

## Feedback — XIII. Clustering

[Help](#)

You submitted this quiz on **Sun 11 May 2014 11:58 PM PDT**. You got a score of **5.00** out of **5.00**.

### Question 1

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

Your Answer	Score	Explanation
<input type="checkbox"/> Given many emails, you want to determine if they are Spam or Non-Spam emails.	✓ 0.25	Classifying input as spam / non-spam requires labels for the data, which K-means does not use.
<input type="checkbox"/> Given historical weather records, predict if tomorrow's weather will be sunny or rainy.	✓ 0.25	K-means cannot make classification predictions, as it does not label its inputs.
<input checked="" type="checkbox"/> From the user usage patterns on a website, figure out what different groups of users exist.	✓ 0.25	We can cluster the users with K-means to find different, distinct groups.
<input checked="" type="checkbox"/> Given a database of information about your users, automatically group them into different market segments.	✓ 0.25	You can use K-means to cluster the database entries, and each cluster will correspond to a different market segment.
Total	1.00 / 1.00	

### Question 2

Suppose we have three cluster centroids  $\mu_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ,  $\mu_2 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}$  and  $\mu_3 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ .

Furthermore, we have a training example  $x^{(i)} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ . After a cluster assignment step, what will  $c^{(i)}$  be?

Your Answer	Score	Explanation
<input type="radio"/> $c^{(i)} = 2$		
<input type="radio"/> $c^{(i)}$ is not assigned		
<input type="radio"/> $c^{(i)} = 3$		
<input checked="" type="radio"/> $c^{(i)} = 1$	✓ 1.00	$x^{(i)}$ is closest to $\mu_1$ , so $c^{(i)} = 1$
Total	1.00 / 1.00	

### Question 3

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

Your Answer	Score	Explanation
<input checked="" type="checkbox"/> Move the cluster centroids, where the centroids $\mu_k$ are updated.	✓ 0.25	The cluster update is the second step of the K-means loop.
<input checked="" type="checkbox"/> The cluster assignment step, where the parameters $c^{(i)}$ are updated.	✓ 0.25	This is the correct first step of the K-means loop.
<input type="checkbox"/> The cluster centroid assignment step, where each cluster centroid $\mu_i$ is assigned (by setting $c^{(i)}$ ) to the closest training example $x^{(i)}$ .	✓ 0.25	This is not a correct description of the cluster assignment step.
<input type="checkbox"/> Randomly initialize the cluster centroids.	✓ 0.25	Initialization is performed only once at the start of K-means, not in the main loop.
Total	1.00 / 1.00	

## Question 4

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?

Your Answer	Score	Explanation
<input type="radio"/> Manually examine the clusterings, and pick the best one.		
<input checked="" type="radio"/> Compute the distortion function $J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_k)$ , and pick the one that minimizes this.	✓ 1.00	A lower value for the distortion function implies a better clustering, so you should choose the clustering with the smallest value for the distortion function.
<input type="radio"/> Use the elbow method.		
<input type="radio"/> The answer is ambiguous, and there is no good way of choosing.		
Total	1.00 / 1.00	

## Question 5

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

Your Answer	Score	Explanation
<input type="checkbox"/> K-Means will always give the same results regardless of the initialization of the centroids.	✓ 0.25	K-means is sensitive to different initializations, which is why you should run it multiple times from different random initializations.
<input checked="" type="checkbox"/> A good way to initialize K-means is to select K (distinct) examples from the training set and set the cluster centroids equal to these selected	✓ 0.25	This is the recommended method of initialization.

examples.

<input checked="" type="checkbox"/> If we are worried about K-means getting stuck in bad local optima, one way to ameliorate (reduce) this problem is if we try using multiple random initializations.	✓	0.25	Since each run of K-means is independent, multiple runs can find different optima, and some should avoid bad local optima.
<input type="checkbox"/> Once an example has been assigned to a particular centroid, it will never be reassigned to another different centroid	✓	0.25	Each iteration of K-means performs a cluster assignment step in which each example may be assigned to a different centroid.
Total		1.00 / 1.00	