06 Oct-2024

OPTIMIZERS

Types of Optimizers:
- Batch Gradient Descent
- Stochastic Gradient Descent
- Mini Batch Gradient Descent

challenges with GD optimizers:
- challenges to select a exact value of learning rate.
- learning rate scheduling (pre defined value)
- finding the optimal value of many weights.
- local minima (may be possible. loss function trapped in local minima)
- saddle point (where slop value are same on all directions)

$$\frac{\partial L}{\partial w} = 0$$ hence no update of weights.

Optimizers

Momentum

Adagrad

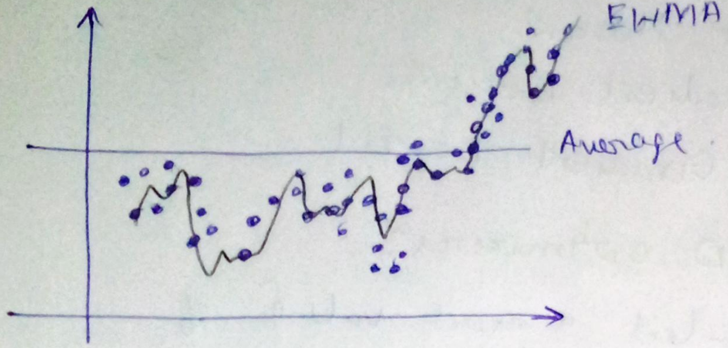
NAG

RMSprop

Adam

Exponential weighted moving Average (EWMA):

It is a technique that is used to find the trends in the time series data.



when it used:
- time series.
- financial.
- signal process.
- used in DL to build optimizers

the upcoming point are most preferable or more weightage. and in the flow of time the previous point's weight reduces.

mathematical formula:

$$V_t = \beta V_{t-1} + (1-\beta)\theta_t$$

Here β is constant rang $[0,1]$

$V_0 = \theta_0$

index	temp (θ)
D1	25
D2	13
D3	17
D4	31
D5	43

$V_1 = 0.9 \times V_0 + 0.1 \times 13$

$= 1.3$

$V_2 = 0.9 \times V_1 + 0.1 \times 17 =$

the weightage of previous point is depends on the value of 'β'

lower value of 'β' current point are more weightage.

mostly $\beta = 0.9$ used

Mathematical Intution:

$$V_t = \beta V_{t-1} + (1-\beta)\,\theta_t$$

where $V_0 = 0$

$$V_1 = (1-\beta)\,\theta_1$$

$$V_2 = \beta V_1 + (1-\beta)\,\theta_2$$

$$= \beta(1-\beta)\,\theta_1 + (1-\beta)\,\theta_2$$

$$V_3 = \beta V_2 + (1-\beta)\,\theta_3$$

$$= \beta^2(1-\beta)\,\theta_1 + \beta(1-\beta)\,\theta_2 + (1-\beta)\,\theta_3$$

$$V_4 = \beta V_3 + (1-\beta)\,\theta_4$$

$$= \beta^3(1-\beta)\,\theta_1 + \beta^2(1-\beta)\,\theta_2 + \beta(1-\beta)\,\theta_3 + (1-\beta)\,\theta_4$$

$$= (1-\beta)\left[\beta^2\theta_1 + \beta^2\theta_2 + \beta\theta_3 + \theta_4\right]$$

in the above equation we see that

because β in range $[0, 1]$

$$\beta^3 < \beta^2 < \beta$$

the coefficients of θ_1, θ_2 and θ_3 respectively