# Machine Learning

# Assignment 1

# Implementing Linear Regression

Vinay Mohan Behara

Wright State University

## 1. Abstract

A Computer Program is said to learn from experience(E) with respect to some task(T) and performance measure(P), if its performance on T as measure by P that improves with experience (E) is termed as a well posed learning problem. -(Tom Mitchell). In a Regression Problem, we will try to predict results within a continuous output where we try to map input variables to some continuous function. The accuracy of hypothesis function is measured by cost function. In order to minimize the Cost Function, we use Gradient Descent Algorithm. In this algorithm, we have to choose theta values so that the hypothesis is close to the training examples.

## Question 1

**Linear Regression with one variable**

**Introduction:**

The data set provided for this is fluML which has some null values so I have added the average of all the values in the column in place of null values considering KnowTrans, Risk and RespEtiq and I have deleted the row where the RespEtiq has a value 9. This gave me a data set containing 410 x 18 matrix.

**Procedure:**

Step1: Import the data from the given data set and load the required KnowTrans and Risk.

Step2: Calculate the length of the matrix using length() function and initialize the theta, iterations and alpha values.

Step3: Plot a graph between the input variable and the Target Variable using Plot function.

Step4: Calculate the Gradient Descent using the algorithm

repeat until convergence:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

where

j=0,1 represents the feature index number.

The Cost Function is calculated using

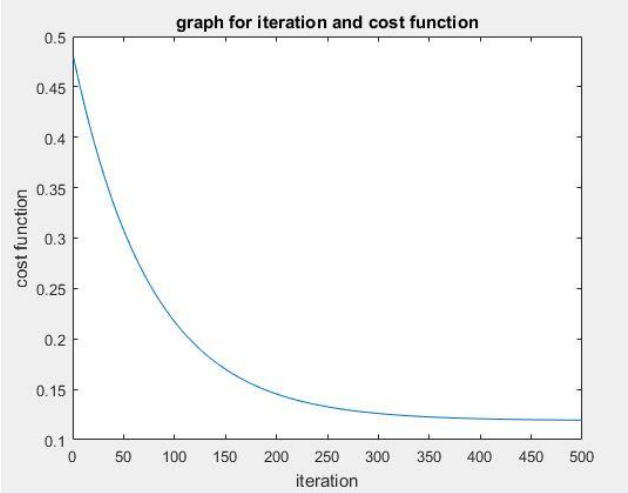
Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

Parameters: $\theta_0, \theta_1$

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

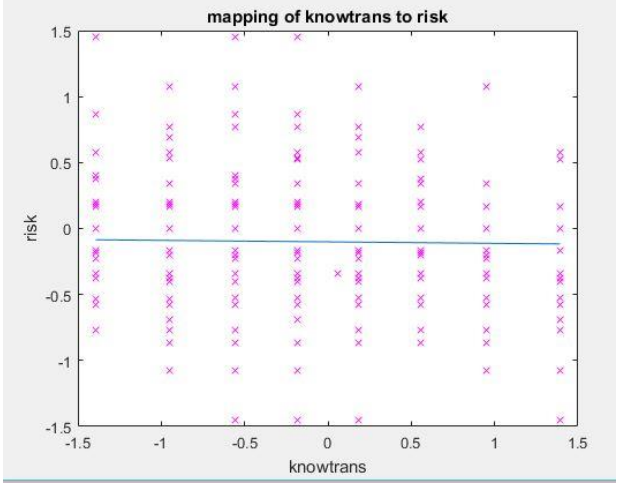
Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

The output from this algorithm is the theta values and the Cost function values.
In the Gradient Descent Algorithm, we start with theta values and keep on changing them to reduce the cost function.
Step5: Plot a graph for the hypothesis and the initial variable.
Step6: Plot a graph for the iterations and the cost values.
Step7: Plot a 3d graph for the theta1, theta2 and the cost values.
**Outputs:**
The graphs for the dataset with 350 observations
The Graph for iterations and the cost function:
The total number of iterations are 500 the value of learning rate (alpha) = 0.01




In the linear regression as the iterations increases, the cost function keeps on decreases and it comes to stable state after some iterations.

