GRADIENT DESCENT

Formula for G.D = $\theta = \theta - \alpha \nabla J(\theta)$

Formula to calculate $\nabla J(\theta) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \theta x_i)^2 x_i$

α is the learning rate.

Dataset:

$(x, y) = (1,2), (2,3), (3,4)$

Initialize the weight parameter θ to 0

$\theta = 0$.

$\nabla J(\theta) = 1/3 \left[(2 - 0(1))^2 + (3 - 0(2))^2 + (4 - 0(3))^2\right] \times 1$

$\nabla J(\theta) = 2$

$\therefore \frac{1}{N} \sum_{i=1}^{N} (y_i - \theta x_i)^2 x_i$

Take $\alpha = 0.1$

$\theta = 0 - 0.1(2) = \boxed{-0.2}$

$\therefore \theta = \theta - \alpha \Delta J(\theta)$

Iteration 1

$\theta = -0.2$

$\nabla J(\theta) = 1/3 \left[(2 - (-0.2)(1))^2 + (3 - (-0.2)(2))^2 + (4 - (-0.2)(3))^2\right] \times 1$

$\nabla J(\theta) = 0.8$

$\theta = -0.2 - 0.1(0.8)$

$\boxed{\theta = -0.28}$

Iteration 2

$\theta = -0.28$

$\nabla J(\theta) = 1/3\left[(2-(-0.28)(1))^2 + (3-(-0.28)(2))^2 + (4-(-0.28)(3))^2\right] \times 1$

$\nabla J(\theta) = 0.48$

$\theta = -0.28 - 0.1(0.48)$

$\boxed{\theta = -0.328}$

Iteration 3

$\theta = -0.328$

$\nabla J(\theta) = 1/3\left[(2-(-0.328)(1))^2 + (3-(-0.328)(2))^2 + 4(-(-0.328)(3))^2\right] \times 1 = 0.248$

$\boxed{\theta = -0.36528}$

Iteration 4

$\theta = -0.36528$

$\nabla J(\theta) = 0.1248$

$\theta = -0.36528 - (0.1248)(0.1)$

$\boxed{\theta = -0.36528}$

We can see that the weight parameter converges to the value $\theta = -0.36528$ after 4 iterations.

Stochastic Gradient Descent

Formula for SGD:- $(y_i - \theta x_i^1)\, x_i = J(\theta)$

Weight Updation :- $\theta_{old} - \alpha J(\theta) = \theta_{new}$

Dataset:

$(x, y) = (1,2), (2,3), (3,4)$

$\therefore \theta = 0$

Iteration 1:- $x = 1$, $y = 2$ $\qquad \alpha = 0.1$

$J(\theta) = (2 - 0(1))1 \quad \therefore J(\theta) = (y_i - \theta x_i) x_i$

$= 2$

$\theta_{new} = 0 - 0.1(2) \quad \therefore \theta_{new} = \theta_{old} - \alpha J(\theta)$

$= -0.2$

Iteration 2:- $x = 2$, $y = 3$

$J(\theta) = (3 + 0.2(2))2$

$= 6.4$

$\theta_{new} = -0.2 - (6.4)0.1$

$= -0.84$

Iteration 3:- $x = 3$, $y = 4$

$J\theta = (4 + 0.84(3))3$

$= 19.5$

$\theta_{new} = -0.84 - 0.1(19.5)$

$= -2.79$

After 3 iterations, the weight parameter converges to the value -2.79

ADAPTIVE GRADIENT DESCENT

Formula for AdaGrad =

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{G_t + \varepsilon}} \cdot g_t$$

Dataset :

$(x, y) = (2,4), (5,7), (6,8)$

We initialize

$\theta = 0$

$\alpha = 0.1$

$G_t = 0$

ε = It's value is so small that it's has negligible impact on the result

Iteration 1

$x = 2$

$y = 4$

$g_t = 2(\theta \cdot x - y)$ $\quad \therefore g_t = 2(\theta_t x_i - y_i)$

$= 2(0 - 4(2))$

$\boxed{g_t = -16}$

$\Rightarrow G_t = G_t + (g_t)^2$

$= 0 + (-16)^2$

$\boxed{G_t = 256}$

$\Rightarrow \theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{G_t + \varepsilon}} \cdot g_t$

