val sparse categorical crossentropy @ memory\_benchmark memory\_augmented\_transformer Istm differentiable\_neural\_computer 10<sup>0</sup> unitary\_rnn matrix exponential unitary rnn transformer recurrent network attention transformer recurrent network augmented transformer gru  $10^{-1}$ neural circuit policies ct rnn ct gru ode Istm unitary ncp  $10^{-2}$  $10^{-3}$  $10^{-4}$ 15 30 120 45 60 75 90 105

epochs