

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 -5

- 1. Perform the following tasks:
 - (a) Import mtcars dataset.
 - (b) Visualize your data using scatter plots
 - (c) Compute the Correlation between mpg and wt variables.
- 2. Perform the following:
 - (a) Define a normalized vector P.
 - (b) Define a normalized vector Q.
 - (c) Combine P and Q as matrix object. Please note that when defining a matrix from vectors, the vectors should be combined as rows.
 - (d) Compute the Euclidean Distance with default parameters
- 3. Compute Manhattan distance and cosine similarity after performing (a),(b),(c) steps of O.2.
- 4. (a) Import the Iris dataset (https://archive.ics.uci.edu/ml/datasets/iris).
 - (b) As you may know, this dataset contains three kinds of flowers: Iris-Setosa, Iris-Versicolor, and Iris-Virginica, having the following four features: sepal length, sepal width, petal length, petal width. choose only two features: petal length, petal width, and plot the data points in a 2-D space where the x-axis and the y-axis represent the petal length and the petal width, respectively.
 - (c) Compute the similarity measure between Iris-Setosa and Iris-Versicolor; and Iris-Versicolor and Iris-Virginica; and Iris-Setosa and Iris-Virginica considering only two features petal length and petal width