University of Regina

ENEL/ENSE 865: Applied Machine Learning

Professor: Dr. Abdul Bais

Instructional Designer/TA: Dr. Muhammad Hamza Asad

Submitted by:

Name: Nandish Bakulkumar Bhatt
Student ID: 200441204

Programming Assignment: 2

→ Now, before going to the derivation of the formula, first we will understand what is actually the gradient descent.

Gradient Descent:-

It is an optimization algorithm to find the minimum of a function. We actually, take any random point on function and move in the negative direction of the gradient of the function to achieve local/global minima.

For example:- we need to find local minima of function, $y = (x+3)^2$, starting at $x = 1$.

→ Now, we know that $y = (x+3)^2$ has minimum value when $x = (-3)$. Thus, $x = (-3)$ is local and global minima of the function.

→ But, now we will see that how this value can be achieved numerically using gradient descent

So,
(1) Assume any arbitrary value of x, let's say $x = 2$.

(2) Find derivative of function. So, $\frac{dy}{dx} = 2(x+3)$

(3) Now, move in direction of negative of the gradient. But, how much to move? which is

actually decided by learning rate (η).

Let's say for our e.g., $\eta = 0.01$

(4) Perform 2 iterations.

The above steps are calculated as below:-

$$x_0 = 2.$$

$$\frac{dy}{dx} = 2(x+3)$$

Learning rate, $\eta = 0.01$

<u>Iteration: 1</u>

$$x_1 = x_0 - \eta \left(\frac{dy}{dx}\right)_{x=x_0}$$

$$= 2 - (0.01)\left(2(2+3)\right)$$

$$= 2 - 0.1$$

$$= 1.9$$

<u>Iteration: 2</u>

$$x_2 = x_1 - \eta \left(\frac{dy}{dx}\right)_{x=x_1}$$

$$= 1.9 - (0.01)\left[2(1.9+3)\right]$$

$$x_2 = 1.8$$

This way, we should iteration and thus x value will slowly decrease and converges to (-3).

- But how many iterations we should perform?

So, we set one precision value, and when the difference between two consecutive iterations is less than the precision value, we will stopp performing iterations.

Q:2 Results:

1. This are the results according to the data of assignment.

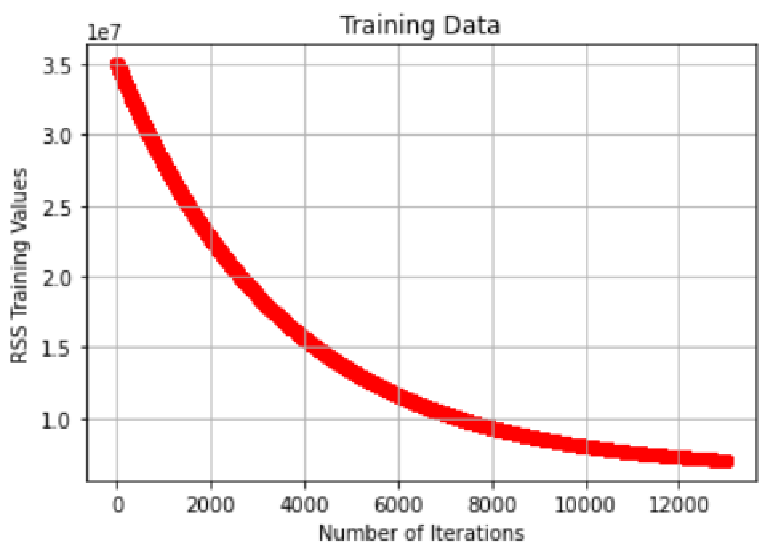
(a) features: ‘Length1’

