

→ Optimized

① G D

- 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 ($B_S = 10$)
 ↳ 1 epoch = 10 iterations

4

Adagrad

- Adaptive GD
 - Can change 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