Problem Set 5 Write Up

3A:

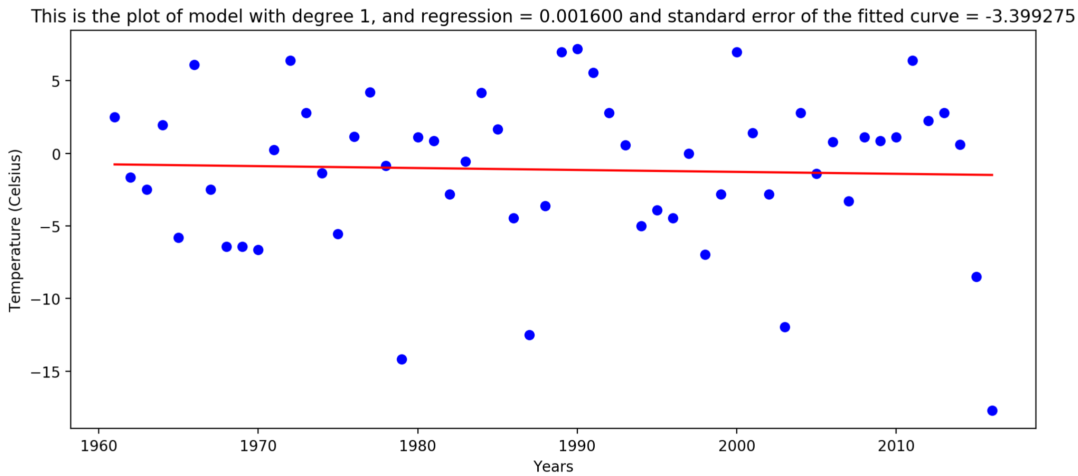

Figure 1- Boston temperature on February 14th over the years. The blue dots are the real temperature and the red line is the fitted curve.

3B:

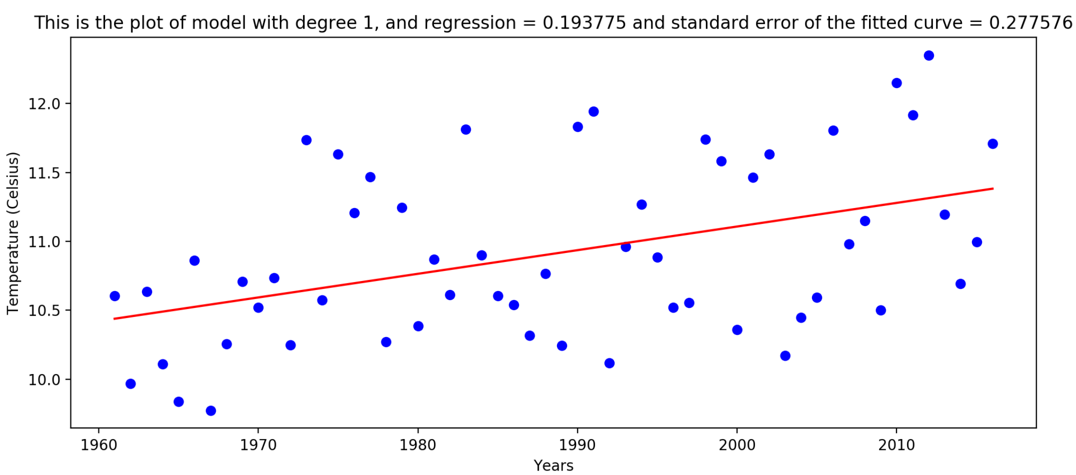

Figure 2- Boston annual average temperature over the years. The blue dots are the real temperature and the red line is the fitted curve.

1. Taking the average over the temperature of the whole year gives us the data of how the temperature was during a whole year rather in one single day the temperature of one city might not follow any specific pattern and is a lot noisier than the average over the whole year. Moreover, by averaging we are decreasing the effect of noise.
2. The regression of the second model is higher; thus the second model is less noisy. Moreover, the SE error of the first one is higher than the second model. The data is noisy because the temperature is a function of many variables and we cannot fit a linear model to it.

