School of Computer Science Engineering and Technology

Course- BCA
Course Code- BCA301L
Year- 2022
Date- 08-08-2022

Type- Core
Course Name-AIML
Semester- ODD
Batch- 5th Sem

1 - Lab Assignment

Objective: To learn basic pre-processing operations involved in a machine learning task.

1. Go to UCI machine learning repository and download the wine-quality dataset (red and/or white) from the link https://archive.ics.uci.edu/ml/datasets/Wine+Quality . (5)

2. Read the dataset in the form of a NumPy matrix and store it in a variable named XY. (10)

3. Perform the following operations on XY that are generally required in a machine learning task (30)

i. Print the shape of matrix XY.
ii. Slicing the matrix: From the matrix XY, create a new variable:
   a. Y, which contains the last column (quality of wine) of XY. Print its shape.
   b. X, all the other columns except last from XY. Print its shape.
   c. Shuffle and take the Transpose of matrix X and print its shape.
   d. Take the maximum and minimum values across the rows as well as columns in X.
   e. Print the number of values which are equal to 5 in Y.

4. Compute the following statistical values using NumPy in-built functions wherever possible. (25)

i. Mean for all columns in X.
ii. Mode of the last column, Y (i.e. quality of wine)
iii. Standard deviation for all columns in X.

5. Plot using Matplotlib library and show different components like the title, axis labels, and legend. (20)

i. A bar graph of unique values contained in variable Y.
ii. Histograms for different columns in X to show the distribution of data.

Following operations from NumPy library functions should also be revised/practiced for acquiring better implementation skills in yet to come labs. You can take your own data for this.

Array Creation: array(), identity(), zeros(), full, copy(), linspace()

Array manipulation: matmul(), multiply(), dot(), append(), concatenate(), insert(), unique(), delete(), reshape(), stack, vstack, hstack etc.