Input: ϵ_0, initial learning rate
Input: α, decay rate of learning rate
Input: β, momentum rate
Input: ρ, discount factor for historical gradient
Input: ζ, small constant to avoid zero division
Input: m, minibatch size
Input: k, epoch size
Input: θ, initial weights
Input: v, initial velocity
Input: $\mathbf{X}$, training dataset inputs
Input: $\mathbf{y}$, training dataset targets
Initialize: $r \leftarrow 0$, accumulation of historical gradient
Initialize: $j \leftarrow 1$, current epoch
while $j \leq k$ **do**
    update learning rate $\epsilon_j \leftarrow \epsilon_0 + \alpha(\epsilon_{j-1} - \epsilon_0)$
    while *stopping criteria is not satisfied* **do**
        $\{\mathbf{x}^1 ... \mathbf{x}^m\}$, $\{\mathbf{y}^1 ... \mathbf{y}^m\} \leftarrow$ get a sample from $\mathbf{X}$ and $\mathbf{y}$ randomly
        calculate estimation of gradient $\hat{g} \leftarrow \frac{1}{m} \sum_{i=1}^{m} L(f(\mathbf{x}^i; \theta), \mathbf{y}^i)$
        accumulate historical graidents $r \leftarrow \rho r + (1 - \rho)\hat{g} \odot \hat{g}$
        calculate step size $v \leftarrow \beta v - \frac{\epsilon_j}{\sqrt{\zeta + r}} \odot \hat{g}$
        update weights $\theta \leftarrow \theta + v$
    $j \leftarrow j + 1$ go to next epoch