3. In the first graph we do not see a pattern of increasing temperature, however in the second model that we averaged the temperature over the whole year we can see a increasing pattern of temperature over the years.

4B:

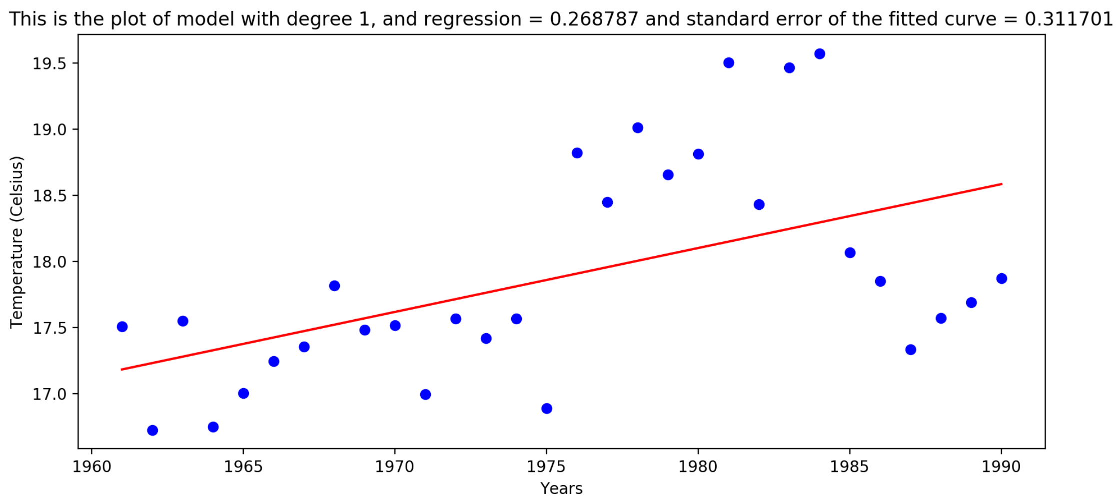

Figure 3- San Diego average annual weather over the years of 1961-1991. The blue dots are the real temperature and the red line is the fitted curve.

1. The best window for years was 1961-1991. And the slope of the one-degree model for this interval is: 4.82627617e-02.
2. From this graph we can conclude that the temperature increased from 1960 to 1990 more drastic than recent years.

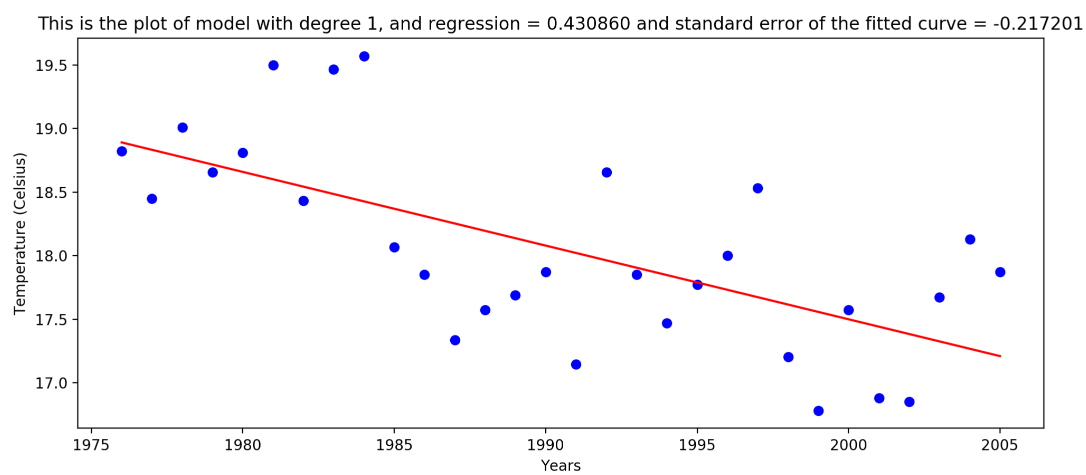

Figure 4- San Diego average annual weather over the years of 1976-2006. The blue dots are the real temperature and the red line is the fitted curve.

