

Data Analytics Course: 18-899

Spring 2019
Carnegie Mellon University

Question 1

- Read the file
- Generate dates/timestamps
- Plot the wind generation against the dates
- Is there any there evidence of intra annual seasonality

Question 2

- Calculate the change in wind generation over time as a percentage of the maximum generation
- Plot it against time
- Is there any evidence of seasonality?

Question 3

- Ramps $\rightarrow r(t,d) = 100 * [x(t+d)x(t)] / \max(x)$
- Calculate the ramps where $d = 1$
- Separate them into positive and negative ramps
- Normalize and sort them
- Plot them
- Similarly plot the normal distribution

Question 4

- Challenge: balancing supply and demand
- investigate the variability in wind generation over different timescales
- Timescales: 1h, 2h, ..., 24h
- Use the percentile analysis on the ramps
- Plot the results
- What did you learn

Question 4 - continued

- Some hints:
 - Calculate the ramps for a timescale eg: $d=1$, $d=2$ or $d=4$
 - Calculate the percentile of the ramps distribution at 1%, 5%, 95%, 99% functions: `prctile(matlab)`, `numpy.percentile(python)`
 - Repeat the process for all timescales
 - Plot all the four percentiles as a function of the timescale

Question 5

- Calculate the autocorrelation
- Remember you still have Nan values
- Hint: xcorr, google the nanautocorr function somebody implemented it , or google any other new function
- Comment on the structure of the autocorrelation function

Question 6

- Calculate Autocorrelation of change in wind generation for lags over 10 days
 - Calculate the change in wind generation
 - Calculate the autocorrelation with lags of 10days(240 hours)
- Plot it
- horizontal lines to detect statistically significance values ($p < 0.05$)
 - Corresponding value can be calculated from the normal distribution
 - plot it for every value
 - question on this post on piazza or skype us
- Is there any evidence of diurnal seasonality?
- Might it be more appropriate to model the change in wind generation than the wind generation?

Question 7

- Remember: You might still be **having NaN** values, use function that handle them
- The variance ratio test will be used to investigate the structure of the wind generation timeseries
- Hint: vratiotest
- What are the returns of the function, doc it
- Can the null hypothesis of a random walk be rejected?
- Test mean reversion
- Is there evidence of either mean-reversion or mean aversion?

Question 8

- For each n
- Calculate the simple moving average(function: `tsmovavg`)
- Calculate the mean absolute error between the simple moving average and the real wind power
- For which n , do you obtain the minimum error
- Is there a simple benchmark that improves on the persistence benchmark?

Question 9

- For each n
- N ranges from 1 to 24 (one hour to one day)
- Calculate the persistence of $n \rightarrow X_{\text{predicted}}(t) = X(t-n)$
- Calculate the mean absolute error between the simple moving average and the real wind power
- Plot MAE as a percentage of the maximum generation for the persistence benchmark.

Question 10

- doc/google the ARIMA model
- Find parameters that it takes
- Add the model to your environment, if you don't have it already
- Loop through a range of parameters to find the optimal parameters
 - pass the parameter to the arima model
 - estimate
 - calculate the AIC and BIC from the estimation
 - find if the current value improve(less) on the previously selected
- What are the optimal parameters?