

# COL772 ASSIGNMENT 1

## MACHINE LEARNING

Saket Dingliwal  
2015CS10254

### Part 1

Normalization of data is done. Linear Regression was done which works well for any number of input feature vectors. Gradient Descent was used for obtaining minima of the error function. Convergence condition used is the difference of error value approaching a small value.

Epsilon used =  $10^{-15}$

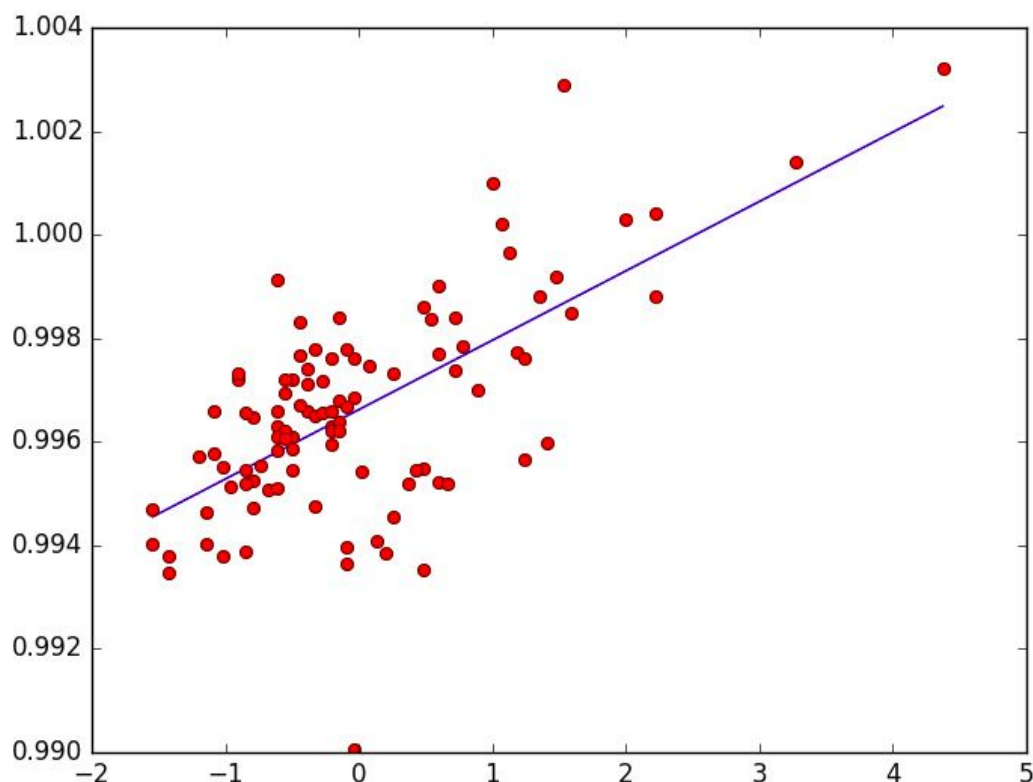
Learning rate = 0.0008

Number of Iterations = 222

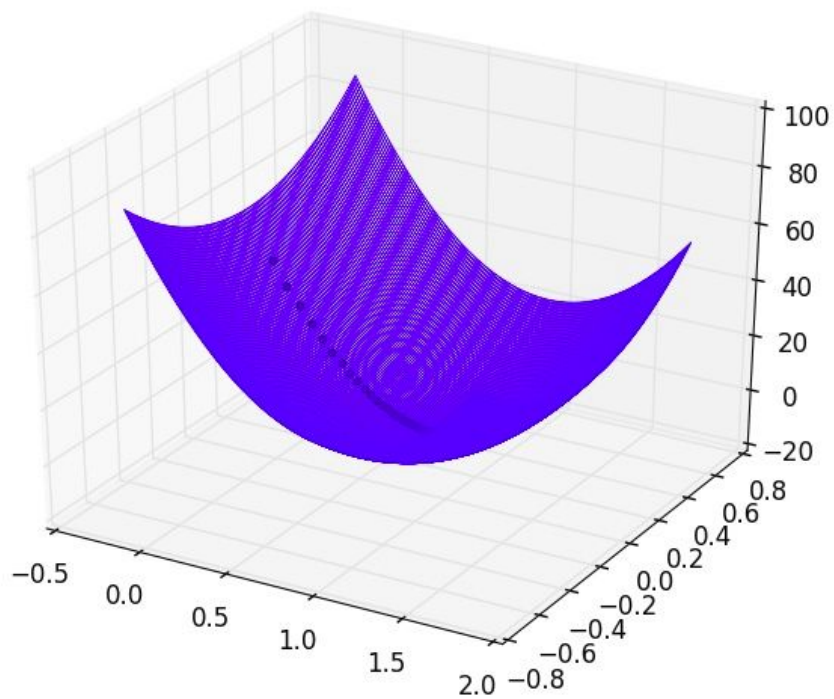
Value of J-theta = 0.00011947898

Result  $\Rightarrow y = 0.00134x + 0.9966$

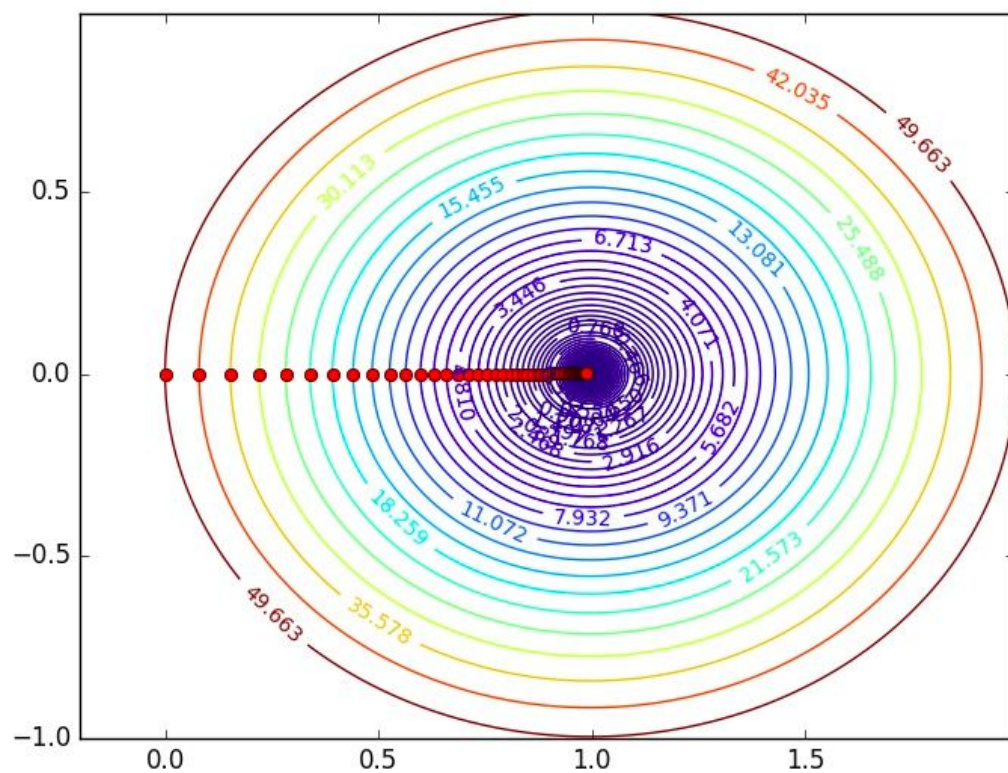
Plot with normalized x on x-axis and y value on y-axis



