Machine Learning Assignment 2 Report
Topic: Ridge Regression
Name: Anirban Nath
Roll No.: 2104101001
Dept.: CSE [MS (by Research)

Implementing Ridge Regression: -

Ridge Regression was encoded into a function that takes training variable, target variable, regularization factor, learning rate and the number of epochs as input. The mode of regression was univariate and gradient descent was used to update the weight variables after each step, in order to reach as close to the global minima of the cost function as possible.

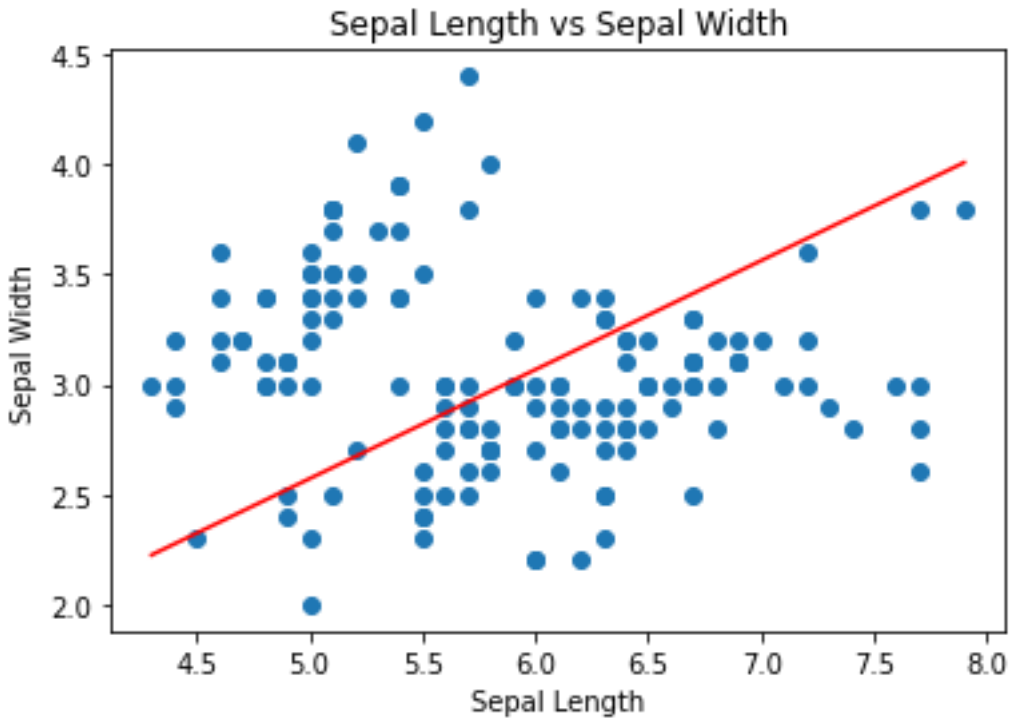
Testing Methodologies: -

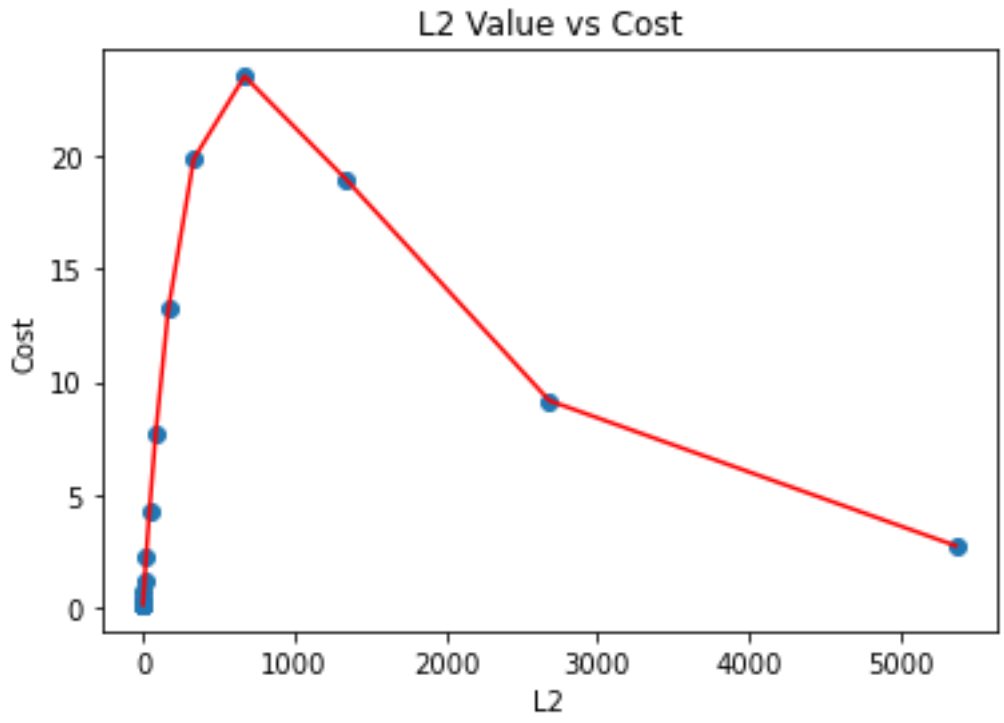
For each of the three datasets used, we recorded four graphical observations: -