1. The best window for decreasing temperature was 1976-2006 with the slope of -5.79693233e-02.
2. From this graph we observe that the average yearly temperature dropped during the years of 1976-2006.

3. Based on these plots we can say that the temperature is oscillating from 16.0 – 20 Celsius.

Average over cities:

Figure 5- National average annual weather over the years of 1990-1993. The blue dots are the real temperature and the red line is the fitted curve.

Figure 5 shows the best graph that has an interval with decrease in the temperature.

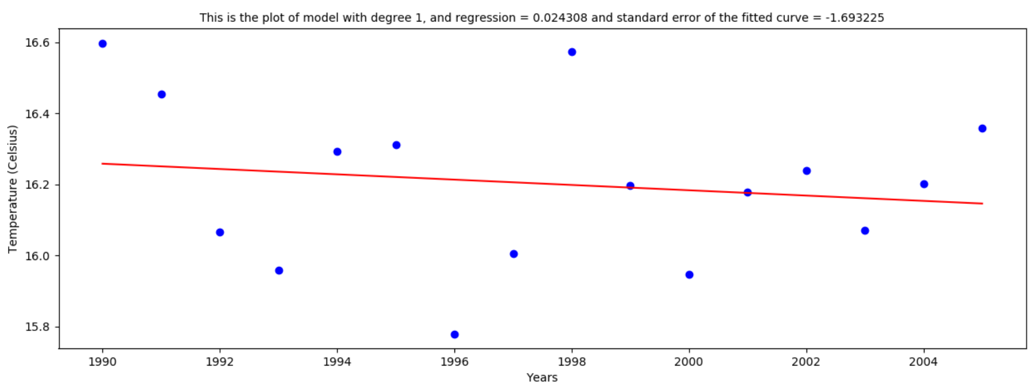

Figure 6- National average annual weather over the years of 1990-2004. The blue dots are the real temperature and the red line is the fitted curve.

Figure 6 is the largest window that still shows a decrease in temperature.
For this largest window the start year is 1990 and the end year is 2004. The national average data will give us an information on how the average temperature in the whole nation has changed over the years. This graph, even though it shows a decrease in temperature, it is not convincing enough as it is noisy.

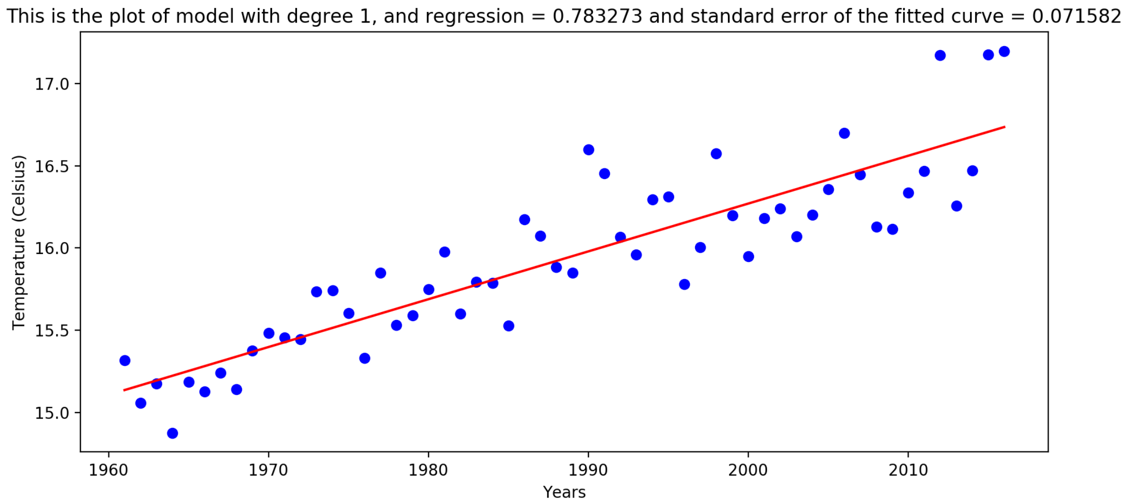

Figure 7- National average annual weather over the years of 1961-2016. The blue dots are the real temperature and the red line is the fitted curve.

