ME 592X: Data Analytics and Machine Learning for Cyber-Physical Systems

Homework 1

Homework Assigned on January 22, 2019 Homework Due on: January 31, 2019
One submission per group as discussed in the class

This homework is about getting everything setup for the theme groups, starting to work together and create a structure for the rest of the semester and finally some brush up of python concepts which might be minimum necessary for the course. So, primarily this homework is about the preliminaries for this course.

1 Cluster usage & Data storage

The usage of cluster is covered extensively in class and the resources are also provided in the canvas code development resources page. There is no mandate to use the cluster but it is expected that the data is finally stored in the HPC Class Cluster at `/ptmp/ME592_2019_GRPNAME`. Replace GRPNAME with the theme group name. The reason to store it there is for access to everyone and for us to reproduce the results for evaluations. So, it is MUST to store the data finally in the above mentioned folder.

2 Git Repository & Code Management

All the homeworks would have to be submitted using git commits containing all the codes. Repositories for each group is already created. By replying to the thread of bitbucket repositories with usernames, we will provide you access to the git repository for writing and editing it. Clone the repository. Each group would be having one git repository. The homework would be finally evaluated by looking at the latest pushed code.

3 Simple Programming & Exploratory Analytics

The following problems must be coded using specified tool only in order to get everyone to use python.

3.1 Images

Task:

1. Load sudoku-original.png image in grayscale mode.
2. Plot the histogram of intensities.

3.2 Time Series

Data: Experimental data used to analyze appliances energy use in an energy efficient building. The data set is collected at a frequency of 10 min for about 4.5 months. The house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. Each wireless node transmitted the temperature and humidity conditions at a period of around 3.3 min. Then, the wireless data was averaged for 10 minute periods.

The energy data was logged every 10 minutes with m-bus energy meters. Weather from the nearest airport weather station (Chievres Airport, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru), and merged together with the experimental data sets using the date and time column. Random variable is included in the data set for testing the regression models and to filter out non predictive attributes (parameters). This data is adopted from the UCI machine learning repositories [1] and several aspects of the data were analyzed [2]. Our motivation is to explore some aspects of this time series data.

Task: Load the data (energydata_complete.csv) and perform the following analysis.

1. Plot the appliances energy consumption for whole period and a closer look at any one week of consumption.
2. Plot heatmap of hourly consumption of appliances for a week. An example heatmap looks like Figure 1.
3. Plot the histogram of energy consumption of appliances.
4. Construct a feature variable NSM (no. of seconds from midnight) and plot energy consumption vs. NSM.
5. Plot appliances energy consumption vs. Press_mm_Hg.
6. It is observed that the major contributing factors for the energy consumption among all other features is NSM and Press_mm_Hg. Comment on it.

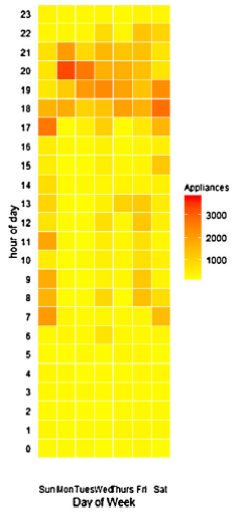

Figure 1: Example heatmap of hourly energy consumption of appliances over a week

3.3 Multi-variate

Data: The NASA data set comprises different sizes of NACA 0012 airfoils at various wind tunnel speeds and angles of attack. The span of the airfoil and the observer position were the same in all of the experiments. This problem has the following inputs:

1. Frequency, in Hz.
2. Angle of attack, in degrees.
3. Chord length, in meters.

4. Free-stream velocity, in meters per second.
5. Suction side displacement thickness, in meters.

The only output is Scaled sound pressure level, in decibels.

Task: Load the data and Compute the following descriptive statistics of the data:

1. Mean
2. Variance (or Standard Deviation)
3. Median
4. Kurtosis
5. Skewness
6. Range

Attachments

1. sudoku-orginal.png
2. energydata-complete.csv
3. airfoil_self_noise.dat

Expected Outcome

Final code committed for each theme group with results for Section 2 should be pushed before the deadline.

References

1. https://github.com/LuisM78/Appliances-energy-prediction-data

2. Data driven prediction models of energy use of appliances in a low-energy house. Luis M. Candanedo, Véronique Feldheim, Dominique Deramaix. Energy and Buildings, Volume 140, 1 April 2017, Pages 81-97, ISSN 0378-7788

3. https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise

4. T.F. Brooks, D.S. Pope, and A.M. Marcolini. Airfoil self-noise and prediction. Technical report, NASA RP-1218, July 1989.