 and  are momentum.

 and  are momentum coefficient.

 are gradients of the parameters we want to update.

are learning rate.

are updated learning rate.

are smoothing term (avoids division by zero).

are weight.