$= 0 - \frac{0.1}{\sqrt{256}} \cdot (-16)$ $\Rightarrow \boxed{\theta_t = 0.1}$

Iteration 2

$x = 5$

$y = 7$

$g_t = 2(0.1 - 7(5))$

$\boxed{g_t = -69.8}$

$G_t = 256 + (-69.8)^2$

$= 256 + 4872.04$

$\boxed{G_t = 5128.04}$

$\theta = 0.1 - \frac{0.1}{\sqrt{5128.04}} \cdot (-69.8)$

$\theta = 0.1 - \frac{0.1}{71.61} (-69.8)$

$\theta = 0.1 - 0.0013 (-69.8)$

$\theta = 0.1 + (0.0974)$

$\boxed{\theta = 0.1974}$

Iteration 3

$x = 6$

$y = 8$

$g_t = 2(0.1974 - 8(6))$

$\boxed{g_t = -95.60}$

$G_t = 5128.04 + (-95.60)^2$

$G_t = 14267.4$

$$\theta = 0.1974 - \frac{0.1}{\sqrt{14267.4}}(-95.60)$$

$$= 0.1974 - \frac{0.1}{119.44}(-95.60)$$

$$= 0.1974 - (0.00083)(-95.60)$$

$$= 0.1974 - (-0.0800)$$

$$\boxed{\theta = 0.2774}$$

After third iteration the parameter vector θ has converged to the value 0.2774.

Adadelta:-

Formula :- $\Delta\theta_t = -\alpha \frac{RMS[g_t]}{RMS[\Delta\theta_{t-1}]}$

$\therefore g_t^i = 2(\theta_t x_i - y_i)$

$\therefore x_i$ = feature

θ_t = Predicted output

Weight Updation :- $\theta_{t+1} = \theta_t + \Delta\theta_t$

Dataset :- $(x, y) = (1, 2), (2, 3), (3, 4)$

Iteration 1 :- $x = 1, y = 2$

$\therefore \theta = 0$

$\therefore g_t = 2(\theta_t x_i - y_i)$

$g_t^1 = 2(0 \times 1 - 2) = -4$

$x = 2, \quad y = 3$

$g_t^2 = 2(0 \times 2 - 3) = -6$

$x = 3, \quad y = 4$

$g_t^3 = 2(0 \times 3 - 4) = -8$

Calculate RMS :-

$$\Rightarrow RMS[g_t] = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[g_t^i]^2}$$

$$= \sqrt{\frac{1}{3}\left((-4)^2 + (-6)^2 + (-8)^2\right)}$$

$$= \sqrt{\frac{1}{3}(16 + 36 + 64)}$$

$$= \sqrt{\frac{116}{3}}$$

$$= \sqrt{38.67}$$

$$RMS[g_1] \approx 6.22$$

Now, we will calculate $\Delta\theta_t$

$$\Rightarrow \Delta\theta_t = -\alpha \frac{RMS[g_t]}{RMS[\Delta\theta_{t-1}]} \qquad \therefore \alpha = 0.1$$