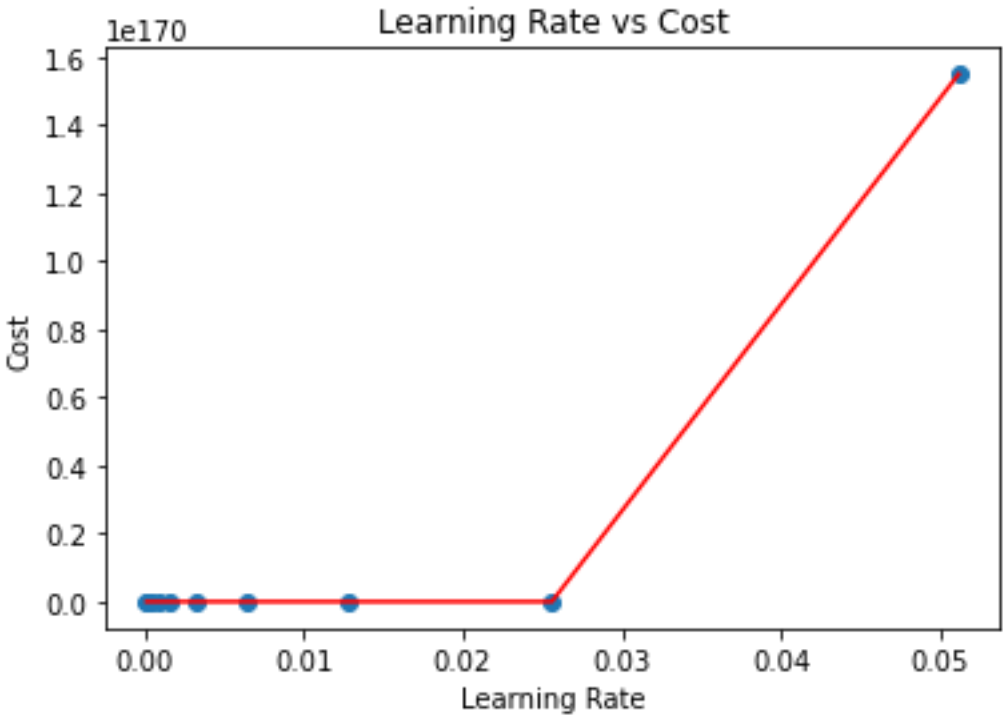
1. The fitting of the best-case regression line.
2. Variance of cost function by taking the initial value of the L2 regularizer as 0.0001 and then multiplying it by 2 at each step. The results from 30 such steps were recorded, keeping learning rate constant at 0.001 and epochs at 200.
3. Variance of cost function with learning rate. The initial value of the learning rate was taken as 0.001 and increased for 10 iterations, while keeping L2 value fixed at 0.0001 and running gradient descent for 300 epochs each time.
4. Variance of cost function with number of epochs. Kept L2 = 0.0001 and learning rate = 0.001. Started with 2 epochs and iterated 10 times for a final run of 1024 epochs.

