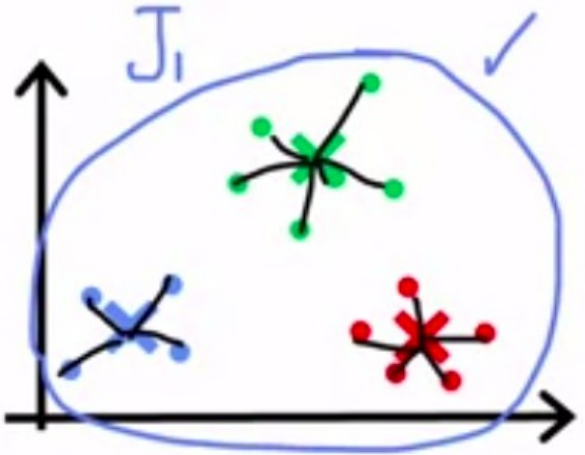


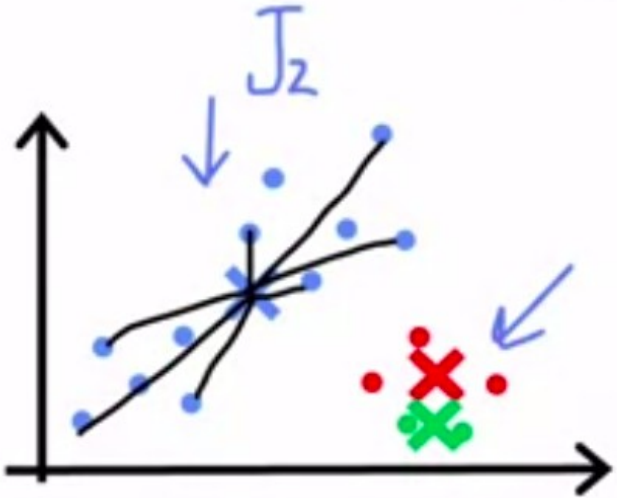

K=3
J1
J2
J3
J(c(1), ..., c(m), μ1, ..., μK)

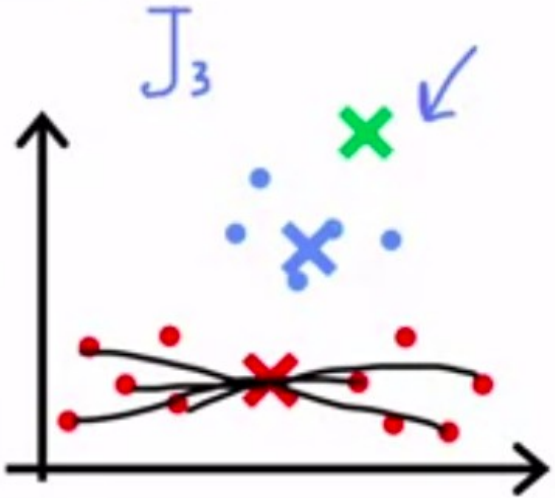
$K=3$



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







Random initialization

50-1000

For $i = 1$ to 100 {

k random examples

Randomly initialize K-means.

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

Computer cost function (distortion)

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

}

Pick set of clusters that gave lowest cost J

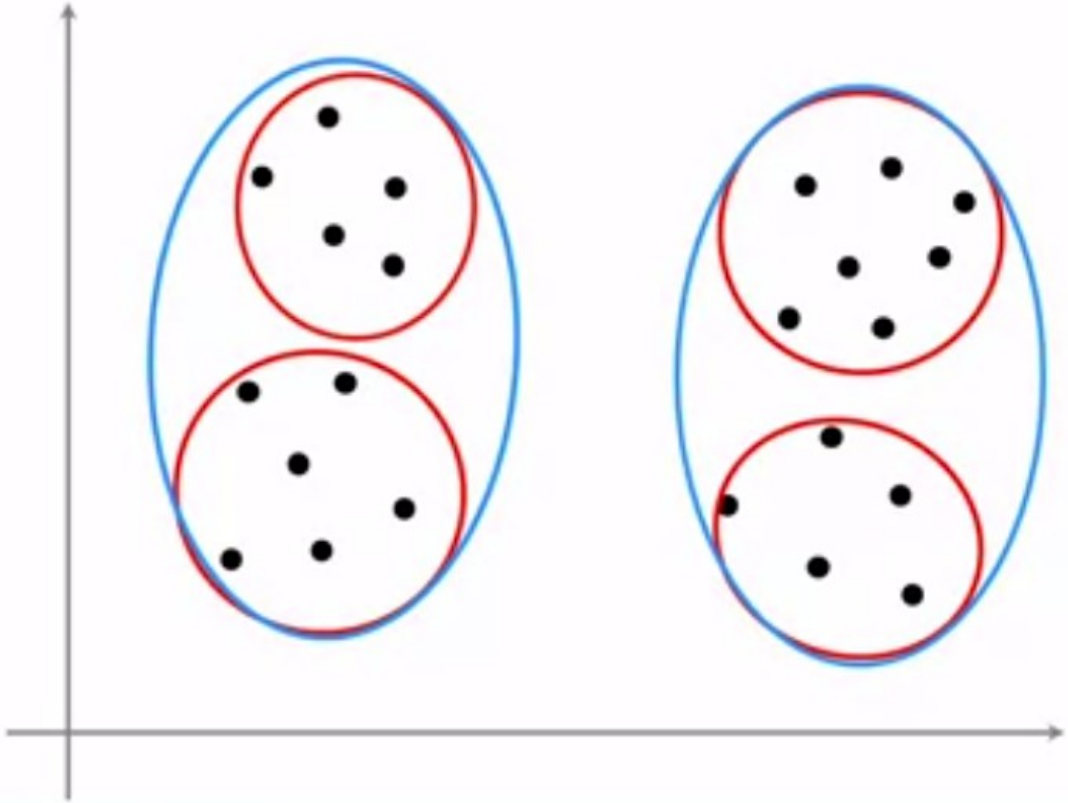
DeepLearning.AI

Stanford
ONLINE

Clustering

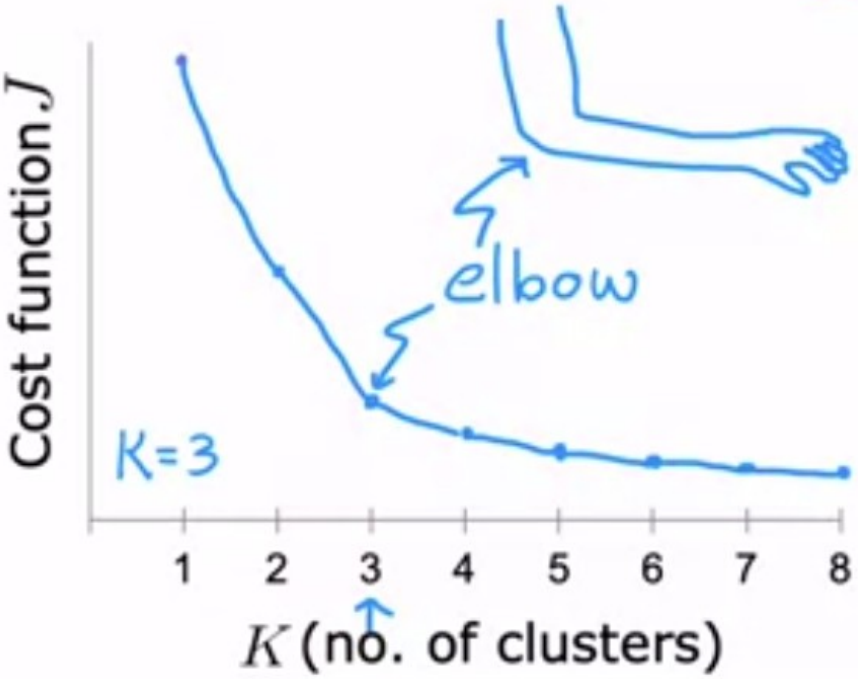
Choosing the Number of Clusters

What is the right value of K?