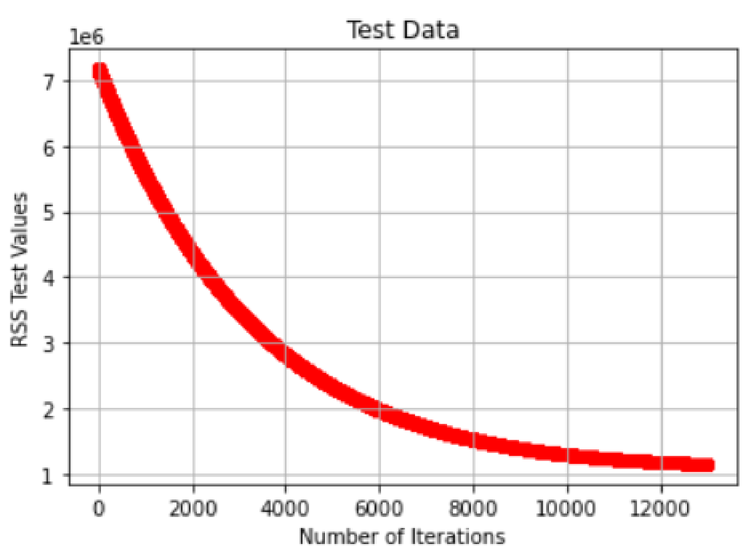
output: ‘Weight’

initial weights: [-7.5, 1] (intercept, Length1 respectively)

step size (learning rate) = 7e-10

tolerance = 1.4e4





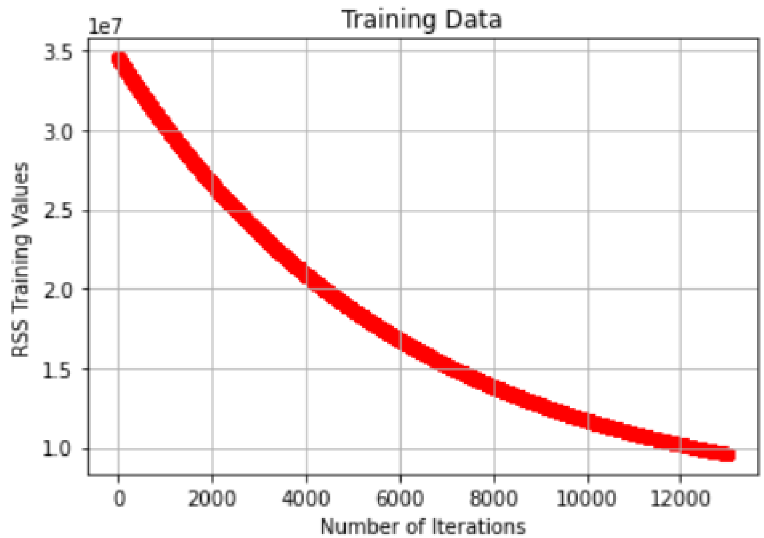
(b) model features = ‘Length1’, ‘Width’

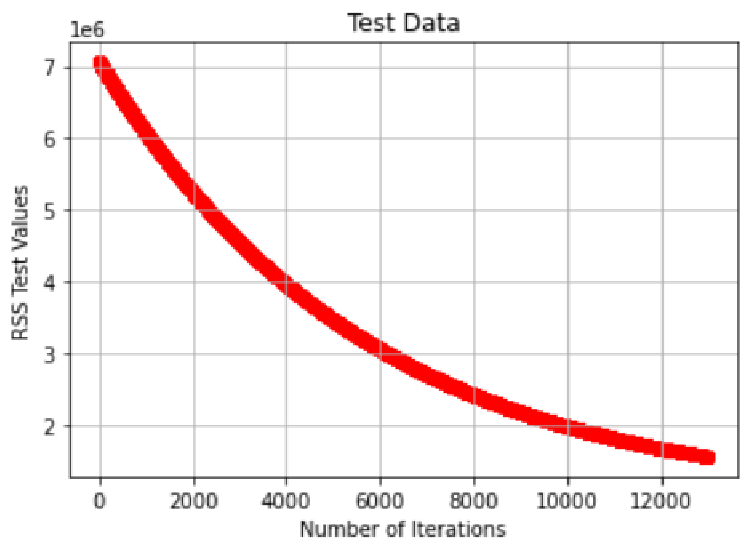
output = ‘Weight’

initial weights = [-8.5, 1, 1] (intercept, Length1 and Width respectively)