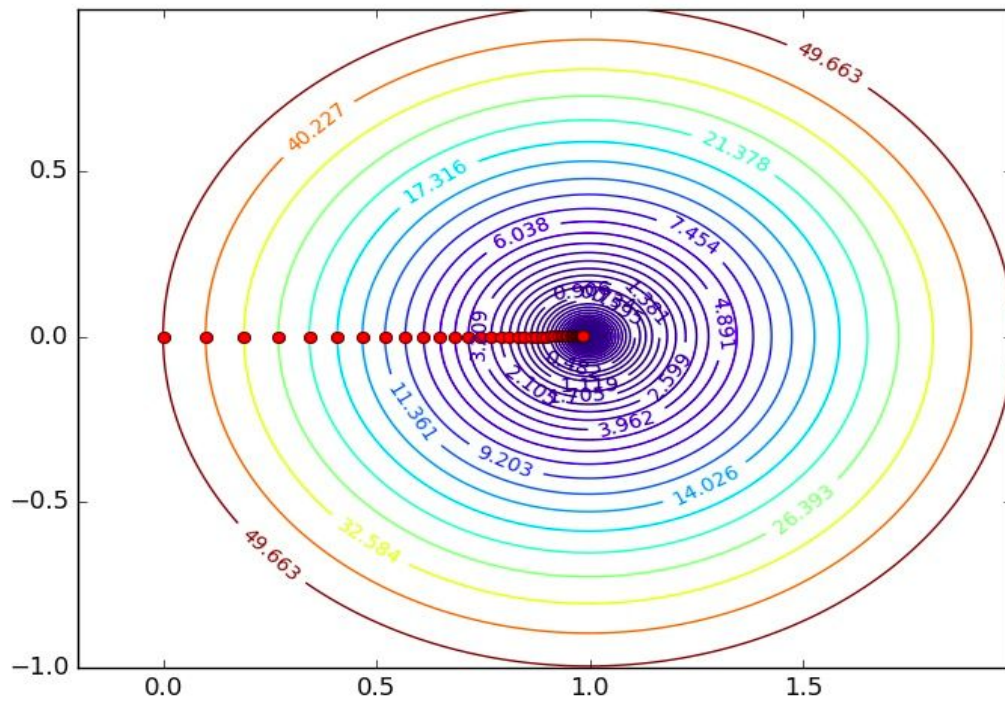
The 3-d plot for error function on Z-axis and theta0,theta1 on x,y-axis



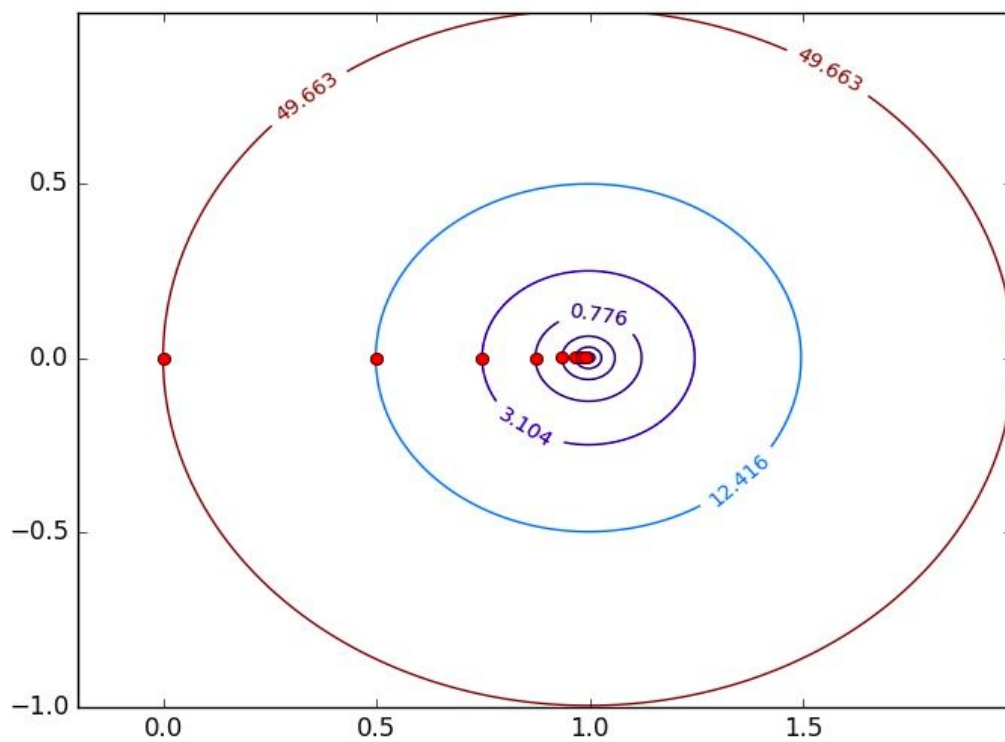
Contour plot for the learning rate = 0.0008 with  $\theta_0, \theta_1$  on x,y-axis



