Cameron Scolari

1. For these, I am choosing k-values that are odd to prevent ties when classifying.

Part 1: The value of k that would minimize training set error is k=1 because each point can see itself so they would all classify themselves correctly. Training set error is not a reasonable estimate for test set error because the model will memorize the training data which is overfitting.

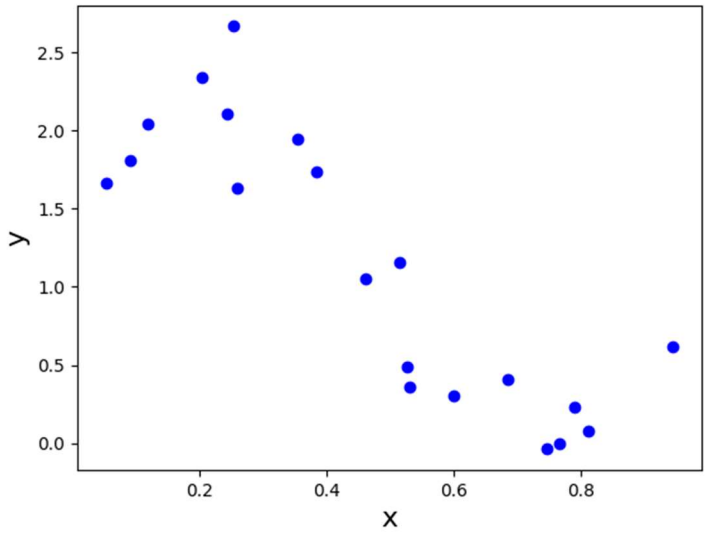
Part 2: For k=5 and k=7, the LOOCV results in a value of 4/14, meaning it incorrectly predicted 4 points out of 14. Cross-validation is a better measure of test performance than training error because it doesn't overfit the training data. The LOOCV for k=1 is 10/14 and the LOOCV for k=13 is 14/14. Using too small a k value could be bad because it can lead to overfitting when the model just memorizes the training data. Using too large a k value could also be bad because it can lead to underfitting because it assumes that every point is important when classifying and does not place enough emphasis on the points nearest the one we are classifying.

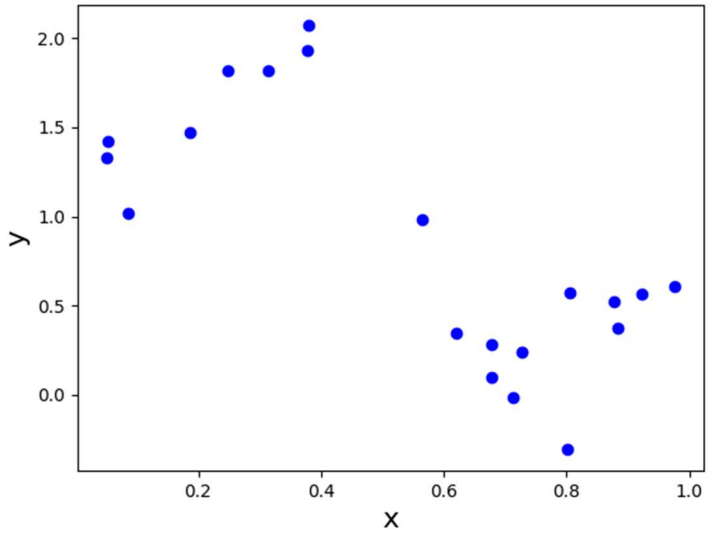
Part 3:

2. I used ChatGPT to clarify how different numpy methods worked. For example, I wanted to see what parameters the np.append method had so I could gain a better understanding of how to implement the code using np.append. It gave me accurate information and I was able to write my code quicker and more efficiently.

3. The training data does not seem to follow a linear pattern which suggests that a polynomial model might be better for this dataset. The test data looks weird because there is no real pattern of the data. If anything, it could be argued that it looks like it follows a polynomial distribution. However, the test data looks like it is clustered into two groups with one outlier. Here are the two graphs (first is train data, second is test data):

Cameron Scolari





4. Part 4:

α	iterations	coefficients
10^{-1}	14	[1.869, -4.015]
10^{-2}	192	[2.384, -2.879]
10^{-3}	1777	[2.442, -2.822]
10^{-4}	15777	[2.446, -2.817]

