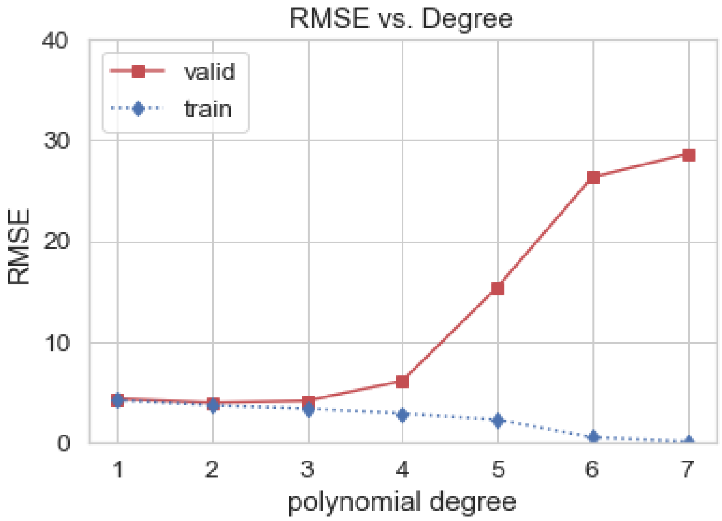
Homework 1

Matthew Carey

Problem 1:

This problem analyzed the RMSE of a linear regression model on a given set of data trying to predict the miles per gallon of a car based on features such as horsepower, weight, cylinders, and displacement. It used validation error analysis with a hyperparameter of the degree of the polynomial function. Below we can see the RMSE for both the train data and the validation data for each polynomial degree, along the x axis:

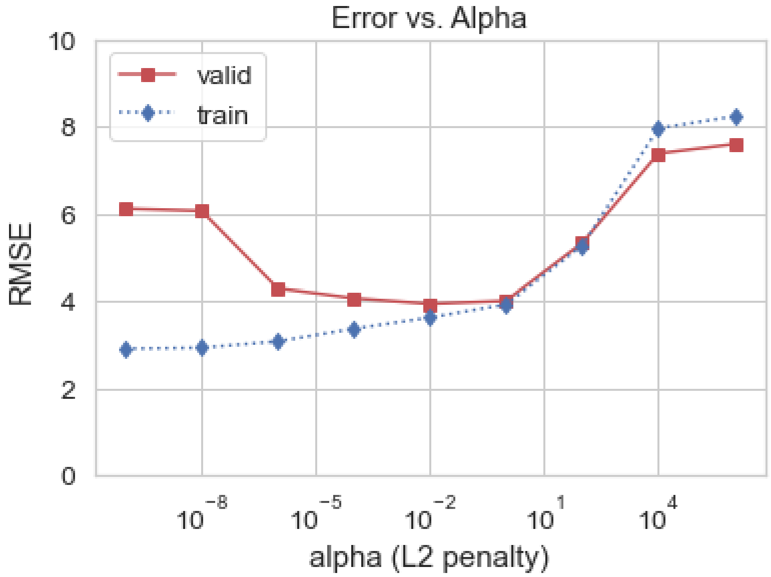


In this graph we can see a clear indication of what we learned in lecture, that while a high polynomial degree may have the capability to fit to the training data with incredible precision, the overfitting causes it to be incredibly sensitive and have significant error when exposed to unseen validation error. We can see that the polynomial degree that fit best when evaluated on unseen data was a degree of 2.

a. It is important to rescale the feature vectors so that they are all on the same order of magnitude, because this way when weights are applied no feature has a disproportionate effect on the outcome.
b. By looking at the weights applied to each feature in degree 1, we see that the weight applied to the weight of the engine is -18.23. This implies that increasing the weight of the engine decreases the mpg of the car as a whole, which makes sense that something that is heavier requires more fuel to move the same distance. On the other hand, displacement has a weight of +0.58, meaning that increasing the displacement of an engine has a positive effect on the mpg of the car, even though that larger displacement means a larger engine typically.
c. The degree 4 model has weights with magnitudes greater than 10,000, and many others in the range of thousands and hundreds. This is a strong indication of overfitting, which we can see because while it fits even better for the training data compared to lower degrees with lower weights, when it is evaluated on new data (the validation set) it does significantly worse, as the model is very sensitive to new data.

