12-752: Data-Driven Building Energy Management Assignment#3

Due: Sunday December 3rd by 11:59pm on Canvas

November 22, 2017

You should implement your work in a Jupyter Notebook and it should contain proper documentation and be suitable to serve as a professional report (e.g., please be diligent in creating professional looking graphs including axis labels and units, as well as writing in a clear and technical language).

You will submit a .ipynb file, as well as a PDF version of the same file, through Canvas. Your notebook (and the PDF) should containing all of your solutions and comments to the tasks below.

1 Piece-Wise Multiple Linear Regression

For the first part of this assignment we will work with Paper #2 (Addy et al., 2015). Specifically, we will fit the same piece-wise linear model that is presented in that paper (Equation 1) to the data we used in Assignment #1 (though you can pick a different meter) and then try to reproduce two things from the paper:

1. One of the figures showing the effects of data sourece, resolution, filtering, etc. In other words, reproduce Figure 1, 2, 3 or 4 (just one of them). If you pick Figure 1 (outdoor air temperature data source) then you'll have to find another source of temperature data than the one I am providnig for this assignment.

2. A single figure similar to Figure 6 but showing only the box plot for the modeling choice that you picked in the previous case. In other words, if you reproduced Figure 1, then you will only produce a single box plot with the base vs. variant mismatch when using different outdoor air temperature data sources.

To remind you, here is the model used in the paper:

$$L_o(t_i, T(t_i)) = \alpha_i + \sum_{j=1}^{k} \beta_j T_{c,j}(t_i) \tag{1}$$

This is the model for the occupied case and though there is a separate (similar) model for the unoccupied case, it is really easy to see that it is an instance of this same model when $k = 1$.

You will need to work efficiently in order to achieve the above goals. Fortunately, there is a good starter kit: Assignment #2 from the 2015 version of the course. In the same folder where this assignment is located (in the GitHub repo) you will find a sub-folder called `starter` that contains those files. Please make sure to read the `README.md` for additional details.

We could finish this section of the assignment here (i.e., you should have sufficient instructions up to this point to get started). However, since you have been asking for more specific instructions (and since it is easier to grade the assignment if we break it down into smaller steps) here are some specific tasks I suggest you follow to complete the abovementioned goals.

Task #1 [10%]: *Load the two CSV files into memory:* `campusDemand.csv` *and* `temperature.csv`

Just as with the last assignment, the first step is to laod the CSV files into memory. There are, however, some important caveats that you should consider when doing this:

- We now have two files, with different number of rows and columns.

- The difference in the number of rows arises due to two facts:
  - The temperature and power measruements are not collected at the same sampling rate, nor for the sampe sampling period.
  - The `campusDemand.csv` file has measurements for many different meters on campus (i.e., even if the temperature and power measurements were done at the same sampling rate and for the same period of time, the power measruement file would be larger).

- You will need to perform some form of interpolation in order to harmonize the two time-series (i.e., to make them have the same time-stamps).

- We will only be focusing on one single meter from the dataset, so pick one and forget about the rest.

Task #2 [20%]: *Generate the Design Matrix A which contains all fo the indicator (dummy) values for the corresponding α_i coefficients, as well as the $T_{c,j}(t_i)$ values that constitute the other independent variables of the model.*

Task #3 [20%]: *Using your method of choice, estimate the coefficients for the regression model (i.e., estimate the α_i and β_i values). This will constitute your baseline model.*

Now you should pick one of the modeling choices whose influence on your results you would like to explore (i.e., pick from the list found in the left columno of Page 5 in Addy et al. 2015).

Task #4 [10%]: *Refit your model to the new modeling choice (i.e., estimate new values for α and β). This will constitute your variant model.*

Finally, generate the figures that match the ones presented in the paper.

Task #5 [10%]: *Using the baseline and variant models, predict the electrical demand of the building with each model and generate a Figure (including caption) like Figures 1 through 4 in the paper (you only need one of these figures).*

Task #6 [5%]: *Generate a box plot with the base vs. variant mismatch, like Figure 6 in the paper.*

2 AC Power

As you saw during Lectures 7 and 8, there is a publicly available dataset of voltage and current measurements for individual appliances called PLAID (available at plaidplug.com). Download this dataset and import one of the CSV files in it (each CSV file corresponds to 1-second of 30kHz measurements for an individual appliance). The first column in the CSV is the timestamp, and the other two columns are voltage and current, respectively.

Pick a single appliance instance from the whole dataset.

> **Task #7 [25%]:** *Using the single appliance measurements you just imported, create three separate plots showing: (a) instantaneous power, (b) real power calculated assuming the period of the signal is $T = 1/60 seconds$, and (c) reactive power (using the same assumptions about the period). What do these plots tell you about the power draw for this appliance?*