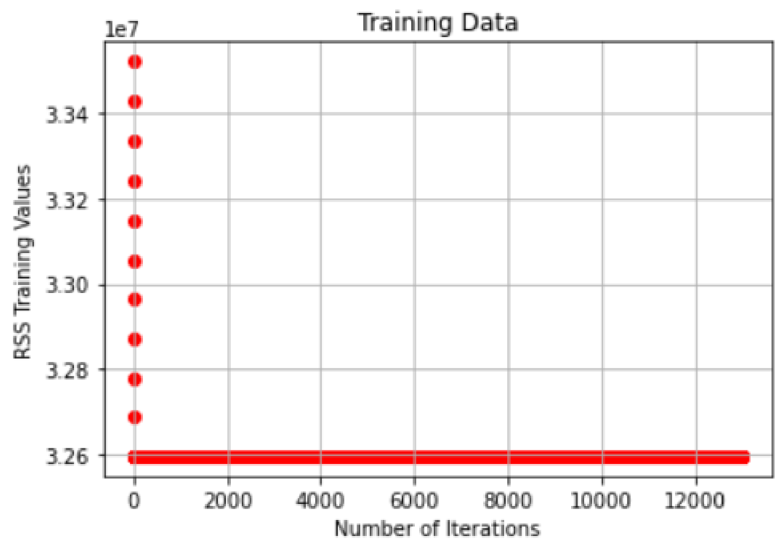
step size (learning rate) = 4e-10

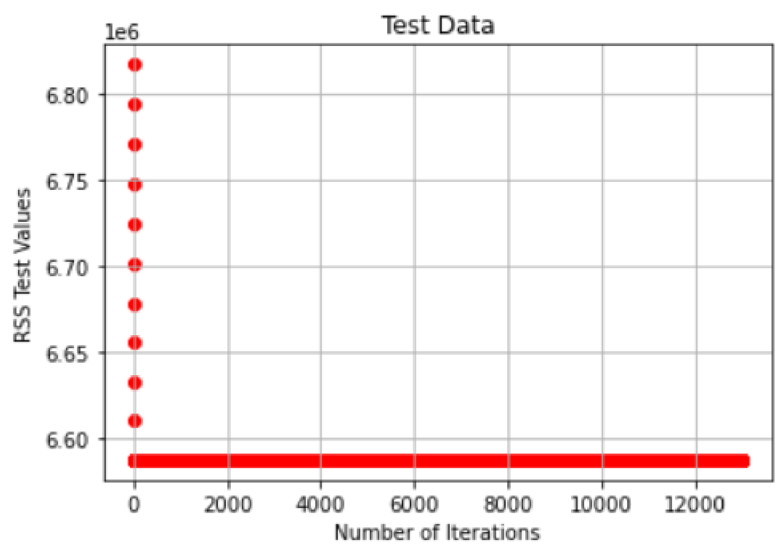
tolerance = 1.4e4



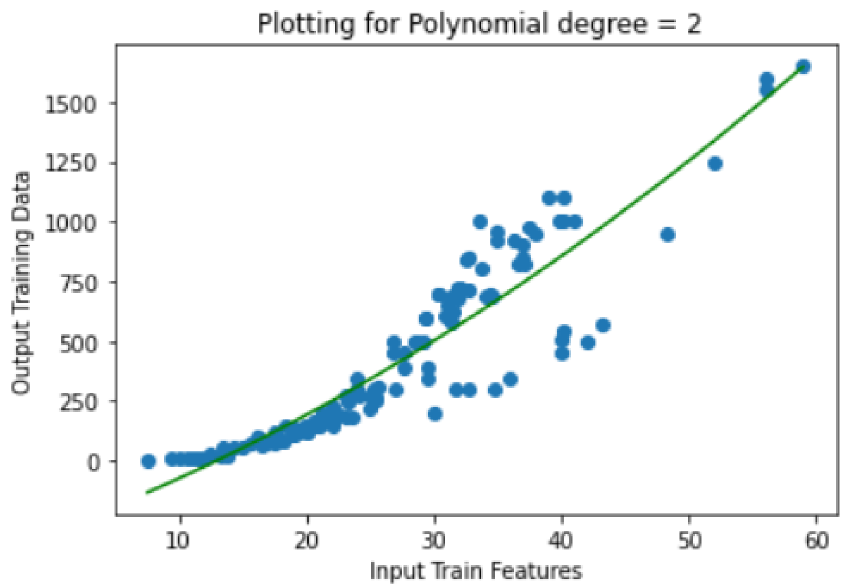


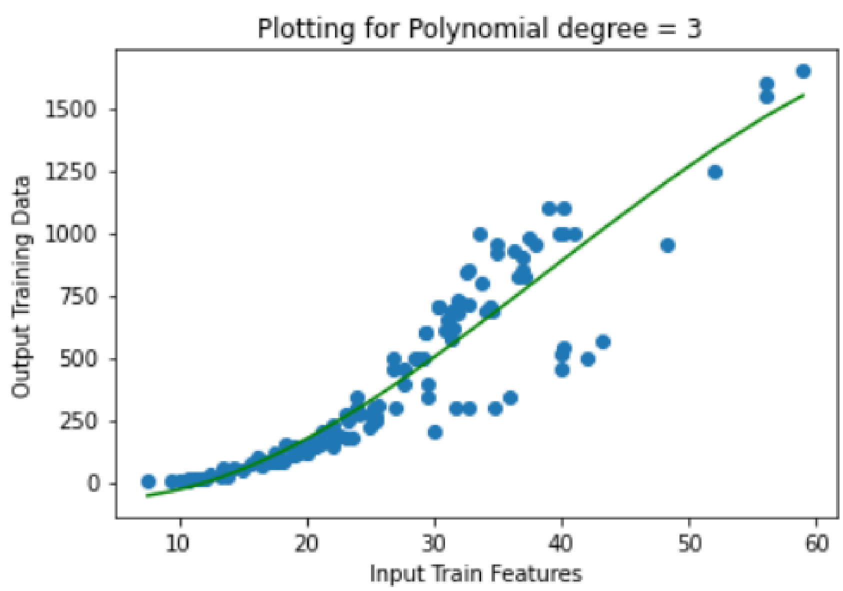
(c) model features = ‘Length1’, 'Width', 'Height'
output = ‘Weight’
initial weights = [-10, 1, 1,1] (intercept, Length1, Width, Height respectively)
step size (learning rate) = 4e-10
tolerance = 1.4e4

