$\nabla L_{in}(\boldsymbol{w}) = \frac{1}{B}\sum_{n=1}^{B} \nabla L_n(\boldsymbol{w})$

$\boldsymbol{m}_{t+1} = \lambda \boldsymbol{m}_t - \eta \nabla L_{in}(\boldsymbol{w}_t)$，　　（$\boldsymbol{m_0} = 0$）

$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t + \boldsymbol{m}_{t+1}$

（6）Adam

$\boldsymbol{m_{t+1}} = \beta_1 \boldsymbol{m}_t - (1-\beta_1)\nabla L_{in}(\boldsymbol{w_t})$，　　（$\boldsymbol{m_0} = 0$）

$\boldsymbol{v_{t+1}} = \beta_2 \boldsymbol{v}_t - (1-\beta_2)\left(\nabla L_{in}(\boldsymbol{w})\right)^2$，　　（$\boldsymbol{v_0} = 0$）

$\widehat{\boldsymbol{m}}_{\boldsymbol{t+1}} = \boldsymbol{m}_{t+1}/(1-\beta_1^{t+1})$

$\widehat{\boldsymbol{v}}_{\boldsymbol{t+1}} = \boldsymbol{v}_{t+1}/(1-\beta_2^{t+1})$

$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t - \eta \widehat{\boldsymbol{m}}_{t+1}/(\sqrt{\widehat{\boldsymbol{v}}_{t+1} + \epsilon})$