

→ Optimizers

① GD

→ Entire dataset

→ 100 samples

↳ 1 epoch = 1 iteration

→ Huge Computation

② SGD

→ Sample by sample

→ 100 samples

↳ 1 epoch = 100 iterations

③ MBSGD

→ Batch by Batch

→ 100 samples (BS = 10)

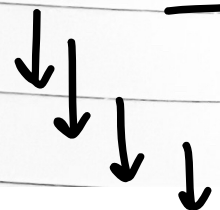
↳ 1 epoch = 10 iterations

④ AdaGrad

→ Adaptive GD

→ Can change LR
(Adaptive LR)

\underline{w}^X



⑤ Adadelta & ⑥ Rmsprop.

Extension of ✓ AdaGrad

Solve the LR problem.

→ here,

'Sum of Squares of G '

instead,

' EA '

→ more Imp → latest/
Recent
' G '

→ lessee Imp to
Past/Old.

⑦ Adam



Adaptive + Momentum