Momentum Gradient Descent vs Vanilla Gradient Descent

Weight	Bias	Learning Rate	Epoch	Vanilla Gradient time	Momentum Gradient time	Vanilla Gradient Loss	Momentum Gradient Loss
-2	-2	5	100	0.015625	0.015625	9.76e^-15	8.71e^-17
-2	-2	5	300	0.0	0.0	4.41e^-16	9.36e^-18
-2	-2	5	500	0.015625	0.0	2.79e^-25	4.47e^-28
-2	-2	5	700	0.0	0.015625	1.00e^-31	3.73e^-32
-2	-2	5	1000	0.015625	0.0	1.00e^-31	3.73e^-32

Conclusion:

- Here we can see that momentum gradient descent is working more efficiently than vanilla gradient descent
- Starting from epoch 100 to 1000 momentum gradient descent error is lesser than vanilla gradient descent
- And even at or after 1000 iterations vanilla gradient descent is unable to achieve the minimum error whereas momentum gradient descent has already achieved at 700 iteration hence we can as that momentum gradient descent not only works efficiently but faster than vanilla gradient descent.