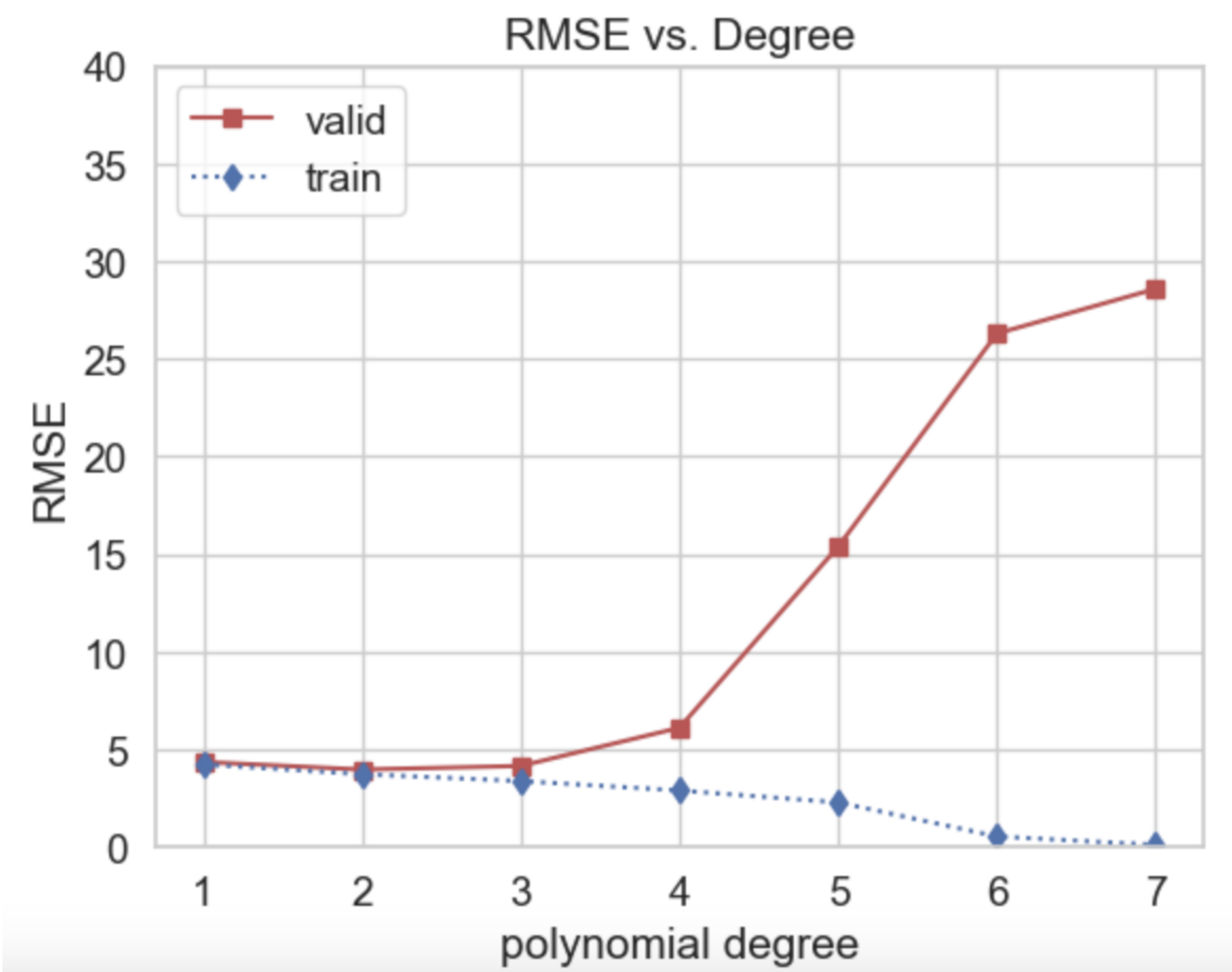
HW 1 PDF Report - Intro to ML (CS 135)

Figure 1 in Report



Caption of Figure 1 in Report:

Provide a 2 sentence caption answering the questions: Does this plot look as you expect based on course concepts? What degree do you recommend based on this plot?

This plot looks as what we expect based on course concepts as the RMSE for the train set should decrease to 0 and the RMSE for the valid set should increase as the polynomial degree becomes more specific to the train set. The degree I recommend based on this plot is 2.

