recurrent network attention transformer @ add benchmark 0.200 0.175 0.150 -0.125 train mean squared error val mean squared error 0.100 test mean squared error 0.075 0.050 0.025 0.000 12 16 24 20 28 32

epochs