Figure 7 is the plot for national average data. As we can see this data shows an increase in temperature over the years of 1961-2016 and it is less noisy than average temperature in the window of 1990-2004. When we analyzed a shorter window of data the best fitted model was different and noisy and not at all precise. Local Climatological Data and global summary of the years so that we can try to find out climate patterns over the years to verify our result from these data.

5B:

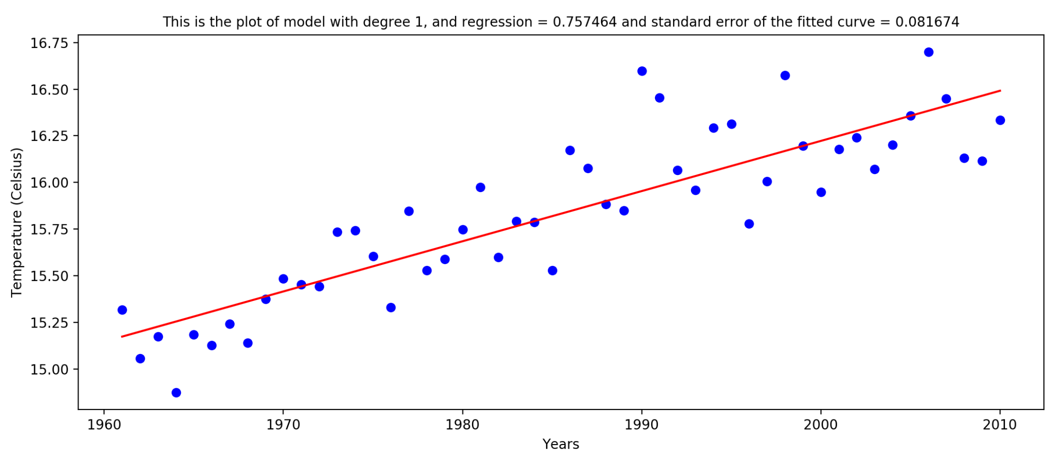

Figure 8- National average annual weather over the years of 1961-2010 as a train set for linear model. The blue dots are the real temperature and the red line is the fitted curve.

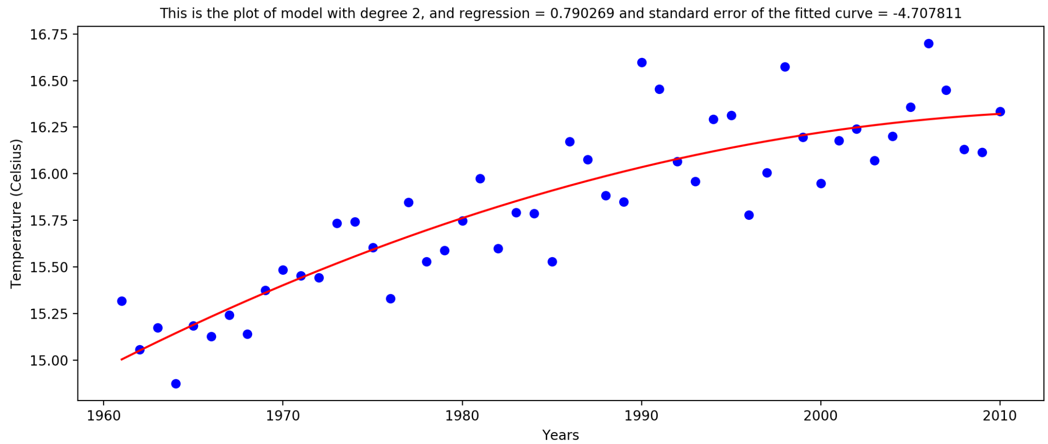

Figure 9- National average annual weather over the years of 1961-2010 as a train set for degree 2 model. The blue dots are the real temperature and the red line is the fitted curve.

