Climate Data Fitting and Local Warming Justification

ECE 532 Final Project Proposal

Group Name: Kirby Machine (#3 members)

Group Member:

1. Junda Chen - jchen693@wisc.edu - 9077653518
2. Haoruo Zhao - hzhao95@wisc.edu - 9075315482
3. Doris Duan - duan35@wisc.edu - 9078286672

§1 Abstract

The project aims to justify local warming by finding a model to fit climate data. This project uses a basis matrix of a linear function and several trigonometric functions rather than using singular value decomposition, and applies LASSO and Ridge Regression to find dominant factors that fit the data perfectly.

In choosing trigonometric functions, students will find periods of sin/cos – functions according to their meanings in reality. For example, they need to consider periodic factors that actually affect the weather, such as seasonal cycles and solar cycles.

After constructing the basis matrix, students will use Ridge Regression and LASSO separately to compare their accuracy, and observe vectors that are chose by LASSO. They will interpret the meaning of those "important" vectors in the context of the reality. Furthermore, they will apply the model they have generated to climate data in other areas to draw conclusions about local warming.

§2 Background

As the advent of Anthropocene and the unprecedented speed of technological development, the emission of greenhouse gases has been surging, rendering the issue of global warming severe. However, the change of climate is so subtle that individuals and industries hardly notice the devastating consequences brought by global warming and still hang on to their destructive behaviors such as factory farming, fossil fuel burning, and driving private cars [1][2]. Therefore, it is significant to arouse people's awareness of the extent of global warming.

In this project, we obtained climate data [3] of McGuire Air Force Base from 1955 to 2010 [4] on *National Oceanic and Atmospheric Administration* (NOAA) website and fit the data with a combination of linear function and trigonometric functions. In order to improve the accuracy of the fitting, we applied LASSO and Ridge Regression on the model to assign different weights to different linear and trigonometric functions. After refining the model, we concluded that the weather in McGuire Air Force Base does suffer a gradual increase. Furthermore, we applied the model to other local areas to determine whether the weather in those places has a trend to increase. (background as a tutorial, build X here, don't introduce new concepts below, add linear term regularizer)

(Refer to these pages:

Local Warming - Construct Basis Matrix - https://beta.observablehq.com/@gindachen/function-slider

Local Warming - Model Visualization - https://beta.observablehq.com/@gindachen/local-warming

)

§3 Warm-Up

1. The file `part1.m` loads the climate data of 55 years in a column vector **temperature** and date of 55 years in a column vector **date**. Section 1.1 displays the overall data in a line chart. Section

1.2 of the code amplifies parts of the data for visualizing. The climate data of 11 years, 5 years and 2 years are displayed. What kind of function do they look like?

a) e^x b) $\sin(x)$ c) $\log(x)$ d) x^2

2. Now that you've already chose an answer, let's further explore functions that can fit the data with high accuracy. Section 1.3 displays the fitting by $\sin(x)$ alone, $\sin(x) + \cos(x)$, $\sin(x) + \sin(2x)$ (sine functions of different frequencies). Which combination is better according to its average error?

a) $\sin(x)$ b) $\sin(x) + \cos(x)$ c) $\sin(x) + \sin(2x)$

3. Now we want to choose trigonometric functions of different frequencies. First, let's assign different weights, phases and offsets to $\sin(x)$. You can find an animation of the fitting to data in this link: https://beta.observablehq.com/@gindachen/function-slider.

4. Now consider what periodic factors affect the climate. Using sin/cos – functions you chose below to build a basis matrix **X**. Hint: the **X** matrix should use each function as a column vector with the offset vector as the first column, that is,

$$\mathbf{X} = \left[1 \ \sin\left(\frac{2\pi}{\mathrm{T}_1}\boldsymbol{x}\right) \ \cos\left(\frac{2\pi}{T_1}\boldsymbol{x}\right)\ldots \ \sin\left(\frac{2\pi}{\mathrm{T}_m}\boldsymbol{x}\right) \ \cos\left(\frac{2\pi}{T_m}\boldsymbol{x}\right)\right], \boldsymbol{x} \text{ is } \textbf{date} \text{ vector.}$$

