

Train loss vs Iteration for different num\_layers (dim=8)

