

# ASSIGNMENT 1

## Statistical Analysis of Linear Regression using Normal Equation Method and the Gradient Descent Method

In this assignment, we have performed linear regression using both Normal Equation method and Gradient Descent Method.

Preprocessing Step: Categorical data was converted from "yes/no" to "1/0"

### 1)Normal Equation Method:

*Results:*

Cost = 255930513.52460164

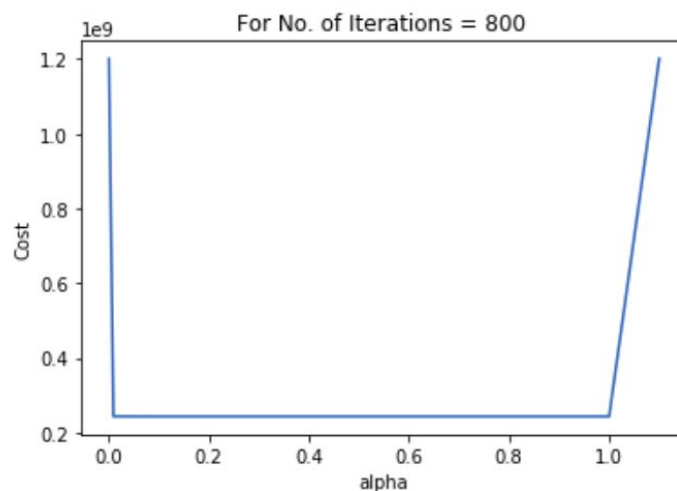
$y = 30000$

$h_x = 43450.157471896244$

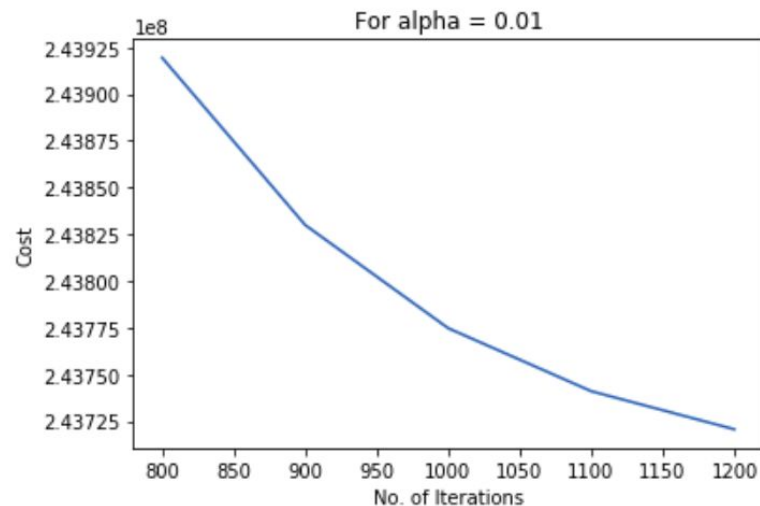
### 2)Gradient Descent:

Preprocessing: Normalization is done according to formula :  $X = (X - X.mean()) / X.std()$

*Results:*



We can see that for a fixed number of iterations, smaller values of  $\alpha$  have a higher cost. If we keep increasing the value of  $\alpha (>1)$  then we ultimately see that the cost becomes extremely high as the optimal values of the parameters  $\theta$  are never obtained due to the step size being very large.



Here we see that for a fixed value of  $\alpha$  the cost decreases with increase in No. of iterations.

For No. of iterations = 800 and  $\alpha = 0.01$

Cost = 243919539.90943313

$y = 30000$

$h_x = 42020.3972814$

Thus by using gradient descent on the same test case we obtained a value of  $h_x$  that is closer to the actual value  $y$  than the value of  $h_x$  obtained using normal equation method.