Exponential Moving Averages
$V_0 = 0$
V1 = β*V0 + (1-β)*Θ1 [Current observation at time 1] [β=0.9]
V2 = β*V1 + (1-β)*Θ2 [Current observation at time 2]

Vt = β*Vt-1 + (1-β)*Θt[Current observation at time t]

Generalized form
V = β*V + (1-β)*Θ

Gradient Descent (mini-batch) with Momentum (SGD - Minibatch)
For each batch:
Compute dw, db

Vdw = β*Vdw + (1-β)*dw
Vdb = β*Vdb + (1-β)*db

w = w - alpha * Vdw
b = b - alpha * Vdb

Root Mean Square Propagation (RMSProp)
Sdw = β2*Sdw + (1-β2)*dw^2
Sdb = β2*Sdb + (1-β2)*db^2

w = w - alpha * dw/**sqrt**(Sdw + epsilon)
b = b - alpha * db/**sqrt**(Sdb + epsilon)
Epsilon = small value to prevent division by 0. Normally 10^(-8)

Adaptive Moment with Estimation (**Adam**)
For each batch:
Vdw = β1*Vdw + (1-β1)*dw
Vdw = Vdw/(1-$\beta1^t$)

Vdb = β1*Vdb + (1-β1)*db
Vdb = Vdb/(1-$\beta1^t$)

Sdw = β2*Sdw + (1-β2)*dw^2
Sdw = Sdw/(1-$\beta2^t$)

Sdb = β2*Sdb + (1-β2)*db^2
Sdb = Sdb/(1-$\beta2^t$)

w = w - alpha * Vdw/[**sqrt**(Sdw)+Epsilon]
b = b - alpha * Vdb/[**sqrt**(Sdb)+Epsilon]
β1 = 0.9
β2 = 0.999
Epsilon = 10^{-8}