NESTEROV ACCELERATED GRADIENT DESCENT

W,b = 2 and learning rate = 10 and gamma=0.1

X = [0.5, 2.5]

Y = [0.2, 0.9]

Max_epoch	**VGD Loss**	**MGD Loss**	**NAG Loss**
100	1.62 e -12	7.34 e -14	1.71 e -15
200	5.67 e -22	1.33 e -24	4.41 e -26
400	2.50 e -32	6.54 e -33	6.16 e -33
800	2.50 e -32	6.54 e -33	6.16 e -33
2000	2.50 e -32	6.54 e -33	6.16 e -33

So, Both momentum based gradient descent and nesterov accelerated gradient descent performs better than vanilla gradient descent.