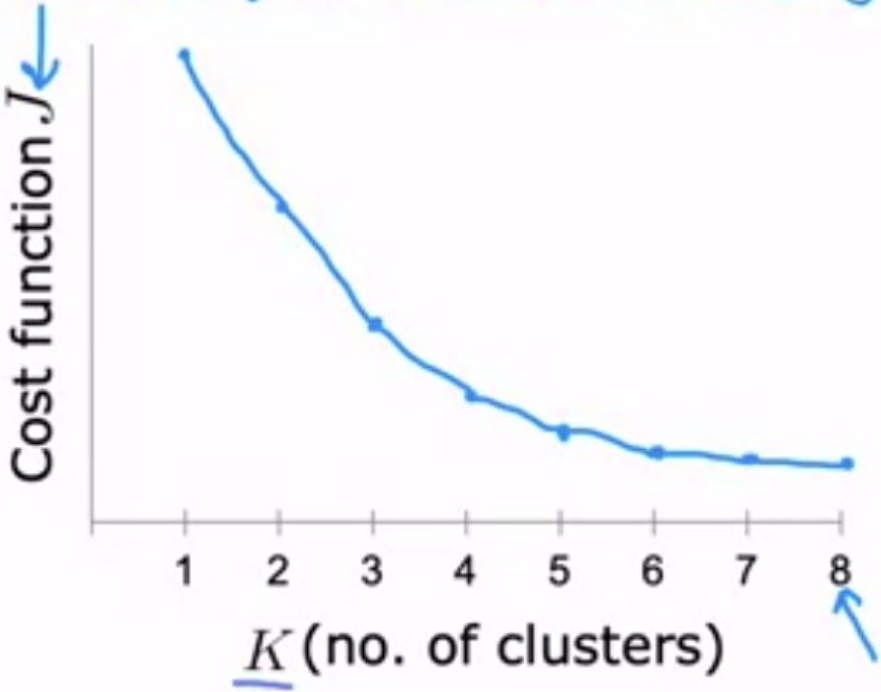
Choosing the value of K

Elbow method



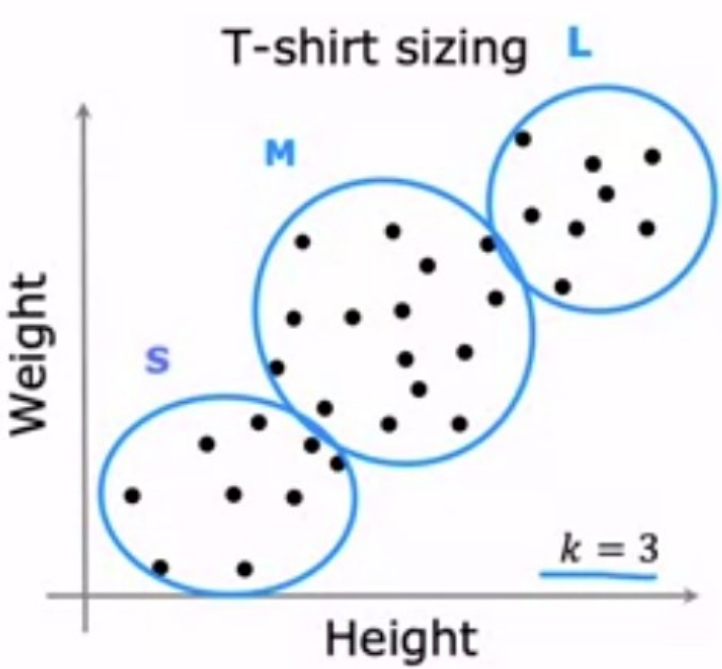
the right "K" is often ambiguous

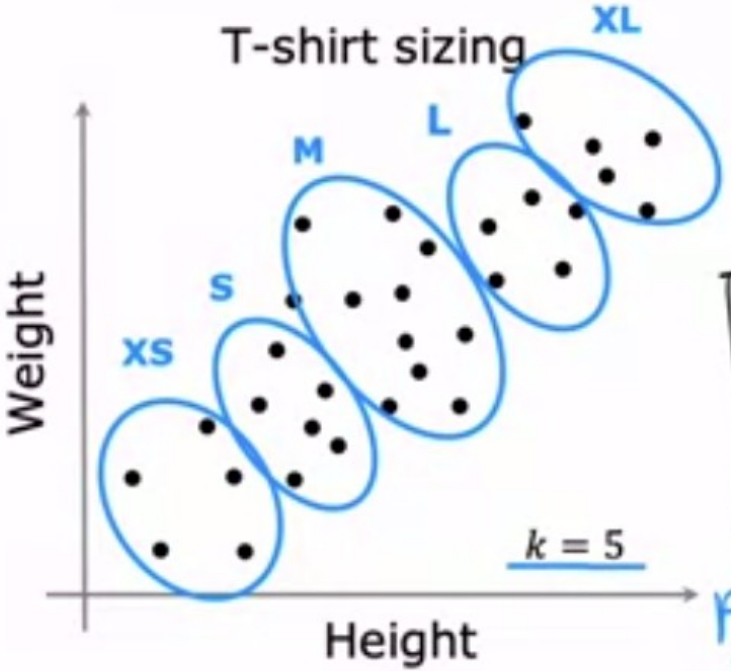
Don't choose K just to minimize cost J

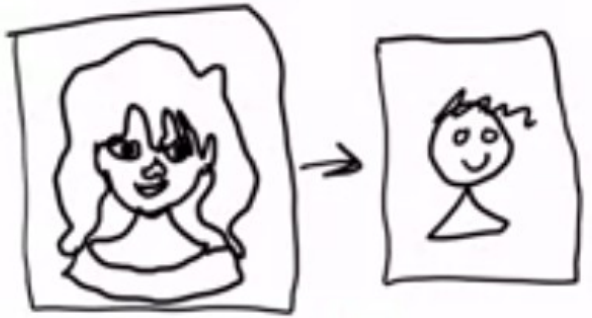


Choosing the value of K

Often, you want to get clusters for some later (downstream) purpose.
Evaluate K-means based on how well it performs on that later purpose.







practice lab:
image compression!

Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to next item

1. Which of these best describes unsupervised learning? **1 / 1 point**

- A form of machine learning that finds patterns using labeled data (x, y)
- A form of machine learning that finds patterns without using a cost function.
- (selected) A form of machine learning that finds patterns using unlabeled data (x).
- A form of machine learning that finds patterns in data using only labels (y) but without any inputs (x) .

