 are momentum.

 are gradients of the parameters we want to update.

are learning rate.

are smoothing term (avoids division by zero).

are weight.