

Which of the following statements are true? Check all that apply.

In unsupervised learning, the training set is of the form  $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$  without labels  $y^{(i)}$ .

Correct

Clustering is an example of unsupervised learning.

Correct

In unsupervised learning, you are given an unlabeled dataset and are asked to find "structure" in the data.

Correct

Clustering is the only unsupervised learning algorithm.

Un-selected is correct

---

Which of the following is the recommended way to initialize k-means?

Pick a random integer  $i$  from  $\{1, \dots, k\}$ . Set  $\mu_1 = \mu_2 = \dots = \mu_k = x^{(i)}$ .

Pick  $k$  distinct random integers  $i_1, \dots, i_k$  from  $\{1, \dots, m\}$ . Set  $\mu_1 = x^{(i_1)}, \mu_2 = x^{(i_2)}, \dots, \mu_k = x^{(i_k)}$ .

Pick  $k$  distinct random integers  $i_1, \dots, i_k$  from  $\{1, \dots, m\}$ . Set  $\mu_1 = x^{(i_1)}, \mu_2 = x^{(i_2)}, \dots, \mu_k = x^{(i_k)}$ .

Correct

Set every element of  $\mu_i \in \mathbb{R}^n$  to a random value between  $-\epsilon$  and  $\epsilon$ , for some small  $\epsilon$ .

---

Suppose you run k-means using  $k = 3$  and  $k = 5$ . You find that the cost function  $J$  is much higher for  $k = 5$  than for  $k = 3$ . What can you conclude?

This is mathematically impossible. There must be a bug in the code.

The correct number of clusters is  $k = 3$ .

In the run with  $k = 5$ , k-means got stuck in a bad local minimum. You should try re-running k-means with multiple random initializations.

Correct

In the run with  $k = 3$ , k-means got lucky. You should try re-running k-means with  $k = 3$  and different random initializations until it performs no better than with  $k = 5$ .

---

For which of the following tasks might K-means clustering be a suitable algorithm? Select all that apply.

Given a database of information about your users, automatically group them into different market segments.

Given sales data from a large number of products in a supermarket, figure out which products tend to form coherent groups (say are frequently purchased together) and thus should be put on the same shelf.

Given historical weather records, predict the amount of rainfall tomorrow (this would be a real-valued output)

Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

---

Suppose we have three cluster centroids  $\mu_1 = [12]$ ,  $\mu_2 = [-30]$  and  $\mu_3 = [42]$ . Furthermore, we have a training example  $x^{(i)} = [-21]$ . After a cluster assignment step, what will  $c^{(i)}$  be?

$c^{(i)}$  is not assigned

$c^{(i)} = 2$

$c^{(i)} = 1$

$c^{(i)} = 3$

---

K-means is an iterative algorithm, and two of the following steps are repeatedly carried out in its inner-loop. Which two?

The cluster centroid assignment step, where each cluster centroid  $\mu_i$  is assigned (by setting  $c^{(i)}$ ) to the closest training example  $x^{(i)}$ .

Move each cluster centroid  $\mu_k$ , by setting it to be equal to the closest training example  $x^{(i)}$

Move the cluster centroids, where the centroids  $\mu_k$  are updated.

The cluster assignment step, where the parameters  $c^{(i)}$  are updated.

---

Suppose you have an unlabeled dataset  $\{x^{(1)}, \dots, x^{(m)}\}$ . You run K-means with 50 different random initializations, and obtain 50 different clusterings of the data. What is the recommended way for choosing which one of these 50 clusterings to use?

Use the elbow method.

Plot the data and the cluster centroids, and pick the clustering that gives the most "coherent" cluster centroids.

Manually examine the clusterings, and pick the best one.

Compute the distortion function  $J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_k)$ , and pick the one that minimizes this.

---

Suppose we apply dimensionality reduction to a dataset of  $m$  examples  $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ , where  $x^{(i)} \in \mathbb{R}^n$ . As a result of this, we will get out:

A lower dimensional dataset  $\{z^{(1)}, z^{(2)}, \dots, z^{(k)}\}$  of  $k$  examples where  $k \leq n$ .

A lower dimensional dataset  $\{z^{(1)}, z^{(2)}, \dots, z^{(k)}\}$  of  $k$  examples where  $k > n$ .

A lower dimensional dataset  $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$  of  $m$  examples where  $z^{(i)} \in \mathbb{R}^k$  for some value of  $k$  and  $k \leq n$ .

Correct

A lower dimensional dataset  $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$  of  $m$  examples where  $z^{(i)} \in \mathbb{R}^k$  for some value of  $k$  and  $k > n$ .

---

Suppose you have a dataset  $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$  where  $x^{(i)} \in \mathbb{R}^n$ . In order to visualize it, we apply dimensionality reduction and get  $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$  where  $z^{(i)} \in \mathbb{R}^k$  is  $k$ -dimensional. In a typical setting, which of the following would you expect to be true? Check all that apply.

k > n

Correct k ≤ n

k ≥ 4

Correct k = 2 or k = 3 (since we can plot 2D or 3D data but don't have ways to visualize higher dimensional data)