

**Unsupervised Machine Learning**

**Assignment**

[support@intellipaat.com](mailto:support@intellipaat.com)

+91-7022374614

US: 1-800-216-8930 (Toll Free)

**Instructions for the Assignment**

1. Use a single file in Jupyter Notebook or Goggle collab for the entire assignment.
2. Download the Assignment file (File> Download > Download as .ipynb), compress it in a Zip folder and submit through LMS only.
3. Submit the Assignment by 13th February 2022 without fail.
4. This Assignment is an Exam kind and is considered for the evaluation and certification. No support is provided by our technical team.
5. If you have any doubts how to work on this assignment (No technical support), please drop an email to [support@intellipaat.com](mailto:support@intellipaat.com)

**FINAL ASSIGNMENT**

1. Match the terms in Group A with the relevant terms in Group B

**Group A                                           Group B**

1. k-means                                     1) unsupervised learning algorithm
2. knn                                            2) k is no. of clusters
3. logistic regression                     3) k is no. of neighbours
4. clustering                                  4) logit function
5. Each centroid in K- means algorithm defines one

1. cluster
2. data point
3. two clusters
4. None of the above
5. The method / metric which is NOT useful to determine the optimal number of clusters in unsupervised clustering algorithms is
6. Dendogram
7. Elbow method
8. Scatter plot
9. None of the above
10. In K-means algorithm, what is the most commonly used distance metric to calculate distance between centroid of each cluster and data points?

1. Chebyshev distance
2. Manhattan
3. Euclidean
4. None of the above
5. Which of the following statement is not correct about k-means?
6. Accuracy of cluster are improved by scaling of attributes.
7. K-means clusters are affected by outliers.
8. K-Means clustering is **NOT** influenced by initial centroids which are called cluster seeds
9. Number of clusters to be built is typically an user input and it impacts the way clusters are created