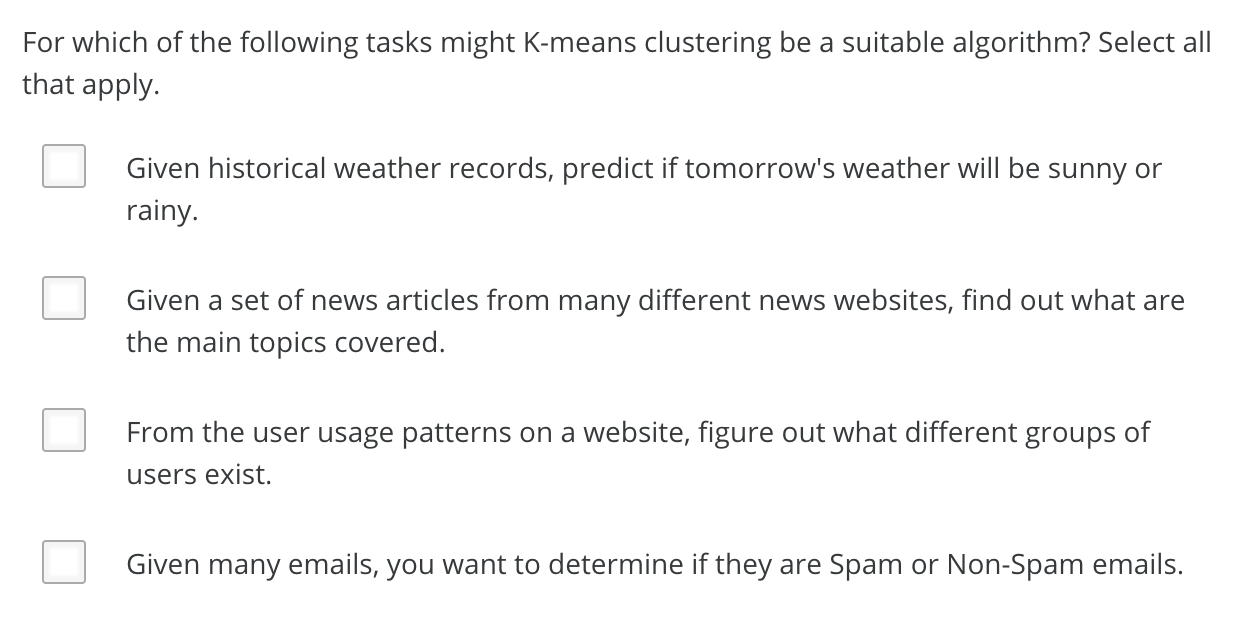
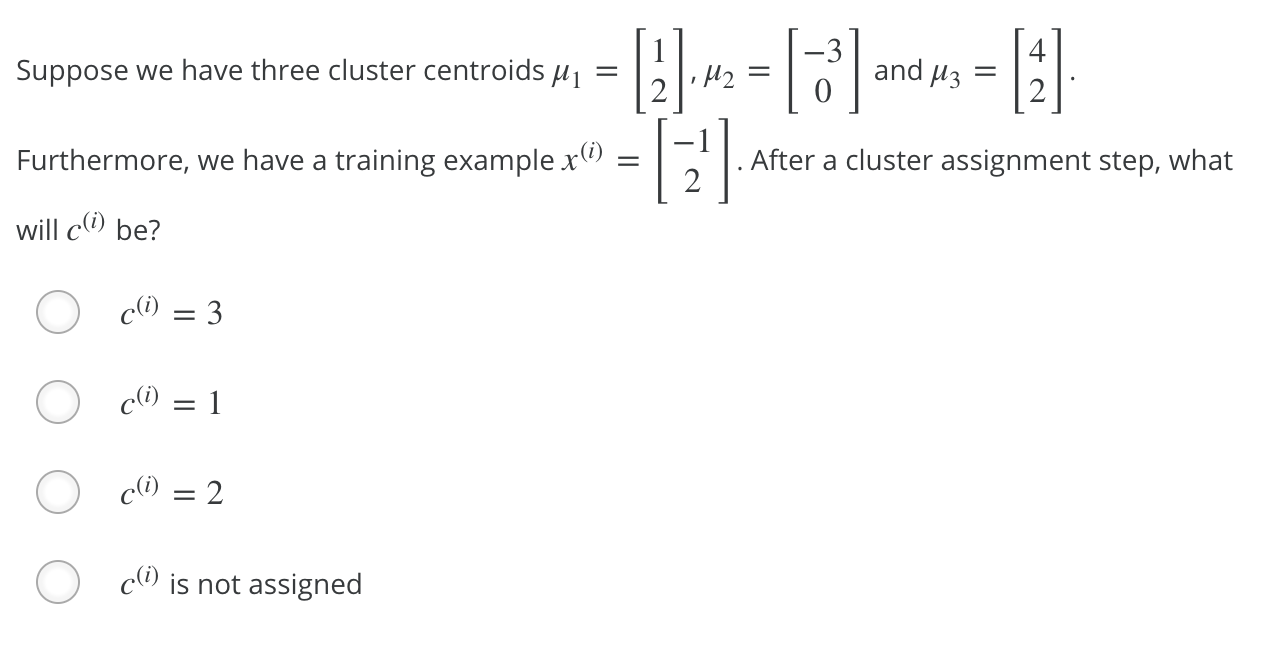
**Machine Learning Week 8 Quiz 1 (Unsupervised Learning) Stanford Coursera**

