Capacity during training process for different data sizes, activation functions, and numbers of layers

Experiment 2:
number of layers, data size
— 1, 50000
— 2, 50000
— 4, 50000
— 1, 70000
— 2, 70000
— 4, 70000
— 1, 100000
— 2, 100000
— 4, 100000

activation function
—— ReLU
---- GELU
······ RReLU
—·—· softmax