Question - 1
Question 1

SCORE: **5 points**

In which of these tasks K-means clustering will be more suitable to use?

- [x] Given a set of news articles, find out what are the main topics of these articles.
- [ ] Given historical weather records, predict if tomorrow's weather will be sunny or rainy.
- [ ] Given a set of emails, you need to predict if they are spam or non-spam emails.
- [x] Given a set of users bios on a website, figure out what types of group users exist.

Question - 2
Question 1

SCORE: **5 points**

In which of these tasks K-means clustering will be more suitable to use?

- [ ] Given a set of reviews, find out what are positive and negative.
- [x] Given wikipedia articles, determine the main tags of the articles.
- [ ] Given historical weather records, predict if tomorrow's weather will be sunny or rainy.
- [x] Given a set of locations to deliver food, determine the most suitable locations for the delivery route.

Question - 3
Question 2

SCORE: **5 points**

Suppose we have four cluster centroids: 1 - [2,2], 2 - [-1,3], 3 - [0,2] and 4 - [3,0], so we have a training example x_i = [7,7]. After a cluster assignment step, what will be the cluster of this training example?

- [x] 1
- [ ] 2
- [ ] 3
- [ ] 4

/

Question - 4
Question 2

SCORE: **5 points**

Suppose we have four cluster centroids: 1 - [0,1], 2 - [-1,2], 3 - [4,2] and 4 - [5,0], so we have a training example x_i = [2,2]. After a cluster assignment step, what will be the cluster of this training example?

- 1
- 2
- 3
- 4

Question - 5
Question 3

SCORE: **5 points**

Which of the following true statements?

- On every iteration of K-means algorithm, the cost function (the distortion function) should either stay the same or decrease, but it should not increase.
- Result of the K-Means algorithm will be the same regardless of the initialization of the centroids.
- A good way to initialize K-means is to select K examples from the training set and set the cluster centroids equal to these selected examples.
- Once a point has been assigned as a centroid, it will never be reassigned to another different centroid.

Question - 6
Question 3

SCORE: **5 points**

Which of the following true statements ?

- On every iteration of the K-means algorithm, the cost function (the distortion function) should increase, but not either decrease or stay the same.
- Result of the K-Means algorithm will not be the same regardless of the initialization of the centroids.
- A good way to initialize K-means is to select K examples from the training set and set the cluster centroids equal to these selected examples.
-

/

Once a point has been assigned as a centroid, it can be reassigned to another different centroid.

Question - 7
Question 4

SCORE: **5 points**

Which of the following is a reasonable way to select the number of principal components k ?

- Use the elbow method.
- Choose k to be the largest value so that at least 99% of the variance is retained.
- Choose k to be the smallest value so that at least 99% of the variance is retained.
- Choose k to be 99% of the number of input examples.

Question - 8
Question 5

SCORE: **5 points**

In what cases is it recommended to use PCA ?

- When we have too much data and we are going to use process-intensive algorithms
- When we need to decrease the number of features to feed into a learning algorithm.
- When you have a high variance of your data.
- When we need to tackle the curse of dimensionality.

Question - 9
Question 5

SCORE: **5 points**

In what cases it is recommended to use PCA?

- Data visualization: Reduce data to 2D, so it can be plotted.
- To increase the number of features to feed into a learning algorithm.
- Avoid overfitting by reducing the number of features.
- To reduce the dimension of your data for less memory or disk space.

Question - 10
Question 6

SCORE: **5 points**

/

Answer yes or no for the following statement about Bag of Words:

Does the word order matter?

- ( ) Yes
- (•) No

Question - 11

Question 6

SCORE: **5 points**

Answer yes or no for the following statement about Bag of Words:

Do long phrases give different input vectors?

- (•) Yes
- ( ) No

Question - 12

Question 6

SCORE: **5 points**

Answer yes or no for the following statement about Bag of Words:

Can we handle complex phrases? (Ex. "Chicago Bulls")

- ( ) Yes
- (•) No

Question - 13

Question 7

SCORE: **5 points**

Write the order of operations in Text learning.

- ( ) 1. Stemming 2. Stopwords removal 3. Bag of Words
- ( ) 1. Bag of Words 2. Stopwords removal 3. Stemming
- ( ) 1. Stopwords removal 2. Bag of Words 3. Stemming
- (•) 1. Stopwords removal 2. Stemming 3. Bag of Words

/