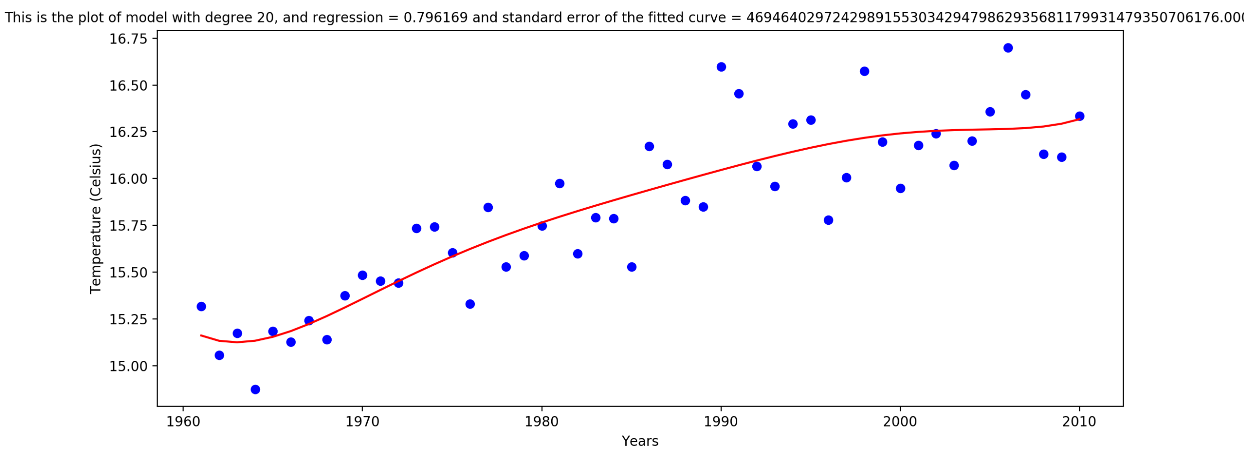

Figure 10- National average annual weather over the years of 1961-2010 as a train set for degree 20 model. The blue dots are the real temperature and the red line is the fitted curve.

The regression of the degree 20 model is the highest and its standard error of the fitted curve is also the highest. The best fitted model is the linear as its standard error is lowest.

ii) Plots of predicted values versus the real data:

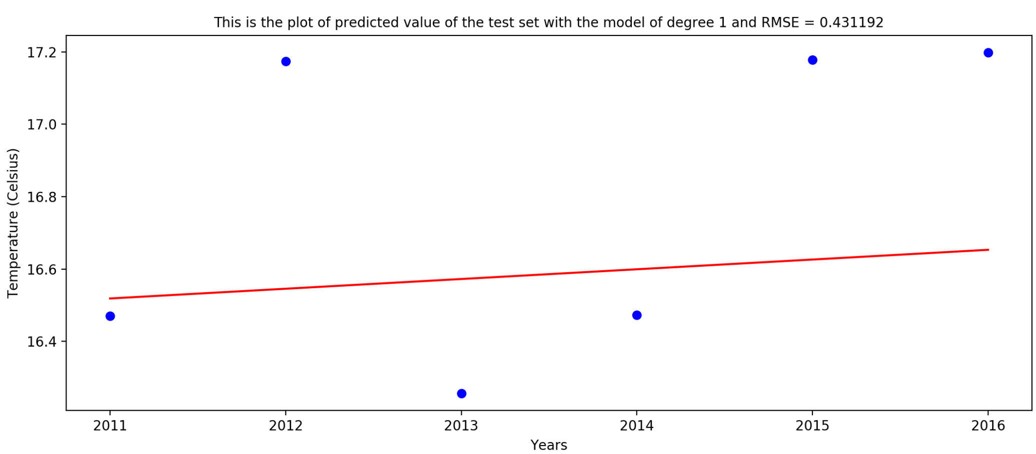

Figure 11- National average annual weather over the years of 2011-2016 as a test set for linear model. The blue dots are the real temperature and the red line is the fitted curve.