Short Answer 1a in Report

The starter code pipelines include a *preprocessing* step that rescales each feature column to be in the unit interval from 0 to 1. Why is this necessary for this particular dataset? What happens (in terms of both training error and test error) if this step is omitted? *Hint: Try removing this step and see.*

```
In polynomial regression, preprocessing is part of the
preparing/cleaning process of machine learning, including feature
scaling and polynomial feature expansion. In the context of this MPG
problem, we have to do some form of min-max  scaling to ensure all
features such as weight (lbs) or horsepower (HP) are measured on
similar scales of 0 to 1. This is to ensure that 2000 lbs for a car
wouldn't be put on the same scale as 2000 cylinders (as 2000 lbs for a
car is normal (closer to 0.5) but 2000 cylinders should be considered
as an outlier or extreme (closer to 1). The 0 to 1 scale allows us to
make the data more understandable and make data with different
measurements comparable to each other. Also seen in this homework,
polynomial feature generation allows us to derive new features based
on the exponentiation of features. This allows us to capture the
nonlinear relationships in the data.
```

Short Answer 1b in Report

Consider the model with degree 1. Following the starter code, print out the values of all the learned weight parameters (aka coefficients). From these values, what can you say about how increasing the engine weight impacts the prediction of mpg? Does this make sense? What about if you increase engine displacement? Recall that displacement refers to the overall volume of air that can move through the engine. Larger engine means larger displacement.

```
-10.43 : x0
 -18.23 : x1
  -1.15 : x2
   0.58 : x3
where
x0 = horsepower
x1 = weight
x2 = cylinders
x3 = displacement

Thus, according to my results, if you increased cylinders, horsepower,
or weight the prediction of mpg should decrease (in increasing order).
However, by increasing engine displacement, the mpg should increase
```

(though this increase is insignificant compared to the total decrease caused by other features).

Short Answer 1c in Report

Consider the models with degree 3 or 4. Inspect the learned weight parameters, including the number of parameters and their relative magnitudes.

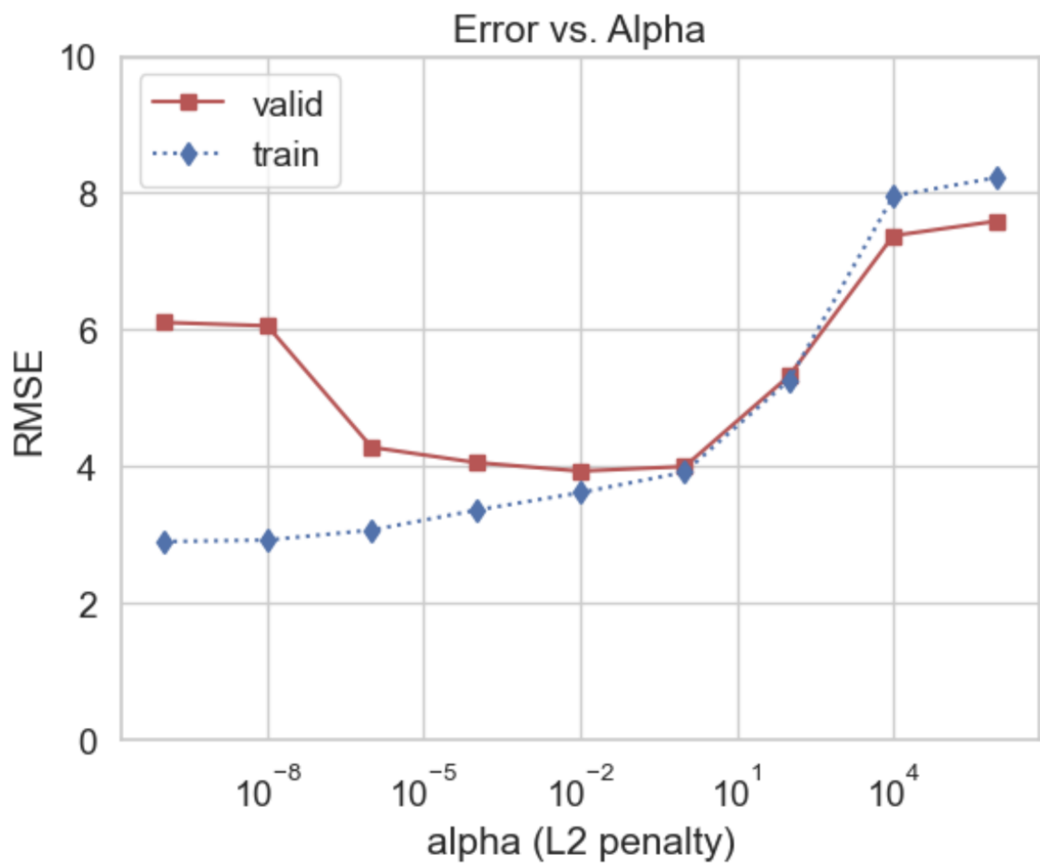
What do you notice about these values compared to the values with degree 1 or 2? How might what you notice be connected to the trends in training and validation set performance you observe in Figure 1?