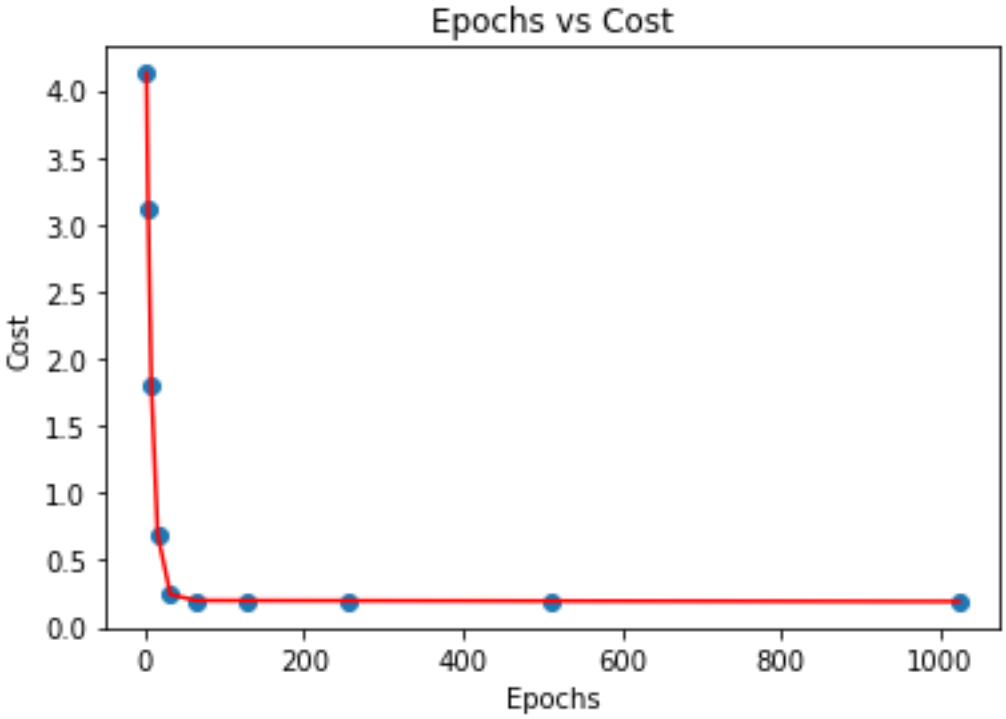
Performance Graphs: -

1. The first dataset used was the iris dataset. For our use, we chose to examine the correlation between sepal length vs sepal width. The following are the resultant graphs as mentioned in the methodologies section.