Correct
Unsupervised learning uses unlabeled data. The training examples do not have targets or labels "y". Recall the T-shirt example. The data was height and weight but no target size.

2. **1 / 1 point**

2. **1 / 1 point**

Which of these statements are true about K-means? Check all that apply.

- [x] If each example x is a vector of 5 numbers, then each cluster centroid μ_k is also going to be a vector of 5 numbers.

Correct
The dimension of μ_k matches the dimension of the examples.

- [x] The number of cluster assignment variables $c^{(i)}$ is equal to the number of training examples.

Correct
$c^{(i)}$ describes which centroid example(i) is assigned to.

- [ ] The number of cluster centroids μ_k is equal to the number of examples.
- [x] If you are running K-means with $K = 3$ clusters, then each $c^{(i)}$ should be 1, 2, or 3.

Correct
$c^{(i)}$ describes which centroid example(i) is assigned to. If $K = 3$, then $c^{(i)}$ would be one of 1,2 or 3 assuming counting starts at 1.

Correct
$c^{(i)}$ describes which centroid example(i) is assigned to. If $K = 3$, then $c^{(i)}$ would be one of 1,2 or 3 assuming counting starts at 1.

3. **1 / 1 point**

You run K-means 100 times with different initializations. How should you pick from the 100 resulting solutions?

- ○ Pick the last one (i.e., the 100th random initialization) because K-means always improves over time
- ○ Pick randomly -- that was the point of random initialization.
- ◉ Pick the one with the lowest cost J
- ○ Average all 100 solutions together.

Correct
K-means can arrive at different solutions depending on initialization. After running repeated trials, choose the solution with the lowest cost.

4. You run K-means and compute the value of the cost function $J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K)$ after each iteration. Which of these statements should be true? **1 / 1 point**

- ○ Because K-means tries to maximize cost, the cost is always greater than or equal to the cost in the previous iteration.

Correct
K-means can arrive at different solutions depending on initialization. After running repeated trials, choose the solution with the lowest cost.

4. You run K-means and compute the value of the cost function $J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K)$ after each iteration. Which of these statements should be true? **1 / 1 point**

- ○ Because K-means tries to maximize cost, the cost is always greater than or equal to the cost in the previous iteration.
- ○ The cost can be greater or smaller than the cost in the previous iteration, but it decreases in the long run.
- ○ There is no cost function for the K-means algorithm.
- ◉ The cost will either decrease or stay the same after each iteration. .

Correct
The cost never increases. K-means always converges.

5. In K-means, the elbow method is a method to **1 / 1 point**

- ○ Choose the best random initialization
- ○ Choose the best number of samples in the dataset

- There is no cost function for the K-means algorithm.
- (•) The cost will either decrease or stay the same after each iteration. .

Correct
The cost never increases. K-means always converges.

5. In K-means, the elbow method is a method to **1 / 1 point**

- Choose the best random initialization
- Choose the best number of samples in the dataset
- (•) Choose the number of clusters K
- Choose the maximum number of examples for each cluster

Correct
The elbow method plots a graph between the number of clusters K and the cost function. The 'bend' in the cost curve can suggest a natural value for K. Note that this feature may not exist or be significant in some data sets.

DeepLearning.AI

Stanford
ONLINE

Anomaly Detection

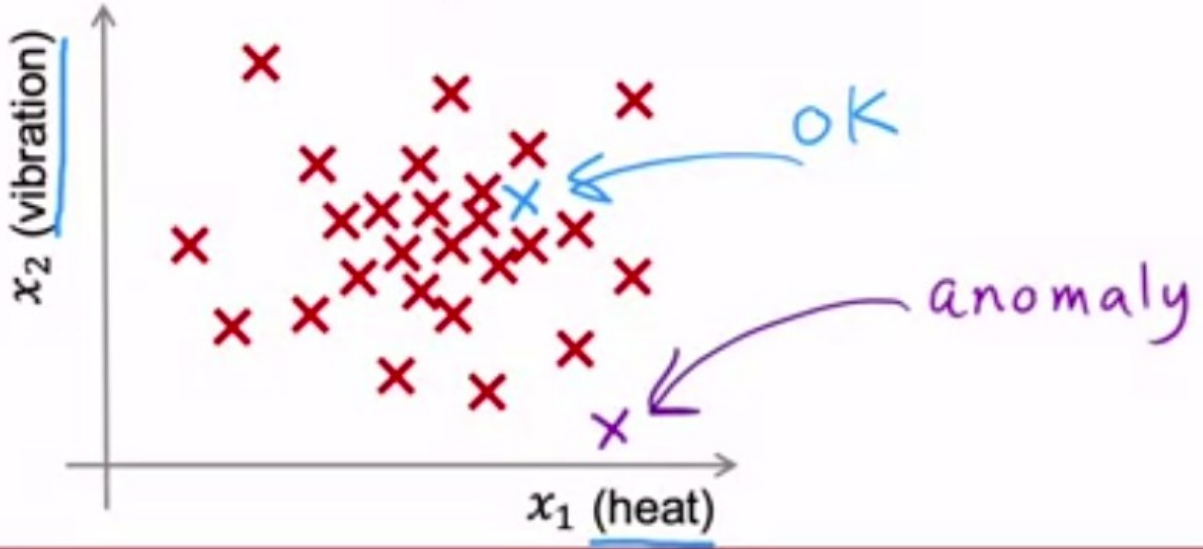
Finding unusual events

Anomaly detection example

Aircraft engine features:
x_1 = heat generated
x_2 = vibration intensity
...

Dataset: $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

New engine: x_{test}



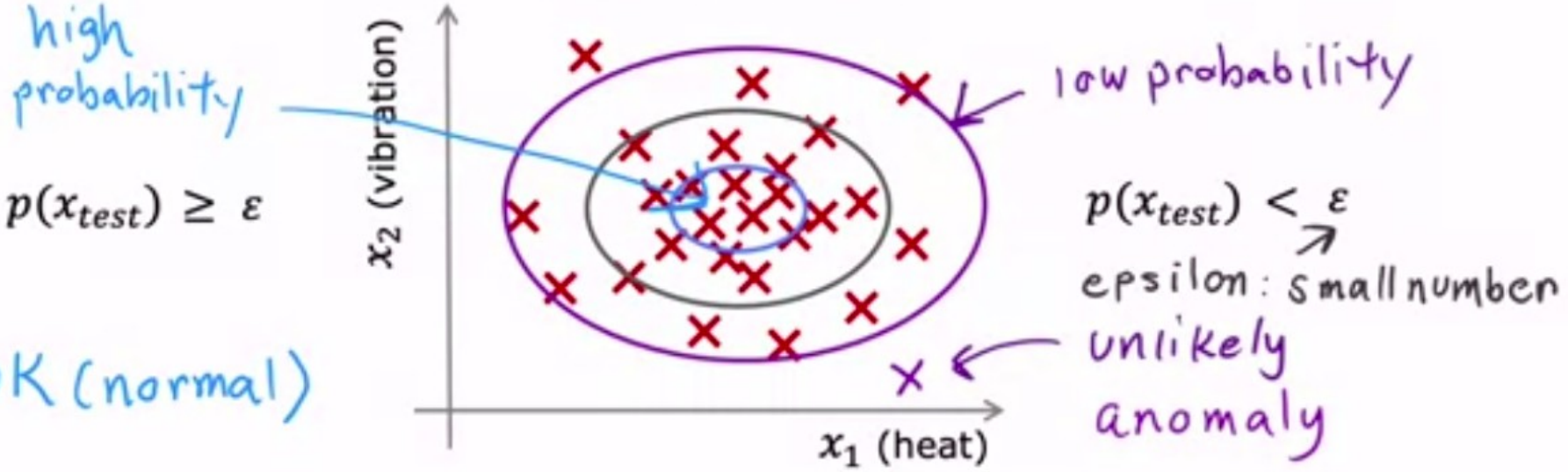
Density estimation

Dataset: $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

probability of x being seen in dataset

Model $p(x)$

Is x_{test} anomalous?



