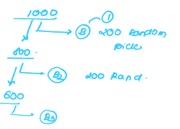
MBSGD

14 September 2023 07:08

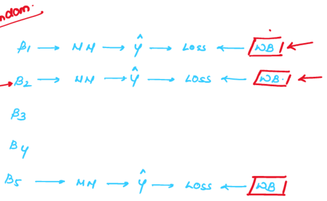
Mini Batch Stochastic Gradient Descent

GD = Entire Dataset

SGD = 1 Record picked.

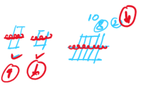


Random.



Iteration = 5
Epoch = 1

Batch size = 200
no. of batches. = 5



R1 R2 R3 R4

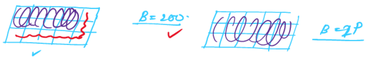
Bad loss ↑ WB ↑
good loss ↓ WB ↓

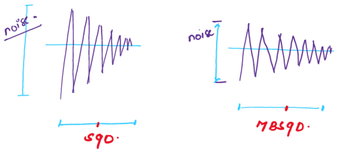
Here we have solve the problem of noise to great extent.

↳ unsolved.

Batch size are often tuned to computational Architecture

P → 2, 4, 8, 16, 32, 64, 128, ...





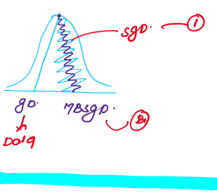
As the batch size increase noise reduces.

MBSGD 1000 → no. of batch = 1
Batch size = 1000.

↓
GD

Batch size ↑ smooth ↑

Smooth GD ↑ SGD ↓



Exponential Moving Avg.

1	2	3	4	5	6	7	8	9	10	11
20	30	20	30	40	30	30	40	40	50	?
				← 0.8 →			← 0.2 →			

MA 33.33
EMA 27.2 ✓

① Avg. (20+30+20)/3 = 60/3 (20)

② Moving Avg. 100/3 = 33.33

EMA.

B = 0.8. ← 5 days significance

$EG = \frac{1}{1-B}$

P = B× previous days. + (1-B)× Today.

= 0.8 × 24 (① 0.8) + (1-0.8) × 40 (② 0.2)

= 19.2 + 8

EMA = 27.2

$5 = \frac{1}{1-B}$

$5 - 5B = 1$

$5 - 1 = 5B$

$4 = 5B$

$B = \frac{4}{5}$ [B = 0.8]

B = 0.7 = 3
B = 0.8 = 5
B = 0.9 = 10 ✓
B = 0.99 = 100 days