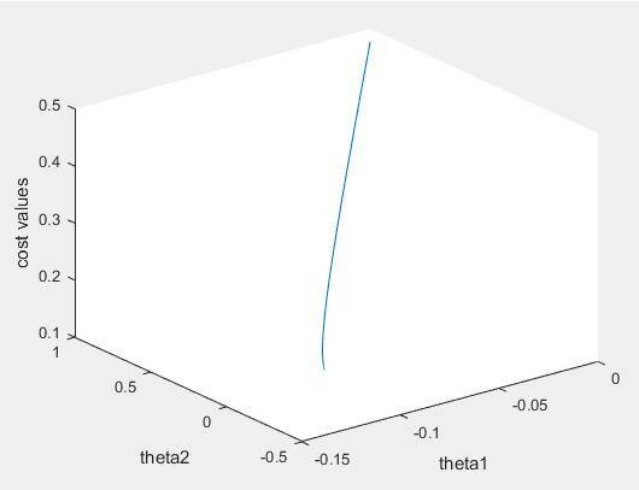
The graph for KnowTrans and the Risk with the obtained hypothesis



The hypothesis is given as `h = -0.1010 -(0.0113 *x);`

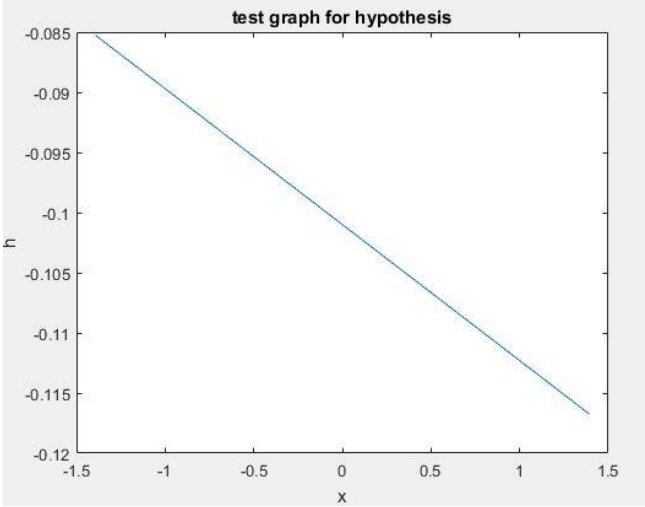
Where the obtained theta values are theta1 = -0.1010 and theta2 = -0.0113

The 3d Graph for theta1, theta, cost:




The obtained minimum cost when I took 350 observations is 0.1194 and the values of theta where I got the global optimum are theta1 = 0.1010 and theta2 = 0.0113.

The graph for the test data of the remaining 60 observations with the hypothesis and the values



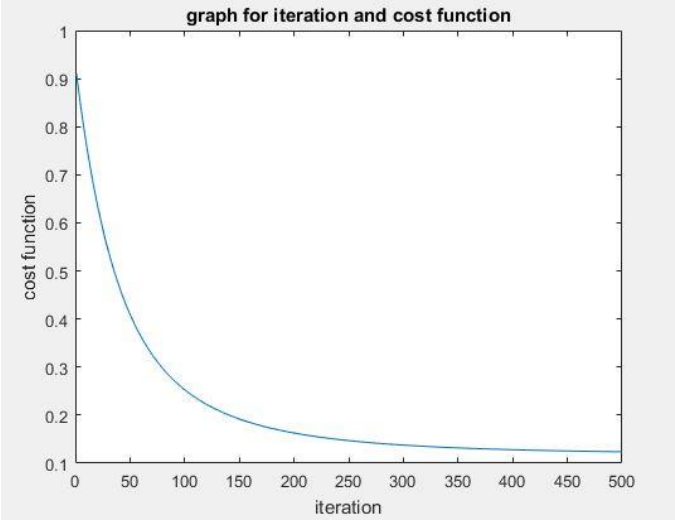

In the test scenario, the cost obtained is 0.1034
So the performance is better since the cost is reduced.

**Question 2**
Quadratic Regression with one variable

In this Model I have taken the square of an input variable and used it as a feature.

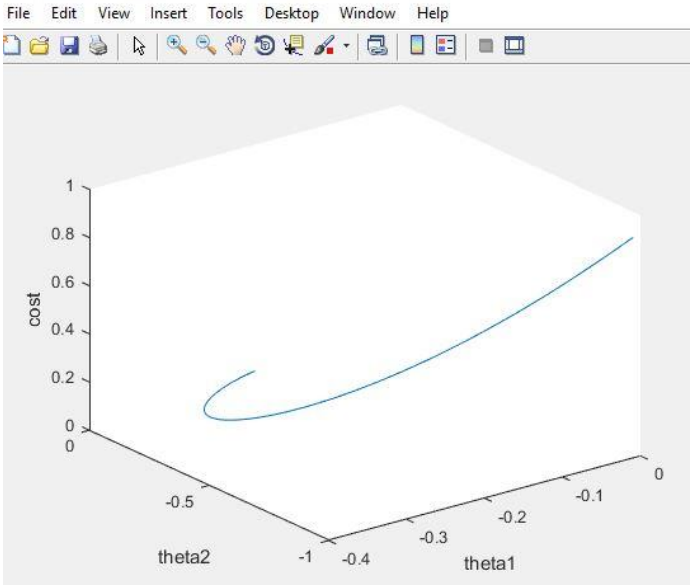
The Graph for iteration and Cost Function is given as follows

The number of iterations are 500 and the learning rate is 0.01.



