Assignment 1 - CS 4375

Aadi Kothari
Andrew Su

I. Preprocessing

Using a fixed training/test split, these are the predicted values based on different preprocessing techniques.

MPG	SGDRegressor	Our Gradient Descent	Unprocessed data (w/o Scaling)	Scaling with mean only	No w0 in the weights vector.
22.3	26.6	27.1	5.0e+15	97.2	3.4
12.0	10.8	11.1	7.44e+15	-89.1	-6.8
25.4	22.7	23.5	6.09e+15	-145.9	3.9
28.0	26.0	26.1	3.7e+15	197.0	-0.4
24.0	28.9	29.4	4.9e+15	-2.6	4.3
38.0	27.7	27.1	5.1e+15	844.3	3.3
18.0	22.4	21.9	4.7e+15	689.6	-2.09
30.7	26.2	26.5	5.5e+15	-1008.9	4.86
38.0	34.5	34.4	3.4e+15	31.51	4.99
18.0	18.9	19.0	5.3e+15	765.7	-3.63
MSE	5.727	5.708	6.1e+31	128412	261

In our initial attempts, we tried to apply gradient descent to raw data except for the instances that had null values (marked with '?'). Our predicted MPG values were off by trillions because our attribute values range from very small values like cylinder count (4-8) to large values like weight (3000-7000). Then we tried scaling by adjusting the features by the mean of

each feature. This resulted in data many orders of magnitude closer, but still, there was a huge error. This can be explained by the fact that some features had relatively low standard deviations like year (all car model years were from 1970 to 1982) to high STD such as weight (2000 to 4000lbs).

After scaling for both standard deviation and mean (normalization), our MPG was off by a factor of 10. We realized that we needed an extra parameter for the bias. This gave us predicted values close to observed values (MSE: ~5.7) which indicated our preprocessing was sufficient.

II. Tuning the Gradient Descent Algorithm

Trial	Iterations	Learning Rate	MSE	R2
1	1000	0.00001	272.103	-7.9256
2	50000	0.00001	6.0372	0.8020
3	50000	0.001	5.72882	0.8120
4	5000	0.1	5.72883	0.8120
5	460	0.1	5.70834	0.8127

Our initial trials used a very low learning rate so we could avoid overshooting the minimum MSE. We had to increase the number of iterations so that our predictions could converge and match the observed values. We eventually got an MSE value of around 5.7 which levelled out after 30,000 iterations.

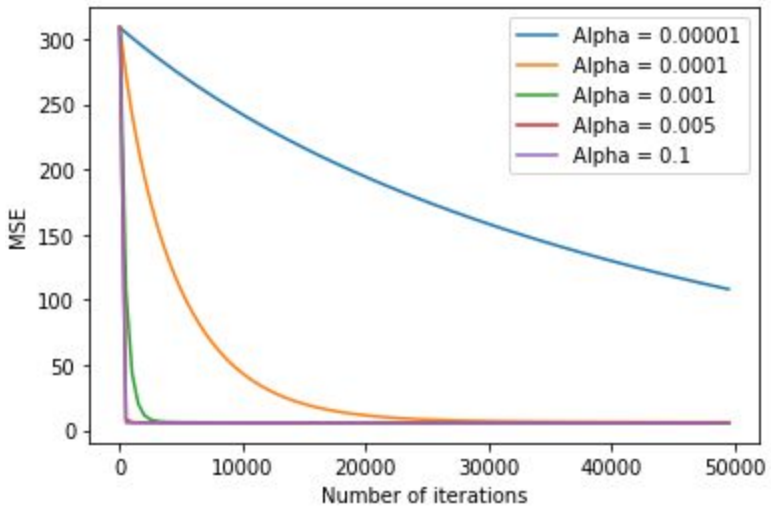

Figure 1: MSE vs Number of iterations plot for different learning rate

We decided to increase the learning rate so that MSE could be minimized earlier and use less iterations.

Our next goal was to optimize the number of iterations to avoid overfitting. Figure 2 shows MSE vs number of iterations, and at a quick glance the MSE seems to constantly to decrease but when we zoom in as in figure 3, we were able to find the optimal number of iterations after which overfitting starts.

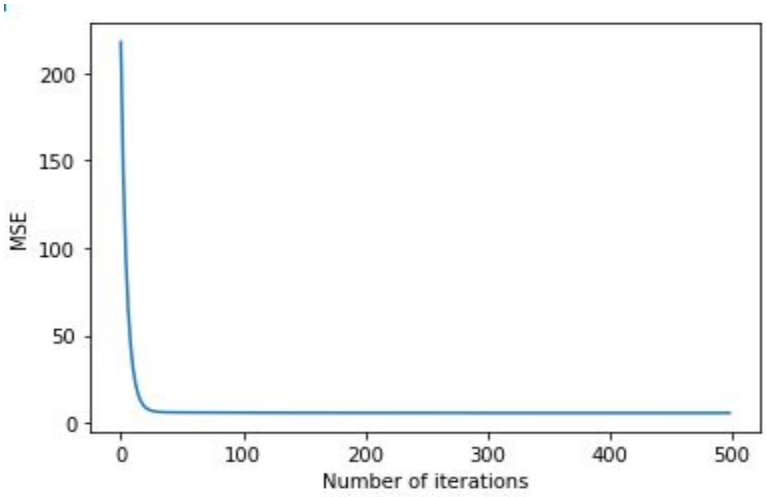

Figure 2: MSE vs number of iterations, plot for optimization

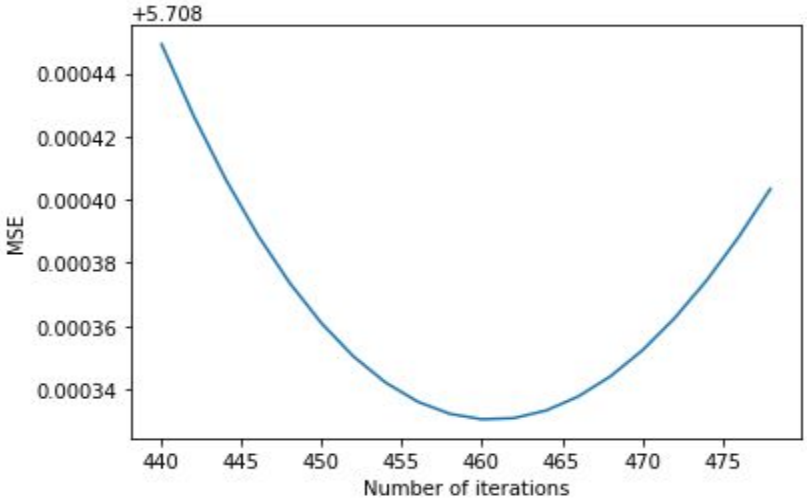

Figure 3: MSE vs number of iterations - zoomed in to find the optimal value

We are satisfied with this result because our MSE was 5.708, which is better than that of the SGDRegressor from the Scikit Learn Package. Also, our r2 value of 0.81 is very high and shows that most of the variation of the predicted MPG is from the variation of the observed MPG. It's very unlikely a predictor could get the MSE close to 0 because other factors could

affect MPG that are not in our dataset such as drag coefficient, transmission type, and engine type.