**WORKSHEET 3 MACHINE LEARNING**

**Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.**

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

**Ans.** d. All of the above

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

**Ans.** D) None

1. **Netflix’s movie recommendation system uses**

**Ans.** C) Reinforcement learning and Unsupervised learning

1. **The final output of Hierarchical clustering is a**

**Ans.** C) The tree represents how close the data points are to each other

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

**Ans.** D) k-nearest neighbor is same as k-means

1. **Which is the following is wrong?**

**Ans.**  D) None

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

**i. Single-link**

**ii. Complete-link**

**iii. Average-link Options:**

**Ans.** D) 1, 2 and 3 .

1. **Which of the following is true?**

**i. Clustering analysis is negatively affected by the multicollinearity of features**

**ii. Clustering analysis is negatively affected by heteroscedasticity Options:**

**Ans.** A) 1 only

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

**Ans.** A) 2

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

**Ans.** B) Given a database of information about your users, automatically group them into different market segments.

1. **Given, six points with the following attributes: Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering**:

**Ans.** A)

1. **Given, six points with the following attributes:**

Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.

**Ans. A)**

**Q13 to Q14 are subjective answers type questions, Answers them in their own words briefly.**

1. **What is the importance of clustering?**

**Ans.** Clustering or cluster analysis is a machine learning technique, which groups the unlabelled dataset. They can cluster different customer types into one group based on different factors, such as purchasing patterns. The factors analyzed through clustering can have a big impact on sales and customer satisfaction, making it an invaluable tool to boost revenue, cut costs, or sometimes even both.

1. **How can I improve my clustering performance?**

**Ans.**  Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance. Surprisingly for some cases, high clustering performance can be achieved by simply performing K-means clustering on the ICA components after PCA dimension reduction on the input data.