

1
point

1.

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.

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

- ☐ $c^{(i)} = 1$
- ☐ $c^{(i)}$ is not assigned
- ☒ $c^{(i)} = 2$

☐ $c^{(i)} = 3$

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?

- ☐ Test on the cross-validation set.
- ☐ Randomly initialize the cluster centroids.
- ☐ The cluster assignment step, where the parameters $c^{(i)}$ are updated.
- ☐ Move the cluster centroids, where the centroids μ_k are updated.

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.
- ☐ 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.

Unsupervised Learning

Quiz, 5 questions

1
point

5.

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

- ☐ 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.
- ☐ 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.
- ☐ Once an example has been assigned to a particular centroid, it will never be reassigned to another different centroid
- ☐ K-Means will always give the same results regardless of the initialization of the centroids.

if you stuck in some weird local minimam, the J can increase, you should use different initial condition and try again !

- ☒ I, **Jun-Chieh Wang**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account. Learn more about Coursera's Honor Code

Submit Quiz



Unsupervised Learning

Quiz, 5 questions