CS325: Linear programming project

Prof. Glencora Borradaile

Due: Thursday, November 21 at 10AM

For this project, you will model the following problems as linear programs 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 GLPK via the Python PuLP package (see `https://projects.coin-or.org/PuLP`). The latter was faster and less cumbersome to use. Just saying.

Warm-up question: Least squares isn't good enough for me

You are given a set of points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ in the plane. You want to find a line $y = ax + b$ that comes close to each point. You probably learnt the method of least squares to find a line of best fit in your past, but we want to find the line of best fit that minimizes the maximum absolute deviation. That is, you want to find the values of a and b that minimizes:

$$\max_{1 \le i \le n} |ax_i + b - y_i|$$

Model this *general problem* as a linear program. Use the linear program to find the line of minimum-maximum-absolute-deviation for *the instance*:

$$(1, 3), (2, 5), (3, 7), (5, 11), (7, 14), (8, 15), (10, 19)$$

Your report must include:

- the linear program for the *general problem* written as an objective and set of constraints
- the best solution for the *specific problem* above
- a plot of the points and your solution for *the instance*

Warming-up question: 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:

$$T(d) = \underbrace{x_0 + x_1 \cdot d}_{\text{linear trend}} + \underbrace{x_2 \cdot \cos\left(\frac{2\pi d}{364.25}\right) + x_3 \cdot \sin\left(\frac{2\pi d}{364.25}\right)}_{\text{seasonal pattern}} + \underbrace{x_4 \cdot \cos\left(\frac{2\pi d}{364.25 \times 10.7}\right) + x_5 \cdot \sin\left(\frac{2\pi d}{364.25 \times 10.7}\right)}_{\text{solar cycle}}$$

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), we can 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 the warm-up question 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.

Your report must include:

- A description for a linear program for finding the best fit curve for temperature data.
- The values of $x_0, x_1, \ldots, x_5$ that your linear program finds for the Corvallis data. Solving this LP may take a while depending on your computer. Be patient.
- A single plot that contains:
  - the raw data plotted as points,
  - your best fit curve, and
  - the linear part of the curve $x_0 + x_1 \cdot d$.
- Based on the value x_1 how many degrees Celcius per century is Corvallis changing and is it a warming or cooling trend?
- **[BONUS]** Repeat this whole process for a 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.

One person from your group should upload all the code you used to solve these LPs to TEACH.