$\therefore RMS[\Delta\theta_{t-1}]$ represents the RMS of previous iteration but for 1st iteration, we will assume it's 1 because there's no previous update

$$\Delta\theta_t = -0.1 \frac{6.22}{1} \qquad \therefore RMS[\Delta\theta_{t-1}] = 1$$

$$\therefore \Delta\theta_t = -\alpha \frac{RMS[g_t]}{RMS[\Delta\theta_{t-1}]}$$

$$\Delta\theta_1 \approx -0.622$$

Now for weight updation:-

$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

$$\theta_{0+1} = \theta_0 + (-0.622)$$

$$\theta_1 = 0 + (-0.622)$$

$$\theta_1 = -0.622$$

Iteration 2:-

Calculate g_t^i

$$g_t^1 = 2(-0.622 \times 1 - 2) = -4.244$$

$$g_t^2 = 2(-0.622 \times 2 - 3) = -8.488$$

$$g_t^3 = 2(-0.622 \times 3 - 4) = -11.732$$

$$RMS[g_2] = \sqrt{\frac{1}{3}\left((-4.244)^2+(-8.488)^2+(-11.732)^2\right)}$$

$$= \sqrt{\frac{1}{3}(18.01+72.04+137.63)}$$

$$= \sqrt{\frac{227.68}{3}}$$

$$= \sqrt{75.89}$$

$$RMS[g_2] \approx 8.71$$

$$\Delta\theta_2 = -0.1\ \frac{8.71}{1}$$

$$\Delta\theta_2 = -0.871$$

Update weight

$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

$$\theta_2 = -0.622 - 0.871$$

$$\theta_2 = -1.493$$

Iteration 3:-

Calculate g_t^i

$$g_t^1 = 2(-1.493 \times 1 - 2) = -6.986$$

$$g_t^2 = 2(-1.493 \times 2 - 3) = -11.972$$

$$g_t^3 = 2(-1.493 \times 3 - 4) = -16.958$$

$$RMS[g_3] = \sqrt{\frac{1}{3}\left((-6.986)^2 + (-11.972)^2 + (-16.958)^2\right)}$$

$$= \sqrt{\frac{1}{3}(48.80 + 143.32 + 287.57)}$$

$$= \sqrt{\frac{479.69}{3}}$$

$$= \sqrt{159.89}$$

$$RMS[g_3] \approx 12.64$$

$$\Delta\theta_3 = -0.1 \frac{12.64}{1}$$

$$\Delta\theta_3 = -1.264$$

Update weight

$$\theta_3 = -1.493 + (-1.264)$$

$$\theta_3 = -2.757$$

Root Mean Square Propagation

Step 1: Initialization

$\theta = 1$

$\alpha = 0.01$

$\gamma = 0.9$

$E[g^2] = 0$

$\varepsilon = 1e\text{-}8$

Dataset:

$(x, y) = (1,2), (2,3), (3,4)$

Step 2: Compute MSE and Gradient for θ

For each data point (x, y)

For θ:

- Compute the predicted value using the current θ: $y_{pred} = \theta . x$
- Calculate the mean squared error (MSE) for that data point: $MSE = 1/2 . (y_{pred} - y)^2$

Compute the gradient with respect to θ:

$$\partial MSE / \partial \theta = (y_{pred} - y) . x$$

Step 3: Update $E[g^2]$ for θ

Update the moving average of squared gradient for θ

$$E[g^2] = \gamma . E[g^2] + (1 - \gamma) . (\partial MSE / \partial \theta)^2$$

Step 4: Update θ

Update θ using the RMSprop formula:

$$\theta = \theta - (\alpha / \sqrt{E[g]^2 + \varepsilon}) . \partial MSE / \partial \theta$$

Iteration 1: For the data point (1, 2)

$y_{pred} = 1 \cdot 1 = 1$

$MSE = 1/2 (1-2)^2 = 0.5$

$\partial MSE / \partial \theta = (1-2) \cdot 1 = -1$

Update $E[g^2] = 0.9(0) + 0.1(-1)^2 = 0.1$.

update θ:

$\theta = 1 - (0.01 / \sqrt{0.1 + 1e-8})(-1) \approx 1.005$.

Iteration 2:

For the datapoint (2, 3)

$y_{pred} = 1.005(2) \approx 2.010$.

