

Data Pre-processing

Deadline: 12 Sept 2021

Assignment1

Q1. Download the classification artificial data from your LMS Account and perform the following operations on this dataset:

- a) Utilizing sklearn to import the libraries needed for Data pre-processing.
- b) Importing the Dataset from the above link.
- c) Consider the concerned or related function to handle the missing data in the artificial datasets.
- d) Calculate the following statistical parameters over the artificial data such as feature mean, mode, median, variance, standard deviation, normal distribution, uniform distribution.
- e) Conduct the data visualizations library functions using matplotlib or seaborn libraries such as density plot, bar diagram, boxplot, histograms, and line chart.
- f) Find the outliers from the artificial data for each feature in the Data visuals.
- g) Change the datatype of the features using astype, to_numeric functions from float to int or vice versa.

Q2. Conduct Data pre-processing on real world classification datasets named mushroom.csv file. (You can download the dataset from <https://archive.ics.uci.edu/ml/datasets/mushroom/> .) Perform the following operations on this dataset:

- a) Utilizing sklearn to import the libraries needed for Data pre-processing.
- b) Importing the Dataset from the above link and convert into the .csv file.
- c) Consider the concerned or related function to handle the missing data in the artificial datasets.
- d) Calculate the following statistical parameters over the artificial data such as feature mean, mode, median, variance, standard deviation, normal distribution, uniform distribution.
- e) Conduct the data visualizations library functions using matplotlib or seaborn libraries such as density plot, bar diagram, boxplot, histograms, and line chart.
- f) Find the outliers from the artificial data for each feature in the Data visuals.
- g) Change the datatype of the features using astype, to_numeric functions from float to int or vice versa.