CS419: Assignment 1

Raghav Singhal, 19D070049

Task 1:

Value of RMSE error for the 3 cases:

Closed form (r1): 260930.36

Gradient descent (r2): 260549.44

Stochastic gradient descent (r3): 282489.00

1a) abs(r1-r2) = 380.91

1b) We stop gradient descent if the RMSE error on the dev set in the previous iteration is lesser than the RMSE error on the dev set in the current iteration.

1c) abs(r2-r3) = 21939.56

Task 2:

pnorm2: 260586.43

pnorm4: 260579.42

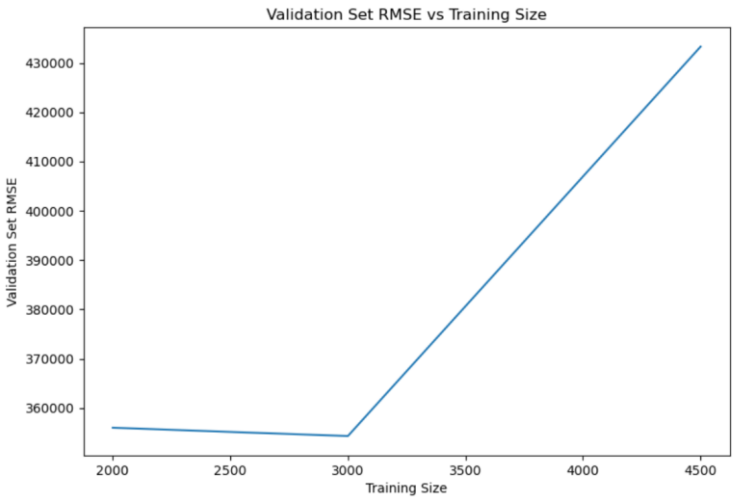
Task 3:

year: year^5. Upon checking the weights, the “year” feature had very high weights, showing that it is an important feature. To accentuate the difference in years, I raised to a higher power, and ^5 gave best results.

km_driven: km_driven^4. km_driven also had corresponding high weights, showing it is also a important feature. To accentuate the difference, I once again raised to a high power, and ^4 gave best results.

RMSE value with basis features: 246305.39

Task 4:



Task 5:

The least useful feature was “torque” since it needed too much pre-processing due to loads of variation in the column and no uniformity in the column.

The second least useful feature was “seats”. This was found by printing the weights and selecting the column with the smallest weights, which showcases a low dependence on the features.

Task 6:

I used p-norm with p=2 with the basis features as described in Task 3 for predicting the outputs.