ISCO630E-ASSIGNEMNT-2

Conclusion

By Nairit Banerjee-IIT2016505

Question 1

We are required to build a Linear Regression model on the Housing Price dataset using regularization.

We define the cost function as

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right]$$

$$\min_\theta J(\theta)$$

We first apply Gradient Descent with **learning rate 0.5 and regularization parameter 0.1**

Repeat{

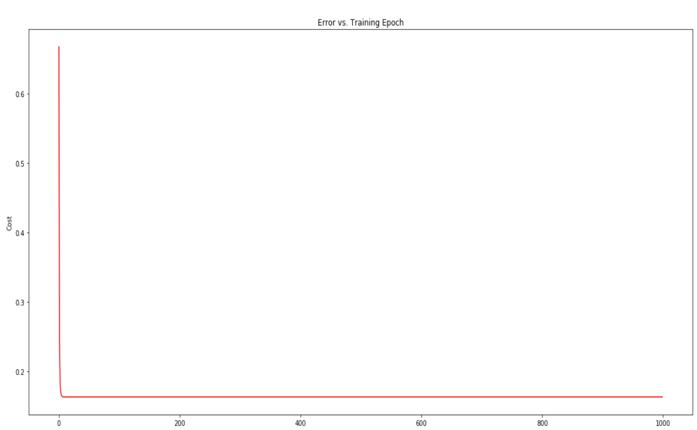
$$\theta_0 := \theta_0 - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m}\theta_j\right]$$

$$(j = \cancel{0}, 1, 2, \cdots, n)$$

}

We observe how the cost decreases with each iteration



After running over 1000 iterations final cost was 0.16313885572854372.

Now applying Normal equations

$$\theta = (X^T X + \lambda \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix})^{-1} X^T \vec{y}$$

We obtain the **cost value as 0.16313916095747125.**

Thus we observe that both the results are pretty close.

Now we experiment with values of learning rate and regularization parameter to further tune the model.

Learning Rate	Reg. Param	Cost(GD)	Cost(Normal Eq.)
0.5	0.1	0.16313885572854372	0.16313916095747125
0.1	10	0.16317406975350043	0.16560467500629283
0.1	0.00001	0.16313885233554	0.1631388523355431

Now we **split the dataset into training and testing dataset in 70:30 ratio**.

We use **learning rate 0.1** and **regularization parameter 0.1** and we get following results.

Cost after training with **GD** on **training data : 0.16909407548278382**
Cost after applying **Normal Equation** on **training data : 0.1690954267457682**

Cost after training with **GD** on **testing data : 0.24850868246565566**
Cost after applying **Normal Equation** on **testing data : 0.24635836733674232**

Question 2

In this question we had to apply Locally Weighted Regression(LWR) on the Housing Price dataset.

We use the following formulas

$$h(x_o) = {x_o}^T \hat{\beta}(x_o)$$
$$\hat{\beta}(x_o) = \text{argmin}_\beta \sum_{x,y} w(x, x_o)(y - x^T \beta)^2$$
$$w(x, x_o) = e^{-\frac{(x-x_o)^2}{2\tau^2}}$$
$$\hat{\beta}(x_o) = (X^T W X)^{-1} X^T W y$$

We then apply Normal Equation to get final values of the coefficients of the hypothesis.

Taking **Tau=1**, or **considering whole batch as neighbourhood**, we get **value of cost** as **0.31795942**

Also we kept aside one tuple of data for testing purpose.

Tuple was **[1, 6000,3 ,1 ,2 ,1 ,0 ,0 ,0 ,1 ,1 ,0]**.

Predicted value : 1.42503106
Original value : 1.38107544

Error: 3.18%

Now we take neighbourhood of size 50, ot tau=50/545

The same tuple now gives the following result

Predicted value : [13.02377364]
Original value : [1.38107544]

Thus we observe that decreasing the value of tau, drastically worsens the prediction error.