Linear Regression, Polynomial Hypothesis and Regularization

TOKA ELMASRI

40-3252

Introduction

In machine learning, linear regression is an approach used to model the relationship between certain features and a target output. A single feature vs the target is called simple linear regression, while more than one feature vs the target is known as multiple linear regression. There are some enhancements that could be done to linear regression to help improve the accuracy. Some of these enhancements include using polynomial hypothesis that could fit the data better than the linear hypothesis, and using regularization where the simplest hypothesis that fits the data is chosen. In this assignment, the linear regression model with different approaches was used on the House Price Dataset to study the relation between the features and the output, calculate the cost function and minimize it as much as possible. The steps used will be discussed in the methodology section. The results achieved will be discussed in the results section and finally a conclusion section will summarize the models used and their results.

Methodology

A. Data Preparation

The first step was preparing the data by removing all the null/empty field if any where present. Next, we visualize the data by plotting each feature against the output which is the “price” to be able to visualize the relation and decide on the hypothesis that will be used in the polynomial hypothesis. Figures 1 to 5 show the plotting of each feature vs the price feature. After plotting and visualizing the relation, we calculate the correlation between each feature and the price feature to analyze the effect of each feature on the price. The features with correlation below 0.5 were removed as they don’t have a strong effect on the price. This step reduced the number of features from 21 to 5 features. These features have strong correlations with the price feature and some of them have linear relation visible from the graph plotting done in the previous steps. The final steps include splitting the data into 60% train, 20% cross validation and 20% test. Afterwards, we normalize the data to scale all features to the same scale, and we add a 1’s column for θ_0.

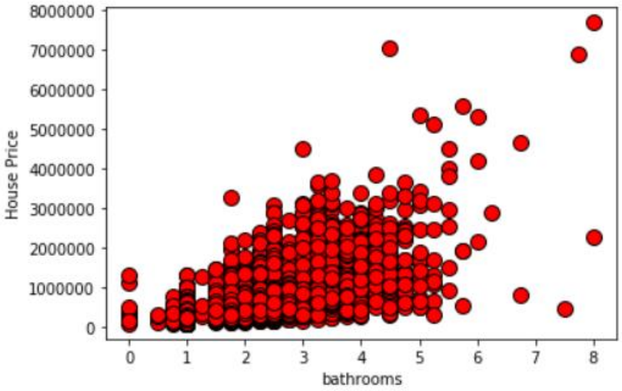

Figure [1]

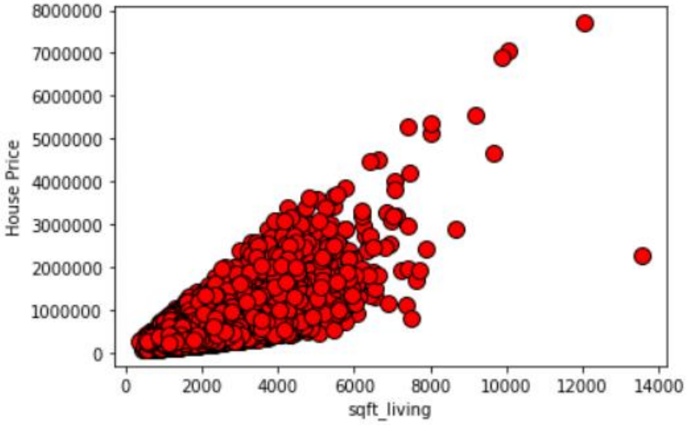

Figure [2]

