Started on	Friday, 27 April 2018, 1:18 PM
State	Finished
Completed on	Friday, 27 April 2018, 1:24 PM
Time taken	5 mins 40 secs
Grade	**9.0** out of 9.0 (**100**%)
Feedback	Well done!

Question 1

Correct

Mark 1.0 out of 1.0

Which of the following statements are true?

Select one or more:

- [x] a. Clustering algorithms are unsupervised learning methods. ✓
  classifications and regressions are supervised learning algorithms
- [ ] b. Clustering algorithms cannot be used to find an outlier.
- [x] c. Clustering is the process of grouping a set of data objects into multiple groups or clusters. ✓
- [x] d. A hierarchical clustering method can be classified as being either agglomerative or divisive. ✓ There are two ways of constructing hierarchical clusters.

Your answer is correct.

The correct answers are: Clustering algorithms are unsupervised learning methods., Clustering is the process of grouping a set of data objects into multiple groups or clusters., A hierarchical clustering method can be classified as being either agglomerative or divisive.

Correct

Marks for this submission: 1.0/1.0.

Question 2

Correct

Mark 1.0 out of 1.0

Consider dataset *D* consisting of 7 data points in two dimensional euclidean space as follows:

	x1	x2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

Our goal is to find 2 clusters using k-means algorithm.

What will be the cluster assignments after the first iteration given two initial cluster means (1.0, 1.0) and (5.0, 7.0)

TIE BREAKING RULE: if there is a point that has the same distance to both cluster means, then we assign that point to (1.0, 1.0).

Select one:

- a. (1, 2, 4, 6), (3, 5, 7)
- b. (1, 2), (3, 4, 5, 6, 7)
- c. (1, 2, 3), (4, 5, 6, 7) ✓
- d. (1, 5, 7), (2, 3, 4, 6)

Your answer is correct.

The correct answer is: (1, 2, 3), (4, 5, 6, 7)

Correct

Marks for this submission: 1.0/1.0.

Question 3

Correct

Mark 1.0 out of 1.0

What will be the clustering result after the second iteration?

Select one:

- a. (1, 2, 3), (4, 5, 6, 7)
- b. (1, 3, 4, 6), (2, 5, 7)
- c. (1, 2), (3, 4, 5, 6, 7) ✓
- d. (1, 4, 5), (2, 3, 6, 7)

Your answer is correct.

The correct answer is: (1, 2), (3, 4, 5, 6, 7)

Correct

Marks for this submission: 1.0/1.0.

Question 4

Correct

Mark 1.0 out of 1.0

Which of the following statements are true?

Select one or more:

- [x] a. The k-medoids method is more robust than k-means in the presence of noise and outliers. ✓
- [ ] b. K-means automatically determines the number of clusters from data.
- [x] c. K-means is a non-deterministic algorithm. ✓
- [ ] d. K-medoid is computationally efficient than K-means.

Your answer is correct.

The correct answers are: K-means is a non-deterministic algorithm., The k-medoids method is more robust than k-means in the presence of noise and outliers.

Correct

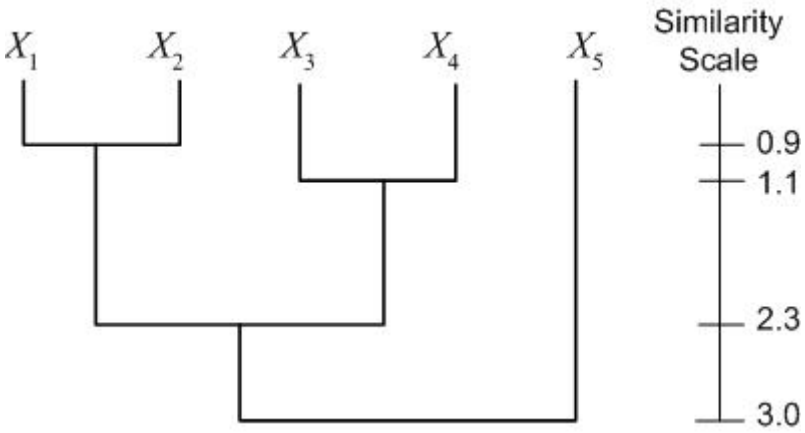
Marks for this submission: 1.0/1.0.

Question 5

Correct

Mark 1.0 out of 1.0

Assume that we obtain the following dendrogram over five items from a hierarchical clustering:



What will be the final clustering configuration if we cut the tree at similarity scale 2.7?

Select one:

- a. (X1), (X2), (X3), (X4), (X5)
- b. (X1, X2, X3, X4), (X5) ✓
- c. (X1), (X2, X3, X4, X5)
- d. (X1, X2), (X3, X4), (X5)
- e. (X1, X2, X3, X4, X5)

Your answer is correct.

The correct answer is: (X1, X2, X3, X4), (X5)

Correct

Marks for this submission: 1.0/1.0.

Question 6

Correct

Mark 1.0 out of 1.0

Assume that we obtain clusters as follows from some clustering algorithm:

C1 = (1, 3)

C2 = (5, 8, 12)

C3 = (18, 25)

C4 = (10, 15)

C5 = (2, 20)

Given these clusters, we'd like to further cluster the clusters using AGNES with euclidean distance.

Which two clusters will be merged if we use **single-link** distance metric to measure the distance between clusters?

Select one:

- a. C4, C5
- b. C2, C5
- c. C1, C5 ✓
- d. C1, C2

Your answer is correct.

The correct answer is: C1, C5

Correct

Marks for this submission: 1.0/1.0.

Question 7

Correct

Mark 1.0 out of 1.0

Which two clusters will be merged if we use **complete link** distance metric to measure the distance between clusters?

Select one:

- a. C1, C4
- b. C4, C5
- c. C3, C4
- d. C2, C4 ✓

Your answer is correct.

The correct answer is: C2, C4

Correct

Marks for this submission: 1.0/1.0.

Question 8

Correct

Mark 1.0 out of 1.0

Let $\epsilon = 0.5$ and *MinPts* = 3.

Given dataset $D = \{0.5, 0.9, 1.2, 1.6, 2.2, 2.4, 2.5\}$ consisting of 7 data points,

which of the following statements are true?

Select one or more:

- [ ] a. 0.9 and 2.2 are density-connected.
- [x] b. 0.5 is density-reachable from 1.2 ✓
- [ ] c. 1.2 is directly density-reachable from 0.5
- [x] d. 2.2 is directly density-reachable from 2.5 ✓

Your answer is correct.

The correct answers are: 2.2 is directly density-reachable from 2.5, 0.5 is density-reachable from 1.2

Correct

Marks for this submission: 1.0/1.0.

Question 9

Correct

Mark 1.0 out of 1.0

Which of the following statements is true?

Select one:

- ( ) a. Hopkins Statistic measures how well the clusters are separated.
- (•) b. Extrinsic evaluation methods can be applied when the ground truth categories are available. ✓
- ( ) c. BCubed precision and recall are intrinsic evaluation methods.
- ( ) d. The optimal number of clusters is square root of n/2 for a dataset of n points.

Your answer is correct.

The correct answer is: Extrinsic evaluation methods can be applied when the ground truth categories are available.

Correct

Marks for this submission: 1.0/1.0.