DSC 275/475: Time Series Analysis and Forecasting

HW #2 (Total points: 40 for students in DSC 275;

50 points for students in DSC 475)

<u>Instructions:</u>

- For each of the questions, please submit a written response that also INCLUDES the results from running the code. In addition, please attach the code. You are welcome to use Jupyter notebook (Python), Rmarkdown (R), or LiveScript (MATLAB) to combine the code and written text in one file, but the notebooks <u>must</u> include the code output. Please attach the original notebook file and its PDF version.
- Please note in this HW, Q2 is only for students in DSC 475. It carries 10 points. Hence, the maximum score for DSC 275 students would be 40/50. This will be weighted in the final grade calculation accordingly. *Unfortunately, no extra credit for undergraduate students for attempting this problem in this homework.*
- For Q1 and Q3, please *do not* directly apply smoothing window functions from native Python or other Time-series libraries/packages.

1. Consider the measurement data provided in the file: *Measurement_Q1.xls*. **(15 pts)**
   (For Q1(a) and Q1(b) below, please do not use an external package/library to implement the moving average window)

   a. **(6 pts)** Use a 11-period *simple moving average* to smooth the data. Plot the result.
   b. **(6 pts)** Repeat the procedure with a 5-period simple moving average. Plot the result.
   c. **(3 pts)** What is the effect of changing the span (length) of the simple moving average window? Explain briefly.

2. (<u>FOR GRADUATE STUDENTS – DSC 475 ONLY</u>) **(10 pts)**

   a. **(3 pts)** Consider the N-span simple moving average applied to data that are uncorrelated with mean μ and variance σ^2. Show that the variance of the moving average is *Var(M_t)* = σ^2/N.
   b. **(7 pts)** Consider an N-span moving average where each observation is weighted by a constant, say, $a_j > 0$. Therefore, the weighted moving average at the end of period T is, $M_T^w = \sum_{t=T-N+1}^{t=T} a_{T+1-t} y_t$. The variance of the original time series y_t is σ^2. Show that *Var(M_T)* = $\sigma^2 \sum_{j=1}^{N} {a_j}^2$.

3. The file: *Yield_Data.xls* presents data on the hourly yield from a chemical process.

   a. **(10 pts)** Use simple (first order) exponential smoothing with λ = 0.2 to smooth the data. Plot the result.
   b. **(5 pts)** Change the smoothing constant (λ) to λ = 0.8. Smooth the data with this new value of λ. Plot the result.

c. **(10 pts)** Compute the Mean Square Difference between the original data and the smoothed data for $\lambda = 0.2$ and $\lambda = 0.8$. Which smoothing constant produced a lower error?