



भारतीय सूचना प्रौद्योगिकी संस्थान गुवाहाटी
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY GUWAHATI

Data Analytics Lab, M.Tech 3rd Semester

Instructions

1. Upload all your codes to Github.
2. You will be called randomly to explain the code based on which marks/grade will be assigned.

Assignment -6

1. Download any two datasets of your choice, read it and perform the following tasks
 - (a) Visualise the data.
 - (b) Perform K-Means Clustering selecting the best value of k and taking Euclidean distance as similarity measure. Check your algorithm with the following three conditions , visualise the clusters and note the difference
 - i. Maximum number of iterations
 - ii. Cluster centroid remains unchanged
 - iii. Highest quality of cluster is reached.
 - (c) Repeat the Q.2 taking Manhattan distance as similarity measure and note the difference between the clusters as compared to that found in Q. 2.