Mean capacity with standard deviation per epoch of training
for different data sizes, activation functions, and layers amount

Configuration
layers amount, 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