ASSIGNMENT-1

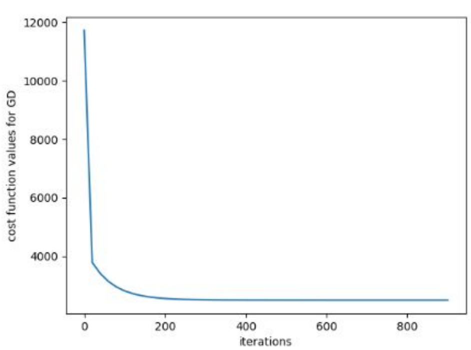
Introduction

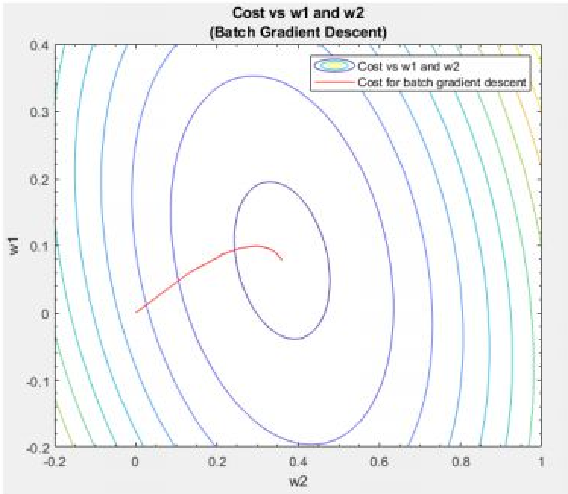
The data consists of 434874 instances and 2 columns corresponding to latitude and longitude of each instance. Before training the model we have normalized the data and used a 70, 20,10 split for training, validation and test. We used a stopping limit of e(i+1)-e(i) < 5 * 10^{-3}.

We have implemented the gradient descent, stochastic gradient descent algorithms and have further used L1 and L2 norms of regularization for the gradient descent algorithms and lastly we implemented the Normal equation method.

Gradient Descent:

- The equation for gradient descent with two features is given by $y = w_0 + w_1x_1 + w_2x_2$. We have used a learning rate around 0.0000004 and the model converged to a minimum squared error of 2500 after 800 iterations.
- The weights we get after training are w0=0.20,w1= -0.099, w2=0.095. The batch size consists of 304113 entries. The R^2 and RMSE error in this case were 0.027 and 19.09

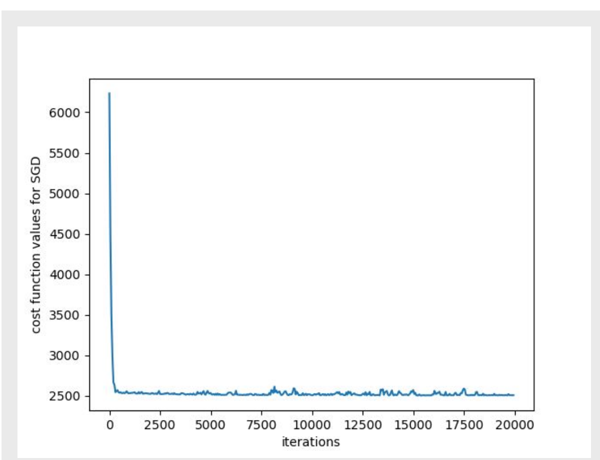




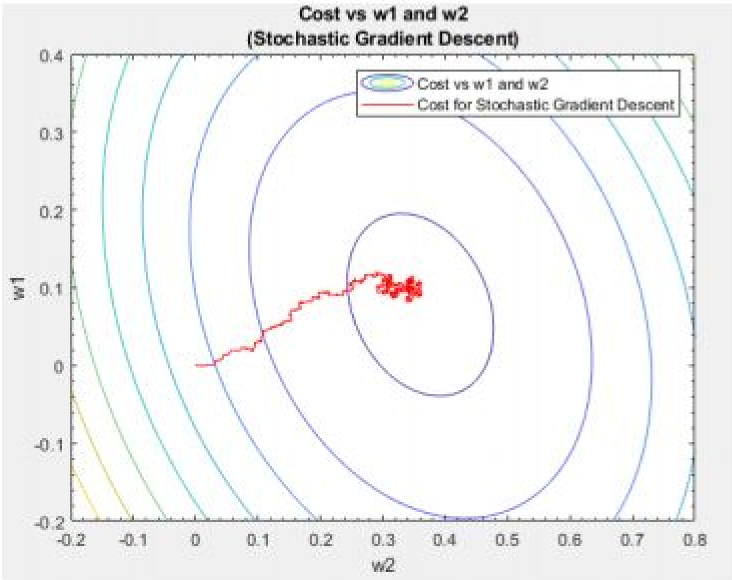
Cost vs w1 and w2 – Contour plot

Stochastic Gradient Descent:

- In SGD the same hypothesis and cost function formulae are used the only difference being that we update weights at every data point. The square error settles to an average minimum of 2500 after approximately 150 iterations.
- The weights we got after training were w0=0.20, w1= -0.11, w2=0.10 and the R^2 and RMSE values were 0.028 and 19.02 respectively.