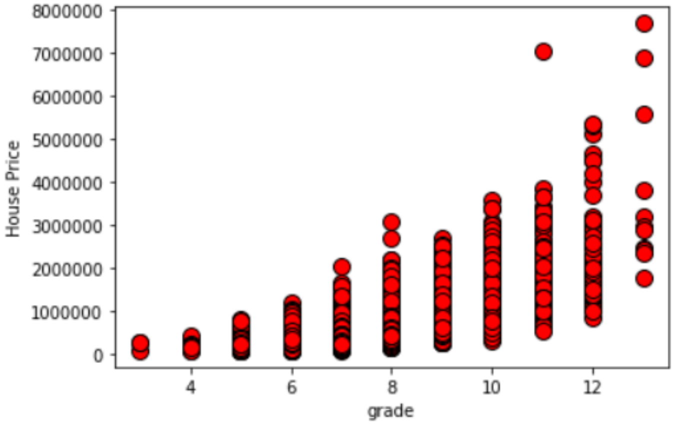

Figure [3]

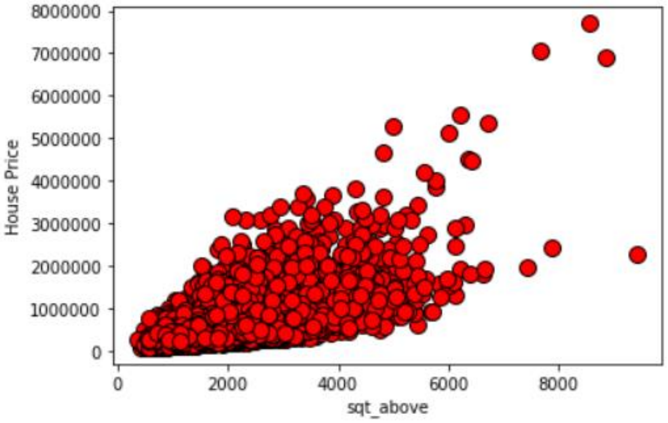

Figure [4]

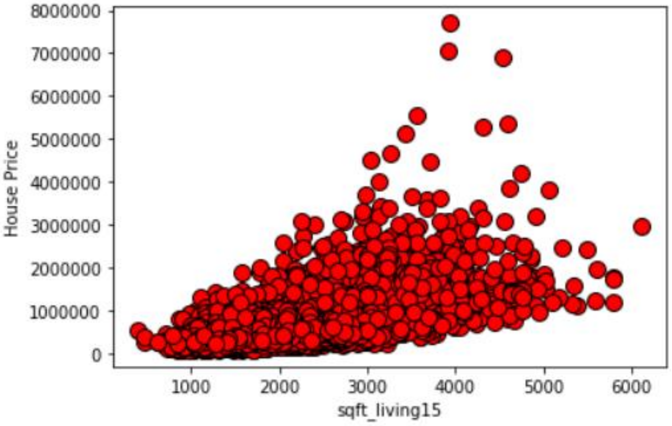

Figure [5]

B. Linear Regression Model

We start the linear regression model by implementing the cost function, gradient descent function and the normal equation function. We then apply the gradient descent and normal equation separately and calculate the cost function for each of these methods. Figure 6 shows the convergence graph for the gradient descent.

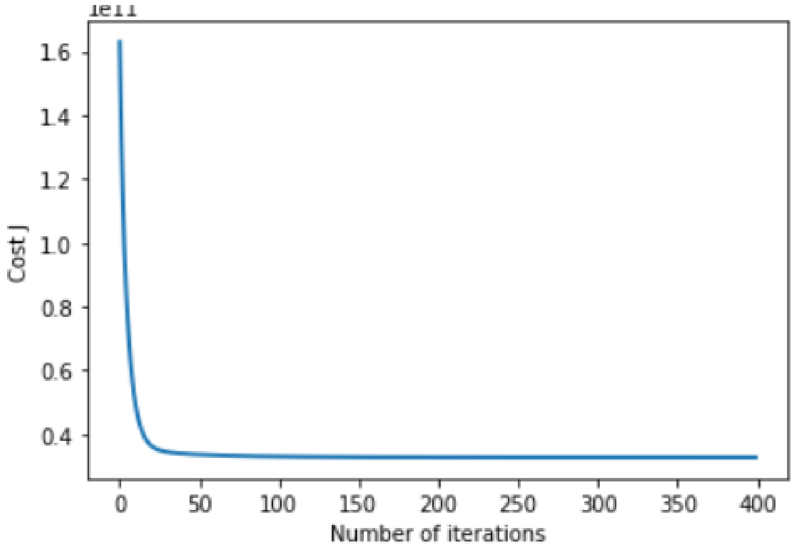

Figure [6]

