

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

Refer: <http://ruder.io/optimizing-gradient-descent/>

Refer: <https://www.slideshare.net/SebastianRuder/optimization-for-deep-learning>