**Question 1**

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/Stanford_Machine_Learning/Week1/data/unsupervisedLearningQ1.png)

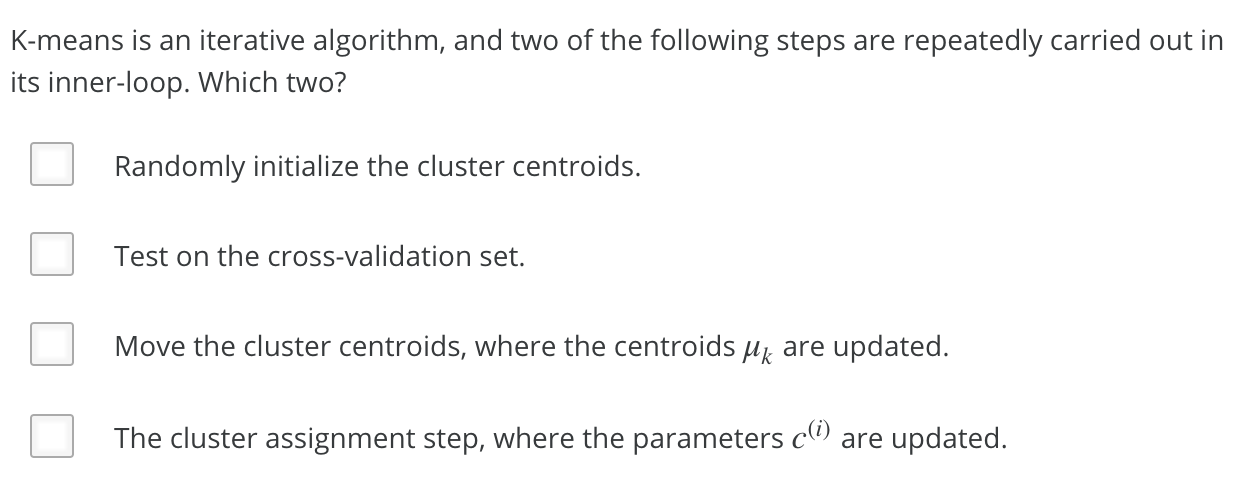
| **True or False** | **Statement** | **Explanation** |
| --- | --- | --- |
| False | Given historical weather records, predict if tomorrow's weather will be sunny or rainy | K-means cannot make classification predictions as it does not label its inputs. |
| True | Given a set of news articles from many different websites, find out what topics are the main topics covered | You can use K-means to cluster, and each cluster will correspond to a different market segment. |
| True | From the user usage patterns on a website, figure out what different groups of users exists. | You can use K-means to cluster users with each cluster corresponding to a different market segment. |
| False | Given many emails, you want to determine if they are Spam or Non-Spam emails. | K-means cannot make classification predictions as it does not label its inputs |
| True | Given a database of information about your users, automatically group them into different market segments. | You can use K-means to cluster the database entries, and each cluster will correspond to a different market segment. |
| True | 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. | Market Segmentation. |
| False | Given sales data from a large number of products in a supermarket, estimate future sales for each of these products. | Such a prediction is a regression problem, and K-means does not use labels on the data, so it cannot perform regression. |

