ECE 4750/6750 + CS 4501/6501: Digital Signal Processing **Due Tu 4/30**

Homework 5: Time Series Prediction with LTI Systems and Neural Networks

Instructions: Submit a single Jupyter notebook (.ipynb) of your work to Canvas by 11:59pm on the due date. All code should be written in Python. **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

You may discuss the concepts with your classmates, but write up the answers entirely on your own. Do not look at another student's answers, do not use answers from the internet or other sources, and do not show your answers to anyone. **Cite any sources you used outside of the class material (webpages, etc.), and list any fellow students with whom you discussed the homework concepts.**

1. Download the data `cho_weather.csv`, which contains hourly weather at Charlottesville Albemarle Airport (CHO) since 2019 [1]. The columns are:

`tmpf`	Temperature (F)
`dwpf`	Dew Point (F)
`relh`	Relative Humidity (%)
`drct`	Wind Direction (deg)
`sped`	Wind Speed (MPH)
`mslp`	Sea Level Pressure (mb)
`p01i`	Precipitation (in)

   Clean the data by filling in missing values with the previous valid value in that column. Missing values are marked with an 'M'. The letter 'T' indicates a trace amount of precipitation. Set these values to 0.

2. Let's use LTI systems to predict the weather. Take the exponential moving average (EMA) system that we discussed in class:

$$y[n] = (1 - g)x[n] + gy[n - 1].$$

   Given a time series $x[n]$ for $n = 0, \dots, L - 1$, we can use this as a prediction model for the next timepoint $x[n + 1]$ (that our model has not seen yet) by taking the last output, $y[n]$, as our prediction for $x[n + 1]$. Implement this exponential averaging system and test it on the provided data of hourly temperatures. Do the following:

   (a) Make predictions for $x[n]$ (using $y[n - 1]$) for $n = 1, \dots, L - 1$. Plot these predictions over a plot of the original data. Do this three times, with gain parameters $g = 0.00.25, 0.5, 0.75$. What difference do you see with the four different parameters?

   (b) Compute the mean absolute error (MAE) of the four different models ($g = 0.0, 0.25, 0.5, 0.75$). The MAE is

$$MAE = \frac{1}{L - 1} \sum_{i=1}^{L-1} |x[n] - y[n - 1]|.$$

   Which choice of g gave the best prediction (lowest MAE)?

[1] Weather data downloaded from `https://mesonet.agron.iastate.edu/request/download.phtml`.

(c) Because this is hourly data, our model is really only predicting one hour into the future, which isn't so hard! Let's predict further into the future. Repeat your exponential moving average model with a delay of 24 (instead of 1). Now predict $x[n+24]$ using $y[n]$. Plot the predictions over the original data again for $g = 0.0, 0.25, 0.5, 0.75$. Report the MAE for each model.

(d) Now try running your EMA system with a delay of 8 hours to predict 8 hours ahead. You don't need to plot, just report the MAE for the same four gain values. What do you notice about how this model performs compared to predicting a full 24 hours ahead? Give a conjecture for why you think the performance difference is they way it is.

3. **Coming soon, look for an announcement!** This will involve predicting the weather using a neural network (RNN or LSTM).