

Transformer: 42 NY Clients, SGD, lr:0.001, steps:60000, features:8 BIG\_BATCH 1024  
Gradient Norm