Add weights, expand matrix

i. $\sin\left(\frac{2\pi}{365.25*0.5}\boldsymbol{x}\right) + \cos(\frac{2\pi}{365.25*0.5}\boldsymbol{x})$ – undergraduate's final exam cycle

ii. $\sin\left(\frac{2\pi}{365.25}\boldsymbol{x}\right) + \cos(\frac{2\pi}{365.25}\boldsymbol{x})$ – earth's seasonal cycle

iii. $\sin\left(\frac{2\pi}{365.25*4}\boldsymbol{x}\right) + \cos(\frac{2\pi}{365.25*4}\boldsymbol{x})$ – U.S. president election periodicity

iv. $\sin\left(\frac{2\pi}{365.25*11.2}\boldsymbol{x}\right) + \cos(\frac{2\pi}{365.25*11.2}\boldsymbol{x})$ – solar cycle

v. $\sin\left(\frac{2\pi}{365.25*18.6}\boldsymbol{x}\right) + \cos(\frac{2\pi}{365.25*18.6}\boldsymbol{x})$ – the moon declination angle changing cycle

§4 Main Activity

1. In `matlabdata/McGuireAFB.data.csv`, daily average temperature from 1955 to 2010 are listed. Let **y** denote this temperature vector. In the above warm-up activity, you built **X** with respect to basis functions you chose. Now, let's try to fit the data using all basis functions listed above. In this way, what is **X**? And what kind of method will you choose to minimize errors?

2. In reality, only a small number of basis functions listed above are significantly contributive to average temperature in McGuire Air Force Base. Now use Ridge (with L-2 regularizer) and

LASSO (with L-1 regularizer) to choose dominant basis functions. Write down each model respectively. Which is better and why?

3. Run the code script in `part2.m`, observe weight vector **w** you get. Which periods have dominant influence on temperature in McGuire Air Force Base? Compare graph generated using basis functions (red) and original temperature (blue). Do they fit well? (Hint: compare tails of the graph).

4. Based on our observation, we found that using trigonometric functions and offset is not enough. Hence, we add linear term **x** into family of basis functions. What's **X** in this case? Repeat above actions to calculate new weight vector **w**. Does it make sense? Why? (Hint: observe **X**) If not, how can you modify it? What's the new? Now what functions play an important role in fitting average temperature?

5. Observe the weight of linear term. What can you say about the trend of average temperature in McGuire Air Force Base from 1955 to 2010?

References

[1]https://www.nrdc.org/stories/global-warming-101?gclid=EAIaIQobChMIm-nc9KKP3wIVyluGCh2uPQKsEAAYASAAEgIp9vD_BwE#causes

[2]https://www.ciwf.org.uk/factory-farming/environmental-damage/?gclid=EAIaIQobChMIm-nc9KKP3wIVyluGCh2uPQKsEAAYAiAAEgKoNPD_BwE

[3]http://www.princeton.edu/~rvdb/ampl/nlmodels/LocalWarming/McGuireAFB/data/McGuireAFB.dat

[4]http://www.princeton.edu/~rvdb/ampl/nlmodels/LocalWarming/McGuireAFB/data/Dates.dat

Appendix

3-1: b) is the correct answer, because the data appears to be periodic and other choices are not periodic functions.

3-2: b) is correct, because it has the smallest average error.

3-4: Suppose we only choose i, iv, v, then **X** built on this is:

$$\mathbf{X} = \begin{bmatrix} 1 \ \sin\left(\frac{2\pi}{365.25}x_1\right) \ \cos\left(\frac{2\pi}{365.25}x_1\right) \ \sin\left(\frac{2\pi}{365.25*11.2}x_1\right) \ \cos\left(\frac{2\pi}{365.25*11.2}x_1\right) \ \sin\left(\frac{2\pi}{365.25*18.6}x_1\right) \ \cos\left(\frac{2\pi}{365.25*18.6}x_1\right) \\ 1 \ \sin\left(\frac{2\pi}{365.25}x_2\right) \ \cos\left(\frac{2\pi}{365.25}x_2\right) \ \sin\left(\frac{2\pi}{365.25*11.2}x_2\right) \ \cos\left(\frac{2\pi}{365.25*11.2}x_2\right) \ \sin\left(\frac{2\pi}{365.25*18.6}x_2\right) \ \cos\left(\frac{2\pi}{365.25*18.6}x_2\right) \\ \cdots \ \cdots \ \cdots \ \cdots \ \cdots \ \cdots \ \cdots \ \cdots \ \cdots \\ 1 \ \sin\left(\frac{2\pi}{365.25}x_n\right) \ \cos\left(\frac{2\pi}{365.25}x_n\right) \ \sin\left(\frac{2\pi}{365.25*11.2}x_n\right) \ \cos\left(\frac{2\pi}{365.25*11.2}x_n\right) \ \sin\left(\frac{2\pi}{365.25*18.6}x_n\right) \ \cos\left(\frac{2\pi}{365.25*18.6}x_n\right) \end{bmatrix}$$

4-1: Least square. We want to minimize $\left\|\mathbf{X}w - y\right\|_2^2$, where **w** is the weight vector.