Anomaly detection example

how often log in?
how many web pages visited?
transactions?
posts? typing speed?

Fraud detection:

- $x^{(i)}$ = features of user i's activities
- Model $p(x)$ from data.
- Identify unusual users by checking which have $p(x) < \varepsilon$

perform additional checks to identify real fraud vs. false alarms

Manufacturing:

$x^{(i)}$ = features of product i

airplane engine
circuit board
smartphone

ratios

Monitoring computers in a data center:

$x^{(i)}$ = features of machine i

- x_1 = memory use,
- x_2 = number of disk accesses/sec,
- x_3 = CPU load,
- x_4 = CPU load/network traffic.

DeepLearning.AI

Stanford
ONLINE

Anomaly Detection

Gaussian (Normal) Distribution

Gaussian (Normal) distribution

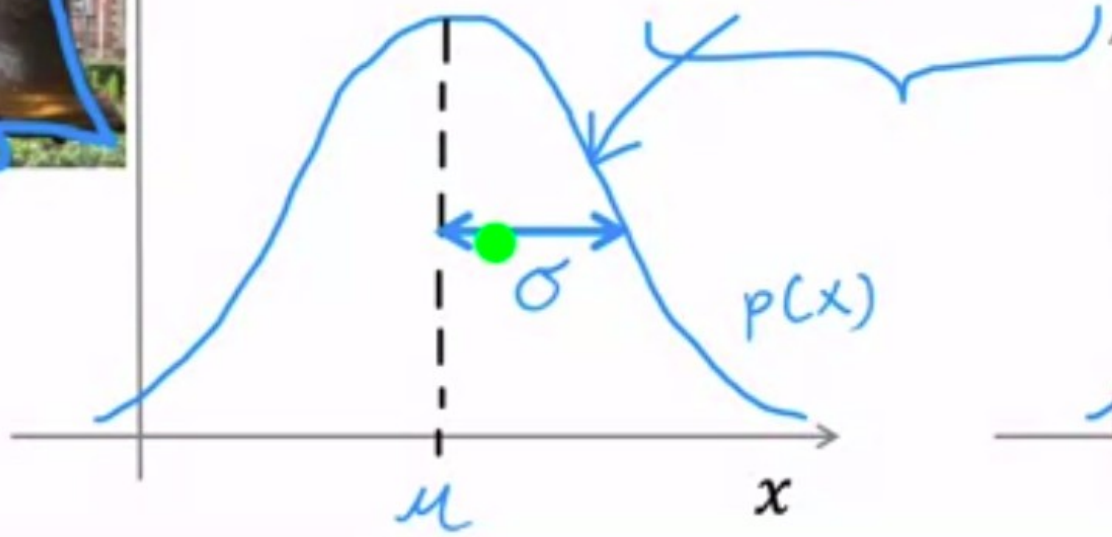
σ standard deviation

σ^2 variance

Say x is a number.
Probability of x is determined by a Gaussian with mean μ, variance σ^2.

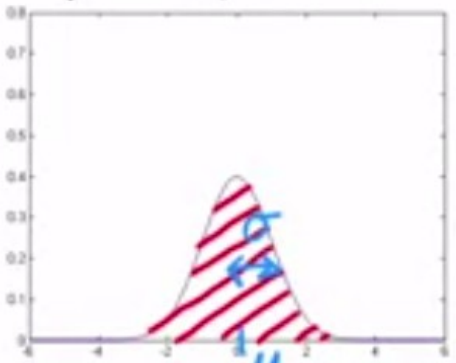
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$\pi = 3.14$