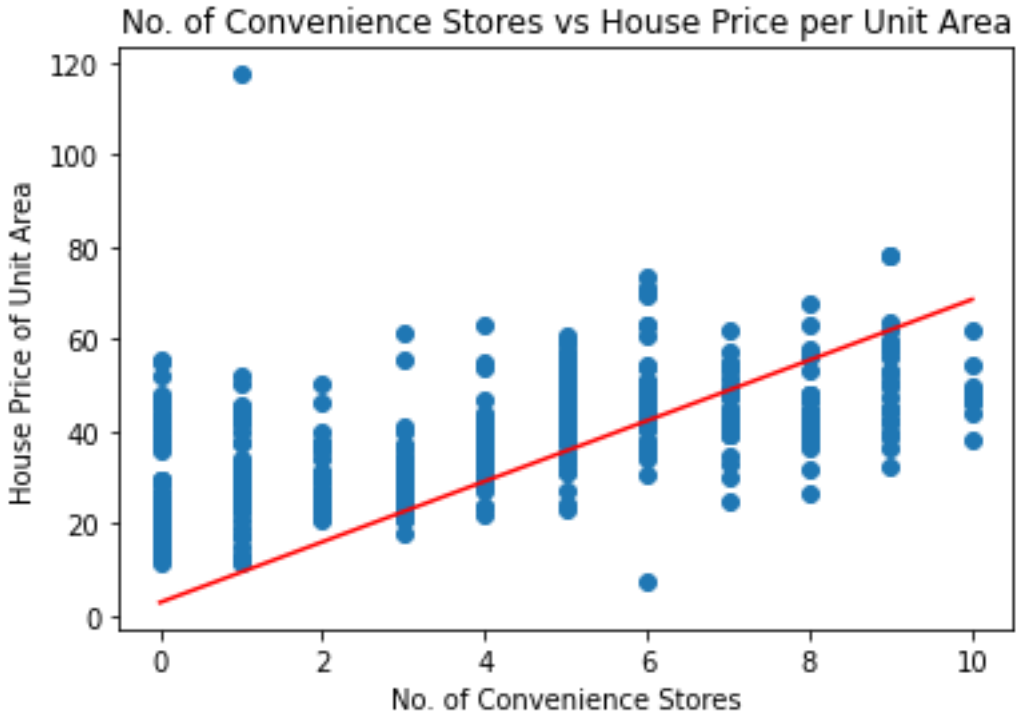


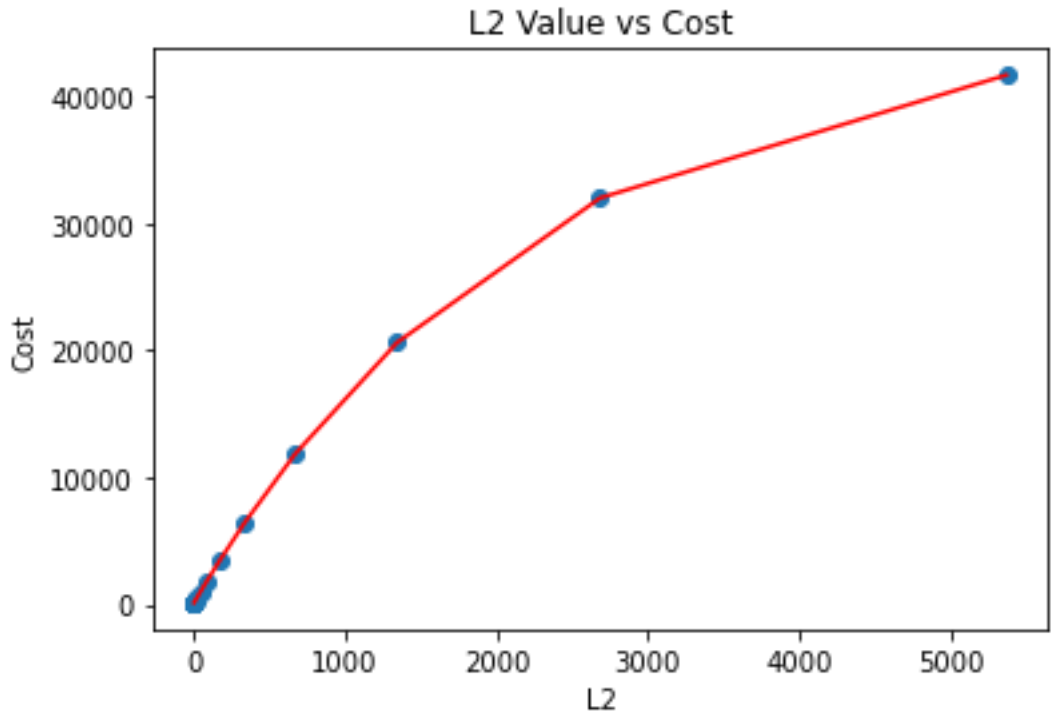


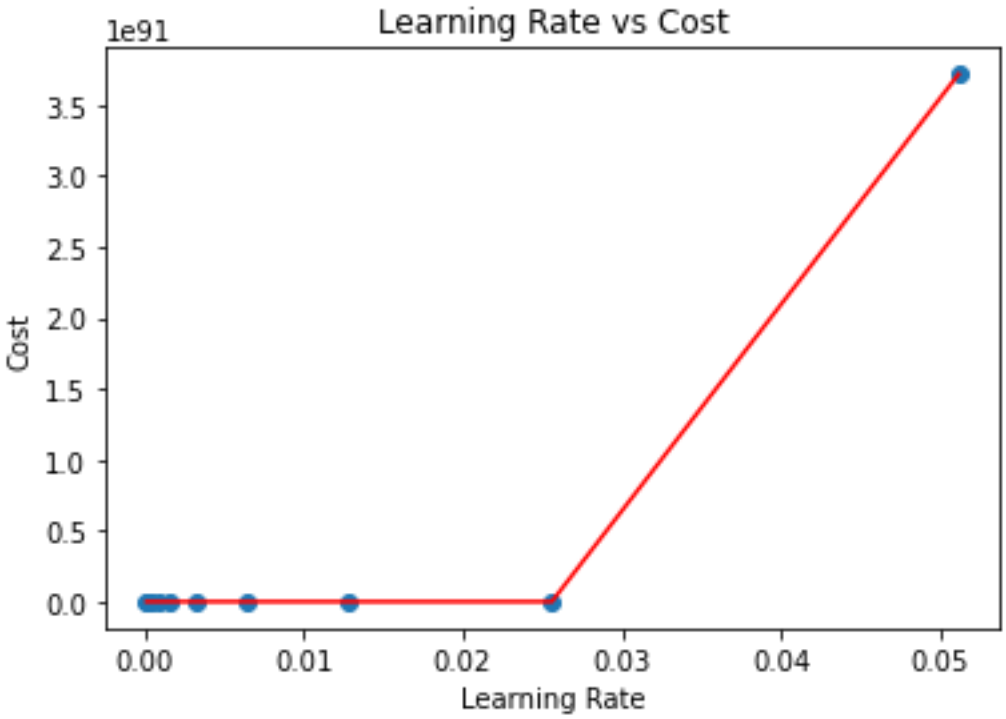

Learning Rate vs Cost
1e170
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Cost
0.00
0.01
0.02
0.03
0.04
0.05
Learning Rate