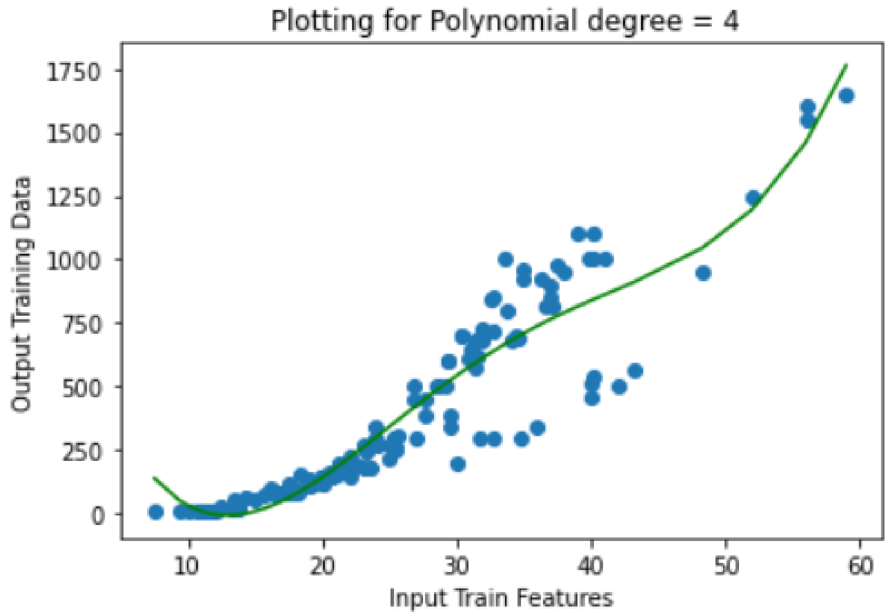


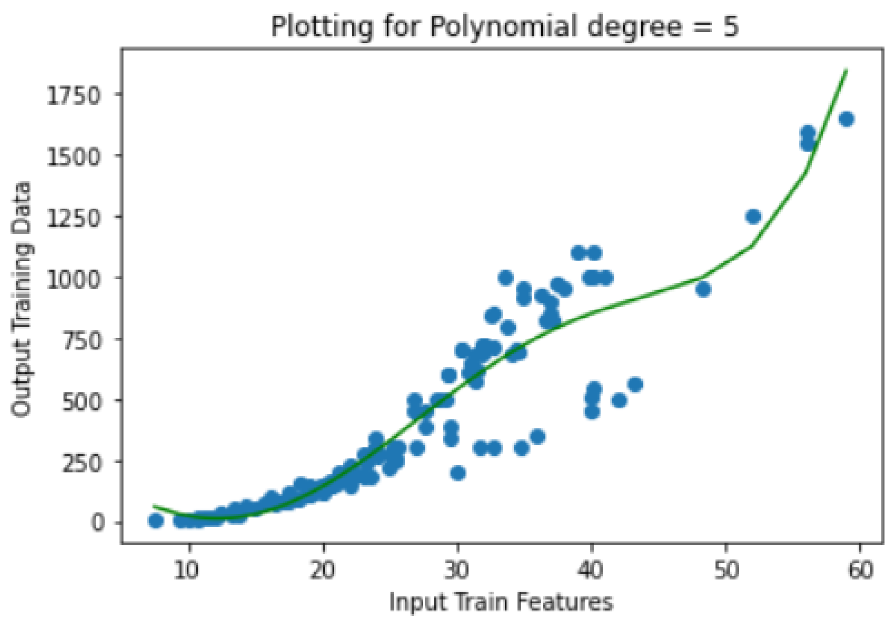


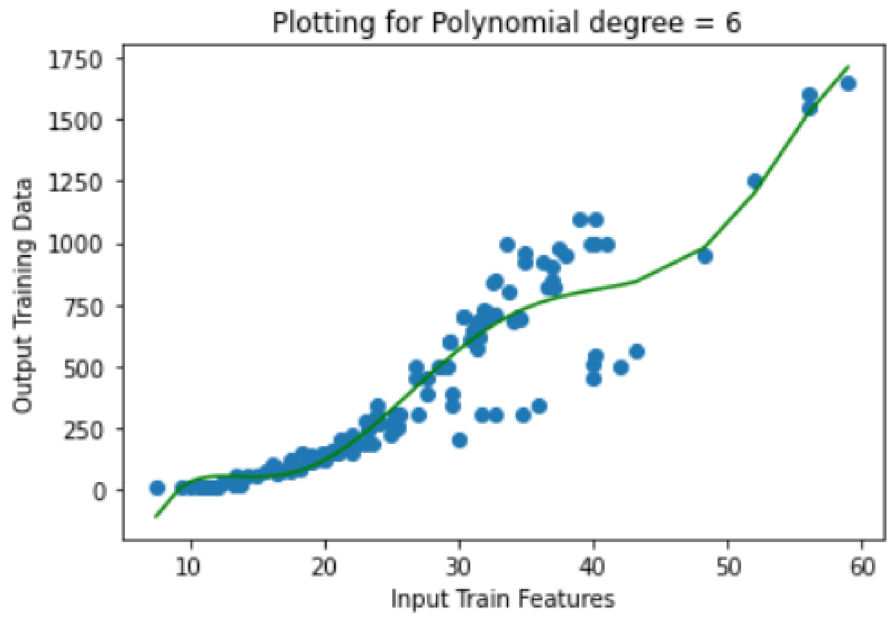
Q:3 Results of Polynomial for Degree=2,3,4,5,6 :






Plotting for Polynomial degree = 4
Output Training Data
Input Train Features
1750
1500
1250
1000
750
500
250
