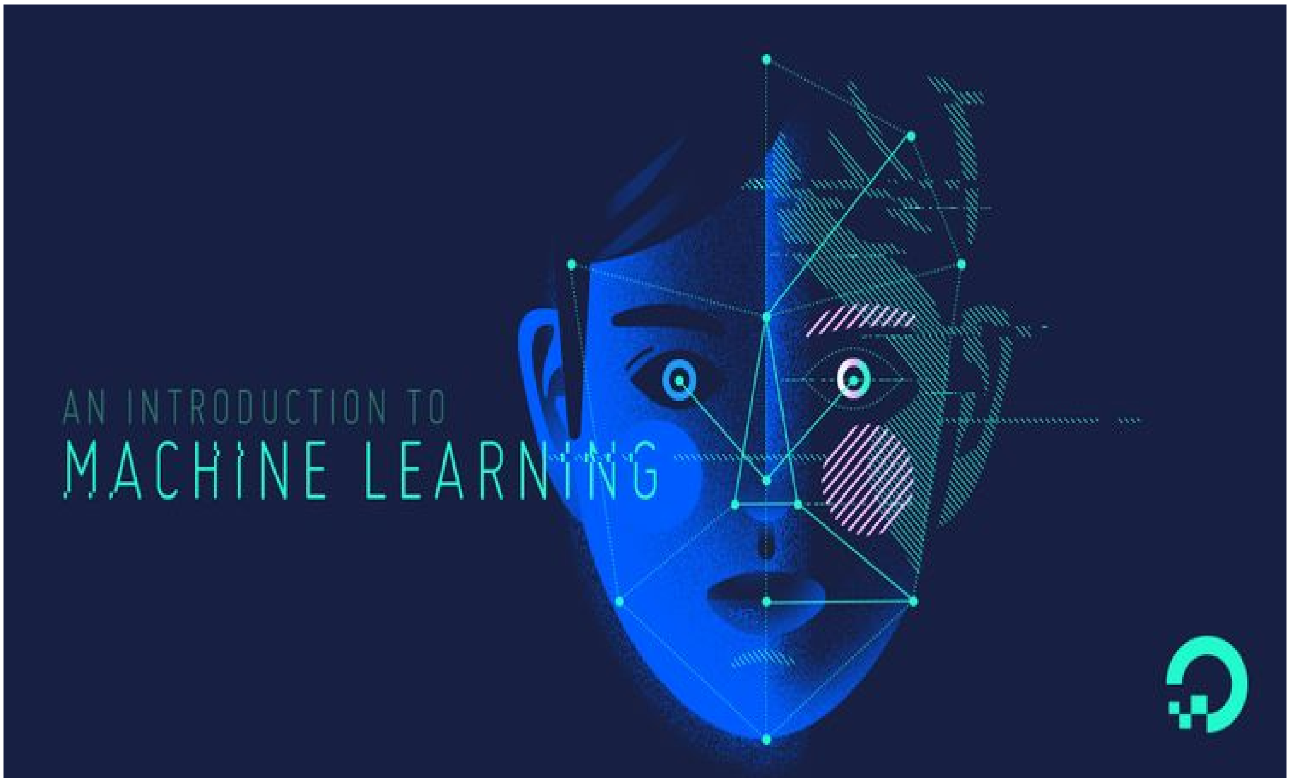
Analysis Of Gradient Descent ,Normal Equation and LWR(Locally Weighted Regression)

(With Regularizer, Assignment # 2)



