Epoch train log loss DY CG-like Adam lr=1e-3 Adam Ir=1e-3 0 CG-like Adam Ir=1e-4 Adam Ir=1e-4 train log loss CG-like Adam Ir=1e-5 Adam Ir=1e-5 CG-like Adam Ir=1e-6 Adam Ir=1e-6 -1050 150 200 25 75 100 125 175 epoch