

21BM681 Machine Learning and Embedded Programming Lab

Exercise 3

Date:16/08/2023

Statistical Measures and Data Visualization

1. For the dataset given,
 - Compute the central tendency measures. (Employ user-defined function for each measure)
 - Compute the measures of dispersion. (Employ user-defined function for each measure)
 - Obtain the scatter plots and boxplots for the given dataset.
 - Give your inferences

2. Create your own dataset.

Specification:

- Four features (f_1, f_2, f_3)
- 100 samples
- Three classes (A, B, C) – Class A 30 Samples; Class B 25 Samples; Class C 45 Samples

and apply question 1 to this dataset. Datatype should be numeric and integers. Give your inferences

Additional Exercises

Split the data set into train and test set for the following train-test ratio

- 90:10
- 80:20
- 70:30
- 60:40

(i) For the given dataset

(ii) For the created dataset

Print the output and analyse the distribution of data in train and test sets

Note: Libraries to be familiarized: Numpy, Pandas, Matplotlib, Sci-kit Learn, Seaborn