



Congratulations! You passed!

Next Item



1 / 1 point

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



Given many emails, you want to determine if they are Spam or Non-Spam emails.

Un-selected is correct



From the user usage patterns on a website, figure out what different groups of users exist.

Correct

We can cluster the users with K-means to find different, distinct groups.



Given historical weather records, predict if tomorrow's weather will be sunny or rainy.

Un-selected is correct



Given a set of news articles from many different news websites, find out what are the main topics covered.

Correct

K-means can cluster the articles and then we can inspect them or use other methods to infer what topic each cluster represents



1 / 1 point

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 $x^{(i)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. After a cluster assignment step, what will $c^{(i)}$ be?



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



$c^{(i)} = 1$



$c^{(i)} = 3$

Correct

$x^{(i)}$ is closest to μ_3 , so $c^{(i)} = 3$



$c^{(i)} = 2$



1 / 1 point

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

Correct

This is the correst first step of the K-means loop.



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

Un-selected is correct



Move the cluster centroids, where the centroids μ_k are updated.

Correct

The cluster update is the second step of the K-means loop.



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

Un-selected is correct



1 / 1 point

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?



Use the elbow method.



Manually examine the clusterings, and pick the best one.



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



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

Correct

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.



1 / 1 point

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



Once an example has been assigned to a particular centroid, it will never be reassigned to another different centroid

Un-selected is correct



On every iteration of K-means, the cost function $J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_k)$ (the distortion function) should either stay the same or decrease; in particular, it should not increase.

Correct

Both the cluster assignment and cluster update steps decrease the cost / distortion function, so it should never increase after an iteration of K-means.



K-Means will always give the same results regardless of the initialization of the centroids.

Un-selected is correct



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.

Correct

This is the recommended method of initialization.