Compared to the values with degrees 1, the number of parameters has increased by 30. In addition to the number of parameters increasing, the relative magnitudes for the model of degree 3 tended to be more exaggerated. For example, x2 x3^2 has magnitude -1696.77 and x2 has a magnitude -71.52, which both show the stark difference in coefficient values (coeff. in hundreds vs. tenths). This increase in parameters (which adds unneeded complexity) and difference in coefficient values largely has to do with what I explained in the caption: the model is overfitting to the training set.

Figure 2 in Report:

Make a line plot of RMSE on y-axis vs. alpha on x-axis. Show two lines, one for error on training set (in blue) and one for error on validation (in red).

Follow the styles in starter code. Be sure to show alpha values on x-axis on log-scale.



Caption of Figure 2 in Report:

Provide a 2 sentence caption answering the questions: Does this plot look as you expect based on course concepts? What alpha value do you recommend based on this plot?

This plot looks as what we expect based on course concepts as the RMSE is expected to create a U-shaped curve as both training and validation decrease with alpha starting off, and then slowly increase due to underfitting. As we did in class, the alpha value that should be recommended should be the point 1.e-04 as it relies on a lower alpha value and is where validation error is minimized and is far from over-regularization.

Short Answer 2a in Report

Inspect the learned weight parameters of your chosen degree-4 model. What do you notice about the relative magnitudes compared to 1c above?

Compared to the relative magnitudes to 1c, the learned weight parameters of my chosen degree-4 model are lower. For example, for parameter x2 x3^3, (degree-4 for 1c), the weight is -56683.53 compared to the parameter of my degree-4 model’s x2 x3^3, 10.32. This similar pattern is shown for most, if not all the values in both models.

Short Answer 2b in Report:

Your colleague suggests that you can determine the regularization strength `alpha` by minimizing the following loss on the *training* set:

What value of α would you pick if you did this? Why is this problematic if your goal is to generalize to new data well?

This is problematic if my goal is to generalize to new data well because by only focusing on reducing the loss on the training set, the model will end up being overfit to the training set (adding unnecessary complication and impractical weights). That said, as shown (above) in part 2, L2 regularization only reasonably trades fitting training data for maintaining simplicity but requires small weights. Most importantly, to generalize well to new data, the model should also maintain a low validation and test error and should be tested on both respective sets. Value of alpha picked would be 1.e-04.

Table 3 in Report:

In one neat table, please compare the test set root-mean-squared-error (RMSE) performance for the following regressors:

- Baseline: A predictor that always guesses the mean y value of the training set, regardless of the new test input
- The best Poly+Linear pipeline, picking degree to minimize val set error (from 1B)
- The best Poly+Ridge pipeline, fixing degree=4 and picking alpha to minimize val set error (from 2B)
- The best Poly+Ridge pipeline, picking degree and alpha to minimize 10-fold cross validation error (from 3B)

	Regressors	Performance
0	Baseline	23.481
1	1B Poly+Linear pipeline	3.992
2	2B Poly+Ridge pipeline	3.837
3	3B Poly+Ridge pipeline	3.781

Ranking of methods in descending order is 3B, 2B, and 1B. This is expected as 3B combines polynomial features with ridge regression based on multiple validation folds. 2B does less in comparison only using ridge regression with a fixed degree. 1B is expected to be the worst model as it works better with more linear relationships. The baseline is just a flat value, not comparable to the test errors.