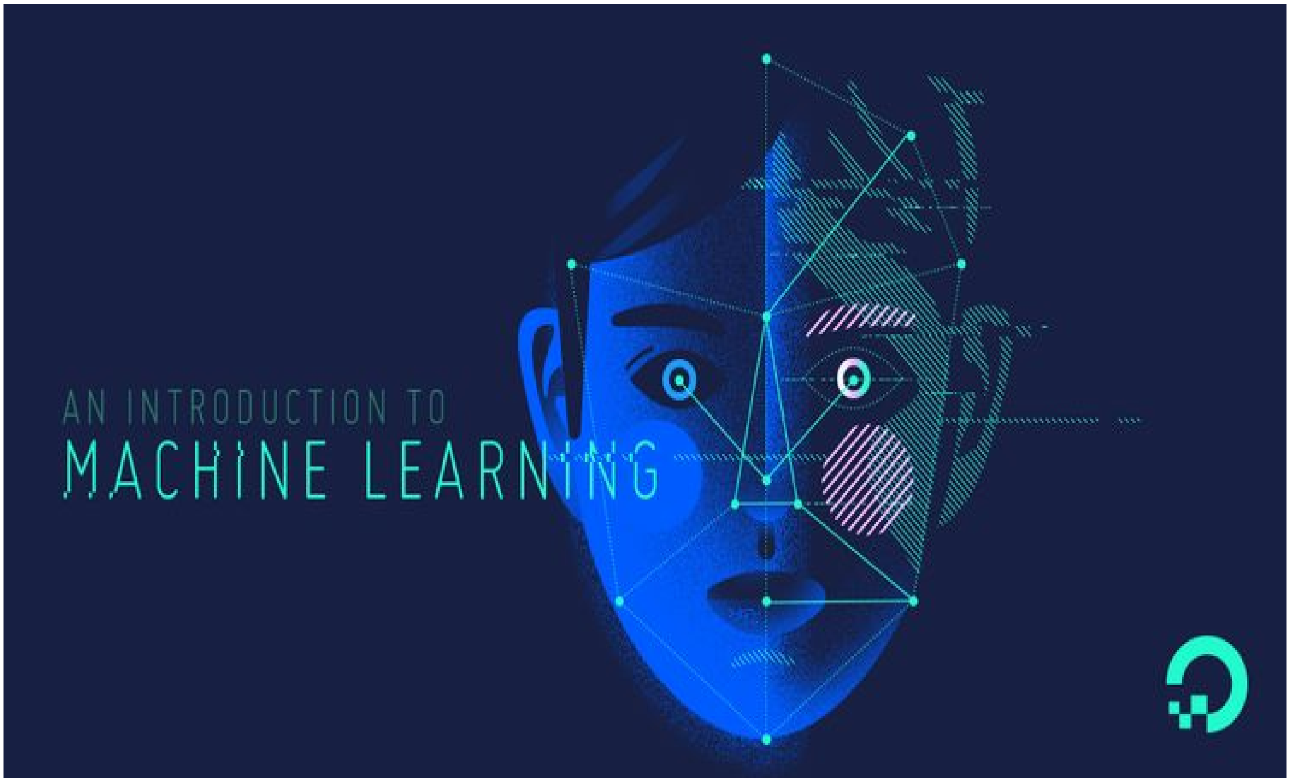
Analysis Of Gradient Descent and Normal Equation (Without Regularizer, Assignment # 1)



