Sheel Shah, 19D070052, CS419_assignment1_report

Note: the phase variable has been set to global in main(), so that other functions can use it

Values of RMSE error for three cases:
      closed solution (r1): 228262.45
      gradient descent (r2): 228440.45
      stochastic gradient descent (r3): 228775.83

1a. abs(r1 – r2) = 177.99
1b. If previous RMSE error on dev set was lesser than current epoch's error for 10 consecutive epochs, we would stop gradient descent. (and use the w with the best validation RMSE)
1c. abs(r2 – r3) = 335.38

2. pnorm2: 228029.56, pnorm4: 228435.39

3. basis: year was log scaled, since the years are all large numbers and 2010/2011 are very similar with respect to selling price. Km_driven was also log scaled because it is the scale and not the exact number that mattered. Furthermore, the km_driven filed had very large numbers hence their differences couldn't be accurately modelled linearly. Finally, the seats variable was exponentiated, in order to emphasise on the difference in number of seats.
RMSE: 259456.62

4.
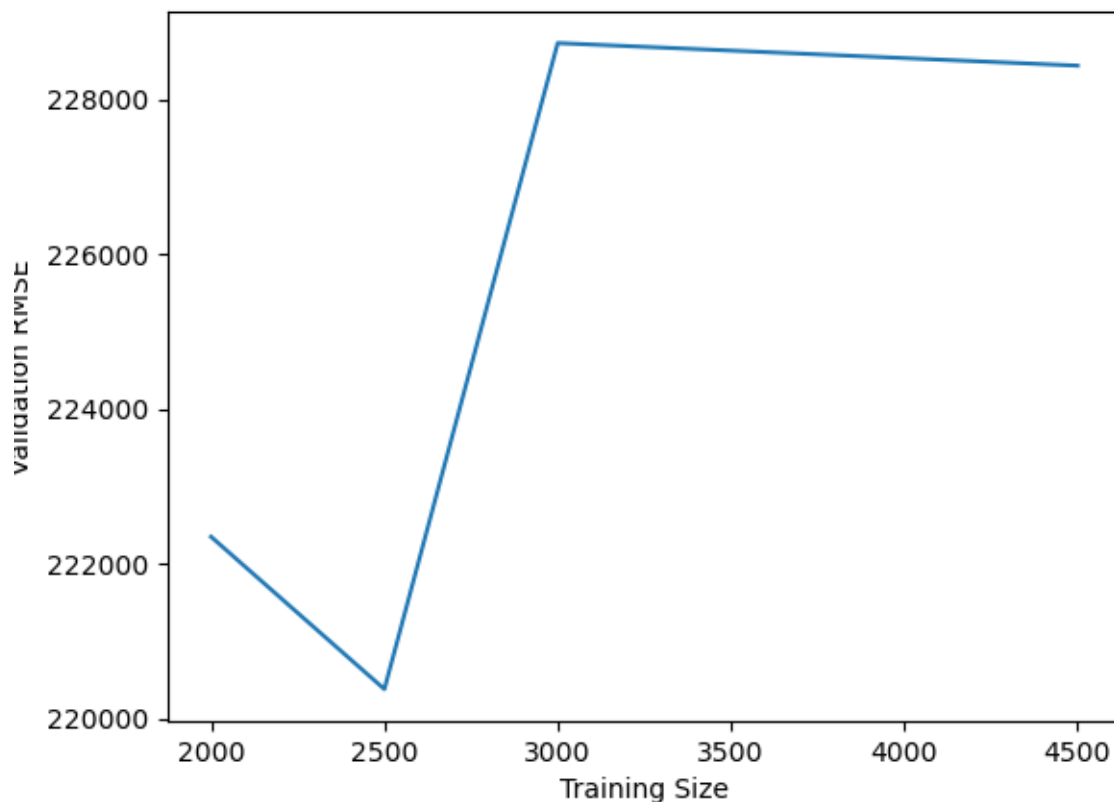
5. The least useful feature is torque because it needed too much pre-processing due to no consistency in the format of the column. Hence torque had to be dropped from the dataset. After normalizing the dataset, the feature with the smallest magnitude of its corresponding weight was decided to be the second least useful. This feature was seller_type. (cumulative weight was seen for features that were one hot encoded, along with plots of features vs price)

6. pnorm with p=2 was used, along with the basis mentioned above. Also, the model tended to predict negative prices, and these were converted to their absolute value. (this isn't exactly ethical, but since the question allowed any enhancement whatsoever, I decided to go ahead with this)

Screenshot of results:

```
D:\drbha\sem4_iitb_material\cs419_ml\assignment1>py lr.py
1a:
177.99274543632055
1c:
335.3798961303837
228262.4535686771 228440.44631411342 228775.8262102438
2: pnorm2
228029.56302513575
2: pnorm4
228435.3866536891
Task 3: basis
227271.75026181273
```