9/17/2019 Unsupervised Learning | Coursera

**Unsupervised Learning** Graded Quiz • 10 min

**Due Nov 11, 1:29 PM IST** 

### Clustering

### Review

- **Reading:** Lecture Slides 10 min
- **Quiz:** Unsupervised Learning 5 questions

#### Motivation

**Principal Component Analysis** 

**Applying PCA** 

**Review** 

**Congratulations! You passed!** TO PASS 80% of higher 10 MIN

**Keep Learning** 

GRADE 100%

# **Unsupervised Learning**

## **Unsupervised Learning**

LATEST SUBMISSION GRADE

100%

Submit your assignment

**DUE** Nov 11, 1:29 PM IST **ATTEMPTS** 3 every 8 hours

1/1 point

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

Receive grade

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

Grade View Feedback 100% We keep your highest score

Correct

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



Try again

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

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

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

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 1/1 point

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

- $\bigcirc \quad c^{(i)} = 3$
- $\bigcirc c^{(i)}=1$
- $\bigcirc \ c^{(i)}=2$
- $\bigcirc \ c^{(i)}$  is not assigned

 $x^{(i)}$  is closest to  $\mu_3$  , so  $c^{(i)}=3$ 

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.

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.

igwedge Move the cluster centroids, where the centroids  $\mu_k$  are updated.

Correct

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

4. Suppose you have an unlabeled dataset  $\{x^{(1)},\dots,x^{(m)}\}$ . 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?

- The answer is ambiguous, and there is no good way of choosing.
- The only way to do so is if we also have labels  $y^{(i)}$  for our data.
- Always pick the final (50th) clustering found, since by that time it is more likely to have converged to a good solution.
- $\bigcirc$  For each of the clusterings, compute  $\frac{1}{m}\sum_{i=1}^m||x^{(i)}-\mu_{c^{(i)}}||^2$ , and pick the one that minimizes this.