assignment3

- Rhea Mae Edwards **(edwardrh)**, Miles Curry **(currymi)** & Rutger Farry **(farryr)**
- CS 325
- Dr. Xiaoli Fern
- 15 March 2016

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

Write summary here The goal of this question was to determine the line of best fit using linear programming. Given the set of points, `(x1, y1), (x2,y2), ... (xN, yN)`, our objective was to find values of `a` and `b` for a linear equation by minimizing:

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

Figure 1: Prob 1 Objective Function

Linear program description

To turn this into a linear program, the variable `m` was created to represent the objective function to be *minimized.* With values: `(1,3), (2,5), (3,7),(5,11), (7,14), (8,15), (10, 19)`.

```
* Objective: Min m

Such that:

* a * x + b - y <= m
* -a * x - b + y <= m
* For all points `(x1, y1), (x2, y2), ... (xN, yN)`.
```

Running the code

The script was written in Python 2 and uses PuLP as its linear program solver as well as numpy. To install dependencies and run the script, just run:

```
pip install -y requirements.txt
python warm_up.py
```

Solution

The best values of `a` and `b`, given the set of points: `(1,3), (2,5), (3,7), (5,11), (7,14), (8,15), (10, 19)` found by the script is:

```
* a: 1.7142857
* b: 1.8571492
```

And the output of PuLP confirmed it was an optimal solution:

```
$ python warm_up.py
Status: Optimal
a: 1.7142857
b: 1.8571492
```

Plot

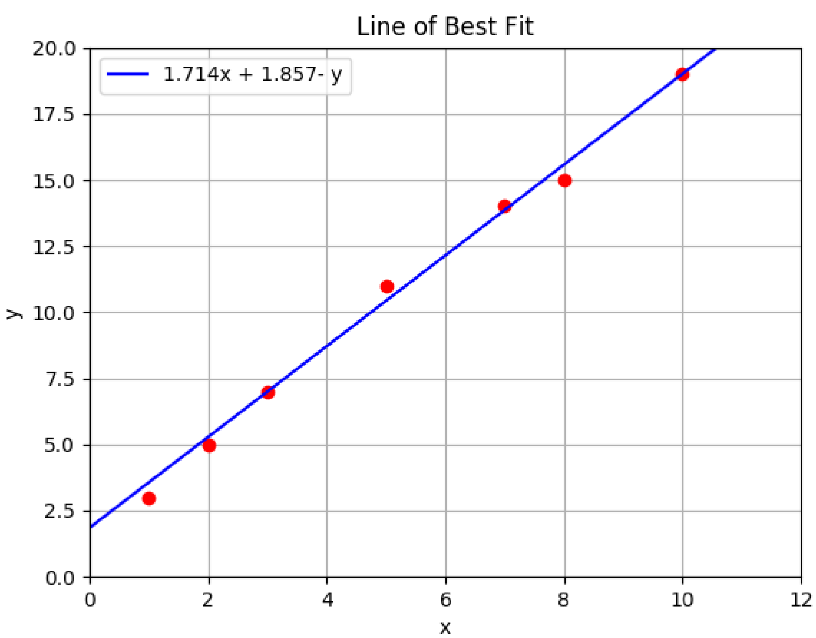

Figure 2: Plot of best fit line

Warming-up question: Local temperature change

This was a fun problem to solve and one of the first times my computer was stressed for a legitimate reason. The program took a little under a minute to solve. We found that the average temperature rise in Corvallis is about **0.66°F** per decade, which is pretty far ahead of the US average of **0.29 to 0.46°F**.

Running the code

The script is written in Python 2, so running it should be simple on most computers. Make sure to install dependencies and pass in the path to a csv file containing weather data like so:

```
pip install -y requirements.txt
python forecast.py {weather_data_csv}
```

Linear program description

This linear program was first presented by Robert Vanderbei in his paper Local Warming. We solve for 6 unknown regression coefficients, `x0...x5` and fill them into the following equation:

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

Figure 3: Climate model

The model is split into three parts, with a *linear trend* indicating an overall rise in temperature over time, a *seasonal pattern* representing the temperature based on day of the year and a *solar cycle* representing the change in temperature based on Earth's position relative to the sun.

`T(d)` will be the predicted temperature for that day after the regression coefficients are calculated. In this case, `d` should equal the number of days since May 1, 1952 (first day in NOAA data), for example, for today, March 15, 2017, we'd plug in `d = 23695`.

Solution

The raw output of our program, using Corvallis data downloaded from Canvas on March 15, 2017, is as follows:

```
$ python forecast.py Corvallis.csv
Status: Optimal
x0: 10.135061
x1: 0.00010095409
x2: 5.2564156
x3: 8.7514693
x4: 1.8864462
x5: -0.38240757
```

This gives us the values of `x0...x5` and the output status from PuLP, showing that the solution was optimal..

Plugging this into the equation, we get the following equation, which should give us an average temperature estimate for any day after May 1, 1952, in degrees Celsius:

$$T(d) \;=\; 10.135 + 0.000100095d \;+\; 5.256cos(\frac{2\pi d}{365.25}) + 8.751sin(\frac{2\pi d}{365.25}) + \;1.886cos(\frac{2\pi d}{365.25\times 10.7}) - 0.382sin(\frac{2\pi d}{365.25\times 10.7})$$

Figure 4: Calculated climate model

Plot

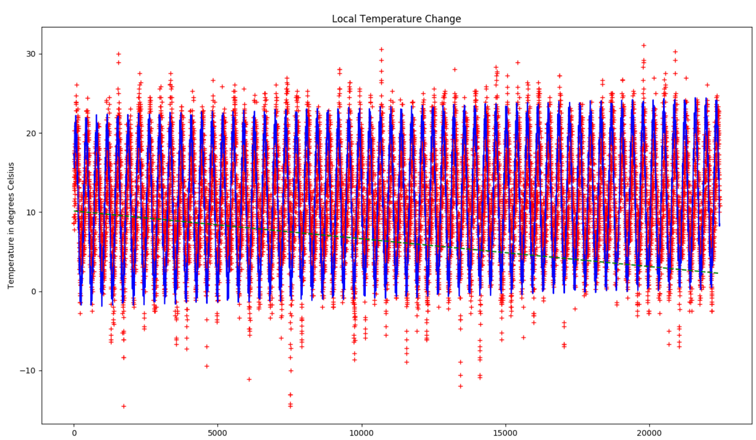

Figure 5: Corvallis Forecasted Data

Reflection

Looking back at the output of the linear equation solver, the most interesting variable is probably `x1`, which is the slope of the linear trend of temperature change in Corvallis, OR. This shows that, on average, the temperature has been rising by 0.00010095409°C per day since May 1, 1952. This translates to a rise of roughly 0.369°C or 0.6642°F per decade. This is a little higher than the EPA's calculations of a rise in US temperatures of 0.29 to 0.46°F per decade since 1979 but is not unreasonable.

Meta

If you're reading the PDF version of this file, we generated it from `README.md` using pandoc. You can update it from the README file using:

```
pandoc README.md --latex-engine=xelatex -o writeup.pdf
```