# **Unsupervised Learning**

Latest Submission Grade 80% 1. Question 1
For which of the following tasks might K-means clustering be a suitable algorithm? Select all that apply.
1 / 1 point
Given many emails, you want to determine if they are Spam or Non-Spam emails.
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
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.
2. Question 2
Suppose we have three cluster centroids $\mu1=[12]\mu_1=$
12 μ1=[12], μ2=[-30]\mu_2 =

```
-30 \mu2=[-30] and \mu3=[42]\mu_3=42 \mu3=[42]. Furthermore, we have a training example x(i)=[-12]x^{(i)}=-12 x(i)=[-12]. After a cluster assignment step, what will c(i)c^{(i)}c(i) be?
```

$$c(i)=2c^{(i)}=2c(i)=2$$

•

$$c(i)=1c^{(i)}=1c(i)=1$$

 $\circ$ 

$$c(i)=3c^{(i)}=3c(i)=3$$

 $\circ$ 

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

Correct

$$x(i)x^{(i)}x(i)$$
 is closest to  $\mu 1 \mu 1$ , so  $c(i)=1c^{(i)}=1c(i)=1$ 

3. Question 3

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

## 0 / 1 point

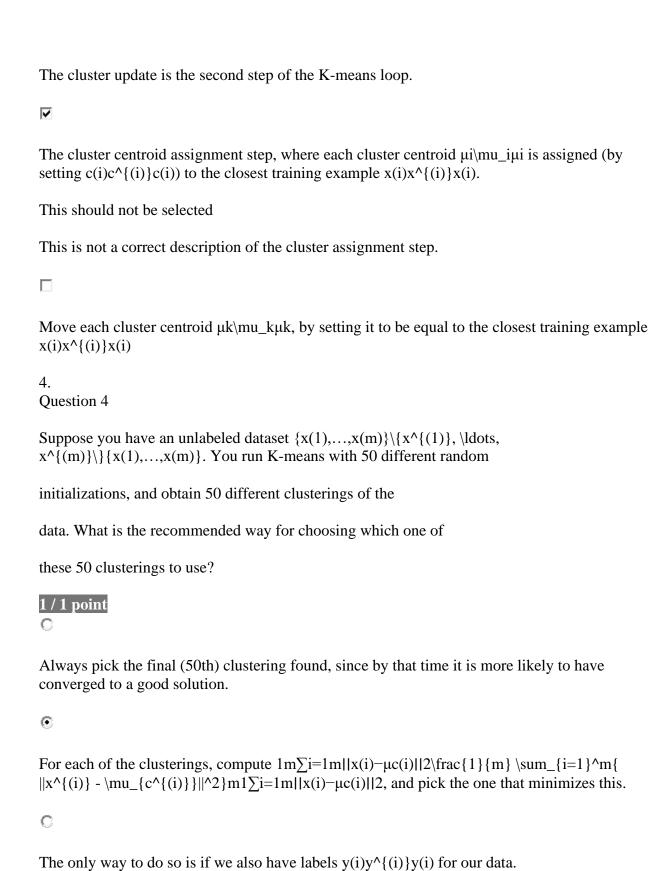
П

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

**V** 

Move the cluster centroids, where the centroids  $\mu k \mu k$  are updated.

Correct



O

The answer is ambiguous, and there is no good way of choosing.

### Correct

This function is the distortion function. Since a lower value for the distortion function implies a better clustering, you should choose the clustering with the smallest value for the distortion function.

5.

Question 5

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



 $\overline{\mathbf{v}}$ 

On every iteration of K-means, the cost function  $J(c(1),...,c(m),\mu 1,...,\mu k)J(c^{(1)})$ , \ldots,  $c^{(m)}$ , \mu\_1, \ldots,\mu\_k) $J(c(1),...,c(m),\mu 1,...,\mu k)$  (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.

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.

✓

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.