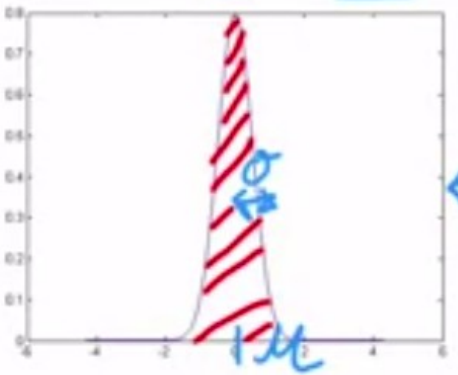


Gaussian distribution example

$\mu = 0, \sigma = 1$

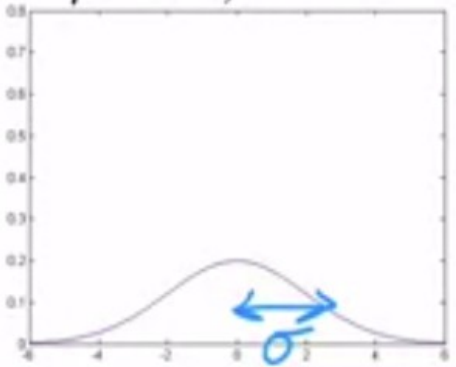


$\mu = 0, \sigma = 0.5$



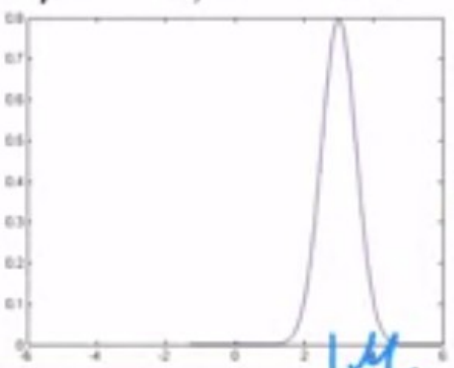
$\sigma^2 = 0.25$

$\mu = 0, \sigma = 2$



$\sigma^2 = 4$

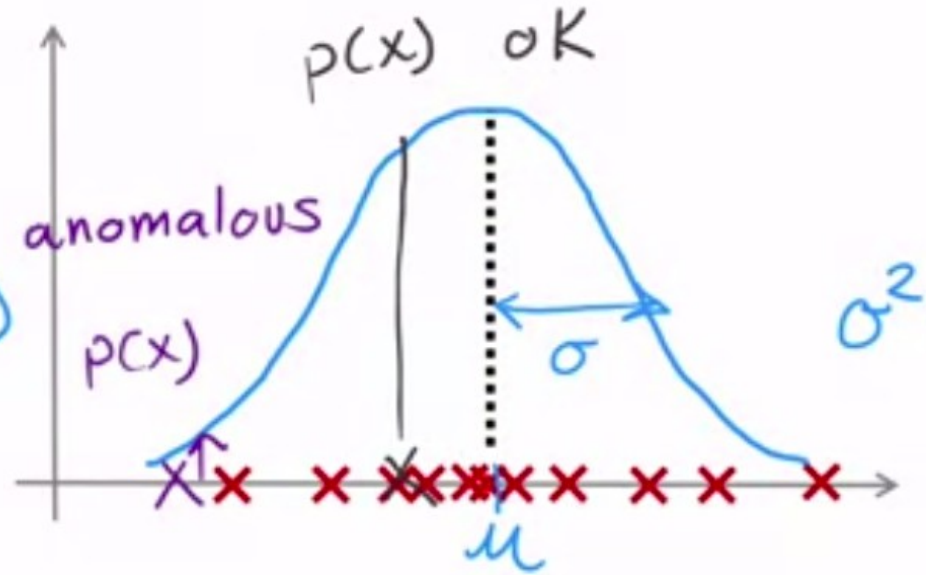
$\mu = 3, \sigma = 0.5$



Parameter estimation

maximum likelihood for μ, σ

Dataset: $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$



$$\mu = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} - \mu)^2$$

$\frac{1}{m-1}$

x is a number (1 feature)

DeepLearning.AI

Stanford
ONLINE

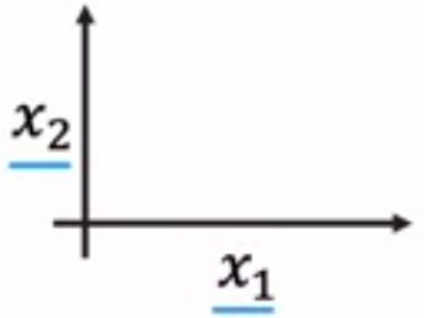
Anomaly Detection

Algorithm

Density estimation

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

Each example $\vec{x}^{(i)}$ has n features



$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$p(\vec{x}) = p(x_1; \mu_1, \sigma_1^2) * p(x_2; \mu_2, \sigma_2^2) * p(x_3; \mu_3, \sigma_3^2) * \cdots * p(x_n; \mu_n, \sigma_n^2)$$

$$= \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2)$$

$\sum$ "add" $\prod$ "multiply"

$$p(x_1 = \text{high temp}) = 1/10$$
$$p(x_2 = \text{high vibra}) = 1/20$$
$$p(x_1, x_2) = p(x_1) * p(x_2)$$
$$= \frac{1}{10} \times \frac{1}{20} = \frac{1}{200}$$

Anomaly detection algorithm

1. Choose n features x_i that you think might be indicative of anomalous examples.
2. Fit parameters $\mu_1, \ldots, \mu_n, \sigma_1^2, \ldots, \sigma_n^2$

$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} x_j^{(i)} \qquad \sigma_j^2 = \frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)} - \mu_j)^2$$

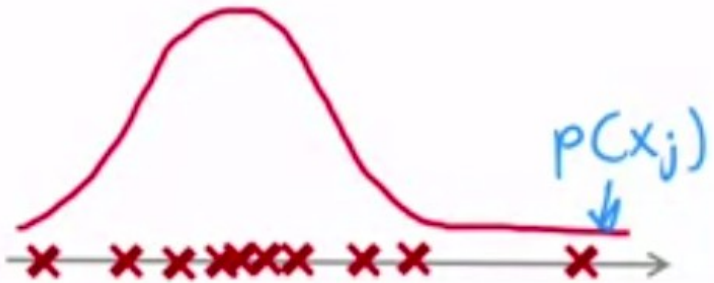
Vectorized formula

$$\vec{\mu} = \frac{1}{m}\sum_{i=1}^{m} \vec{x}^{(i)} \qquad \vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_n \end{bmatrix}$$

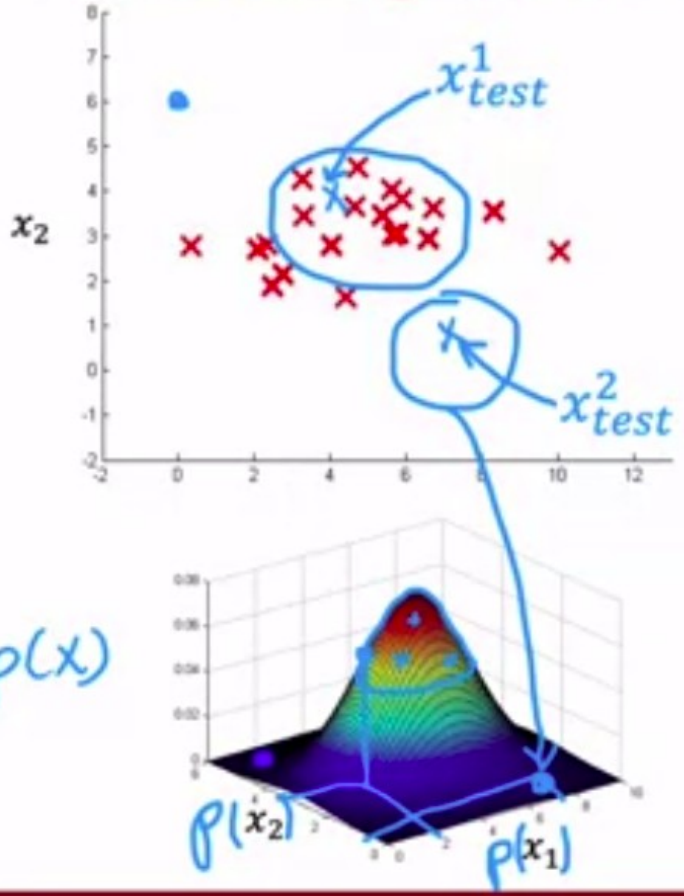
3. Given new example x, compute $p(x)$:

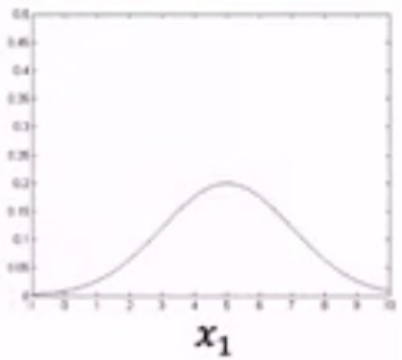
$$p(x) = \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_j} exp(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2})$$

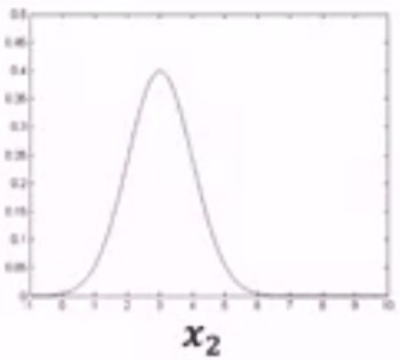
Anomaly if $p(x) < \varepsilon$



Anomaly detection example







$\mu_1 = 5,\ \sigma_1 = 2$

$p(x_1; \mu_1, \sigma_1^2)$

$\mu_2 = 3,\ \sigma_2 = 1$

$p(x_2; \mu_2, \sigma_2^2)$

$\varepsilon = 0.02$

$p\left(x^{(1)}_{test}\right) = 0.0426 \longrightarrow$ "ok"

$p\left(x^{(2)}_{test}\right) = 0.0021 \longrightarrow$ anomaly

DeepLearning.AI

Stanford
ONLINE

Anomaly Detection

Developing and evaluating an anomaly detection system

The importance of real-number evaluation

When developing a learning algorithm (choosing features, etc.),
making decisions is much easier if we have a way of evaluating our learning algorithm.

