12-752: Data-Driven Energy Management of Buildings Assignment#2

Mario Bergés
Assistant Professor
Department of Civil and Environmental Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
marioberges@cmu.edu

November 18, 2015

Some notes before you begin:

- Make sure you document everything you do, and not just write down the answer to the question. This will both help during grading as well as improving your learning process.
- Do not write down any solution or process that you do not understand. If you feel that you do not understand how to do something, seek some help: e-mail the TAs or the instructor
- To submit your assignment, please do so using Blackboard. Two files should be uploaded via Blackboard: (a) the IPython Notebook (i.e. a .ipynb file) documenting all the tasks found in the assignment and all of your answers (including the output of your code); and (b) a PDF copy of this notebook
- Please upload a single compressed ZIP file containing the above, and name it as follows: *andrewID_assignment_#.zip* (where *andrewID* is your AndrewID and # is the assignment number

1 Preface

This assignment builds on the 2014 version of the same. In other words, you should first make sure you understand the steps taken in the solution to Assignment #2 from 2014 before continuing with this assignment. The solutions are provided in Blackboard under the Lectures forlder (Lecture #5, to be precise). You should also make sure to have read Paper #2, which discusses the piece-wise linear regression model that we will be implementing in this assignment.

1.1 Loading the Dataset

In this assignment we will be utilizing a dataset of historical electric power measurements collected for a number of buildings on campus during the last 12 months, as well as a dataset of temperature measurements collected from a weather station located on campus, during the same period of time. The datasets can be found on Blackboard under the Datasets folder.

Just as with the last assignment, the first step will be 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 #1 [20%]: Load the CSV files into memory. Specifically, store the contents of the files into a single variable named `data`, which should be an `ndarray`, or a similar data structure, containing 3 columns (time stamp, power measurement, temperature) and as many rows as required after harmonizing the datasets.

2 Preparing the Data

Task #2 [10%]: Create an array to store the indeces for the rows in `data` that correspond to measurements collected during the weekends (i.e., between 12:00am on Saturday until 11:59pm on Sunday).

Task #3 [10%]: Create another similar array to store the indices for the rows in `data` that correspond to "occupied" times (i.e., times during which the building is likely to be occupied)

Note: Please update your repository to get an updated copy of this file, as I will be expanding it as we go.