Problem 2:

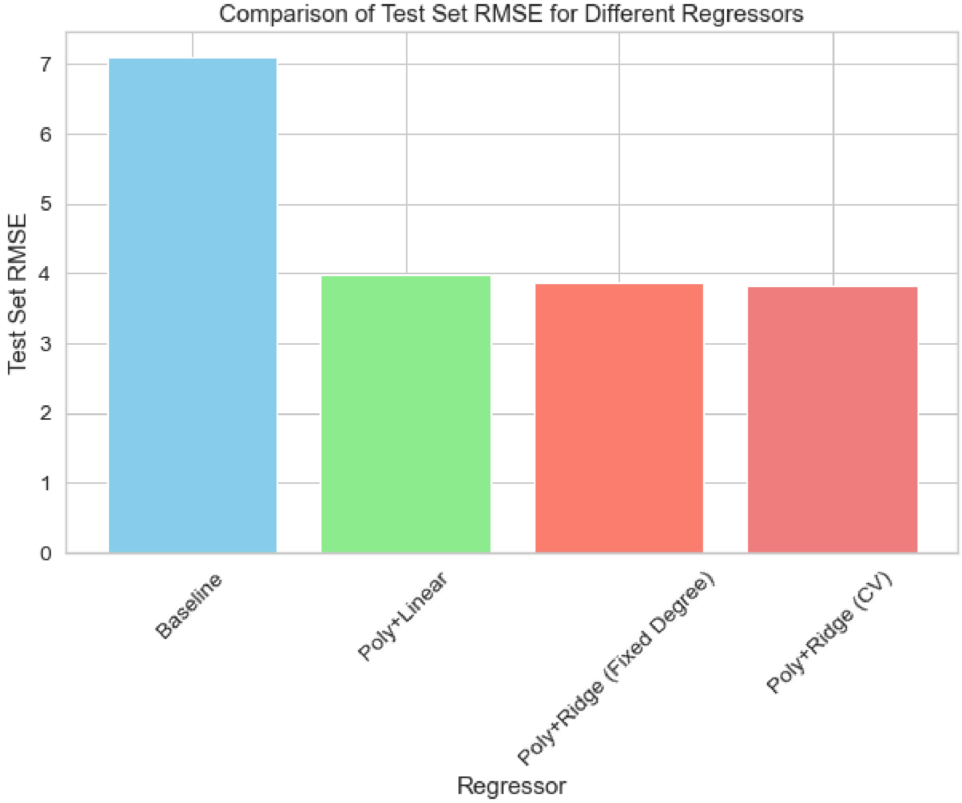
This problem aims to analyze the effects of penalizing large weights in models that are over fit. It fixes the polynomial degree value at 4, where we started getting overfitting in the last problem, and instead varies the value of alpha, which is a constant multiplier applied to the penalty term (A term that's size is proportional to the weights of the model, so penalizes large weights when we try to minimize the function). Below we can see the graph of the RMSE for each value of alpha in the studied range:

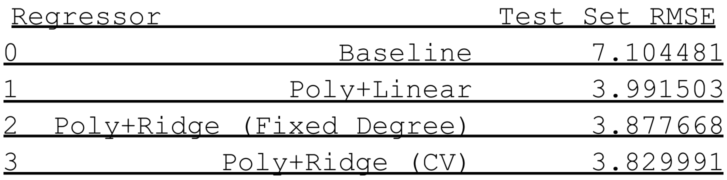


Looking at the train data, we see that increasing alpha continuously increases the error. This makes sense as the model is being fit to the train data, so when it is allowed to have large weights, it will overfit and fit almost perfectly to the train data. However by looking at the validation data we see that a larger penalty initially decreases the error, as the model is prevented from over fitting, though as we continue to increase the penalty the error increases again. This is because the model is likely underfitting to the data, as a result of it now preferring smaller weights that just do not map the data.

a. When we set alpha to the ideal value then analyze the weights of the model, we see that they are significantly less than they were for degree 4 in the previous model. This shows that the penalty term is effectively making our model prefer smaller weights and prevent overfitting.
b. The problematic aspect of using the training set loss to determine the value of α is that it may lead to overly complex models with high variance. The regularization term penalizes large weights, encouraging simpler models and reducing overfitting. However, if the regularization strength α is chosen solely based on the training set loss without considering generalization performance, the resulting model may not generalize well to new data.

Problem 3:





```
 Regressor                          Test Set RMSE
0                        Baseline        7.104481
1                     Poly+Linear        3.991503
2   Poly+Ridge (Fixed Degree)            3.877668
3                 Poly+Ridge (CV)        3.829991
```

We can see that the model that performs best with testing data when using penalty terms in conjunction with validation techniques, and specifically performs the best with cross validation and without a fixed degree. This is consistent with what we expect, as it has the most freedom to vary parameters and hyperparameters but has validation techniques in place to prevent the overfitting that comes with freedom.