Unit 12 - Week 11

Assignment 11

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2020-04-15, 23:59 IST.

1) Select the correct option for a data set with 7 objects and an interval-scaled variable 'f 'we have the following measurements: *1 point*

f = (1, 2, 3, 4, 5, 8, 50)

containing one outlying value.

- Std deviation(std_f) is more affected by the outlier
- Mean absolute deviation (s_f) is more affected by the outlier
- Std deviation (std_f) is less affected by the outlier
- Std deviation (std_f) and mean absolute deviation (s_f) are having the same effect of the outlier

No, the answer is incorrect.
Score: 0

Accepted Answers:
Std deviation(std_f) is more affected by the outlier

2) Euclidean and Manhattan distances between the objects P, Q and R are- *1 point*

(1, 2, 3), (2, 1, 0)

- 3.32, 4 respectively
- 3.30, 5 respectively
- 5, 3.32 respectively
- 3.32, 5 respectively

No, the answer is incorrect.
Score: 0

Accepted Answers:
3.32, 5 respectively

3) Discriminant analysis is – *1 point*

- Unsupervised learning technique
- Supervised learning technique
- Both (a) and (b)
- None of these

No, the answer is incorrect.
Score: 0

Accepted Answers:
Supervised learning technique

4) Select the correct statement about the standardization in the following options- *1 point*

- Standardizing the data always gives inefficient result while making clusters
- Standardizing the data always beneficial during clustering analysis
- The variables having an absolute value may not efficient after standardization during clustering analysis
- Outliers can not be detected by standardized data

No, the answer is incorrect.
Score: 0

Accepted Answers:
The variables having an absolute value may not efficient after standardization during clustering analysis

5) Dissimilarity can be defined as – *1 point*

- How much certain objects differ from each other
- How much certain objects similar from each other
- dissimilarities are non-negative numbers d(i, j) that are small (close to zero) when i and j are "near" to each other and that become large when i and j are very different
- both (a) and (c)

No, the answer is incorrect.
Score: 0

Accepted Answers:
both (a) and (c)

6) Select incorrect statement about similarity – *1 point*

- Similarity s(i, j) typically takes on values between 0 and 1, where 0 means that i and j are not similar at all and 1 reflects the maximal similarity
- S(i, i) = 1
- Similarities between variables can be defined using Pearson or the Spearman correlation coefficient
- It is not necessary to transform the similarities into dissimilarities while data consists of the similarity matrix

No, the answer is incorrect.
Score: 0

Accepted Answers:
It is not necessary to transform the similarities into dissimilarities while data consists of the similarity matrix

7) Formula for dissimilarity computation between two objects for categorical variable is – *1 point*

Here p is a categorical variable and m denotes number of matches

- D(i, j) = p-m / p
- D(i, j) = p-m / m
- D(i, j) = m-p / p
- D(i, j) = m-p / m

No, the answer is incorrect.
Score: 0

Accepted Answers:
D(i, j) = p-m / p

8) Which library is used for calculating distance measure in clustering using python – *0 points*

- distance_matrix
- scipy.spatial
- scipy_spatial
- distance.matrix

No, the answer is incorrect.
Score: 0

Accepted Answers:
distance_matrix

9) Selects some facts about k-means clustering – *1 point*

- It comes under the partitioning method
- Number of clusters are predefined for this method
- Cluster similarity is measured in regard to the mean value of the objects in a cluster
- All of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
All of the above

10) Selects some facts about hierarchical clustering – *1 point*

- A hierarchical method comes under either agglomerative or divisive algorithms
- Hierarchical methods suffer from the fact that once a step (merge or split) is done, it can never be undone
- A tree structure called a dendrogram is commonly used to represent the process of hierarchical clustering
- All of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
All of the above