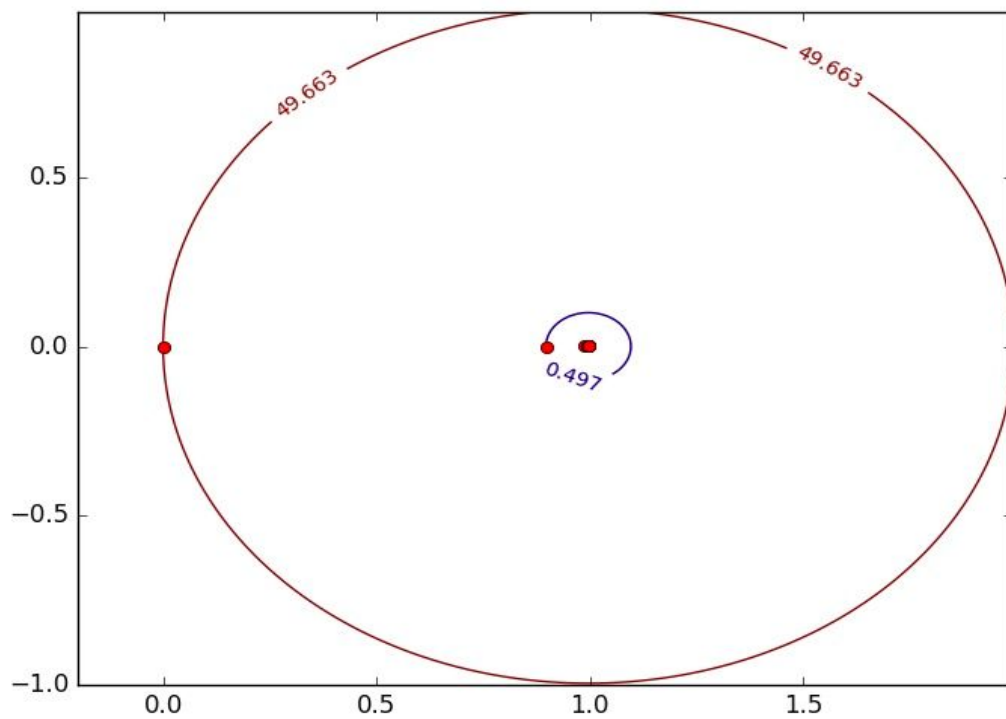
### Contour Plots for different learning rates with theta0,theta1 on x,y-axis