Assume we have some labeled data, of anomalous and non-anomalous examples.

$y=1$ $y=0$

Training set: $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$ (assume normal examples/not anomalous)

$y=0$ for all training examples

Cross validation set: $\left(x_{cv}^{(1)}, y_{cv}^{(1)}\right), \ldots, \left(x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})}\right)$

Test set: $\left(x_{test}^{(1)}, y_{test}^{(1)}\right), \ldots, \left(x_{test}^{(m_{test})}, y_{test}^{(m_{test})}\right)$

include a few anomalous examples $y=1$

mostly normal examples $y=0$

Aircraft engines monitoring example

10000 good (normal) engines
20 flawed engines (anomalous)

2 to 50

2

$y=0$

$y=1$

Training set: 6000 good engines

train algorithm on training set

CV: 2000 good engines ($y = 0$) 10 anomalous ($y = 1$)

use cross validation set

tune ε tune x_j

Test: 2000 good engines ($y = 0$), 10 anomalous ($y = 1$)

Alternative: No test set

Use if very few labeled anomalous examples

higher risk of overfitting

Training set: 6000 good engines

2

CV: 4000 good engines ($y = 0$), 20 anomalous ($y = 1$)

tune ε tune x_j

Algorithm evaluation

course 2 week 3

skewed datasets

Fit model $p(x)$ on training set $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$

On a cross validation/test example x, predict

$$y = \begin{cases} 1 & if\ p(x) < \varepsilon\ (anomaly) \\ 0 & if\ p(x) \geq \varepsilon\ (normal) \end{cases}$$

10

2000

Possible evaluation metrics:

- True positive, false positive, false negative, true negative
- Precision/Recall
- F_1-score

Use cross validation set to choose parameter ε

DeepLearning.AI

Stanford
ONLINE

Anomaly Detection

Anomaly detection vs. supervised learning

Anomaly detection vs. Supervised learning

Very small number of positive examples ($y = 1$). (0-20 is common). Large number of negative ($y = 0$) examples.

$p(x)$

$y = 1$

Many different "types" of anomalies. Hard for any algorithm to learn from positive examples what the anomalies look like; future anomalies may look nothing like any of the anomalous examples we've seen so far.

Fraud

Large number of positive and negative examples.

20 positive examples

Enough positive examples for algorithm to get a sense of what positive examples are like, future positive examples likely to be similar to ones in training set.

Spam

Anomaly detection vs. Supervised learning

- Fraud detection
- Manufacturing - Finding new previously unseen defects in manufacturing.(e.g. aircraft engines)
- Monitoring machines in a data center

⋮

- Email spam classification
- Manufacturing - Finding known, previously seen defects $y=1$ scratches
- Weather prediction (sunny/rainy/etc.)
- Diseases classification

⋮

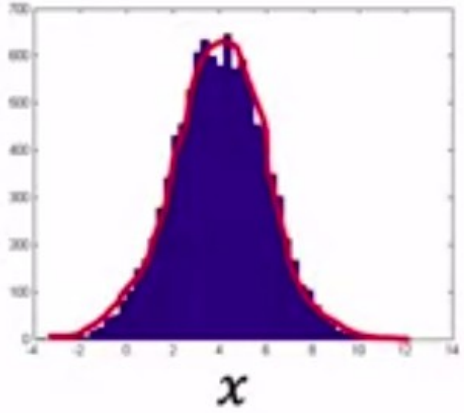
DeepLearning.AI

Stanford
ONLINE

Anomaly Detection

Choosing what features to use

Non-gaussian features



$p(x_1; \mu_1, \sigma_1^2)$

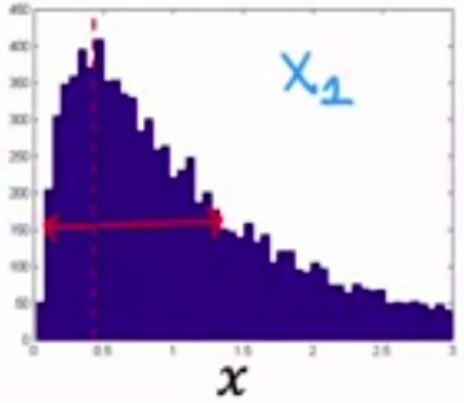
`plt.hist(x)`

$x_1 \leftarrow \log(x_1)$

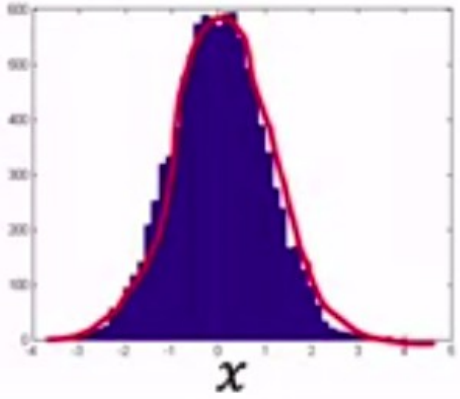
$x_2 \leftarrow \log(x_2 + 1)$ $\log(x_2 + c)$

$x_3 \leftarrow \sqrt{x_3} = x_3^{1/2}$

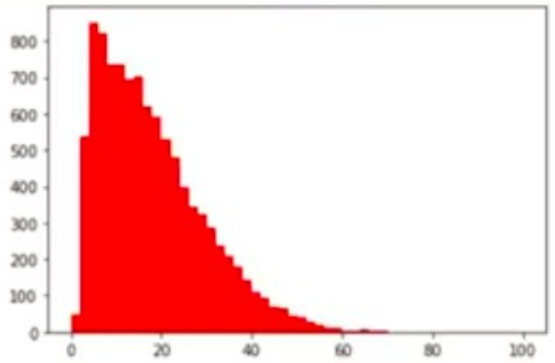
$x_4 \leftarrow x_4^{1/3}$



`np.log(x)`

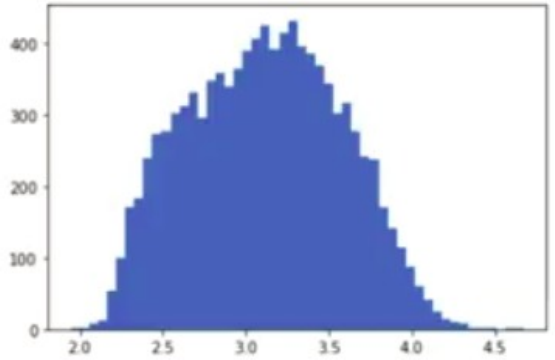


```
#Plot histogram to check skewness
plt.hist(x, bins=50, color='r');
```



In [16]:

