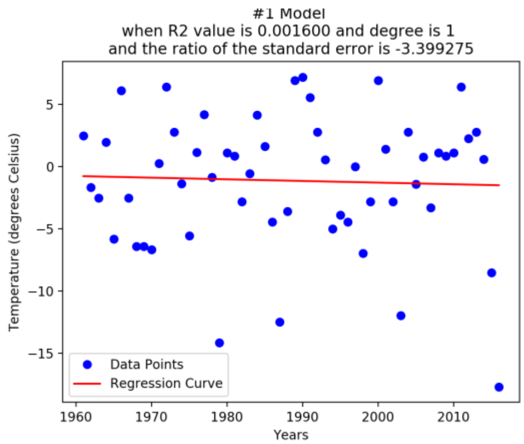
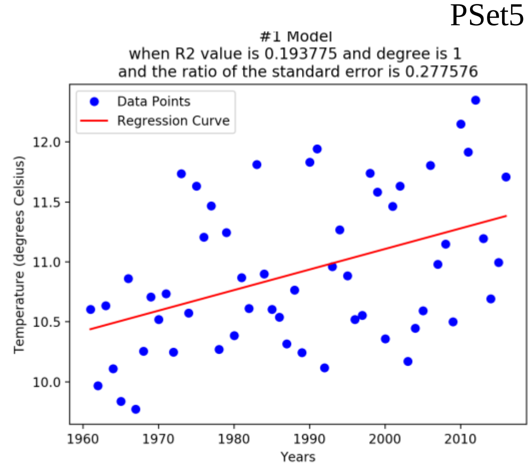
Tuo Sun
December 3, 2017
sunt@mit.edu
PSet5 Write Up





PROBLEM 3A (left):
From 1961 to 2016, temperature measurements in Boston on February 14th for each year fits the data to a degree-one polynomial.

PROBLEM 3B (right):
From 1961 to 2016, the average annual temperature in Boston.

1. What difference does choosing a specific day to plot the data versus calculating the yearly average have on the goodness of fit of the model? Interpret the results.

- The ratio of the standard error to slope is much bigger in a specific day and the regression curve can hardly fit the model. As we can see, the data of a specific day goes from 7 to -15 with the interval of 23 degree when that of the average one is only 12.5 to 9.5.

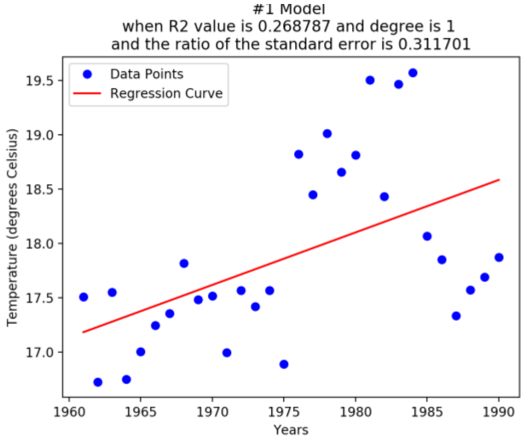
2. Why do you think these graphs are so noisy? Which one is noisier? How do outliers affect the generated model?

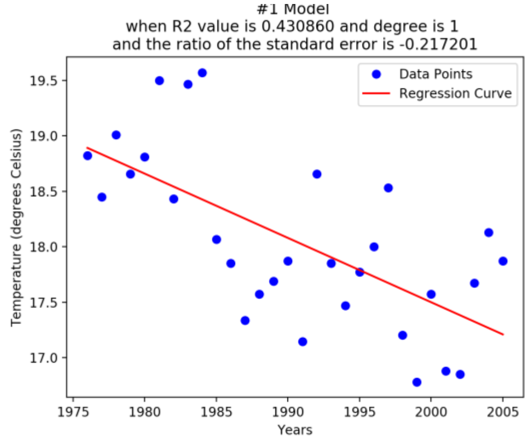
- Because the temperature can be influenced by various reasons. The graph of the specific day is noisier. The outliers drop down the regression curve seriously and make the R2 and ratio worse.

3. How do these graphs support or contradict the claim that temperature has been increasing over time? The slope and the standard error-to-slope ratio could be helpful in thinking about this.

- The average graph (right one) can be used to indicate the warming issue, but either the specific-day one does. The slope and the standard error-to-slope ratio means how the regression curve fits the trend, and the left one has a negative ratio.

Tuo Sun
December 3, 2017
sunt@mit.edu
PSet5 Write Up





PROBLEM 4B

30 years to show that the average annual temperature in San Diego is rising (left) and decreasing (right).

