✓ **Congratulations! You passed!**

Next Item

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

1 / 1 point

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

Correct
You can use K-means to cluster the database entries, and each cluster will correspond to a different market segment.
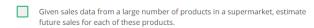
☑ Given sales data from a large number of products in a supermarket, figure out which products tend to form coherent groups (say are frequently purchased together) and thus should be put on the same shelf.

Correct
If you cluster the sales data with K-means, each cluster should correspond to coherent groups of items.

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

Un-selected is correct

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

Un-selected is correct

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

1 / 1 point

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

◉ $c^{(i)} = 3$

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

○ $c^{(i)} = 1$

○ $c^{(i)} = 2$

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

1 / 1 point

☐ Test on the cross-validation set.

Un-selected is correct

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

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

☐ Randomly initialize the cluster centroids.

Un-selected is correct

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

4. Suppose you have an unlabeled dataset $\{x^{(1)}, \ldots, x^{(m)}\}$. You run K-means with 50 different random

   1 / 1 point

   initializations, and obtain 50 different clusterings of the

   data. What is the recommended way for choosing which one of

   these 50 clusterings to use?

   ○ Manually examine the clusterings, and pick the best one.

   ⦿ Compute the distortion function $J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_k)$, 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.

   ○ Use the elbow method.

   ○ Plot the data and the cluster centroids, and pick the clustering that gives the most "coherent" cluster centroids.
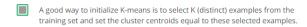
5. Which of the following statements are true? Select all that apply.

   1 / 1 point

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

   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.

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

   Un-selected is correct

   ☑ On every iteration of K-means, the cost function $J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_k)$ (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 decrese the cost / distortion function, so it should never increase after an iteration of K-means.