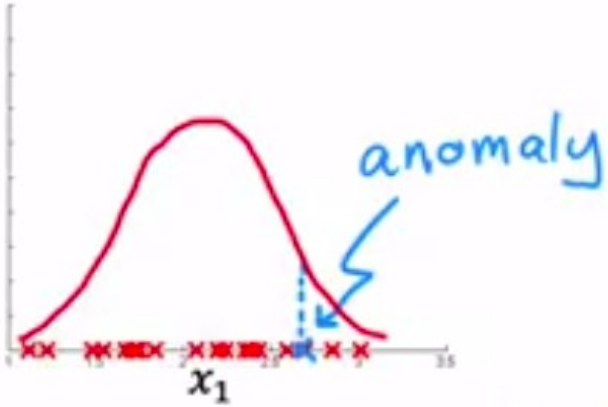
```
plt.hist(np.log(x+7), bins=50);
```

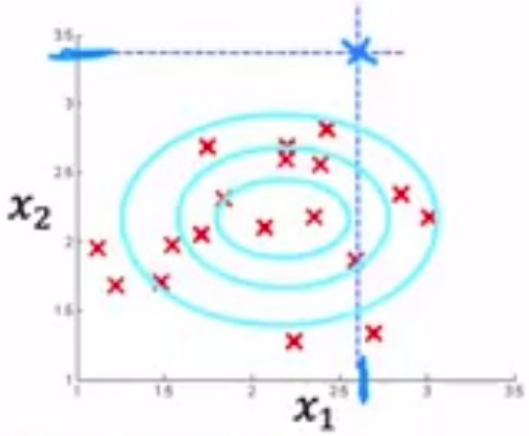


Error analysis for anomaly detection

Want $p(x) \geq \varepsilon$ large for normal examples x.
$p(x) < \varepsilon$ small for anomalous examples x.

Most common problem:
$p(x)$ is comparable for normal and anomalous examples.
($p(x)$ is large for both)





x_1 num transactions x_2 typing speed

Monitoring computers in a data center

Choose features that might take on unusually large or small values in the event of an anomaly.

x_1 = memory use of computer
x_2 = number of disk accesses/sec
x_3 = CPU load
x_4 = network traffic

high
low

not unusual

$$x_5 = \frac{\text{CPU load}}{\text{network traffic}} \qquad x_6 = \frac{(\text{CPU load})^2}{\text{network traffic}}$$

Deciding feature choice based on $p(x)$
Large for normal examples;
Becomes small for anomaly in the cross validation set

Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to next item

1. 1 / 1 point

You are building a system to detect if computers in a data center are malfunctioning. You have 10,000 data points of computers functioning well, and no data from computers malfunctioning. What type of algorithm should you use?

- (•) Anomaly detection
- ( ) Supervised learning

Correct
Creating an anomaly detection model does not require labeled data.

2. 1 / 1 point

You are building a system to detect if computers in a data center are malfunctioning. You have 10,000 data points of computers functioning well, and 10,000 data points of computers malfunctioning. What type of algorithm should you use?

Correct
Creating an anomaly detection model does not require labeled data.

2. **1 / 1 point**

You are building a system to detect if computers in a data center are malfunctioning. You have 10,000 data points of computers functioning well, and 10,000 data points of computers malfunctioning. What type of algorithm should you use?

- ○ Anomaly detection
- ◉ Supervised learning

Correct
You have a sufficient number of anomalous examples to build a supervised learning model.

3. **1 / 1 point**

Say you have 5,000 examples of normal airplane engines, and 15 examples of anomalous engines. How would you use the 15 examples of anomalous engines to evaluate your anomaly detection algorithm?

- ○ You cannot evaluate an anomaly detection algorithm because it is an unsupervised learning algorithm.

3. 1 / 1 point

Say you have 5,000 examples of normal airplane engines, and 15 examples of anomalous engines. How would you use the 15 examples of anomalous engines to evaluate your anomaly detection algorithm?

- ○ You cannot evaluate an anomaly detection algorithm because it is an unsupervised learning algorithm.
- ○ Because you have data of both normal and anomalous engines, don't use anomaly detection. Use supervised learning instead.
- ○ Use it during training by fitting one Gaussian model to the normal engines, and a different Gaussian model to the anomalous engines.
- ◉ Put the data of anomalous engines (together with some normal engines) in the cross-validation and/or test sets to measure if the learned model can correctly detect anomalous engines.

Correct
Anomalous examples are used to evaluate rather than train the model.

4. Anomaly detection flags a new input x as an anomaly if $p(x) < \epsilon$. If we reduce the value of ϵ, what happens? 1 / 1 point

- ○ The algorithm is more likely to classify new examples as an anomaly.

Correct
Anomalous examples are used to evaluate rather than train the model.

4. Anomaly detection flags a new input x as an anomaly if $p(x) < \epsilon$. If we reduce the value of ϵ, what happens? **1 / 1 point**

- ○ The algorithm is more likely to classify new examples as an anomaly.
- ◉ The algorithm is less likely to classify new examples as an anomaly.
- ○ The algorithm is more likely to classify some examples as an anomaly, and less likely to classify some examples as an anomaly. It depends on the example x.
- ○ The algorithm will automatically choose parameters μ and σ to decrease $p(x)$ and compensate.

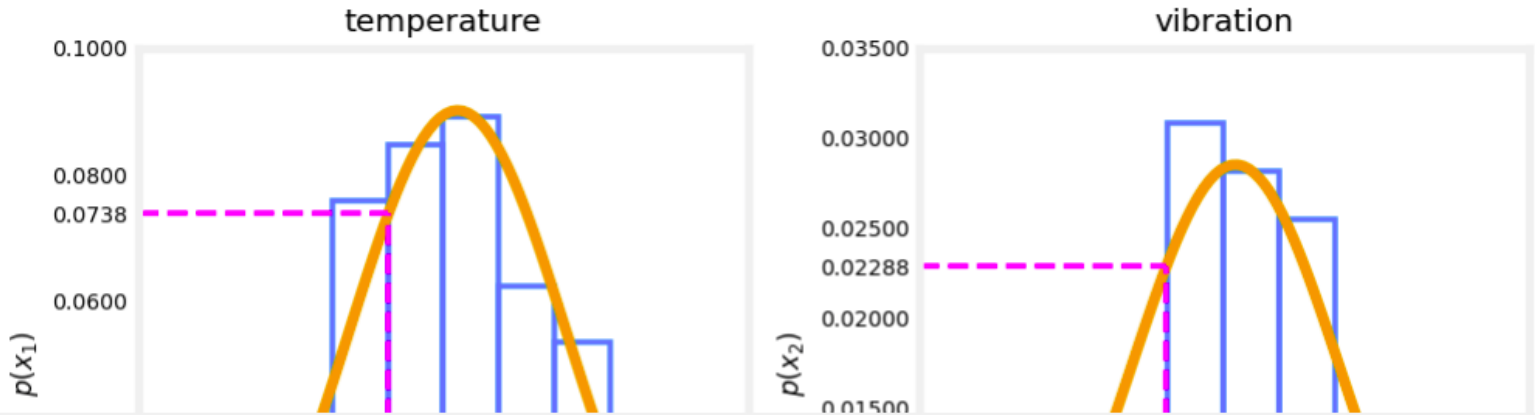
Correct
When ϵ is reduced, the probability of an event being classified as an anomaly is reduced.

