

TO TRAIN NETWORK USING GRADIENT DESCENT

W, b = -2

Learning rate	Max epoch	Loss
0.01	1000	0.414
0.1	1000	0.01644
1	1000	2.4 e -11
10	1000	2.5 e -32
100	1000	0.42499
0.01	10000	0.016
0.08	10000	1.6 e -9
0.09	10000	2.04 e -10
0.1	10000	2.5 e -11
1	10000	2.5 e -30
10	10000	2.5 e -32
20	10000	6.5 e -33
50	10000	0.0004

So, as we can see from the table

- If we keep max. epoch 1000 and increase learning rate starting from 0.01 then it will decrease error (near to zero). But after increasing learning rate to 100, error will start increasing again.
- If we keep max. epoch 10,000 then we will observe same behaviour. Error will decrease until we keep learning rate below 20. After that, it will start increasing again.
- So, Learning rate should be kept in 20 and max. epoch 10,000 for minimum loss.