X=[0.5,2.5]

Y=[0.2,0.9]

Gamma=0.9

W=-2, B=-2, Learning rate=0.1, Epoch=1000

|  |  |
| --- | --- |
| Vanilla GD Time | 0.03590703010559082 |
| Momentum GD Time | 0.04286837577819824 |
| Nesterov Accelerated GD Time | 0.04091167449951172 |
| Vanilla GD Error | 0.01644945958207577 |
| Momentum GD Error | 0.12027135917909929 |
| Nesterov Accelerated GD Error | 3.4249204911514565e-12 |

W=0, b=0, learning rate=0.1, epoch=100

|  |  |
| --- | --- |
| Vanilla GD Time | 0.0019807815551757812 |
| Momentum GD Time | 0.00698089599609375 |
| Nesterov Accelerated GD Time | 0.0019943714141845703 |
| Vanilla GD Error | 0.04925120442438629 |
| Momentum GD Error | 0.12027135917909929 |
| Nesterov Accelerated GD Error | 0.0008486015234387917 |

W=1, b=1, learning rate=0.2, epoch=100

|  |  |
| --- | --- |
| Vanilla GD Time | 0.001996278762817383 |
| Momentum GD Time | 0.0059888362884521484 |
| Nesterov Accelerated GD Time | 0.0009925365447998047 |
| Vanilla GD Error | 0.040598927882388114 |
| Momentum GD Error | 0.11618198660988774 |
| Nesterov Accelerated GD Error | 1.6705687495761948e-06 |

W=10, B=10, Learning Rate = 0.2 ,Epochs=1000

|  |  |
| --- | --- |
| Vanilla GD Time | 0.03989243507385254 |
| Momentum GD Time | 0.040891408920288086 |
| Nesterov Accelerated GD Time | 0.041887521743774414 |
| Vanilla GD Error | 0.324999755263292 |
| Momentum GD Error | 0.11618198660988771 |
| Nesterov Accelerated GD Error | 0.32499975512980184 |

W=20, B=10, Learning rate=0.4, Epochs=100

|  |  |
| --- | --- |
| Vanilla GD Time | 0.002967357635498047 |
| Momentum GD Time | 0.0019927024841308594 |
| Nesterov Accelerated GD Time | 0.0019948482513427734 |
| Vanilla GD Error | 0.32499999835107707 |
| Momentum GD Error | 0.10951047538568884 |
| Nesterov Accelerated GD Error | 0.32499999835107585 |

**Conclusion :**

From the above tables, we can see that if we initialize weight and bias values near around its optimal value then in that case Nesterov Accelerated Gradient Descent performs way better than the momentum and vanilla gradient descent. But if we initialize weight and bias value very far from its optimal value then in that case momentum gradient descent perform better than the nesterov accelerated and vanilla gradient descent.