**Question 2**

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/Stanford_Machine_Learning/Week1/data/unsupervisedLearningQ2.png)

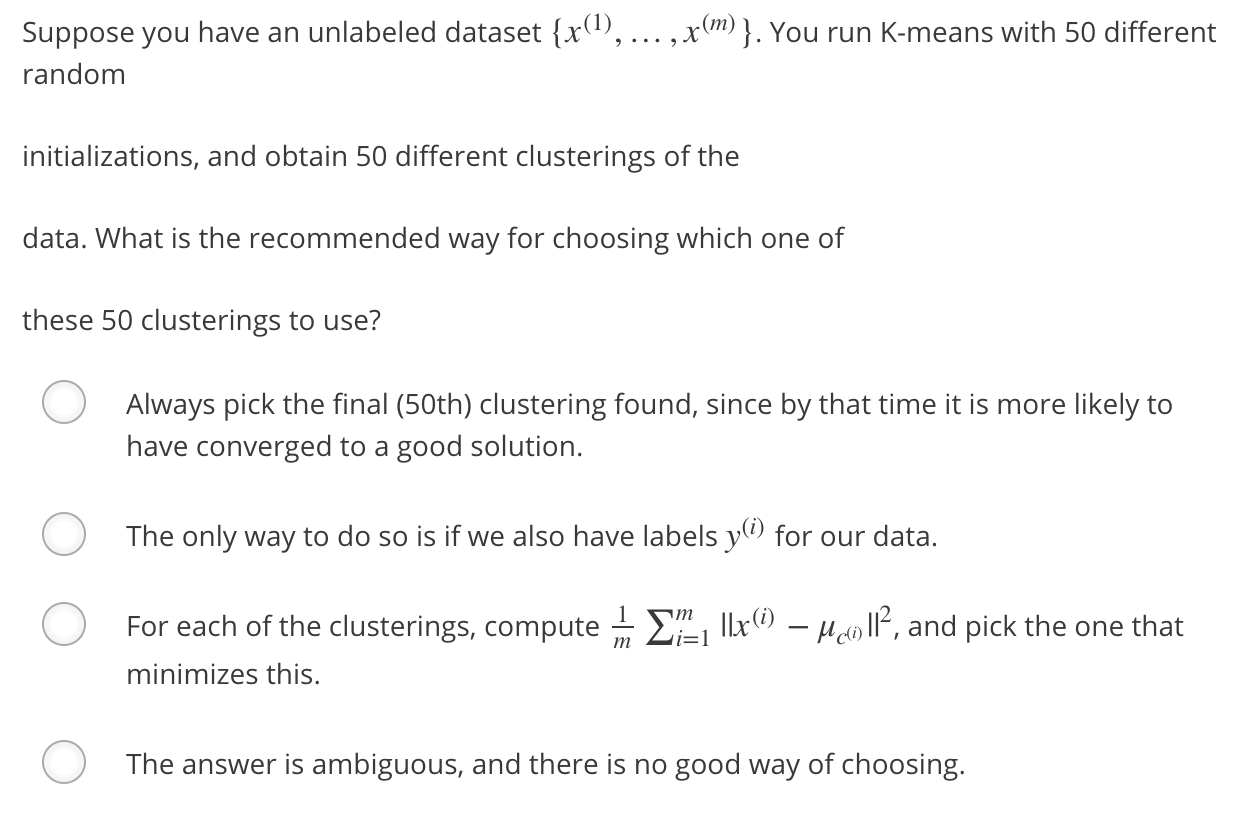
| **Answer** | **Explanation** |
| --- | --- |
| c(i) = 1 | x(i) is closest to μ1, so c(i) = 1 |

**Question 3**

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/Stanford_Machine_Learning/Week1/data/unsupervisedLearningQ3.png)