Amarjeet Yadav

MIT2018053

RESULTS

1. Gradient Descent

Parameters Obtained : For Alpha = 0.1, Epochs = 1000

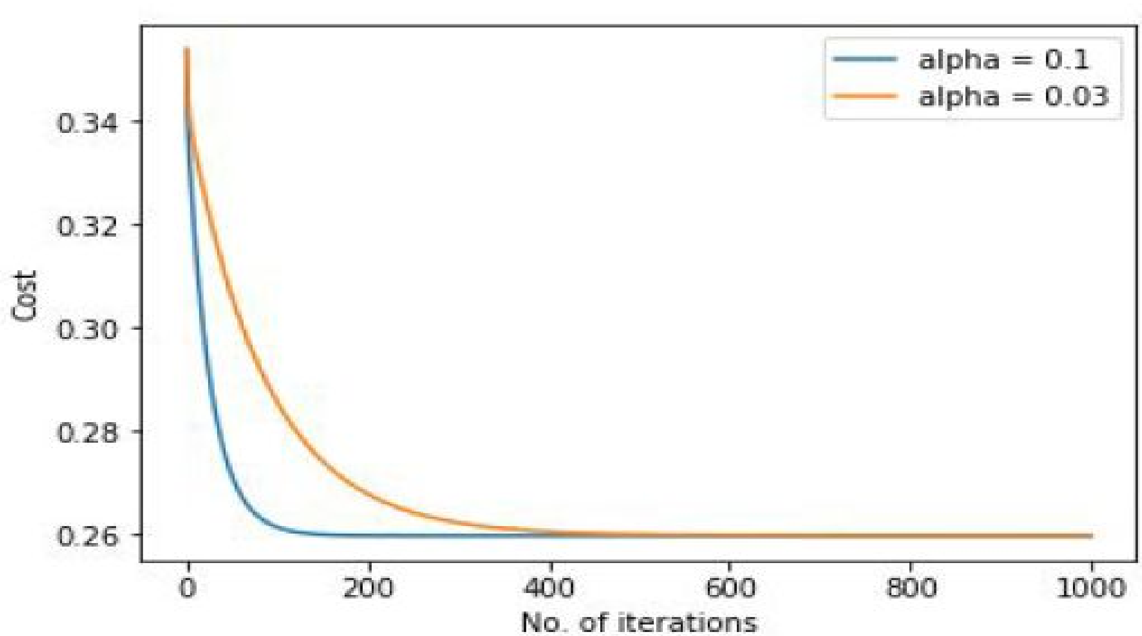
[[-0.61091344 0.39245601 0.18560983 0.18659032 0.19251586 0.17920666 0.18069675 0.17988942 0.17798045 0.18438004 0.18434518 0.18103322]]

Parameters Obtained : For Alpha = 0.03, Epochs = 1000

[[-0.61606843 0.39144044 0.18042739 0.18116219 0.18277396 0.17905148 0.17965533 0.17931572 0.17883456 0.18075424 0.1807737 0.17975492]]

The Squared error for (alpha = 0.1) : 0.35637255518437705

The Squared error for (alpha = 0.03) : 0.3563890015813656



We can see that here when alpha large then it converge faster as compare to when alpha is small.

Accuracy : 65%

2. Normal Equation

Parameters Obtained : [[1.37658095e+03] [2.15081916e-01] [1.11110315e+02] [8.69446180e+02][3.97676252e+02] [4.05611325e+02] [2.73607675e+02] [3.30685115e+02] [7.78220072e+02] [7.66180033e+02] [2.57447260e+02] [5.68257433e+02]]

Total squared error is : 0.32687637935855013

Accuracy Obtained: 68%

CONCLUSION

Let we have m training examples and n features.

Disadvantages of gradient descent:

- we need to choose the learning rate, so we may need to run the algorithm at least a few times to figure that out.
- It needs many more iterations, so, that could make it slower

Compared to the normal equation:

- we don't need to choose any learning rate
- we don't need to iterate

Disadvantages of the normal equation:

- Normal Equation is computationally expensive when we have a very large number of features (n features), because we will ultimately need to take the inverse of a n x n matrix in order to solve for the parameters data.TC O(n^3).

Compared to gradient descent:

- It will be reasonably efficient and will do something acceptable when we have a very large number (millions) of features.

So if n is large then use gradient descent.

If n is relatively small (on the order of a hundred ~ ten thousand), then the normal equation