Amarjeet Yadav

MIT2018053

RESULTS

1. Gradient Descent + Regularizer

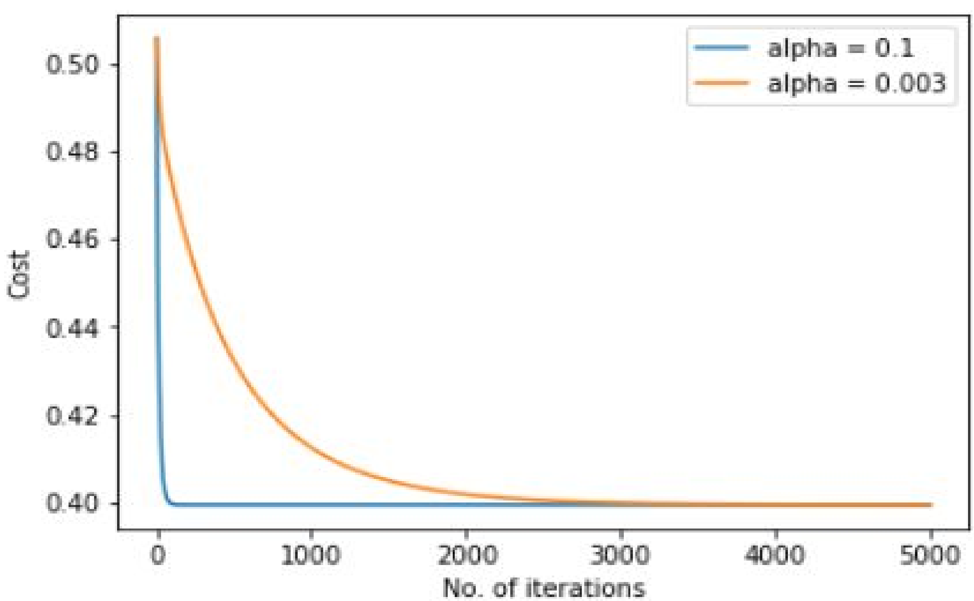
1. I . 70 % Training and 30 % Testing

Parameters Obtained : For Alpha = 0.1, Epochs = 5000
parameters after gradient descent= [[-0.31201252 0.22385437 0.0904532   0.09079023 0.09100656 0.09045408 0.09063603 0.0905283   0.09048319 0.09083362 0.0908532 0.09065369]]

Parameters Obtained : For Alpha = 0.003, Epochs = 5000
parameters after gradient descent= [[-0.3119227   0.22379977 0.09041708 0.09075363 0.09096338 0.09042555 0.0906061   0.09049911 0.09045618 0.09079973 0.09081938 0.0906234]]

The Squared error for (alpha = 0.1) : 0.357868382700491

The Squared error for (alpha = 0.003) :0.3578898032959659

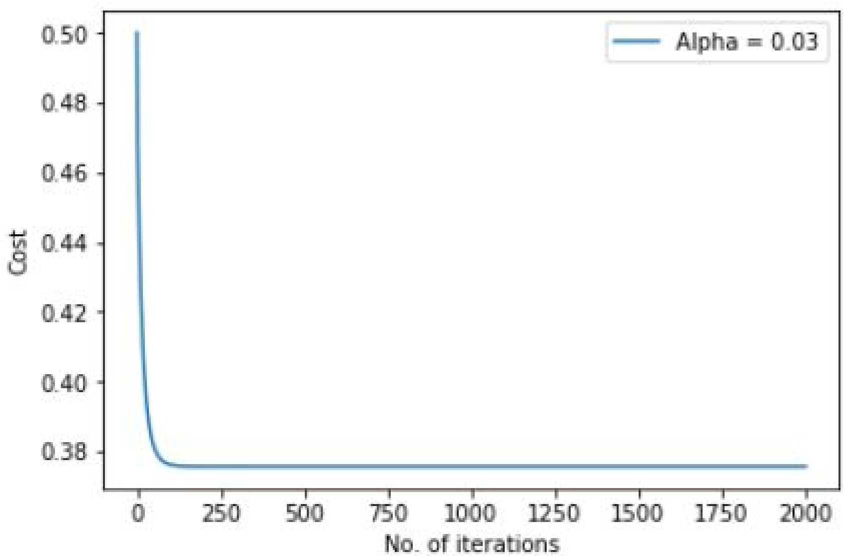


We can see that here when alpha large then it converge faster as compare to when alpha is small.(But carefully choose alpha it may overshoot for large values)

