

Gradient norm

10^0

10^{-1}

-0.002

0.000

0.002

Learning rate

