Data Analytics

Course: 18-899

Recitation 4

Question 1

- Download intraday 15-minute energy demand data
- Save data as CSV file
- Create a datetime
- Use linear interpolation to Fix day-light saving issues
- Plot and carefully label the time series of energy demand during 2014.

Question 2

- Estimate the autocorrelation coefficients for 10 days
- EirGrid energy data is sampled every 15 minutes
- Lags for 10 days: 10 x 24 x 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

Average demand for every month

- Loop through data to calculate the average demand for each month of the year

for every month of the year

  get average demand

  store the average demands for the 12 months of the year in an array

- Provide a bar graph of the monthly average demand

Question 5

Daily demand profile

- Loop through data to calculate the demand for each hour

for every hour of the day

  get average demand

  store the average demands for the 24 hours of the day in an array

- Provide a bar graph of the average demand for each hour

Question 6

Day of week

- Loop through data to calculate the demand for each day

for every day of the week

  get average demand

  store the average demands for the 7 days of the week in an array

- Provide a bar graph of the average daily 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
- Use statistical hypothesis test to test whether there is statistically difference between the weekend and weekday energy demand.
- Select a significance level (α), a probability threshold below which the null hypothesis will be rejected.
- Common values are 5% and 1%.
- Perform a statistical hypothesis test to reach a conclusion

Question 9

- Divide your data into two dataset(training and testing)
  → hint: using indices is better for our case.
- Benchmark forecasting a.k.a persistence
  → $\bar{y}(t+k|t) = y(t)$ seen as $\bar{y}(t) = y(t-k)$ in codes where t start from the testing index sets to the end
- Calculate the MAE between the predicted and actual in test set
  → mae($\bar{y}_{test}$, y_{test})
- Do for k up to 1 day
- Plot MAE against time leads
- Interpret the plot

Question 10

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

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

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

Questions