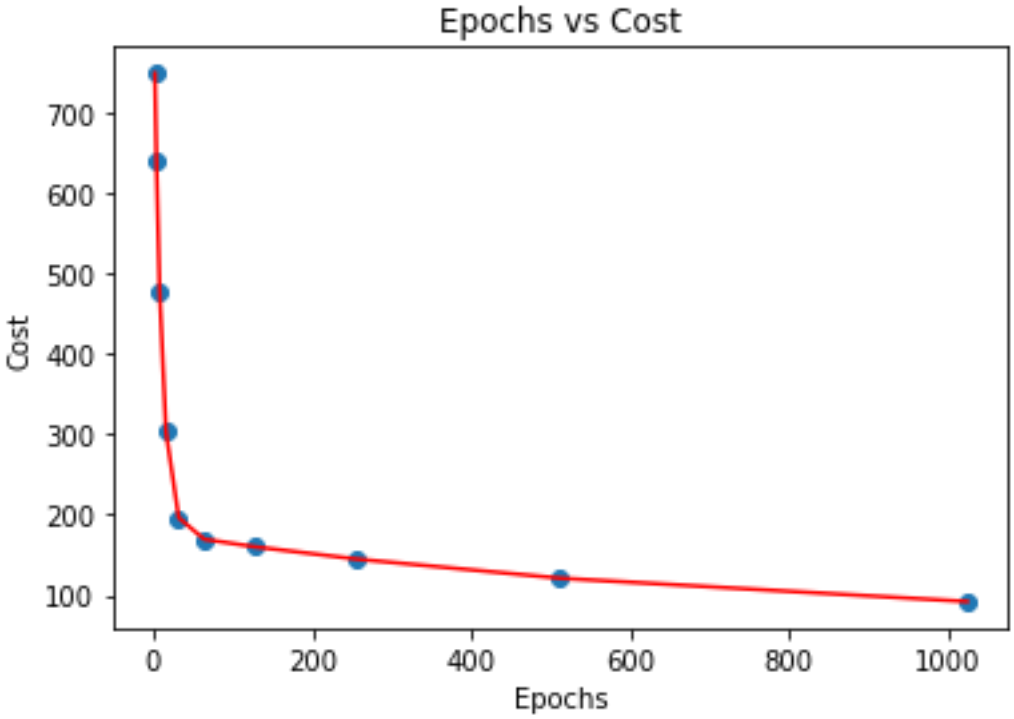

Epochs vs Cost
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Cost
0
200
400
600
800
1000
Epochs

2. The second dataset used was a housing price dataset. For our use, we chose to examine the correlation between the number of convenience stores in the vicinity of a given house vs the price per square unit of area. The following are the resultant graphs as mentioned in the methodologies section.