C. Polynomial Hypothesis

Several polynomial hypothesis were constructed and evaluated using the gradient descent to be able to investigate the effect on the cost function and observe any increase or decrease. Polynomials of degree 2, 3, and 4 were used and a combination of degrees was used to test the cost function. The convergence graph for the polynomial degrees 2, 3, and 4 are shown in figures 7, 8, and 9 respectively. The combinations of degrees are [2,1,2,1,1], [3,1,3,1,1], and [4,1,4,1,1] for features 1, 2, 3, 4, and 5 respectively. Their convergence graphs are shown in figures 10, 11 and 12. The convergence graphs are similar; however, the cost function differs slightly between each polynomial hypothesis. This will be further discussed in the results section.

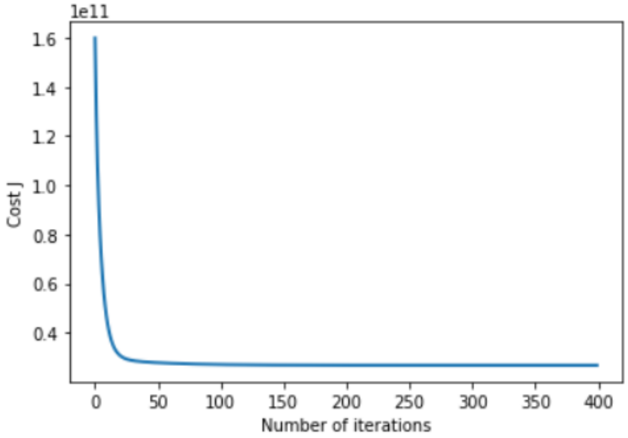

Figure [7]

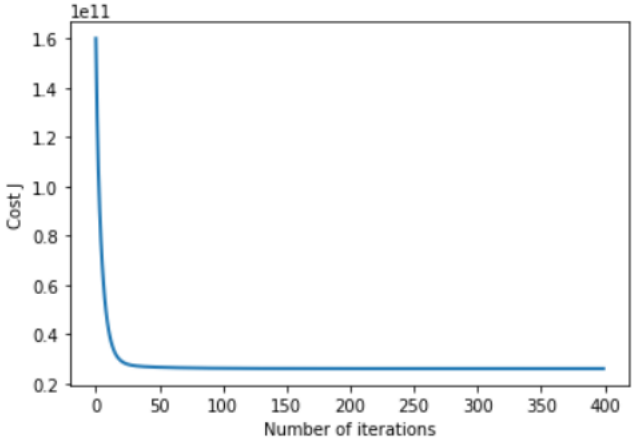

Figure [8]

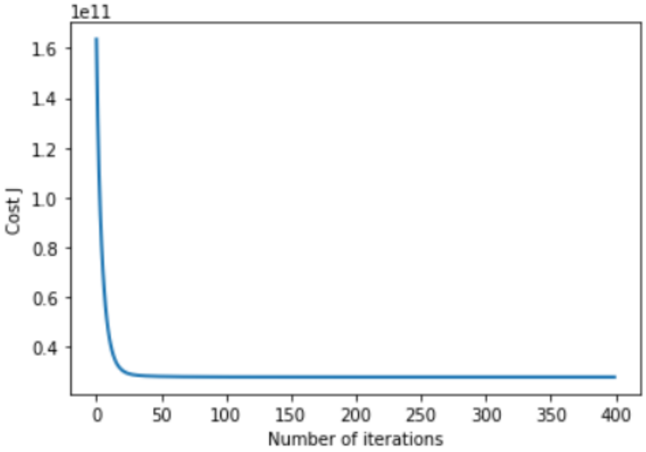

Figure [9]