$MSE = 1/2 (2.010 - 3)^2 \approx 0.505$

$\partial MSE / \partial \theta \approx (2.010 - 3) 2 \approx -1.979$

Update $E[g^2]$:

$E[g^2] \approx 0.9(0.1) + 0.1(-1.979)^2 \approx 0.2979$

Update θ:

$\theta \approx 1.005 - (0.01 / \sqrt{0.2979 + 1e-8})(-1.979) \approx 1.008$

Iteration 3:

For the datapoint (3, 4)

$y_{pred} \approx 1.005 \quad (3) \approx 3.025$

$MSE \approx 1/2 (3.025 - 4)^2 \approx 0460$

$\partial MSE / \partial \theta \approx (3.025 - 4) 3 \approx -2.924$

update $E[g^2]$:

$E[g^2] \approx 0.9(0.2979) + 0.1(-2.924)^2 \approx 0.517$.

Update θ:

$\theta \approx 1.008 - (0.01 / \sqrt{0.517 + 1e-8})(-2.924) \approx 1.017$

Adam:-

Formulas:-

1^{st} Moment = $m_t = \beta_1 \times m_{t-1} + (1-\beta_1) \times g_t$ $\Rightarrow$ Mean

2^{nd} Moment = $v_t = \beta_2 \times v_{t-1} + (1-\beta_2)(g_t)^2$ $\Rightarrow$ Varience

Bias = $\hat{m}_t = \frac{m_t}{1-\beta_1^t}$

$\hat{v}_t = \frac{v_t}{1-\beta_2^t}$

$$\theta_{t+1} = \theta_t - \left(\alpha / \left(\sqrt{\hat{v}_t + \varepsilon}\right)\right) \times \hat{m}_t$$

$\therefore w, b \Rightarrow \theta_{t+1} = \theta_t - \frac{\alpha \hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}}$ $\Rightarrow$ Parameter Update

Initialization:

$\alpha = 0.01$

$\beta_1 = 0.9$

$\beta_2 = 0.999$

$\varepsilon = 1e-8$

$w = 0$ (slope)

$b = 0$ (intercept)

$m_w = 0$ (1^{st} moment for slope)

$m_b = 0$ (1^{st} moment for intercept)

$v_w = 0$ (2^{nd} moment for slope)

$v_b = 0$ (2^{nd} moment for intercept)

$t = 0$

Iteration 1:- $(x, y) = (1, 2)$

$$y_{pred} = 0.1 + 0 = 0$$

$$\frac{\partial MSE}{\partial w} = (y_{pred} - 2)\,1 = -2$$

$$\frac{\partial MSE}{\partial b} = (y_{pred} - 2) = -2$$

Update 1st Moments for w & b

$$m_w = 0.9 \times 0 + 0.1 \times (-2) = -0.2$$

$$m_b = 0.9 \times 0 + 0.1 \times (-2) = -0.2$$

Update 2nd Moments for w & b

$$v_w = 0.999 \times 0 + 0.001 \times (-2)^2 = 0.002$$

$$v_b = 0.999 \times 0 + 0.001 \times (-2)^2 = 0.002$$

Calculate bias in 1st Moment

$$\hat{m}_w = \frac{m_w}{1 - \beta_1^t} = -0.2$$

$$\hat{m}_b = \frac{m_b}{1 - \beta_1^t} = -0.2$$

Calculate bias in 2nd moments

$$\hat{v}_w = \frac{v_w}{1 - \beta_2^t} = 0.002$$

$$\hat{v}_b = \frac{v_b}{1 - \beta_2^t} = 0.002$$

Update parameters w & b:

$$w = w - \frac{\alpha}{\sqrt{\hat{v}_w + \varepsilon}} \cdot \hat{m}_w$$

$$= 0 - \frac{0.01}{\sqrt{0.002 + 1e-8}} (-0.2)$$

$$w \approx 0.141$$

$$b = b - \frac{\alpha}{\sqrt{\hat{v}_b + \varepsilon}} \cdot \hat{m}_b$$

$$= 0 - \frac{0.01}{\sqrt{0.002 + 1e+8}} (-0.2)$$

$$b \approx 0.141$$

Iteration 2:- $(x, y) = (2, 3)$

Initialization-

$w = 0.141$, $b = 0.141$

$m_w = -0.2$, $m_b = -0.2$

$v_w = 0.002$, $v_b = 0.002$

$t = 1$ (incremented from iteration 1)

