Assignment 1

Advanced Machine Learning (CS566)

Department of CSE, IIT Patna

(Read all the instructions carefully and adhere to them.)

Date:13-Feb-2020 Deadline:-20-Feb-2020

Instructions:

1. Markings will be based on the correctness and soundness of the outputs. Marks will be deducted in case of plagiarism.
2. Proper indentation and appropriate comments (if necessary) are mandatory.
3. You should zip all the required files and name the zip file as roll_no_of_all_group_members .zip , eg. 1501cs11_1201cs03_1621cs05.zip.
4. Upload your assignment (the zip file) in the following link:
   https://www.dropbox.com/request/jxTo873seo8SZUJtYk5Q

For any queries regarding this assignment contact:

Apoorva(apoorva_1921cs19@iitp.ac.in)

Questions:

1. Implement baseline prediction for time series forecasting known as Naive forecast, or Persistence.
2. Use summary statistics and plots of the data to learn about the structure of the prediction problem:
   a) Line Plot
   b) Density Plot
   c) Box and Whisker Plot
3. Develop the Autoregressive Integrated Moving Average (ARIMA) model for the problem.

Dataset:-

i. Given the **"*TimeSeries.csv*"** dataset for the experiment .

ii. The dataset provides the annual water usage of someplace over a span of 79 years. The values are given in units of liters per capita per day, and there are 79 observations.

Evaluation:-

1. **Validation Dataset:-** This final decade of data will be used to validate the final model.
2. **Model Evaluation:-** Model evaluation will only be performed on the observations from 1885 to 1953.
   The model evaluation will involve two elements:

a)Performance Measure- Evaluate the performance of predictions using the root mean squared error (RMSE).

b)Test Strategy- Candidate models will be evaluated using walk-forward validation.