

Refer: <https://towardsdatascience.com/stochastic-gradient-descent-with-momentum-a84097641a5d>

Correction: @ 6:20 we wrote,  $w' = w_t - \gamma * v_{t-1}$  but it should be  $w' = w_{t-1} - \gamma * v_{t-1}$