* The observation by running the gradient-descent program is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Weight(w) | Bias(b) | Learning Rate(eta) | Epoch(epoch) | Error(err) |
| 0.9 | 1 | 0.04 | 700 | 0.0273 |
| 0.9 | 1 | 0.09 | 700 | 0.0059 |
| 0.9 | 1 | 0.09 | 1000 | 0.0023 |
| 0.9 | 0.3 | 0.09 | 1000 | 0.0017 |
| 0.9 | 0.3 | 0.09 | 2000 | 0.00017 |
| 0.9 | 0.2 | 0.09 | 2000 | 0.00016 |
| 0.9 | 0.01 | 0.09 | 2200 | 0.00010 |

* At the end of the observation, I got the error=0.00010 which was the minimum out of all the errors.
* Conclusion:
* The weight and bias affect the error the most.
* The learning rate make the change till a certain point. Here if it is decreased or increased it will increase the error.
* The epoch should be less than 2500, if we take epoch greater than 2500, there will be increase in the error. So the epoch must be between (0-2500).
* The weight and bias makes great difference as we can see loss decreasing as we change the values for 1000 epochs.