



Recitation 3

Data Analytics (18-899)

Friday 18 Feb, 2022



Assignment Objectives

- Explore and visualize time series data
- Investigate variability in time series data over different time periods
- Perform statistical hypothesis test
- Evaluate model performance based on benchmarking



Question 1 & 2

- Download intraday 15-minute energy demand data from EirGrid for 2014
- Create a datetime column by concatenating the Date and Time columns
- Using linear interpolation to fix daylight saving issues
- Estimate the autocorrelation coefficients for 10 days
- EirGrid energy data is sampled every 15 minutes
- Lags for 10 days: $10 \times 24 \times 4$



Question 3

- Create a time of year variable that ranges between 0 and 1
- Show how the demand varies over the course of the year using a graphic.



Question 4

- Calculate the average demand for each month of the year
- Provide a bar graph of the monthly average demand



Question 5

- Calculate the average demand for each hour of the day for the whole year
- Provide a bar graph of the hourly average demand or daily demand profile



Question 6

- Calculate the average demand for each day of the week for the whole year
- Provide a bar graph of the daily average demand



Question 7

- For each hour of the day calculate the average daily demand
- Get daily demand profiles for the 7 days
- Plot the variations of the daily demand profiles over 24 hours



Question 8

- Separate the energy demand for weekends from weekdays
- Perform a statistical hypothesis test to determine whether there is a statistically significant difference between energy demand during the weekend and during the weekday
- Use a significance level $\alpha = 0.05$

Question 9

- Divide the data into two halves and use the second half for evaluation purposes
- Use the benchmark forecasting approach also known as persistence to forecast the energy demand for the second half of the data

$$\mathbf{y_hat(t+k) = y(t)}$$

- Calculate the MAE between the predicted energy demand and the actual energy demand of the second half of the data for the forecast horizons with lead times up to one day ahead
- Plot MAE against time leads and interpret the plot

Question 10

- Repeat the same process as 9 but instead of using MAE use MAPE

$$MAE = \frac{1}{n} \sum_{t=1}^n |\bar{y} - y|$$

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{\bar{y} - y}{y} \right|$$

Submission Files (MATLAB/Python)



- Single MATLAB code file (.m) - **AndrewID-Name-DAassignment3.m**
OR
- Single PYTHON code file (.ipynb) - **AndrewID-Name-DAassignment3.ipynb**
- Assignment report(.pdf) - **AndrewID-Name-DAassignment3.pdf**
- Data files (as given)

Submission process:

- Put code files and data files in a **single folder**
- The name of the folder should be your **Andrew ID** and zip this folder
- Submit the the zipped file (**.zip**) and the assignment report to CANVAS:
 - After attaching the zipped file, click on "Add Another File" from the assignment submission page
 - Attach your report



Q&A