

→ Optimized

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) ↓ ↓ ↓

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