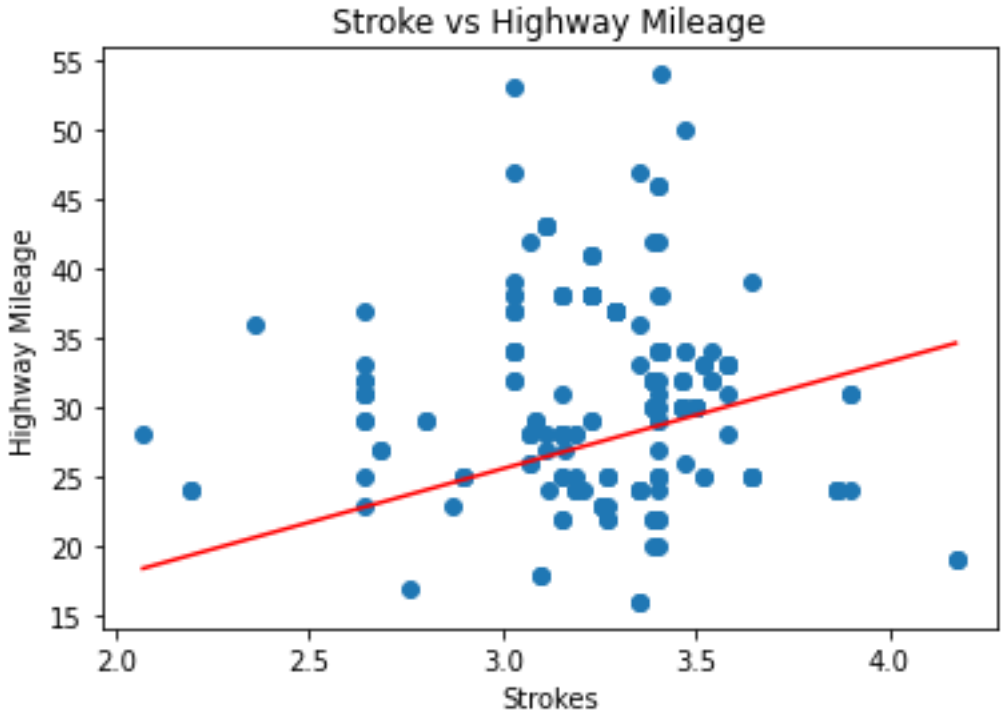


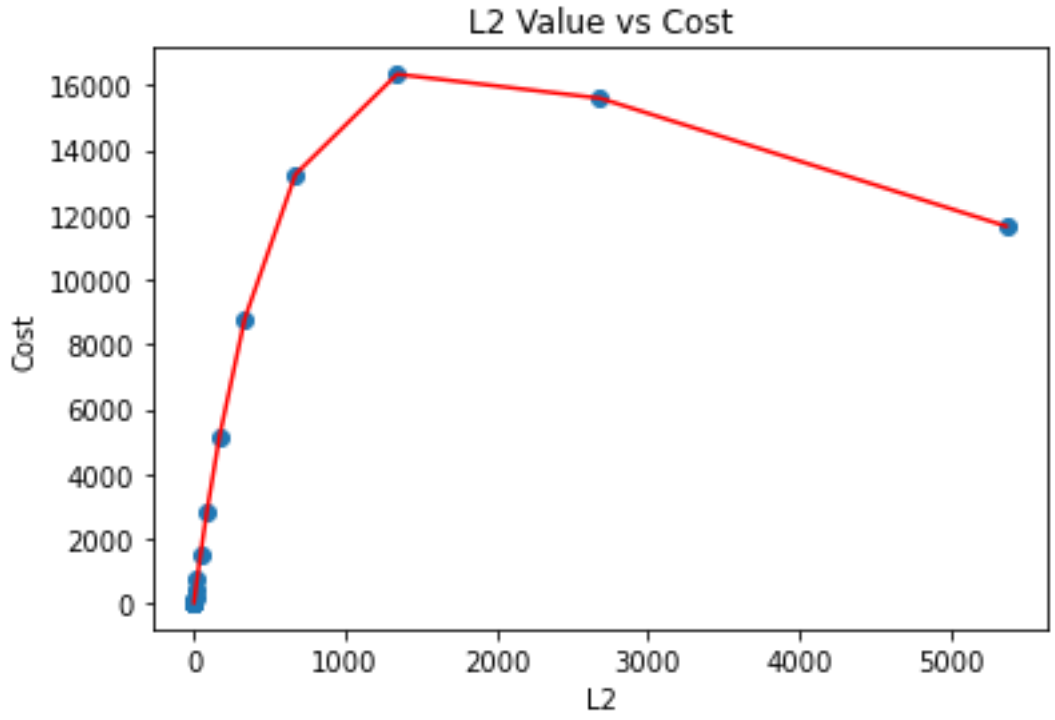


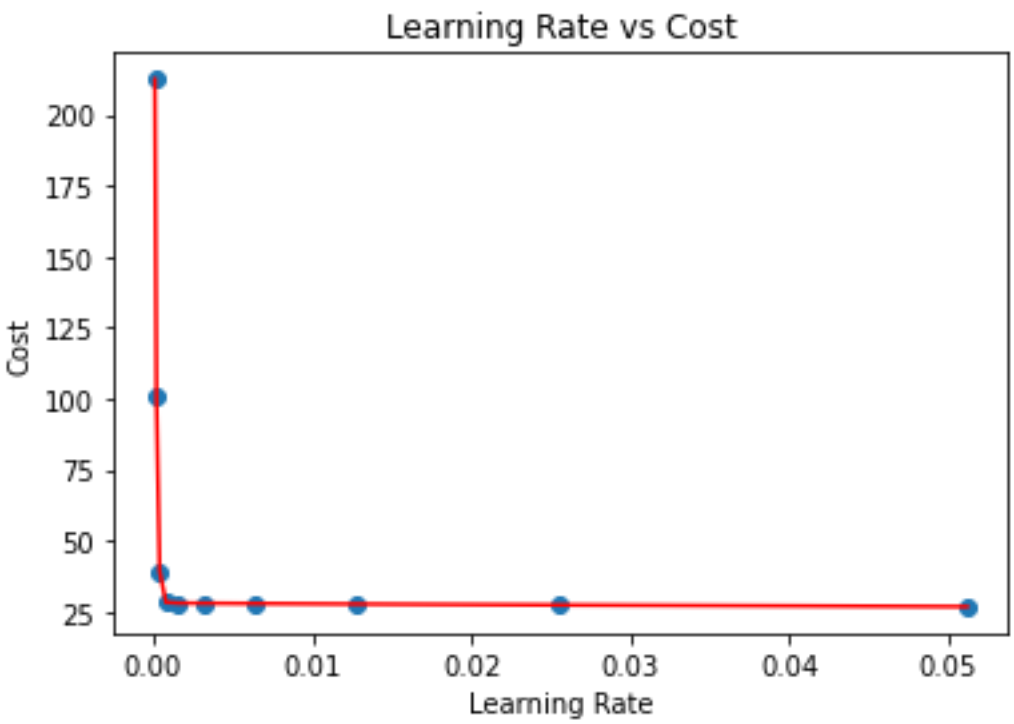

Learning Rate vs Cost
1e91
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Cost
0.00
0.01
0.02
0.03
0.04
0.05
Learning Rate

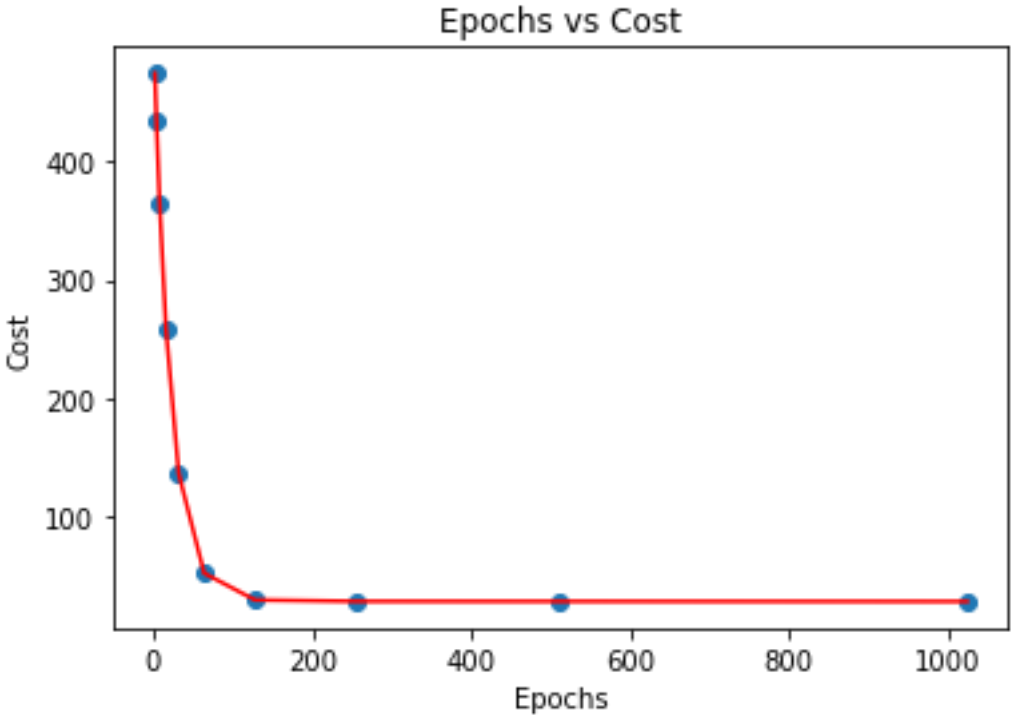

Epochs vs Cost
700
600
500
400
300
200
100
Cost
0
200
400
600
800
1000
Epochs

3. The third dataset used was a car features dataset. For our use, we chose to examine the correlation between the number of strokes vs the highway mileage of given cars. The following are the resultant graphs as mentioned in the methodologies section.






Learning Rate vs Cost
Cost
200
175
150
125
100
75
50
25
0.00
0.01
0.02
0.03
0.04
0.05
Learning Rate


Epochs vs Cost
Cost
400
300
200
100
0
200
400
600
800
1000
Epochs