1   **Input:** ϵ_0, initial learning rate
2   **Input:** α, decay rate of learning rate
3   **Input:** β, momentum rate
4   **Input:** m, minibatch size
5   **Input:** k, epoch size
6   **Input:** θ, initial weights
7   **Input:** v, initial velocity
8   **Input:** $\mathbf{X}$, training dataset inputs
9   **Input:** $\mathbf{y}$, training dataset targets
10 $j \leftarrow 1$
11 **while** $j \leq k$ **do**
12   update learning rate $\epsilon_j \leftarrow \epsilon_0 + \alpha(\epsilon_{j-1} - \epsilon_0)$
13   **while** *stopping criteria is not satisfied* **do**
14       $\{\mathbf{x}^1...\mathbf{x}^m\}, \{\mathbf{y}^1...\mathbf{y}^m\} \leftarrow$ get a sample from $\mathbf{X}$ and $\mathbf{y}$ randomly
15   calculate estimation of gradient $\hat{g} \leftarrow \frac{1}{m} \sum_{i=1}^{m} L(f(\mathbf{x}^i; \theta), \mathbf{y}^i)$
16   update velocity $v \leftarrow \beta v - \epsilon_j \hat{g}$
17   update weights $\theta \leftarrow \theta + v$
18     $j \leftarrow j + 1$ go to next epoch