

Batch gradient descent (calculated on entire dataset)

SGDM

Stochastic gradient descent

Gradient Descent

Quasi-hyperbolic momentum

Momentum

Aggregated momentum

QHAdam

Adam

AMSGrad

Adaptive learning rate

AdamW (weight decay)

DemonAdam