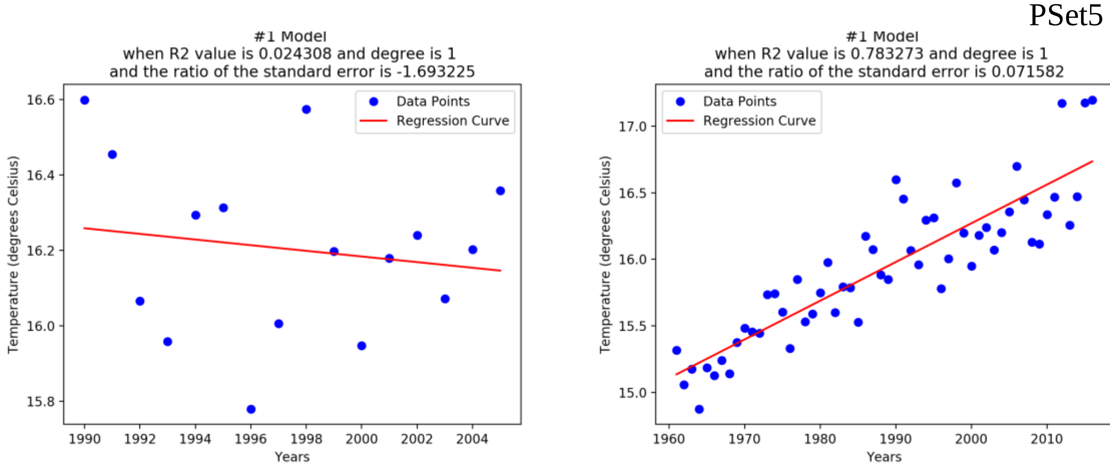
RISING:

1. What was the start and end year for your window? What was the slope?
   - From 1960 to 1990; the scope is 0.0482627617032;
2. What conclusions might you make from this plot with respect to how temperature is changing over time?
   - The climate in San Diego is warming;
3. How recent is the window? What might this imply about relevance?
   - 27 years ago; There is a great temperature rising until 27 years ago;

DECREASING:

1. What was the start and end year for your window? What was the slope?
   - From 1975 to 2005; the scope is -0.0579693233185;
2. How recent is the window? What might this imply about relevance?
   - 12 years ago; There is a great temperature decline until 12 years ago;
3. Considering both plots, what conclusions might you make with respect to how temperature is changing over time?
   - The temperature is changing over time stochastically, which means we cannot find the obvious trend from the data. Or we can argue the average temperature went up and down in a longer period.

Tuo Sun
December 3, 2017
sunt@mit.edu
PSet5 Write Up



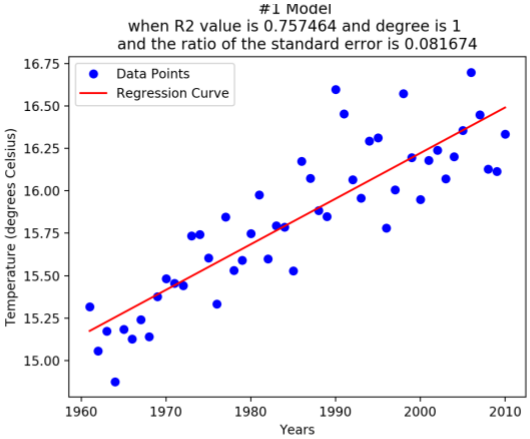
the national average annual temperature from 1990 to 2004 (left) and from 1960 to 2016 (right)

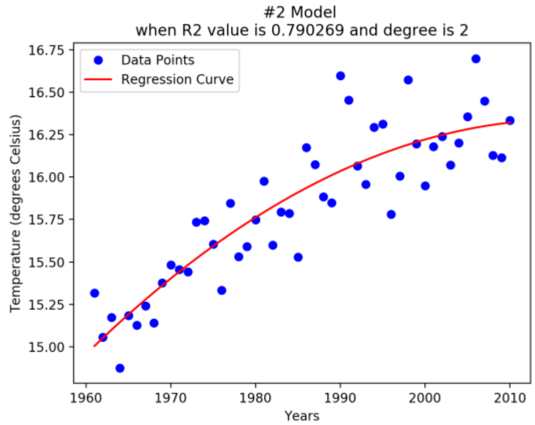
1. What was the start and end year for your window? What was the slope?
   - From 1990 to 2005; the scope is -0.00747592855502;
2. Do you think your plot is convincing? Why or why not? How does using the national average change your previous conclusions from the San Diego data?
   - The plot is not convincing since the R2 value is very low and the curve doesn't fit the data at all. It will not become anything of the conclusions, as the sample is using small size and different factors.
3. Compare your plot to the plot of the national average annual temperatures across the entire range of 1961-2016 (include this plot in your write-up). How does using a shorter interval affect our analysis of this data?
   - Under-fitting without seeing the big trend; samples are not large enough for seeing it
4. Senator Bigday Da accuses you of conflating the implications of temperature data in connection with climate change. Of the 22 enumerated datasets available on the NCEI website, pick two datasets that would help improve your analysis of climate change. Briefly explain your reasoning.
   - Global Summary of the Year; which is a big data set without trapping into a local climate change or irregular phenomenon;
   - Individual Annual/Dec. Issue; which can be used for correcting temperature changing with the precipitation variable.

