MACHINE-LEARNING-WORKSHEET

1.Which of the following is an application of clustering?

b. Market trend prediction

2. On which data type, we cannot perform cluster analysis?

a. Time series data

3. Netflix’s movie recommendation system uses

b. Unsupervised learning

4. The final output of Hierarchical clustering is

d. All of the above

5. Which of the step is not required for K-means clustering?

a. a distance metric

6. Which is the following is wrong?

c. k-nearest neighbor is same as k-means

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

a. 1 and 2

8. Which of the following are true?

a. 1 only

9. In the figure above, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed?

a. 2

10. For which of the following tasks might clustering be a suitable approach?

b. Given a database of information about your users, automatically group them into different market segments.

11. Given, six points with the following attributes:

D.

12. Given, six points with the following attributes:

A.

13. What is the importance of clustering?

Clustering is useful for exploring data. If there are many cases and no obvious groupings, clustering algorithms can be used to find natural groupings. Clustering can also serve as a useful data-preprocessing step to identify homogeneous groups on which to build supervised models. It is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups. A good clustering algorithm is able to identity clusters irrespective of their shapes. It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields.

14. How do you cluster a profile?

Cluster profiling involves generating descriptions of the clusters with reference to the input variables you used for the cluster analysis. Profiling acts as a class descriptor for the clusters and will help you to ‘tell a story’ so that you can understand this information and use it across your business. The goal of cluster profiling or segmentation is to achieve a clear description of who your customer is in each cluster, how you can meet their needs, and what you can do to achieve a return on your investment.

To profile, interpret and understand your clusters, you can follow the steps below:

* Graphically represent your clusters according to your input variables.
* Score your clusters in a table so that you can measure and compare them on each input variable with regards to numerical or descriptive values.
* Variables should be described in a type of ‘story’ about the category or customer base. This will help buyers and marketers to use this information strategically with an in-depth understanding of the differences between each cluster and which variables define the groupings. This step sets cluster profiling apart from traditional segmentation.

15. How can I improve my clustering performance?

Clustering is an unsupervised machine learning methodology that aims to partition data into distinct groups, or clusters. There are a few different forms including hierarchical, density, and similarity based. Each have a few different algorithms associated with it as well. The guiding principle of similarity-based clustering is that similar objects are within the same cluster and dissimilar objects are in different clusters. This is not different than the goal of most conventional clustering algorithms.

K-means clustering algorithm can be significantly improved by using a better initialization technique, and by repeating (re-starting) the algorithm. When the data has overlapping clusters, k-means can improve the results of the initialization technique.

* Number of clusters: Linear dependency
* Cluster overlap: Overlap is good
* Dimension: No direct effect