0
10
20
30
40
50
60


Plotting for Polynomial degree = 5
Output Training Data
Input Train Features
1750
1500
1250
1000
750
500
250
0
10
20
30
40
50
60



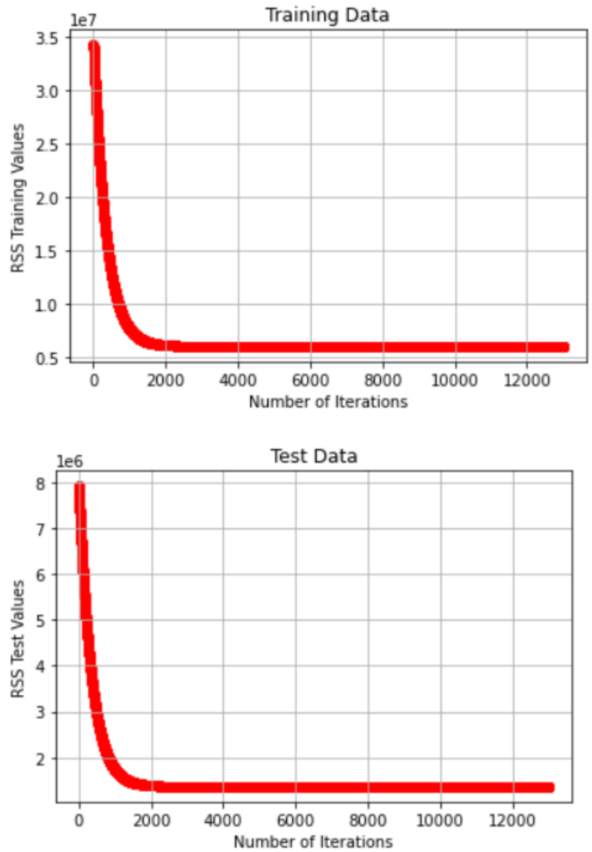
Now, if I change the value of eta(η) that is Step Size to all current values but raise to (-9).