$$y_{pred} = 0.141 \times 2 + 0.141$$

$$y_{pred} \approx 0.424$$

$$\frac{\partial MSE}{\partial w} \approx (0.424 - 3)\, 2$$

$$\frac{\partial MSE}{\partial w} \approx -4.351$$

$$\frac{\partial MSE}{\partial b} \approx (0.424 - 3)$$

$$\frac{\partial MSE}{\partial b} \approx -2.575$$

Update 1st Moment for w & b.

$$m_w = 0.9 \times (+0.2) + 0.1 \times (-4.351)$$

$$m_w \approx -0.385$$

$$m_b = 0.9 \times (-0.2) + 0.1 \times (-2.575)$$

$$m_b \approx -0.298$$

Update 2nd Moment for w & b.

$$v_w = 0.999 \times 0.002 + 0.001 \times (-4.351)^2$$

$$v_w \approx 0.004$$

$$v_b = 0.999 \times 0.002 + 0.001 \times (-2.575)^2$$

$$v_b \approx -0.298$$

Calculate bias in 1st Moment:

$$\hat{m}_w \approx \frac{-0.385}{1 - 0.9^2} \approx -0.053$$

$$\hat{m}_b \approx \frac{-0.298}{1 - 0.9^2} \approx -0.041$$

Calculate bias for 2nd Moment

$$v_w \approx \frac{0.004}{1-0.999^2} \approx 0.111$$

$$v_b \approx \frac{0.0023}{1-0.999^2} \approx 0.058$$

Update w & b.

$$w \approx 0.141 - \frac{0.01}{\sqrt{0.111+1e-8}} \cdot (-0.053)$$

$$w \approx 0.283$$

$$b \approx 0.141 - \frac{0.01}{\sqrt{0.058+1e-8}} \cdot (-0.041)$$

$$b \approx 0.184$$

Iteration 3: $(x, y) = (3, 4)$

$$y_{pred} \approx 0.283 \times 3 + 0.184$$

$$y_{pred} \approx 0.917$$

$$\frac{\partial MSE}{\partial w} \approx (0.917-4) \times 3$$

$$\frac{\partial MSE}{\partial w} \approx -9.748$$

$$\frac{\partial MSE}{\partial b} \approx (0.917-4) \approx -3.082$$

Update 1st Moment for w & b

$$m_w \approx 0.9 \times (-0.053) + 0.1(-9.748)$$
$$\approx -0.964$$

$$m_b \approx 0.9 \times (-0.041) + 0.1(-3.082)$$
$$\approx -0.434$$

Update 2nd Moment for w & b

$$v_w \approx 0.999\,(0.111) + 0.001(-9.748)^2$$
$$\approx 0.166$$

$$v_b \approx 0.999\,(0.058) + 0.001(-3.082)^2$$
$$\approx 0.084$$

Calculate bias for 1st Moment

$$\hat{m}_w \approx \frac{-0.964}{1 - 0.9^3} \approx -0.268$$

$$\hat{m}_b \approx \frac{-0.434}{1 - 0.9^3} \approx -0.119$$

Calculate bias for 2nd Moment

$$\hat{v}_w \approx \frac{0.166}{1 - 0.999^3} \approx 0.333$$

$$\hat{v}_b \approx \frac{0.084}{1 - 0.999^3} \approx 0.167$$

Update w & b

$$w \approx 0.283 - \frac{0.01}{\sqrt{0.333 + 1e-8}} \cdot (-0.268)$$

$$w \approx 0.043$$

$$b \approx 0.184 - \frac{0.01}{\sqrt{0.167 + 1e-8}} \cdot (-0.119)$$

$$b \approx 0.259$$