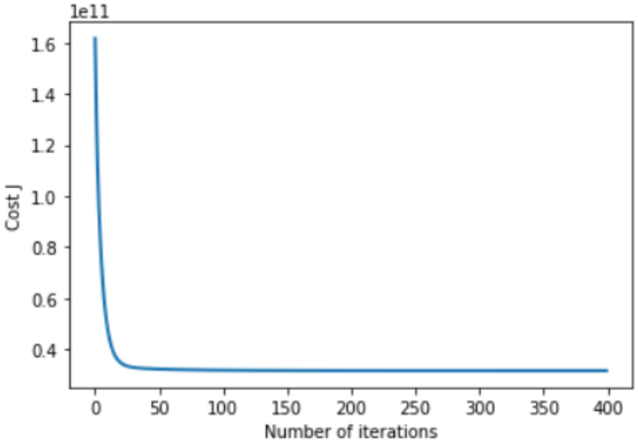

Figure [10]

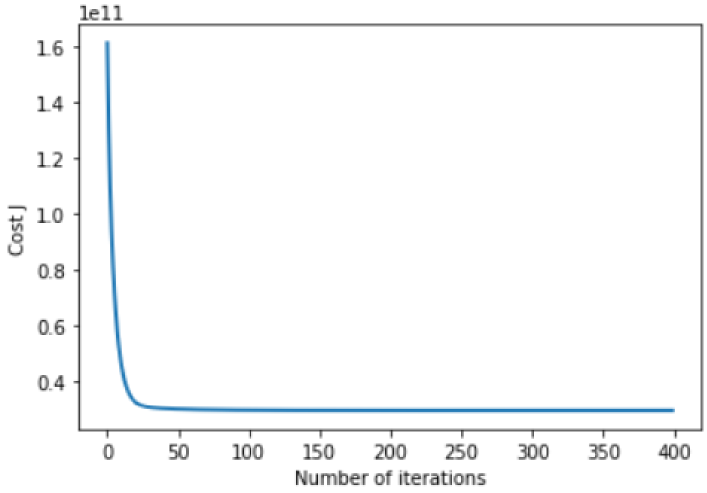

Figure [11]

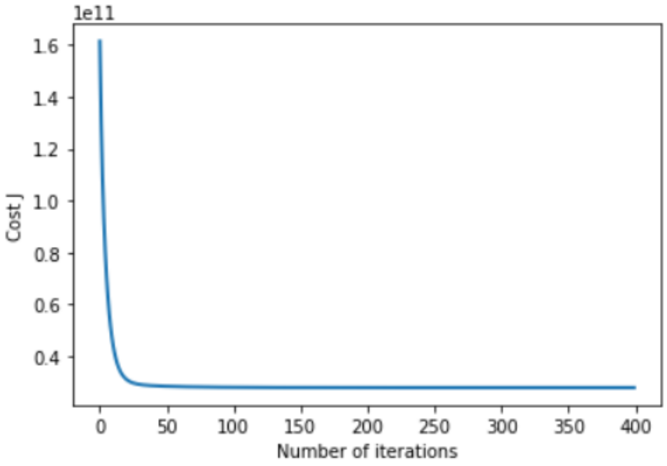

Figure [12]

D. K Fold

The K Fold method was used to sample the data and shuffle it for different iterations in order to get an average cost function which might help reduce the cost function. The K parameter chosen is equal to 10. The gradient descent and cost function were applied to each set in each iteration and the test cost functions of the 10 iterations were averaged. For simplicity, only a train and test set were used and a linear hypothesis was applied. The results will be discussed in the results section.

F. Regularization

Similarly, the steps applied before such as the linear hypothesis, polynomial hypothesis and K-fold were applied to the regularized gradient descent and cost function. This was done to examine any difference that could appear in the cost function. The convergence graph for the linear hypothesis is shown in figure 13 and for the 2nd degree polynomial is shown in figure 14. The same conclusion can be made, that the graphs look similar, however the cost function values differ slightly.