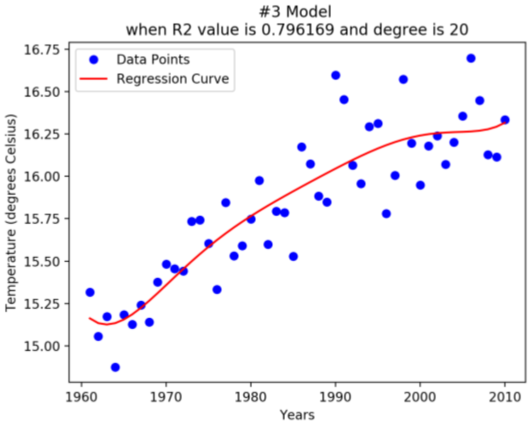
Tuo Sun
December 3, 2017
sunt@mit.edu
PSet5 Write Up

PROBLEM 5B

i) From 1960 to 2010, the 1, 2 and 20-degree polynomial fitted curve of the national annual average temperature:







1. How do these models compare to each other?

- 2-degree curve looks having the balance between r2 value and conciseness. The linear is under-fitting and 20-degree curve is overfitting.

2. Which one has the best R2? Why?

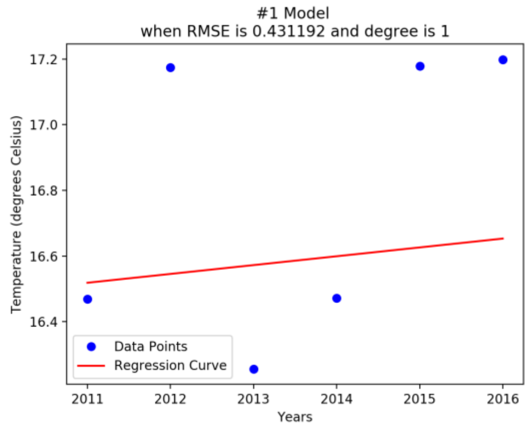
- The 20-degree curve has; it can link to more trends so the most estimated variables fit the data.

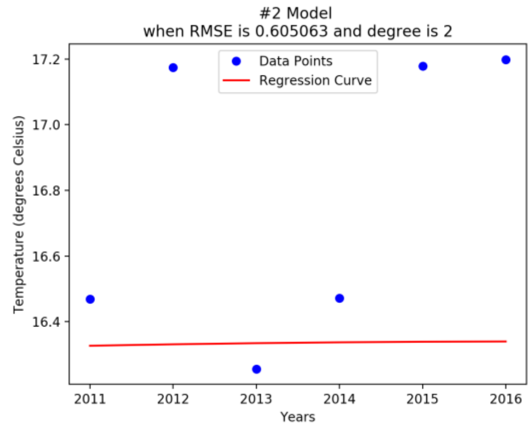
3. Which model best fits the data? Why?

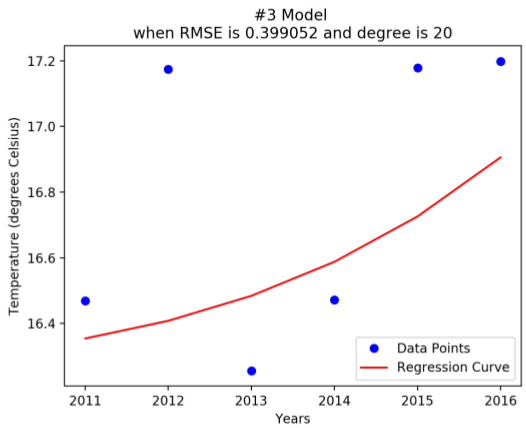
- The parabolic curve; Occam's Razor, once two models can fit a phenomenon, the more concise one is the better.

Tuo Sun
December 3, 2017
sunt@mit.edu
PSet5 Write Up

ii) From 2010 to 2016, the comparison between the prediction from the 1, 2 and 20-degree polynomial fitted curve and the real data







1. How did the different models perform? How did their RMSEs compare?
   - None of them performs very well. The 20-degree one has the smallest RMSE.
2. Which model performed the best? Which model performed the worst? Is this different from the training performance in the previous section? Why?
   - The 20-degree one has the best RMSE and parabolic model has the worst. This is quite different than the previous section. Because parabolic model can only change the trend by once and the future will be stable, but the temperature is varying hugely in future.
3. If we had generated the models using the data from Problem 3B (i.e. the average annual temperature of Boston) instead of the national annual average over the 22 cities, how would the prediction results 2011-2016 have changed?
   - The RMSEs of predictions will be very high, which means the highly error rates of 3 models. The data of a city is not big enough to learn the trend.