Accuracy : 65%

II. I . 100 % Training

Parameters Obtained : For Alpha = 0.03, Epochs = 2000
parameters after gradient descent= [[-0.37511052 0.26244665 0.10882016 0.10919075 0.10945691 0.10876014 0.10895348 0.10892115 0.10878663 0.10924533 0.10932175 0.10902757]]



Total squared error is : 0.3757259053169172

Accuracy : 63%

1. II. Normal Equation + Regularizer

Parameters Obtained :

[[-5.24518300e-01]
[1.78179352e-01]
[7.61790706e-05]
[7.70854509e-05]
[1.16510337e-04]
[2.42095558e-05]

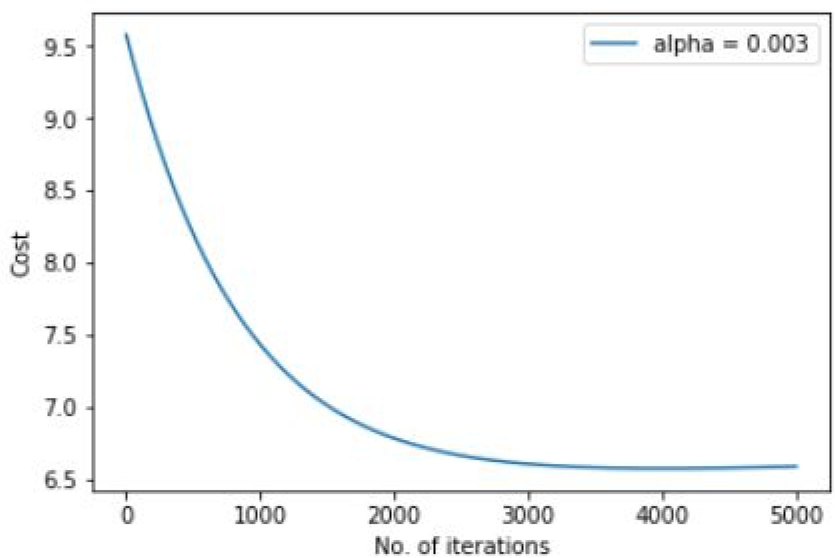
[3.02381446e-05]
[1.93311642e-05]
[6.70909955e-06]
[5.81742479e-05]
[7.27457291e-05]
[3.36204449e-05]]

Total squared error is : 0.37573213271663414

Accuracy Obtained: 63%

2. LWR (Locally Weighted Regression)

Parameters Obtained : For Alpha = 0.03, Epochs = 5000
[[1.28025788] [1.16606709][1.28025807] [1.280258] [1.28025695] [1.28025792] [1.28025772]
[1.28025781] [1.28025765] [1.28025753] [1.2802579] [1.28025773]]



Predicted value for given point is : [[0.7038954]

The original value for given point is : [0.5114314]

Accuracy : 79%

It is suitable when hypothesis is too much curvy then it give good prediction

CONCLUSION

1 . Regularizer is a technique applied to Cost Function J(θ) in order to avoid Overfitting.

2. Regularizer help to remove the problem of non-invertibility

3.The core idea in Regularization is to keep more important features and ignore unimportant ones. The importance of feature is measured by the value of its parameter θj.

In linear regression, we modify its cost function by adding regularization term. The value of θj is controlled by regularization parameter λ .

4. If λ is too large, then all the values of θ may be near to zero and this may cause Underfitting. In other words, this model has both large training error and large prediction error. (Note that the regularization term starts from θ1)

- If λ is zero or too small, its effect on parameters θ is little. This may cause Overfitting.

  To sum up, there are two advantages of using regularization.

- The prediction error of the regularized model is lesser, that is, it works well in testing data .
- The regularization model is simpler since it has less features (parameters).

5. In LWR there is no training phase. Hypothesis is developed for each query . Therefore costly if too many prediction we have to done.

What if Tau is small ?

The parameter tau controls how quickly the weight of a training example falls off with its distance the query point xx and is called the bandwidth parameter. In this case, increasing tau increases the "width" of the bell shaped curve and makes further points have more weight.