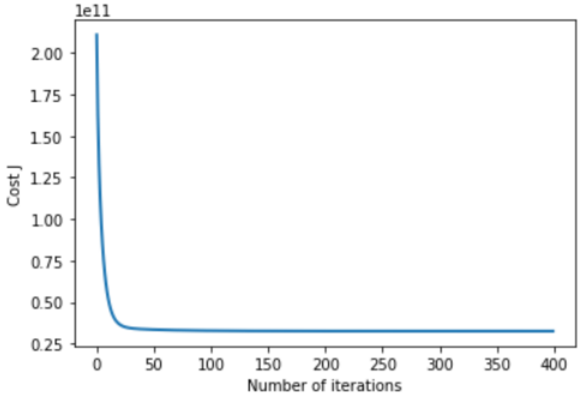

Figure [13]

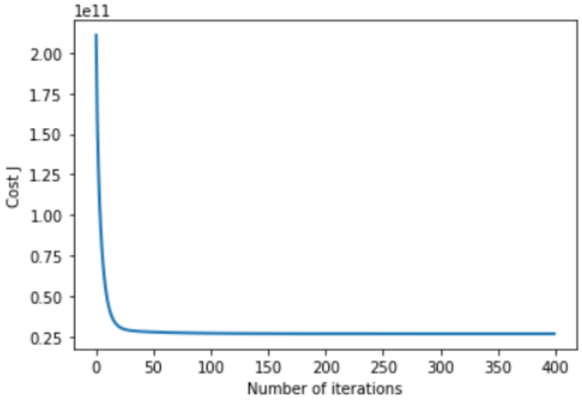

Figure [14]

G. Parameters

The parameters used for the linear regression, polynomials and K-Fold are ***alpha = 0.1*** and ***number of iterations = 400.*** The parameters used for the regularization model are ***alpha = 0.1, number of iterations = 400,*** and ***lambda = 0.5***. These parameters were chosen by trying different alphas, iterations and lambdas, and choosing the ones that give the minimum cost function.

Results

The cost functions of the train set, cv set and test set for all the model approaches used are shown in tables 1-17.

Set	Cost Function
Train	32629699005.660954
CV	27984019629.227573
Test	27449967371.79753

Table [1] Linear Hypothesis Gradient Descent

Set	Cost Function
Train	32629488121.98338
CV	27987073492.49158
Test	27456738713.7441

Table [2] Linear Hypothesis Normal Equation

The gradient descent and normal equation display the same range of cost function for the linear hypothesis, however the gradient descent has a smaller cost function.

Set	Cost Function
Train	26758129957.2129
CV	29956612688.111473
Test	25344094108.386795

Table [3] Polynomial Hypothesis Degree 2 Gradient Descent

Set	Cost Function
Train	26092785778.789345
CV	41700970940.653435
Test	25709983108.1827

Table [4] Polynomial Hypothesis Degree 3 Gradient Descent

Set	Cost Function
Train	28034797795.831837
CV	49806354624.24752
Test	27505652818.17719

Table [5] Polynomial Hypothesis Degree 4 Gradient Descent

Set	Cost Function
Train	31588003509.2547
CV	27346400878.555122
Test	26559846882.020172

Table [6] Polynomial Hypothesis combination 1 ([2,1,2,1,1]) Gradient Descent

Set	Cost Function
Train	29427065539.117912
CV	27112680113.681103
Test	25612575529.1467

Table [7] Polynomial Hypothesis combination 2 ([3,1,3,1,1]) Gradient Descent

Set	Cost Function
Train	27998327902.978558
CV	27592451233.004025
Test	25072028662.74415

Table [8] Polynomial Hypothesis combination 3 ([4,1,4,1,1]) Gradient Descent

Regarding all the cost functions for the polynomial hypothesis achieved above, the lowest cross validation cost function reached was for the polynomial combination 2 which is [3,1,3,1,1] with cost function equal to 27112680113.681103. However, the polynomial hypothesis with combination 3 [4,1,4,1,1] has the lowest test cost function. It also has a lower train cost function that the polynomial hypothesis 2. Therefore, these two combinations reduced the cost function in comparison with the linear hypothesis.