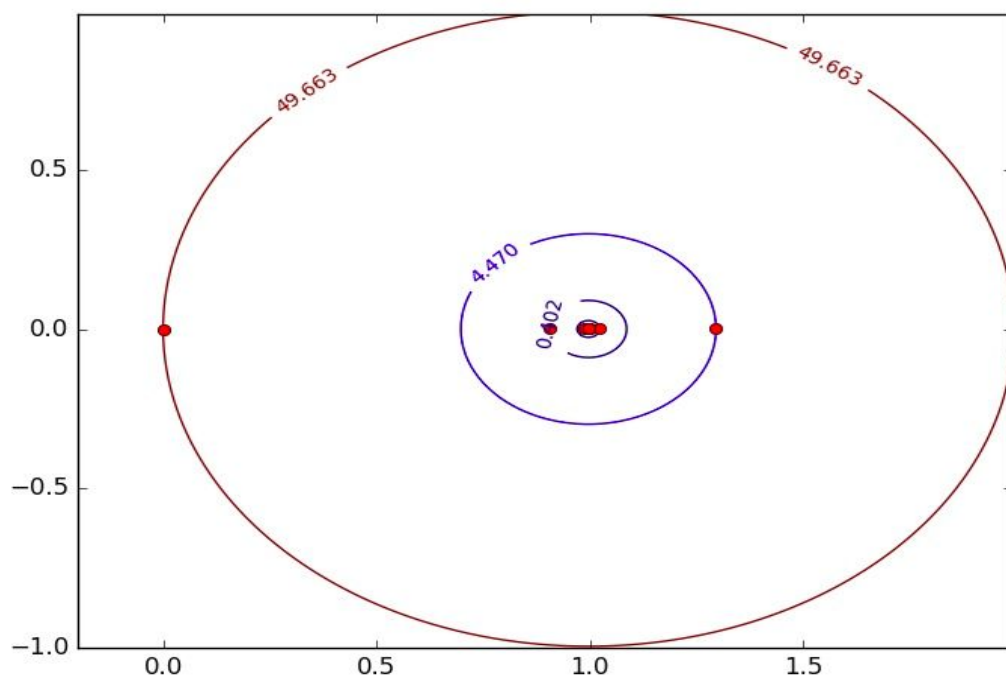
Alpha = 0.001 There is no oscillation in the value of theta



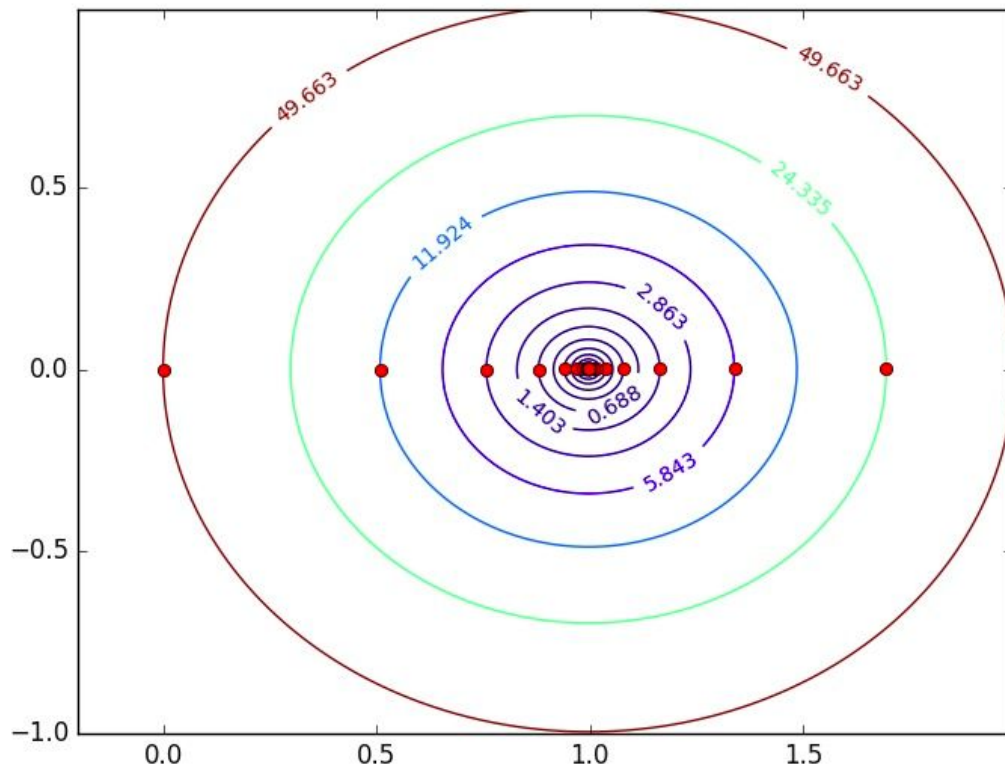
Alpha = 0.005 The number of steps for convergence decreases