i.e. new values of eta(η) for sub-questions of Q:2 are as follows:

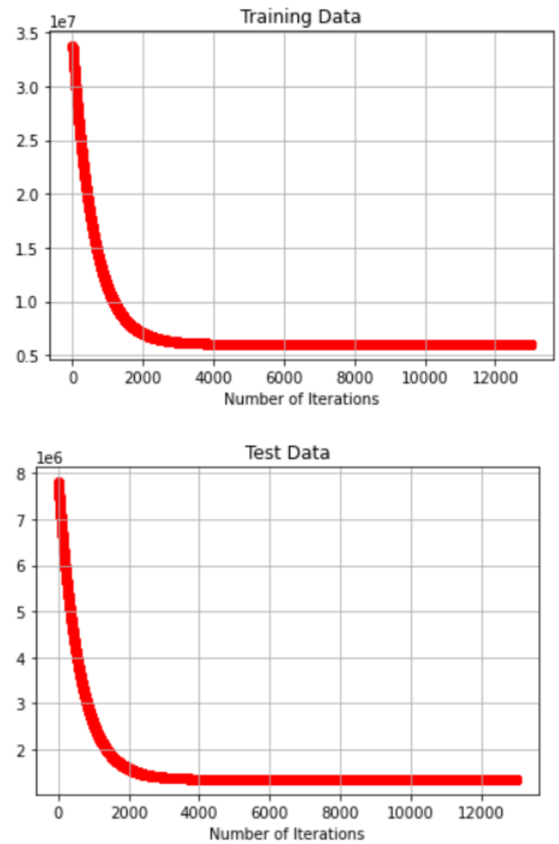
(a) η=7e-9 = $7*(10^{-9})$
(b) η=4e-9 = $4*(10^{-9})$
(c) η=4e-9 = $4*(10^{-9})$

Then we will get perfect graph of Gradient Descent Algorithm, which is as below:

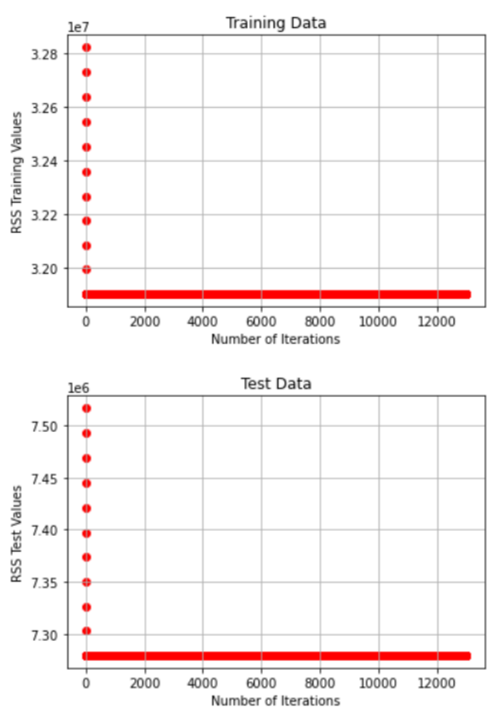
Thus, from below graphs, we can say that with this value of η, it converges faster.



Q:3(a) With Modified η



Q:3(b) With Modified η



Q:3(c) With Modified η

Acknowledgment:

I would like to express my gratitude to Dr. Abdul Bais and Dr. Muhammad Hamza Asad for their continuous guidance and encouragement during this course. Without their support, this work would not been possible. The notes prepared by them is really comprehensive and has a deep learning.