Set	Cost Function
J Test Average	30678990022.922386

Table [9] K-Fold Linear Hypothesis Gradient Descent

The K-fold with linear hypothesis used did not improve the cost function as it produced a cost function higher than the linear hypothesis and polynomial hypothesis.

Set	Cost Function
Train	32632880368.06238
CV	27984019629.227573
Test	27459510575.29002

Table [10] Linear Hypothesis Regularized Gradient Descent

Set	Cost Function
Train	26759956850.23836
CV	29956612688.111473
Test	25349574279.992893

Table [11] Polynomial Hypothesis Degree 2 Regularized Gradient Descent

Set	Cost Function
Train	26094120303.232925
CV	41700970940.653435
Test	25713986310.812214

Table [12] Polynomial Hypothesis Degree 3 Regularized Gradient Descent

Set	Cost Function
Train	28035930191.365376
CV	49806354624.24752
Test	27509049690.22349

Table [13] Polynomial Hypothesis Degree 4 Regularized Gradient Descent

Set	Cost Function
Train	31589286529.93702
CV	27346400878.555122
Test	26563695587.672493

Table [14] Polynomial Hypothesis combination 1 ([2,1,2,1,1]) Regularized Gradient Descent

Set	Cost Function
Train	29428336256.01244
CV	27112680113.681103
Test	25616387326.85336

Table [15] Polynomial Hypothesis combination 2 ([3,1,3,1,1]) Regularized Gradient Descent

Set	Cost Function
Train	27999675819.900352
CV	27592451233.004025
Test	25076072039.088154

Table [16] Polynomial Hypothesis combination 3 ([4,1,4,1,1]) Regularized Gradient Descent

Regarding the regularized gradient descent, the cost functions achieved for all the hypothesis are in the same range as the normal gradient descent for the same hypothesis.

However, the normal gradient descent achieved lower cost functions. The lowest cross validation cost function was for the polynomial combination 2 ([3,1,3,1,1]) with cost function equal to 27112680113.681103. Similar to the normal gradient descent, the lowest test cost function as achieved by the polynomial combination 3 ([4,1,4,1,1]) and was equal to 25076072039.088154. The train cost function is lower than the polynomial 3 combination. Therefore, polynomial combination 2 and 3 achieve the best results.

Set	Cost Function
J Test Average	30687907653.54081

Table [17] K-Fold Linear Hypothesis Regularized Gradient Descent

The K-Fold for linear hypothesis with regularized gradient descent achieved a higher cost function than the K-Fold with the normal gradient descent. Moreover, it has a higher cost function than all the other hypotheses tested with the regularized gradient descent.

Conclusion

To conclude this assignment, the linear regression model with normal gradient descent and regularized gradient descent was implemented using linear and polynomial hypothesis. The K-Fold approach was also applied using the linear hypothesis with normal gradient descent and regularization. These approaches/methods were implemented to observe how the cost function varies and to be able to evaluate the model without underfitting or overfitting the model, in parallel to minimizing the cost function. In addition, the lowest cost functions were achieved by the polynomial hypothesis with the normal gradient descent. The polynomial combination 2 achieved the lowest cross validation cost function, while the polynomial combination 3 achieved the lowest test and train cost functions. Therefore, these two hypotheses would be a good evaluation for the model depending on which cost function we want to minimize the most. The regularized gradient descent showed a similar pattern to the normal gradient descent; however, it achieved higher cost functions than the normal gradient descent. The K-Fold achieved a higher cost function for both the gradient descent and regularized gradient descent than the model without K-Fold.