Alpha = 0.009 the number of iterations for convergence decrease further



Alpha = 0.013 there is oscillations in theta values on left and right side and hence the number of iterations increased a little



Alpha = 0.017 The oscillations increases and so does the number of iterations

The gradient descent do not converge for alpha values = 0.021 and 0.025. The values of J-theta shoots to infinity after a few iterations. We observe that the number of iterations required are minimum for alpha = 0.009 as there is no oscillations and value of j-theta decreases very quickly. So too small learning rate may increase the number of steps to be taken, while too large learning rates may lead to oscillations or failure of gradient descent by taking j-theta to infinity. Hence it s value must be optimum between the two extremes.

## Part 2

The values of theta obtained without using weights are

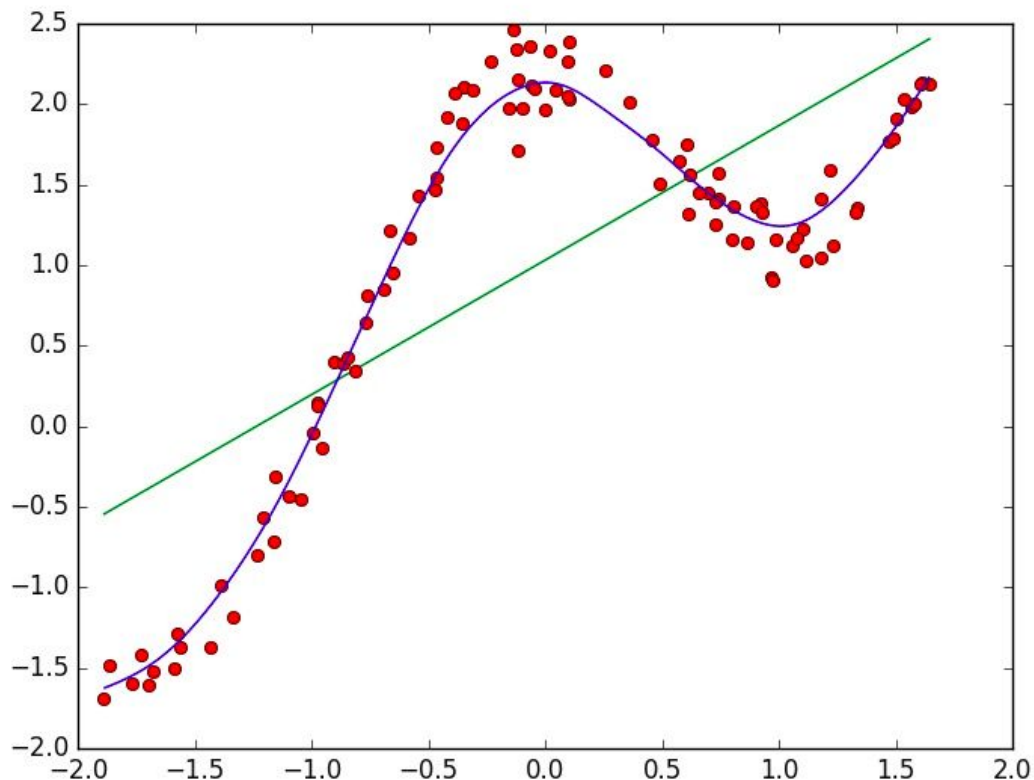
theta0 = 1.03128116

Theta1 = 0.83519315

Resulting line =>  $y = 0.83x + 1.031$

The plot for weighted linear model is done by choosing 1000 random points between maximum and minimum of x values and plotting them on graph.

T value used is 0.2 for the normalized data



It can be observed that if smaller values of  $t$  ie 0.03, 0.1 etc are used there is overfitting of the training data. The curve tries to pass through each and every training point and hence might not perform well on new data. For  $t = 0.2$  a better fit is obtained. On increasing the value further, underfitting is observed. The obtained curve do not pass close to the training points. At  $t = 2, 10$  the curve is almost linear and is mostly same as unweighted linear regression.

Plots for different  $t$  values are shown