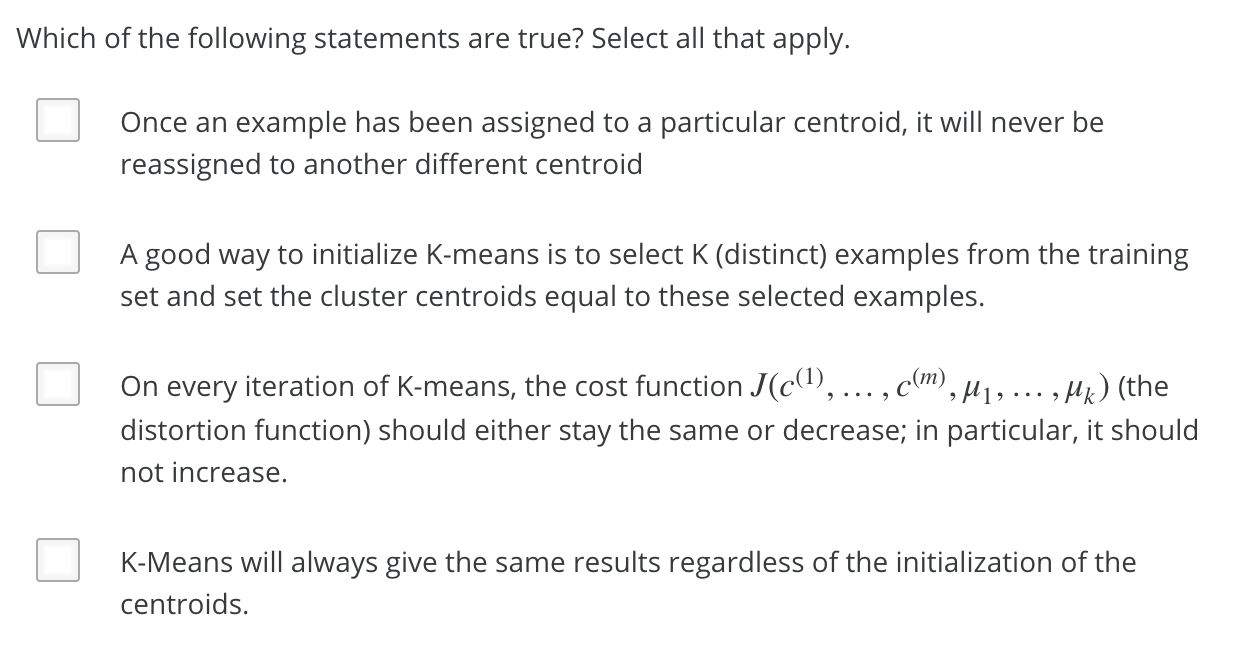
| **True or False** | **Statement** | **Explanation** |
| --- | --- | --- |
| False | Randomly initialize the cluster centroids | Done earlier |
| False | Test on the cross-validation set | Any sort of testing is outside the scope of K-means algorithm itself |
| True | Move the cluster centroids, where the centroids, μk are updated | The cluster update is the second step of the K-means loop |
| True | The cluster assignment step, where the parameters c(i) are updated | This is the correst first step of the Kmeans loop |

**Question 4**

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/Stanford_Machine_Learning/Week1/data/unsupervisedLearningQ4.png)

| **Answer** | **Explanation** |
| --- | --- |
| [Distortion Cost FUnction](https://github.com/mGalarnyk/datasciencecoursera/blob/master/Stanford_Machine_Learning/Week1/data/ClusteringDistortionFunction.png) | This is the distortion cost function which we seek to minimize |

**Question 5**

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/Stanford_Machine_Learning/Week1/data/unsupervisedLearningQ5.png)

| **True or False** | **Statement** | **Explanation** |
| --- | --- | --- |
| False | Once an example has been assigned to a particular centroid, it will never be reassigned to another centroid | Not sure yet |
| True | 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 examples. | This is the recommended method of initialization. |
| True | On every iteration of K-means, the cost funtion J(c(1), ..., c(m), μ1, ..., μk (the distortion function) should either stay the same or decrease; in particular, it should not increase | True |
| False | K-Means will always give the same results regardless of the initialization of the centroids. | K-means is sensitive to different initializations, which is why you should run it multiple times from different random initializations |
| True | For some datasets, the "right" or "correct" value of K (the number of clusters) can be ambiguous, and hard even for a human expert looking carefully at the data to decide. | Look at an elbow curve for an example. It can often be ambiguous. |
| True | 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. | None needed |