# SGD

## Hàm mục tiêu:

f(w1, w2) = 0.1 \* w1^2 + 2 \* w2^2

## Gradient:

∂f/∂w1 = 0.2 \* w1, ∂f/∂w2 = 4 \* w2

## Khởi tạo ban đầu:

W = [-5, -2]

Learning rate (lr) = 0.4

## Epoch 1:

Gradient:

dW = [0.2 \* (-5), 4 \* (-2)] = [-1.0, -8.0]

Cập nhật W:

W = W - lr \* dW = [-5, -2] - 0.4 \* [-1.0, -8.0]

W = [-5 + 0.4, -2 + 3.2] = [-4.6, 1.2]

## Epoch 2:

Gradient:

dW = [0.2 \* (-4.6), 4 \* (1.2)] = [-0.92, 4.8]

Cập nhật W:

W = W - lr \* dW = [-4.6, 1.2] - 0.4 \* [-0.92, 4.8]

W = [-4.6 + 0.368, 1.2 - 1.92] = [-4.232, -0.72]

# SGD +Momentum

## Hàm mục tiêu:

f(w1, w2) = 0.1 \* w1^2 + 2 \* w2^2

## Gradient:

∂f/∂w1 = 0.2 \* w1, ∂f/∂w2 = 4 \* w2

## Khởi tạo ban đầu:

W = [-5, -2]

## Epoch 1:

dW = [0.2 \* (-5), 4 \* (-2)] = [-1.0, -8.0]

v = beta \* v + (1-beta)\*dw = [0.5 \* 0 + (1-0.5) \* (-1), 0.5 \* 0 + (1-0.5) \* (-8)] = [-0.5,-4]

w = w – lr\*v = [-5 – 0.6 \* (-0.5), -2 – 0.6 \* (-4)] = [-4.7, 0.4]

## Epoch 2:

dW = [0.2 \* (-4.7), 4 \* (0.4)] = [-0.94, 1.6]

v = beta \* v + (1-beta)\*dw = [0.5 \* (-0.5) + (1-0.5) \* (-0.94), 0.5 \* (-4) + (1-0.5) \* (1.6)] = [-0.72,-1.2]

w = w – lr\*v = [-4.268,1.12]

# RMSProp

## Hàm mục tiêu:

f(w1, w2) = 0.1 \* w1^2 + 2 \* w2^2

## Gradient:

∂f/∂w1 = 0.2 \* w1, ∂f/∂w2 = 4 \* w2

## Khởi tạo ban đầu:

W = [-5, -2]

## Epoch 1:

dW = [0.2 \* (-5), 4 \* (-2)] = [-1.0, -8.0]

s = gama \* s1 + (1-gama)\*dW^2 = [0.1, 6.4]

w = w – lr \* dW / (sqrt(s1 + epsilon)) = [-4.051, -1.051]

## Epoch 2:

dW = [0.2 \* (-4.051), 4 \* (-1.051)] = [-0.8102, -4.204]

s = gama \* s1 + (1-gama)\*dW^2 = [0.156, 7.527]

w = w – lr \* dW / (sqrt(s1 + epsilon)) = [-3.436, -0.591]

# ADAM

## Hàm mục tiêu:

f(w1, w2) = 0.1 \* w1^2 + 2 \* w2^2

## Gradient:

∂f/∂w1 = 0.2 \* w1, ∂f/∂w2 = 4 \* w2

## Khởi tạo ban đầu:

W = [-5, -2]

## Epoch 1:

dW = [0.2 \* (-5), 4 \* (-2)] = [-1.0, -8.0]

v = beta1\* v + (1-beta1)\*dw = [-0.1, -0.8]

s = beta2 \* s1 + (1-beta2)\*dW^2 = [0.001, 0.064]

v\_corr = v / 1 – beta1^t = [-1, -8]

s\_corr = s / 1 – beta2^t = [1, 64]

w = w – lr \* dW / (sqrt(s1 + epsilon)) = [-4.8, -1.8]

## Epoch 2:

dW = [0.2 \* (-4.8), 4 \* (--1.8)] = [-0.96, -7.2]

v = beta1\* v + (1-beta1)\*dw = [-0.186, -1.44]

s = beta2 \* s1 + (1-beta2)\*dW^2 = [0.0019, 0.1158]

v\_corr = v / 1 – beta1^t = [-0.9789, -7.5789]

s\_corr = s / 1 – beta2^t = [0.96, 57.9]

w = w – lr \* dW / (sqrt(s1 + epsilon)) = [-4.6, -1.6]