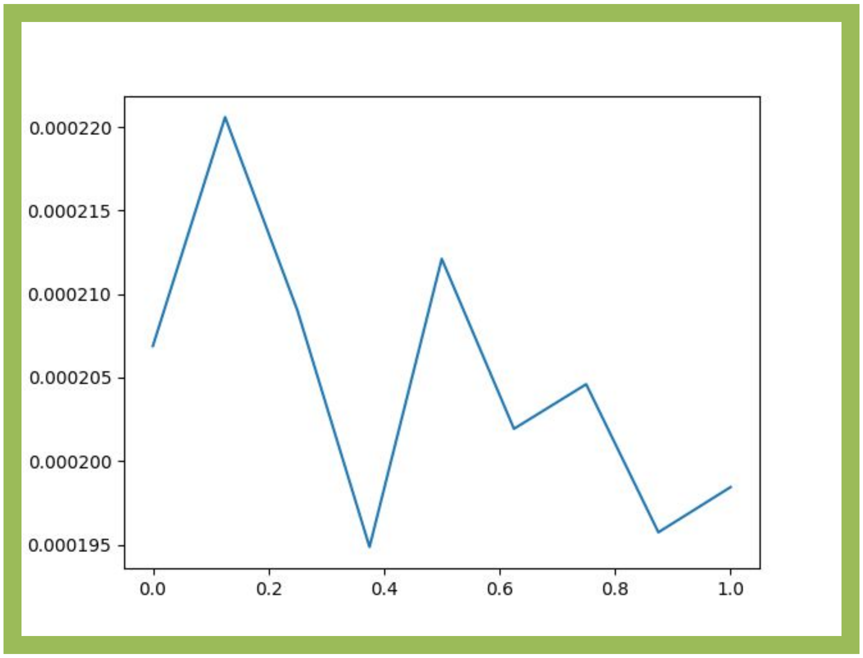
Cost vs w1 and w2 3D plot for stochastic gradient descent



Cost vs w1 and w2 – Contour plot

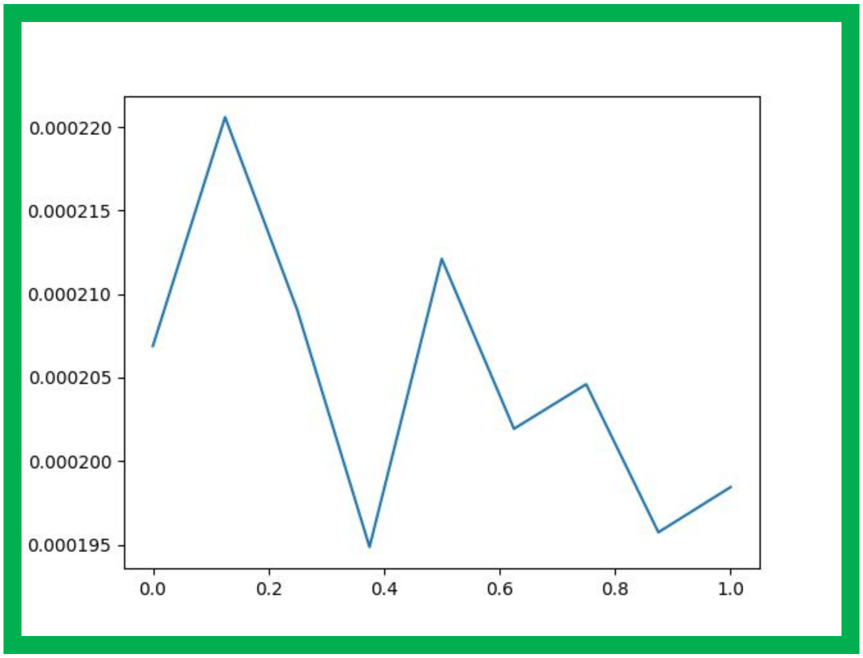
Gradient Descent with L1 regularization :

- In this method the cost function formula is altered to include Beta where β is the L1 regularization coefficient. This method ensures that higher values of weights are avoided by placing a penalty factor.
- The L1 values we used are 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95 and we obtain the L1 parameter which gave the least validation loss. The mean minimum error obtained was 2510.24 with a mean validation loss of 717.6. The final weights and β value obtained were w0=0.2086843,w1= -0.1008773, w2=0.09468503 and 0.35 . The R^2 and RMSE values obtained finally were 0.026 and 18.98.



Gradient Descent with L2 regularization :

- In this method the cost function formula is altered to include Beta where β is the L2 regularization coefficient .It is similar to the L1 regularization by ensuring that higher values of weights are avoided by placing a penalty factor.
- The values of L2 we used are 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95 and the L2 parameter which gave us the least validation loss was taken. A mean minimum error of 2506.20 was recorded with a mean validation loss of 718.26.
- The final weights and β value recorded are w0=0.20889902, w1=-0.101212, w2=0.090636 and 0.05 . The R^2 and RMSE values obtained finally were 0.02822 and 18.9209 respectively.



***Normal Equations*:**

- In this method, we obtain our final weights by solving linear equations and using linear algebra properties.
- The final weights obtained were w0=0.208, w1=-0.098,w2= 0.094. The R^2 and RMSE values obtained finally were 0.027 and 18.81 respectively. This method gave us a very fast and accurate prediction of the parameters.