**Name:**

**Kerberos:**

Problem Set 5 Write-up

Modeling Temperature Change

|  |
| --- |
| *Reminder, all of your plots must have:*   * *Axis labels* * *Title including all of the relevant information* |

# Problem 4

Plot 4A: Average Daily Temperature for San Francisco on 12/25 (1961-2016)

|  |
| --- |
|  |

Plot 4B: Average Yearly Temperature for San Francisco (1961-2016)

|  |
| --- |
|  |

4.1 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? Select all that apply

▢ Using the average yearly temperature results in higher R^2.

▢ Using the specific day temperature results in higher R^2.

▢ Using the specific day temperature results in a better fit for the linear model.

▢ Using the average yearly temperature results in a better fit for the linear model.

▢ Both models have similar goodness of fit. It is impossible to say which performed better than the other.

4.2 Why do you think these graphs are so noisy?

|  |
| --- |
|  |

# 

# Problem 5

Plot 5B: Increasing Interval (Tampa, length=30)

|  |
| --- |
|  |

5.1 What were the start and end years for your window? What was the slope? *Note that the end year should be the last year whose temperatures are included in the model.*

|  |
| --- |
|  |

Plot 5C: Decreasing Interval (Tampa, length=15)

|  |
| --- |
|  |

5.2 What were the start and end years for your window? What was the slope? *Note that the end year should be the last year whose temperatures are included in the model.*

|  |
| --- |
|  |

5.3 Considering *both* plot 5B and 5C, what conclusions, if any, can you make with respect to how temperature is changing over time based on your models and their goodness of fit? Select all that apply.

▢ R^2 of increasing plot is higher than R^2 of decreasing plot

▢ SE plots are inconclusive because they do not show a particular trend

▢ Results from SE and R^2 suggest a significant increase in temperatures

▢ R^2 of increasing plot is lower than R^2 of decreasing plot

▢ R^2 suggests increase in temperature but SE does not

# 

# Problem 6

Plot 6B: Training Data (1961-1999), Degree 2

|  |
| --- |
|  |

Plot 6B: Training Data (1961-1999), Degree 10

|  |
| --- |
|  |

6.1 How do these models compare to each other in terms of R^2 and fitting the data? Select all the apply

▢ Degree 10 has lower R^2 values.

▢ Degree 10 has higher R^2 values.

▢ Degree 10 is clearly overfitting the data and should not be utilized.

▢ Degree 2 is clearly underfitting the data and should not be utilized.

▢ Degree 10 shows some sign of possible overfit but more analysis on test data is necessary to determine if it is actually overfitting.

Plot 6B: Test Data (2000-2016), Degree 2

|  |
| --- |
|  |

Plot 6B: Test Data (2000-2016), Degree 10

|  |
| --- |
|  |

6.4 Which model performed the best and how do you know? If this is different from the training performance in the previous section, explain why this occurred. Select all that apply.

▢ Degree 2 performed better visually

▢ Degree 10 performed better visually

▢ Degree 2 performed better in terms of RMSE value

▢ Degree 10 performed better in terms of RMSE value

▢ It is impossible to say whether degree 2 or 10 performed better.

6.5 If we had generated the models using the data from Problem 4B (i.e. the average annual temperature of Portland) instead of the national annual average over the 22 cities, how would the prediction results on the national data have changed? Select all that apply

▢ Resulting model will be similar to test data and will result in similar/better predictions

▢ Resulting model will have lower RMSE than the test data

▢ Resulting model will be far dissimilar from the test data and will result in worse predictions

▢ Resulting model will have higher RMSE than the test data

▢ It is impossible to say because it is highly contingent on the data