CS325: Group Assignment 4 – Linear programming

Prof. Glencora Borradaile

Due: Tuesday, December 1 at 10AM

You are encouraged to work in groups of up to three students. Only one member of each group should submit the group's work to TEACH, including your typeset ***project report as a pdf*** *and the* **code** *you wrote. The report should have* ***all member's names included****. You may use any language you choose for implementation.* ***No questions about this assignment will be answered after the due date above.***[1]

For this project, you will model a problem as a linear program and solve them using a language and linear programming solver of your choice. For a (non-comprehensive) list of freely available LP solvers, see this wikipedia page: `http://en.wikipedia.org/wiki/Linear_programming` These problems were tested using Matlab and Matlab's linear programming solver, `linprog` (which you have access to through the College of Engineering) as well as CBC and GLPK solvers via the Python PuLP package (see `https://projects.coin-or.org/PuLP`). The latter was faster and less cumbersome to use. Just saying. We will only be able to provide useful help with use of these two solvers.

Local temperature change

The daily average temperature at a given location can be modelled by a linear function plus two sinusoidal functions; the first sinusoidal function has a period of one year (modelling the rise and fall offind temperature with the seasons) and the second sinusoidal function has a period of 10.7 years (modelling the solar cycle). That is, a decent model of average temperature T on day d is given by:

$$T(d) = \underbrace{x_0 + x_1 \cdot d}_{\text{linear trend}} + \underbrace{x_2 \cdot \cos\left(\frac{2\pi d}{365.25}\right) + x_3 \cdot \sin\left(\frac{2\pi d}{365.25}\right)}_{\text{seasonal pattern}} + \underbrace{x_4 \cdot \cos\left(\frac{2\pi d}{365.25 \times 10.7}\right) + x_5 \cdot \sin\left(\frac{2\pi d}{365.25 \times 10.7}\right)}_{\text{solar cycle}}$$

The values of $x_0, x_1, \ldots, x_5$ depend on the location. For example, the amplitude of the seasonal change (x_2 and x_3) would be much greater in Chicago, IL than in San Diego, CA. Given daily temperature recordings (pairs (d_i, T_i): the average temperature T_i on day d_i), your goal is to find values for $x_0, x_1, \ldots, x_5$ that result in an equation $T(d)$ that best fits the data. You will use what you learned in class and in the practice assignment to find the curve $T(d)$ of best fit that minimizes the maximum absolute deviation for a given set of daily average temperatures.

Data from Corvallis can be found here: `http://web.engr.oregonstate.edu/~glencora/cs325/Corvallis.csv`. The first four columns are the raw data downloaded from NOAA. Raw minimum and maximum temperature recordings are given in *tenths* of degrees Celcius. Average daily temperature (column *average*) is in degrees Celcius and is simply the average of the maximum and minimum temperatures on a given day. The number of days since May 1, 1952 is given in the last column (*day*). Note that you need to take the day number into account because several days (and a few entire months) were missing from the data set. These last two columns give you the (d_i, T_i) data pairs.

Your best fit curve (defined by the linear program you use to find the values of $x_0, x_1, \ldots, x_5$ that minimize the maximum absolute deviation of $T(d)$ from your data points), gives you a value x_1 that describes the linear drift of the temperature as degrees per day.

[1] The due date in TEACH is 24 hours later, as submissions submitted within 24 hours of the deadline above will not be penalized.

Your report must include:

1. A description for a linear program for finding the best fit curve for temperature data.
2. The values of all of the variables to your linear program in the optimal solution that your linear program solver finds for the Corvallis data. Solving this LP may take a while depending on your computer. Be patient. You may want to do testing on a small part of the data set. Include the output of the LP solver that you use (showing that an optimal solution was found).
3. A single plot that contains:
   - the raw data plotted as points,
   - your best fit curve, and
   - the linear part of the curve $x_0 + x_1 \cdot d$.
4. Based on the value x_1 how many degrees Celcius per century is Corvallis changing and is it a warming or cooling trend?
5. Repeat steps 2-4 for a *different* location of your choice. You can download this from NOAA (`http://www.ncdc.noaa.gov/cdo-web/search`) for many locations. Be sure that your data covers at least 50 years to get a good fit. You will need to plan ahead as NOAA can take several hours to return the data to you given a request. Also note that the data is not necessarily *clean*: it may miss some measurements or include nonsense measurements (like -9999) that should be removed. The number of elapsed days should be carefully calculated from the date stamp. We have made the code available that we used to compute the *day* number for each data entry available here: `http://web.engr.oregonstate.edu/~glencora/cs325/computeDate.cpp` Note that you will need to adapt this code to your own needs.