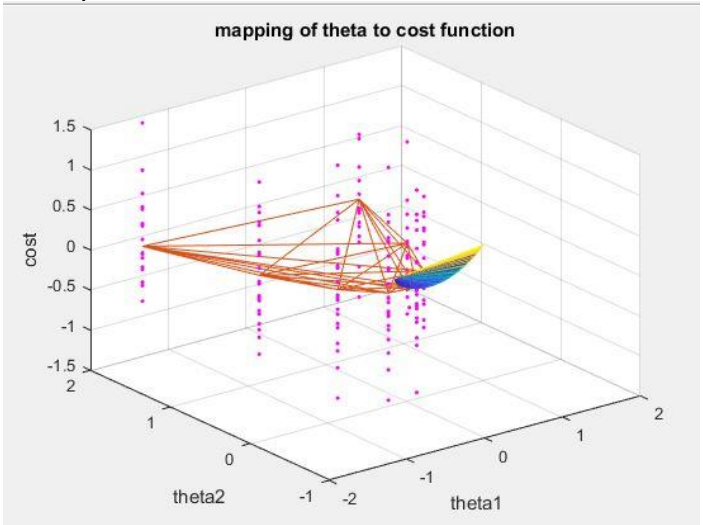

In this case as the iterations increases, the cost function decreases and obtains an optimum value and then becomes constant with the increase of iterations.

The 3d line plot for theta values and the cost

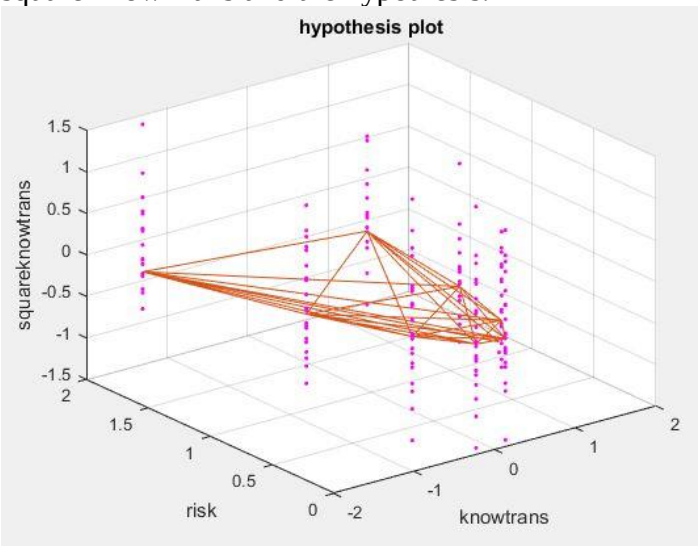



The theta values are taken as theta1=-0.2144 theta2 = -0.0879 and theta3 = 0.1192 .

The 3d mesh plot for theta values and the cost function

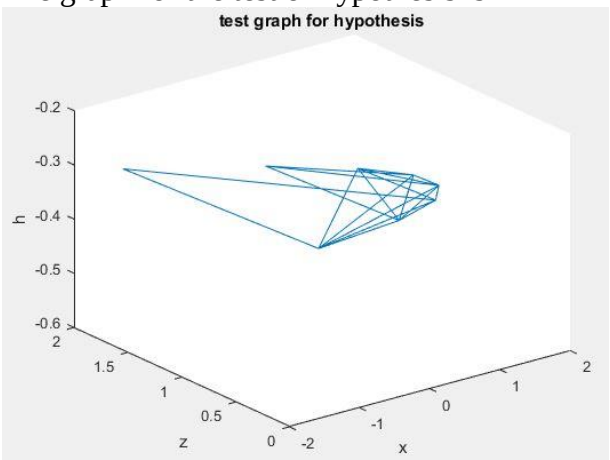



This is a 3d plot of the KnowTrans Risk and the other input variable squareKnowTrans and the hypothesis.




