Momentum Gradient vs Vanilla Gradient vs Nesterov Gradient

Weight	Bias	Learning Rate	Epoch	Vanilla Gradient time	Momentum Gradient time	Vanilla Gradient Loss	Momentum Gradient Loss	Nesterov Gradient Loss	Nesterov Gradient Time
-2	-2	5	100	0.015625	0.015625	7.11e^-07	1.99e^-07	1.88e^-07	0.015625
-2	-2	5	300	0.0	0.0	4.41e^-16	9.36e^-18	1.02e^-17	0.015625
-2	-2	5	500	0.015625	0.0	2.79e^-25	4.47e^-28	5.64e^-28	0.03125
-2	-2	5	700	0.0	0.015625	1.00e^-31	3.73e^-32	3.73e^-32	0.03125
-2	-2	5	1000	0.015625	0.0	1.00e^-31	3.73e^-32	3.73e^-32	0.03125

Conclusion:

- Here we I have taken epoch range from 100 to 1000 and least error achieved is 3.73e^-32
- From above calculations we can see that both momentum and Nesterov gradient descent are working more efficiently than vanilla gradient descent
- Further comparing between Nesterov and Momentum gradient descent we can see that at some points Nesterov is more efficient than Momentum.
- Here we have taken only two points so major difference has is not been identified between Nesterov and Momentum while with more data difference can be seen.