5. You are monitoring the temperature and vibration intensity on newly manufactured aircraft engines. You have measured 100 engines and fit the Gaussian model described in the video lectures to the data. The 100 examples and the resulting distributions are shown in the figure below. **1 / 1 point**

Correct
When ϵ is reduced, the probability of an event being classified as an anomaly is reduced.

5. You are monitoring the temperature and vibration intensity on newly manufactured aircraft engines. You have measured 100 engines and fit the Gaussian model described in the video lectures to the data. The 100 examples and the resulting distributions are shown in the figure below. **1 / 1 point**

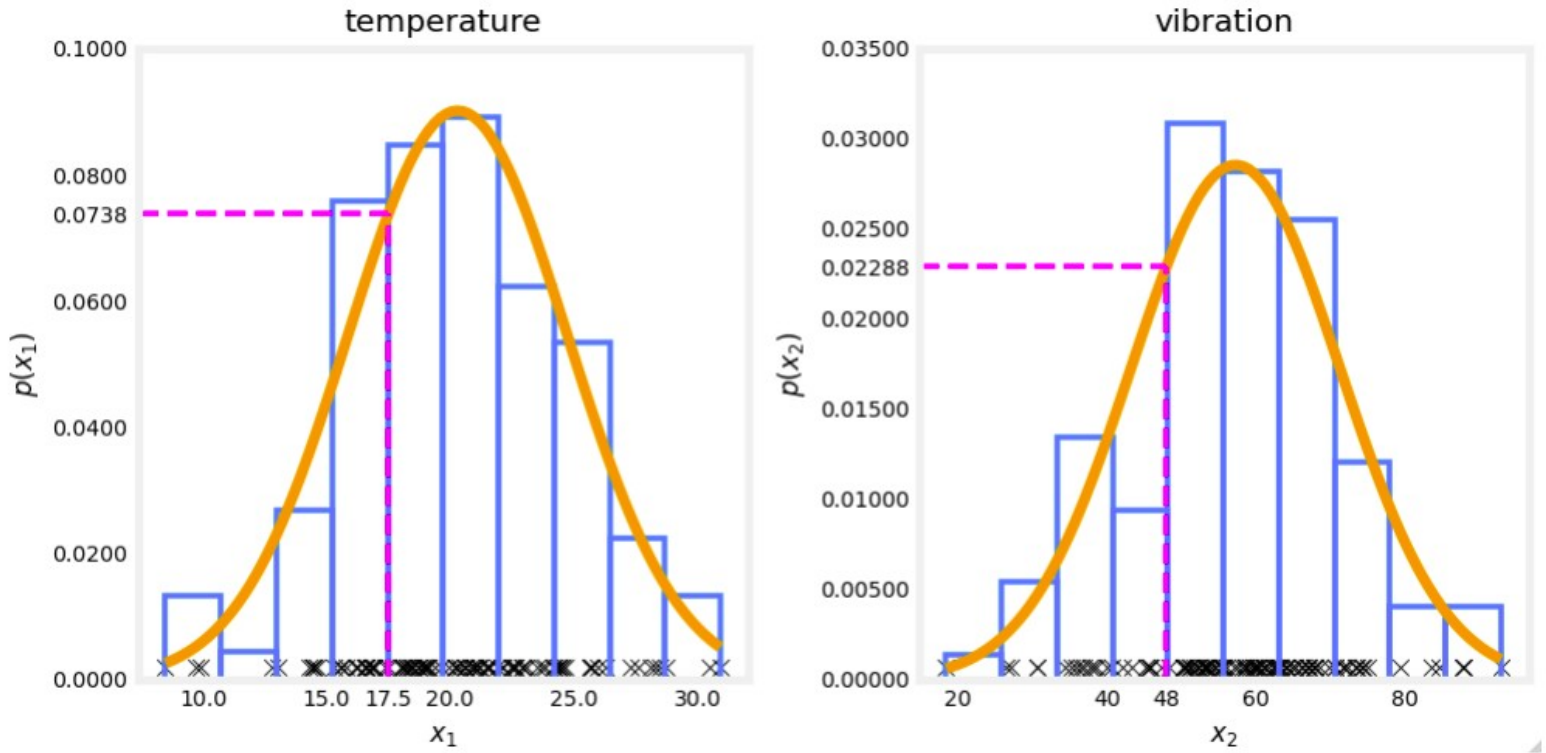
The measurements on the latest engine you are testing have a temperature of 17.5 and a vibration intensity of 48. These are shown in magenta on the figure below. What is the probability of an engine having these two measurements?

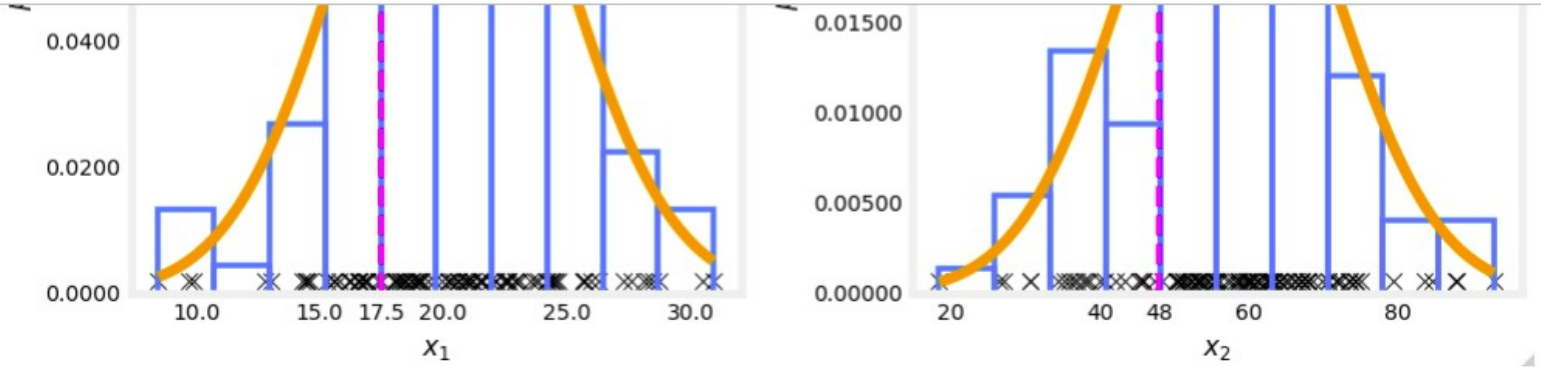


Back **Anomaly detection**
Graded Quiz • 30 min

Due Jul 30, 11:59 PM IST

The measurements on the latest engine you are testing have a temperature of 17.5 and a vibration intensity of 48. These are shown in magenta on the figure below.
What is the probability of an engine having these two measurements?





- ○ 0.0738 + 0.02288 = 0.0966
- ◉ 0.0738 * 0.02288 = 0.00169
- ○ 17.5 + 48 = 65.5
- ○ 17.5 * 48 = 840

Correct
According to the model described in lecture, p(A, B) = p(A) * p(B). .