The value of hypothesis is `h= -0.2144+(-0.0879*x) +(0.1192 *(z));`
Where the theta values are theta1 = -0.2144 theta2 = -0.0879 theta3 = 0.1192
The minimum cost value for the 350 observations is 0.1238.

The graph for the test of hypothesis is



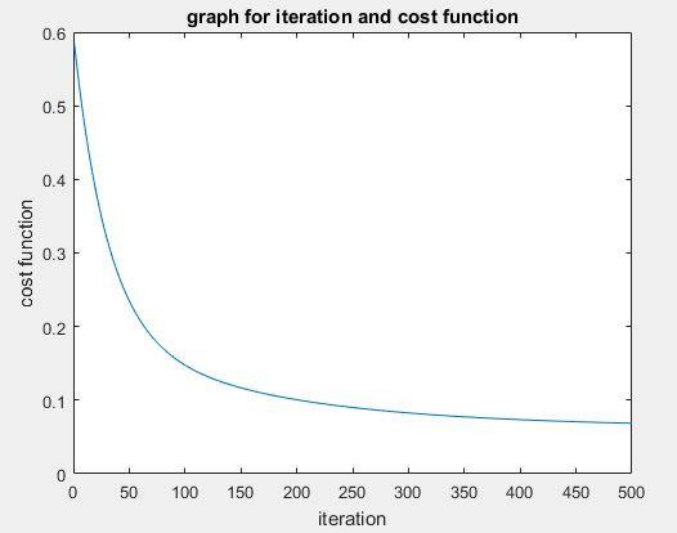
The obtained cost for the rest of 60 observations with respect to the hypothesis is 0.1463 which is greater than that of the above cost for 350 observations.
So this is a proof that Linear Regression with one variable is better than Quadratic Regression with one variable.

**Question 3**

**Linear Regression with two variables**

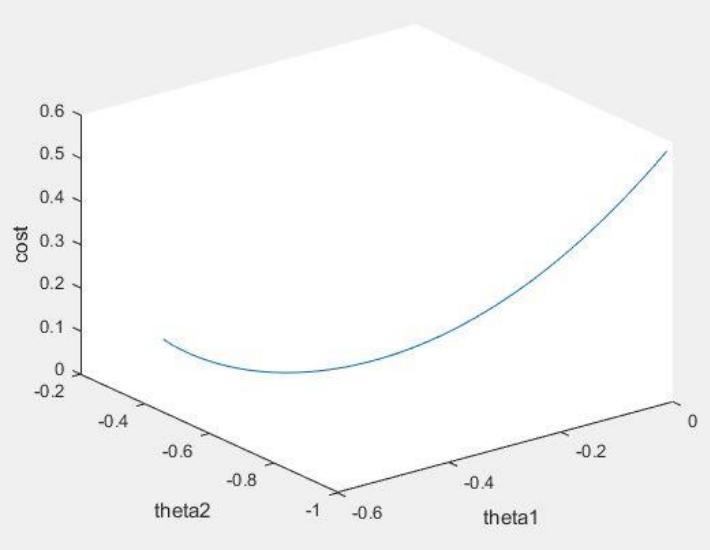
In linear regression with two variables, since a new variable is added and its ranges are different from that of the previous ones, I have normalized the values of the dataset and made the range in between -1 and 1.
In this Regression, we take another variable named respEtiq and plot the values based on the gradient descent algorithm. The alpha value is 0.01 and the iterations are 500.

The graph for iterations and cost function



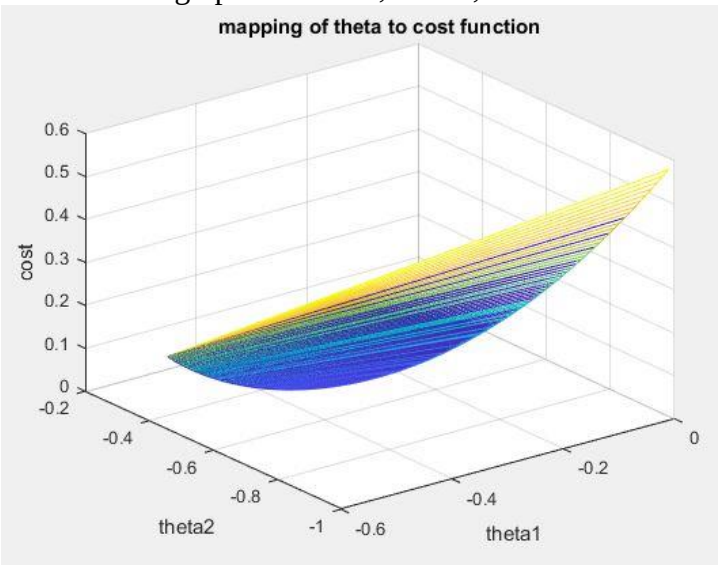

The total iterations are 500 and the cost decreases as we increase the number of iterations and reach the global value for a particular theta values.