The general form of **X** is:

$$\begin{bmatrix} 1 & sin(2\pi x_1/T_1) & cos(2\pi x_1/T_1) & sin(2\pi x_1/T_2) & cos(2\pi x_1/T_2) & \dots & sin(2\pi x_1/T_m) & cos(2\pi x_1/T_m) \\ 1 & sin(2\pi x_2/T_1) & cos(2\pi x_2/T_1) & sin(2\pi x_2/T_2) & cos(2\pi x_2/T_2) & \dots & sin(2\pi x_2/T_m) & cos(2\pi x_2/T_m) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & sin(2\pi x_n/T_1) & cos(2\pi x_n/T_1) & sin(2\pi x_n/T_2) & cos(2\pi x_n/T_2) & \dots & sin(2\pi x_n/T_m) & cos(2\pi x_n/T_m) \end{bmatrix}$$

When we consider all basis functions listed in part1, **X** should be:

$$\mathbf{X} = \begin{bmatrix} 1 & sin(2\pi x_1/T_1) & cos(2\pi x_1/T_1) & sin(2\pi x_1/T_2) & cos(2\pi x_1/T_2) & \dots & sin(2\pi x_1/T_5) & cos(2\pi x_1/T_5) \\ 1 & sin(2\pi x_2/T_1) & cos(2\pi x_2/T_1) & sin(2\pi x_2/T_2) & cos(2\pi x_2/T_2) & \dots & sin(2\pi x_2/T_5) & cos(2\pi x_2/T_5) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & sin(2\pi x_n/T_1) & cos(2\pi x_n/T_1) & sin(2\pi x_n/T_2) & cos(2\pi x_n/T_2) & \dots & sin(2\pi x_n/T_5) & cos(2\pi x_n/T_5) \end{bmatrix}$$

$$T_1 = 365.25*0.5, T_2 = 365.25, T_3 = 365.25*4, T_4 = 365.25*11.2, T_5 = 365.25*18.6$$

4-2: LASSO: $min\left\|\mathbf{X}w - y\right\|_2^2 + \lambda\|w\|_1$, Ridge Regression: $min\left\|\mathbf{X}w - y\right\|_2^2 + \lambda\|w\|_2^2$

LASSO is better. Because it gives sparsity estimate.

4-3: One-year period (T_1 = 365). No, the blue graph located a little above red graph at the tail.

4-4: **w** doesn't make sense. Because linear terms are too big.

We can regularize linear term in **X** by dividing length of **y** entry wise. i.e. **d** = **d**./length(**y**).

That is, we change **X** from

$$\begin{bmatrix} 1 & x_1 & sin(2\pi x_1/T_1) & cos(2\pi x_1/T_1) & sin(2\pi x_1/T_2) & cos(2\pi x_1/T_2) & \dots & sin(2\pi x_1/T_5) & cos(2\pi x_1/T_5) \\ 1 & x_2 & sin(2\pi x_2/T_1) & cos(2\pi x_2/T_1) & sin(2\pi x_2/T_2) & cos(2\pi x_2/T_2) & \dots & sin(2\pi x_2/T_5) & cos(2\pi x_2/T_5) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & sin(2\pi x_n/T_1) & cos(2\pi x_n/T_1) & sin(2\pi x_n/T_2) & cos(2\pi x_n/T_2) & \dots & sin(2\pi x_n/T_5) & cos(2\pi x_n/T_5) \end{bmatrix}$$

to

$$\begin{bmatrix} 1 & x_1/L & sin(2\pi x_1/T_1) & cos(2\pi x_1/T_1) & sin(2\pi x_1/T_2) & cos(2\pi x_1/T_2) & \dots & sin(2\pi x_1/T_5) & cos(2\pi x_1/T_5) \\ 1 & x_2/L & sin(2\pi x_2/T_1) & cos(2\pi x_2/T_1) & sin(2\pi x_2/T_2) & cos(2\pi x_2/T_2) & \dots & sin(2\pi x_2/T_5) & cos(2\pi x_2/T_5) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_n/L & sin(2\pi x_n/T_1) & cos(2\pi x_n/T_1) & sin(2\pi x_n/T_2) & cos(2\pi x_n/T_2) & \dots & sin(2\pi x_n/T_5) & cos(2\pi x_n/T_5) \end{bmatrix}$$

$$L = length(y)$$

4-5: Temperature rises gradually, indicating that there is local warming in McGuire Air Force Base.