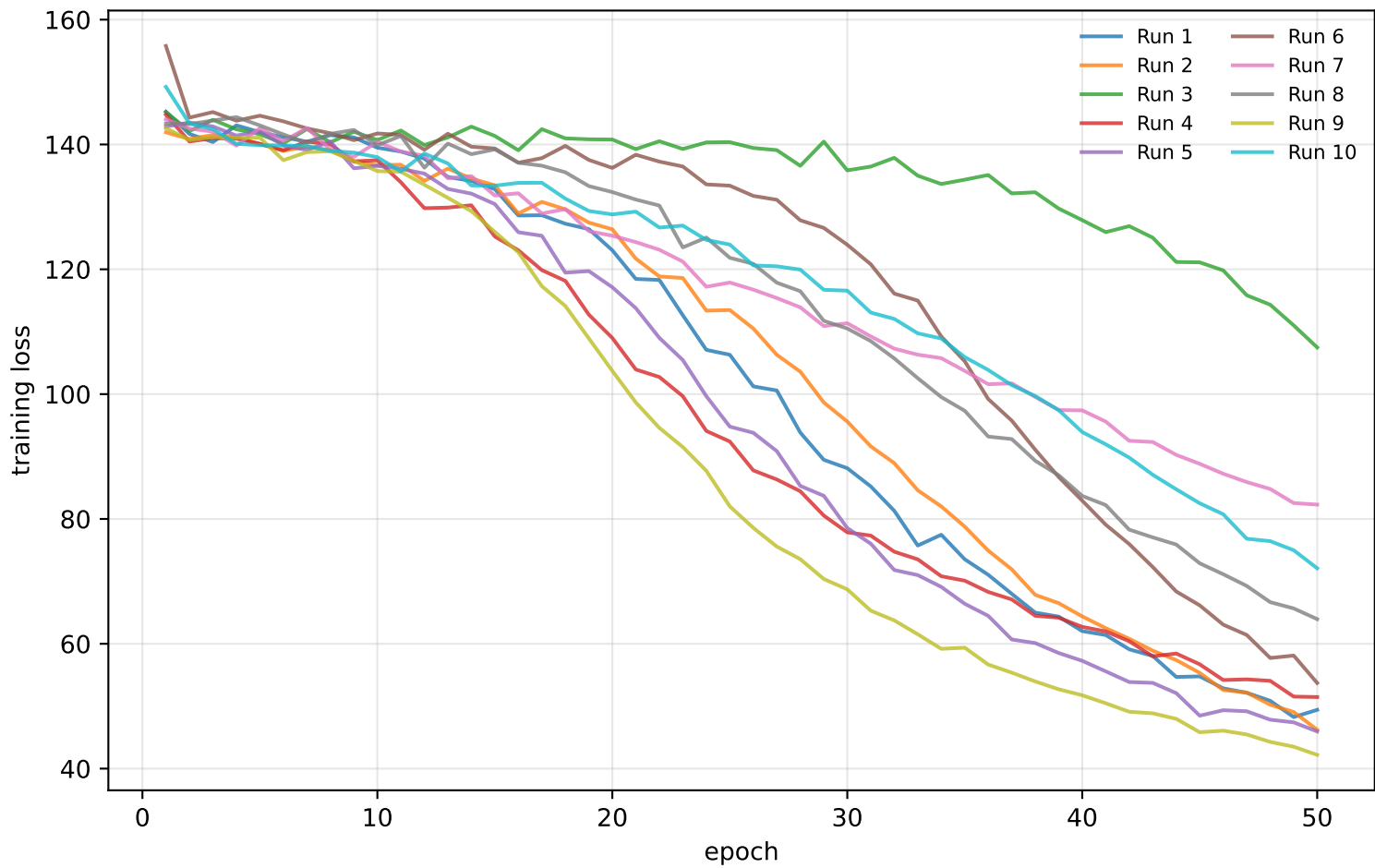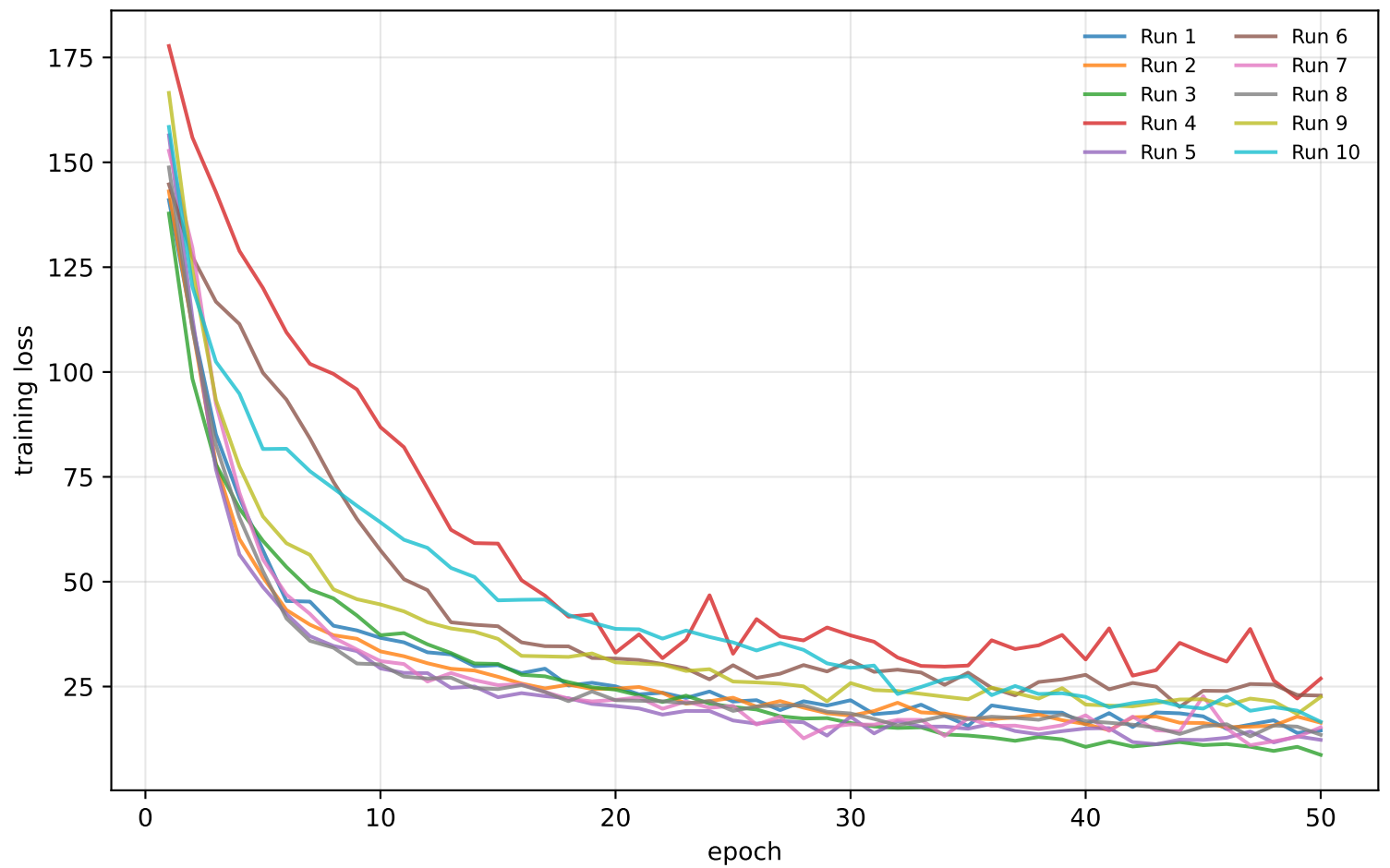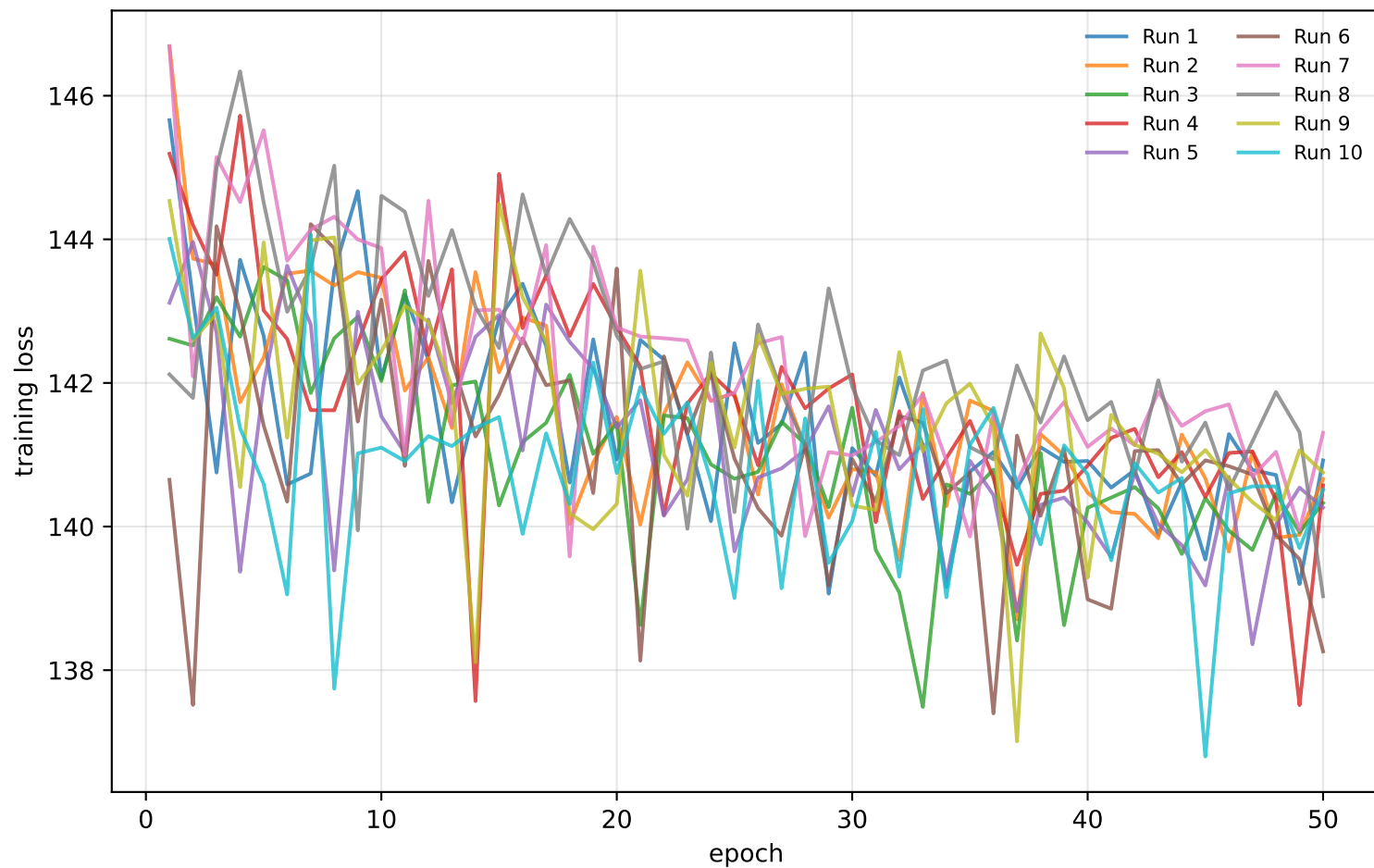Learning curves for 3-layer MLP with sigmoid (10 runs). Each line is a run; either activation is effective for a 3-layer network.
Learning rate: 0.05

Learning curves for 3-layer MLP with relu (10 runs). Each line is a run; either activation is effective for a 3-layer network.
Learning rate: 0.05

Learning curves for 5-hidden-layer MLP with sigmoid (10 units per layer, 10 runs). Learning question: does relu learn better in deeper networks?
Learning rate: 0.05

Learning curves for 5-hidden-layer MLP with relu (10 units per layer, 10 runs). Learning question: does relu learn better in deeper networks?
Learning rate: 0.05