Adam optimization algorithm

$v_{dW} = 0, s_{dW} = 0, v_{db} = 0, s_{db} = 0$

$On\ iteration\ t:$

$compute\ dW, db\ using\ current\ mini-batch$

$v_{dW} = \beta_1 v_{dW} + (1-\beta_1)dW,\ v_{db} = \beta_1 v_{db} + (1-\beta_1)db \leftarrow\ 'mementum'$

$s_{dW} = \beta_2 s_{dW} + (1-\beta_2)dW^2,\ s_{db} = \beta_2 s_{db} + (1-\beta_2)db \leftarrow\ 'RMSprop'$

$v_{dW}^{correct} = \frac{v_{dW}}{(1-\beta_1^t)},\ v_{db}^{correct} = \frac{v_{db}}{(1-\beta_1^t)}$

$s_{dW}^{correct} = \frac{s_{dW}}{(1-\beta_2^t)},\ s_{db}^{correct} = \frac{s_{db}}{(1-\beta_2^t)}$

$W := W - \alpha \frac{v_{dW}^{correct}}{\sqrt{s_{dW}^{correct}}+\epsilon}$

$b := b - \alpha \frac{v_{db}^{correct}}{\sqrt{s_{db}^{correct}}+\epsilon}$

$Hyperparameters\ choice:$

$\alpha:\ needs\ to\ be\ tuned$

$\beta_1: 0.9$

$\beta_2 = 0.999$

$\epsilon: 10^{-8}$

$Adam:\ adaptive\ moment\ estimation$