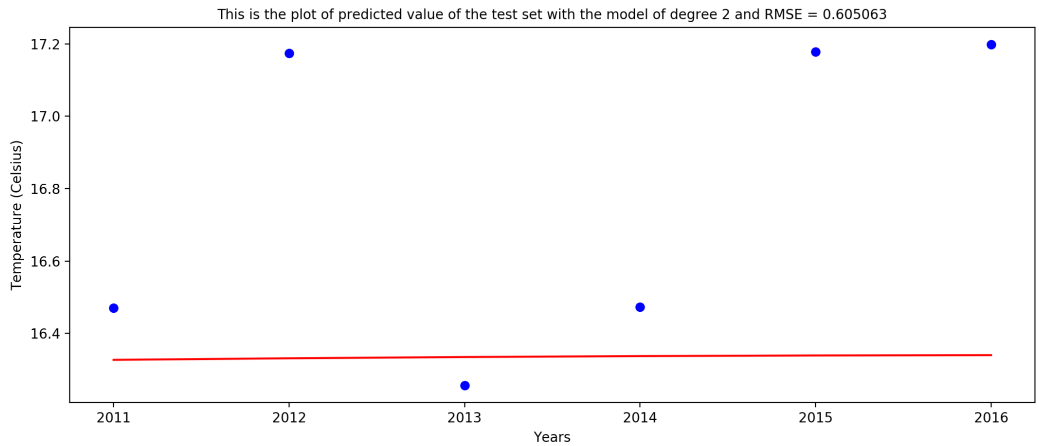

Figure 12- National average annual weather over the years of 2011-2017 as a test set for degree 2 model. The blue dots are the real temperature and the red line is the fitted curve.

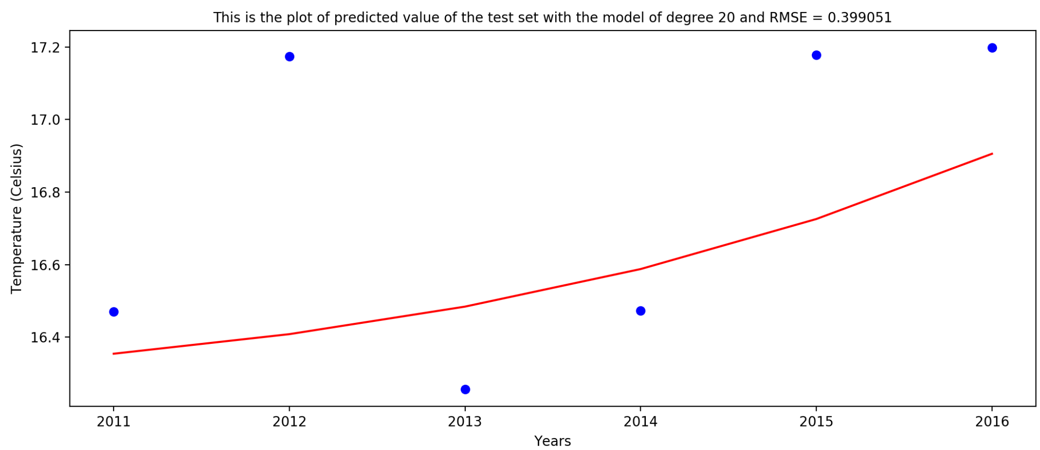

Figure 13- National average annual weather over the years of 2011-2017 as a test set for degree 20 model. The blue dots are the real temperature and the red line is the fitted curve.

The degree 20 model had the lowest RMSE and the degree 2 model had the highest. The degree 20 model is the best model here. This is different from the training set in a way that, linear model had a lower error than degree 20 model. However, here we observe that degree 20 model has a lower error in predicting the data.

If we used the data in 3B the model would have been worst in finding a good model. As the data is noisier and it makes it harder to fit a curve.