The 3d line plot for theta1 theta2 and cost is given as

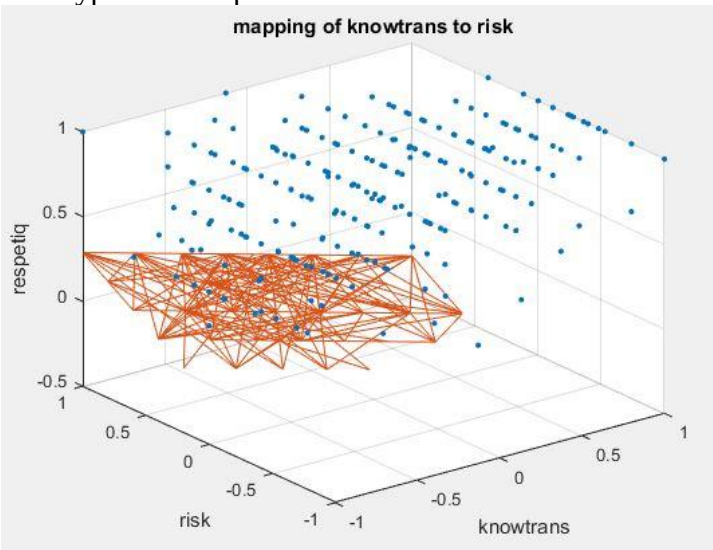


The theta values for optimum cost are theta1 = -0.4937 theta2 = -0.2733 theta3 = 0.5089
The obtained minimum cost is 0.0686 for 350 observations.

The 3d mesh graph for theta1, theta2, cost




The hypothesis is plotted as follows

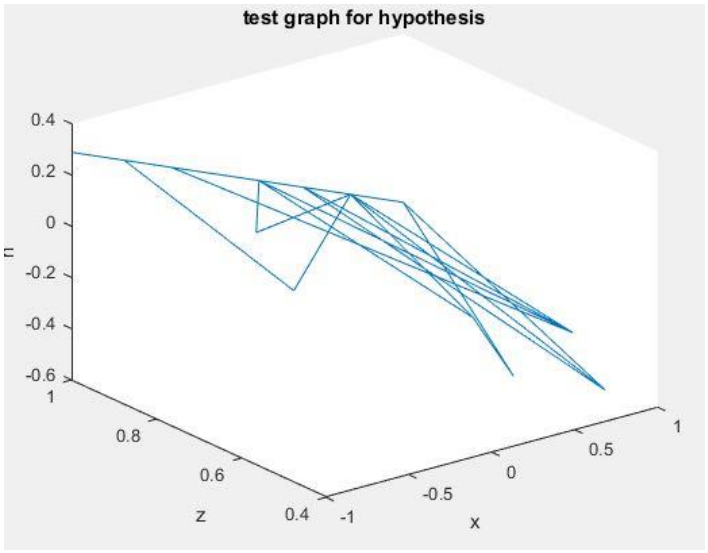



Where the hypothesis is

```
h= -0.4937+(-0.2733*x)+ (0.5089 *z);
```

The obtained minimum cost with 350 observations is 0.0686

The test graph for the remaining 60 observations




The obtained cost is 0.0181 for 60 observations and this shows the obtained cost function for 350 observations is less than the cost obtained by the linear regression with one variable.
So it is proved that in some cases Linear Regression with two variables will best fit the hypothesis and gives us with an optimum cost when compared to the Linear Regression with a single variable.

**Question 4**

Performance Tested for the different proportions are as follows

| Regression type | Observations taken 350,60 | Observations taken 250,160 | Observations taken 230,180 |
|---|---|---|---|
| Linear Regression with one variable | 0.1194 | 0.1369 | 0.1050 |
| Quadratic Regression with one variable | 0.1238 | 0.1492 | 0.1105 |
| Linear Regression with two variables | 0.0686 | 0.0623 | 0.0611 |

Having two variables in the linear Regression gives the best result when compared to the one with one variable in linear regression and the quadratic regression with one variable. The linear regression with one variable is better than the Quadratic regression with one variable as shown above the cost increases when its quadratic.

The best proportion from the set of experiments I have conducted is 230,180.