Cameron Scolari

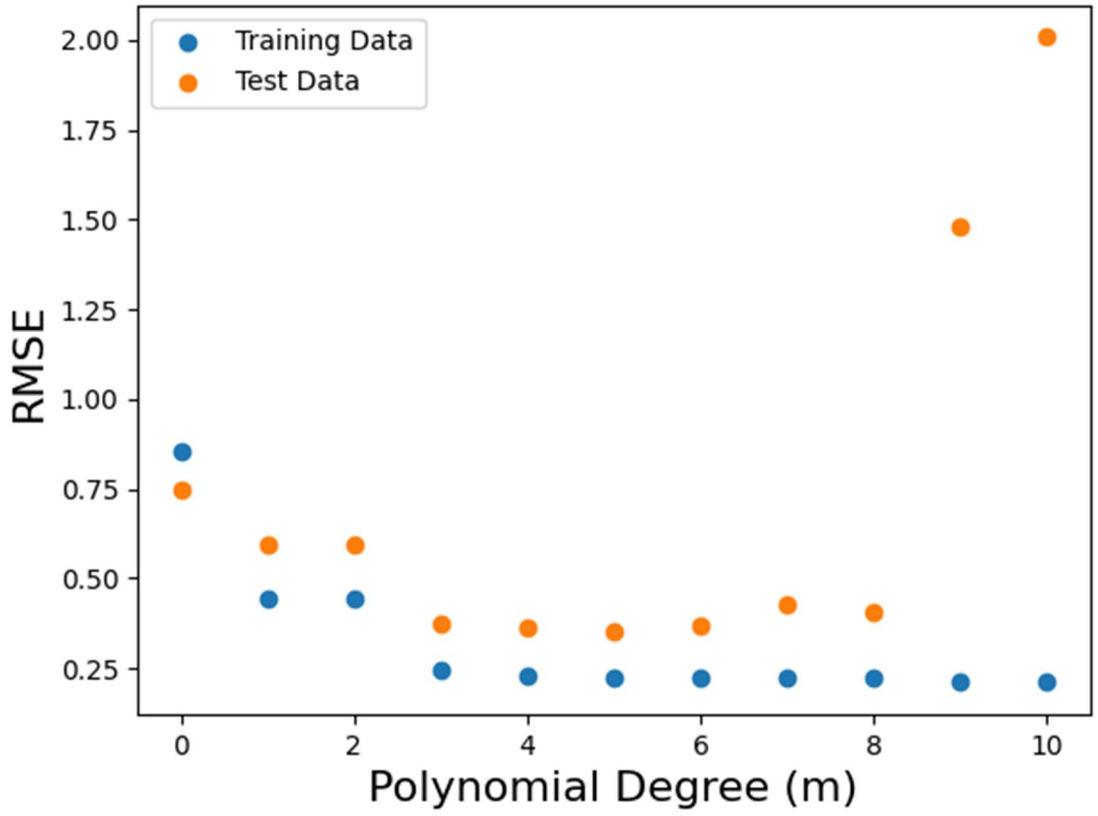
As the alpha value gets smaller, the number of iterations needed to converge increases. The coefficients all share the same signs. The first coefficient increases as the alpha level decreases. The second coefficient also increases as the alpha level decreases. The first alpha value has coefficients that are very different from those of the other alpha levels. The other alpha values have coefficients that are relatively close to each other.

Part 5: The coefficients obtained using SGD are [2.44248226 -2.82210644] and those obtained using the closed-form solution are [2.44640709 -2.81635359]. They look very similar and do not differ if rounding to one decimal place. After that, they differ slightly. Using Python's time module, we can see that: the SGD takes longer than the closed-form method. The closed-form method took 0 seconds when tested for multiple trials. The SGD method took somewhere around .3 seconds for each trial.

Part 6: I tried multiple equations for alpha and a lot of them took the maximum number of iterations. One algorithm I found that did not take the maximum number of iterations was 1/(1+(0.l*t)). It takes the algorithm 298767 to converge to [2.44639538 -2.81637388]. Another one I found was just 1/(1+t) and this one took 92454 iterations to converge to [2.44641303 -2.81638034]. This is only the case when we are using a normal linear regression model and not a polynomial regression model. With a polynomial regression model, the algorithm takes 1000000 iterations to converge.

Cameron Scolari

5. Part 9:



There are signs of overfitting at m=9 and m=10. This is because the model performs extremely well on training data but performs poorly on the test data. I would say there is underfitting at m=0 because the model performs poorly for both the training data and test data. This could also be argued for m=1 and m=2. Interestingly, the model performed better on the test data than the training data for m=0. I believe that the m=5 degree polynomial best fits the data because the test data's RMSE appears to be the lowest at this m value.

Cameron Scolari

0. 3 hours
1. I liked the ample comments in this ps that helped guide me on what to do. I also enjoyed the useful hints. The hardest part of this homework was understanding some of the grammar.