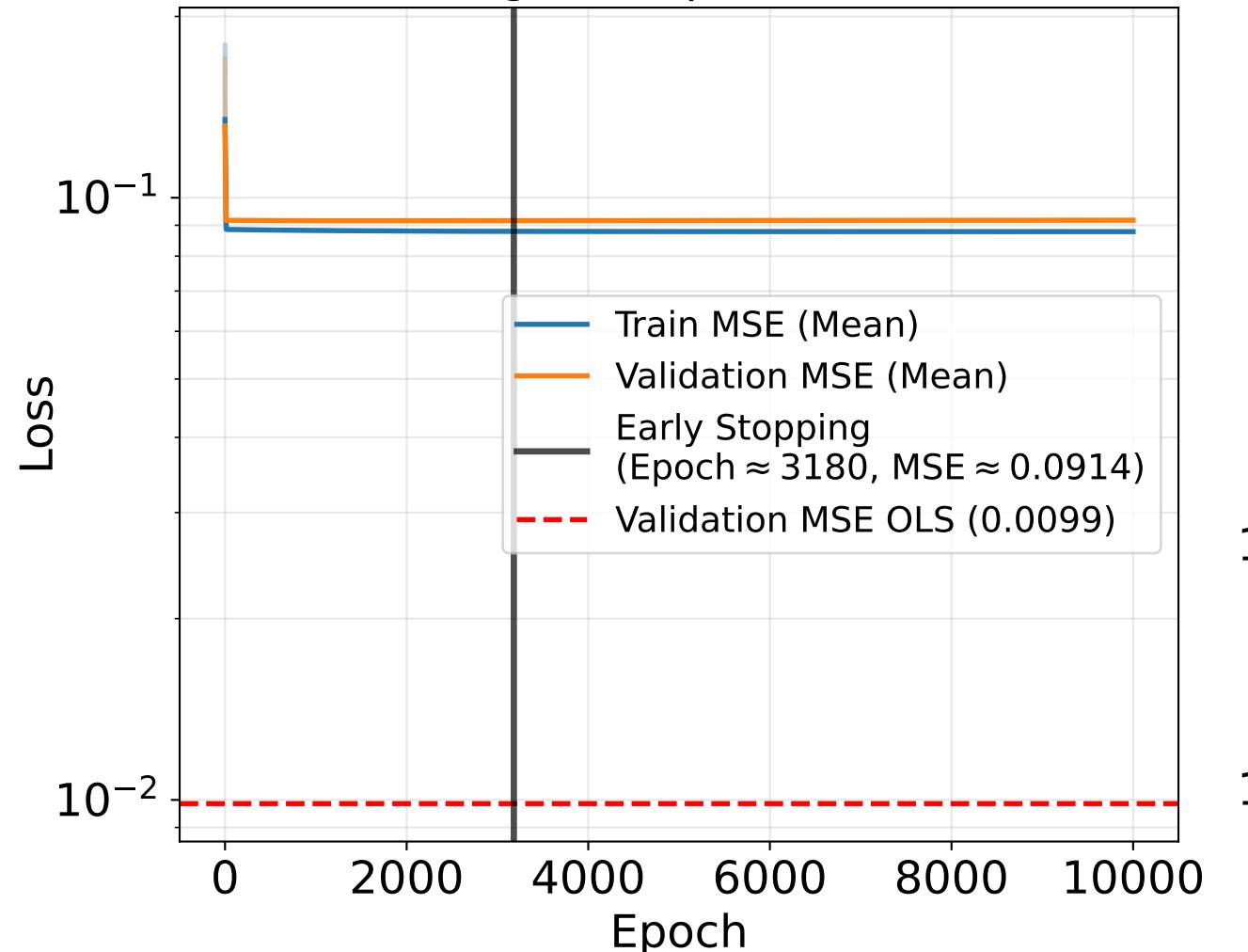


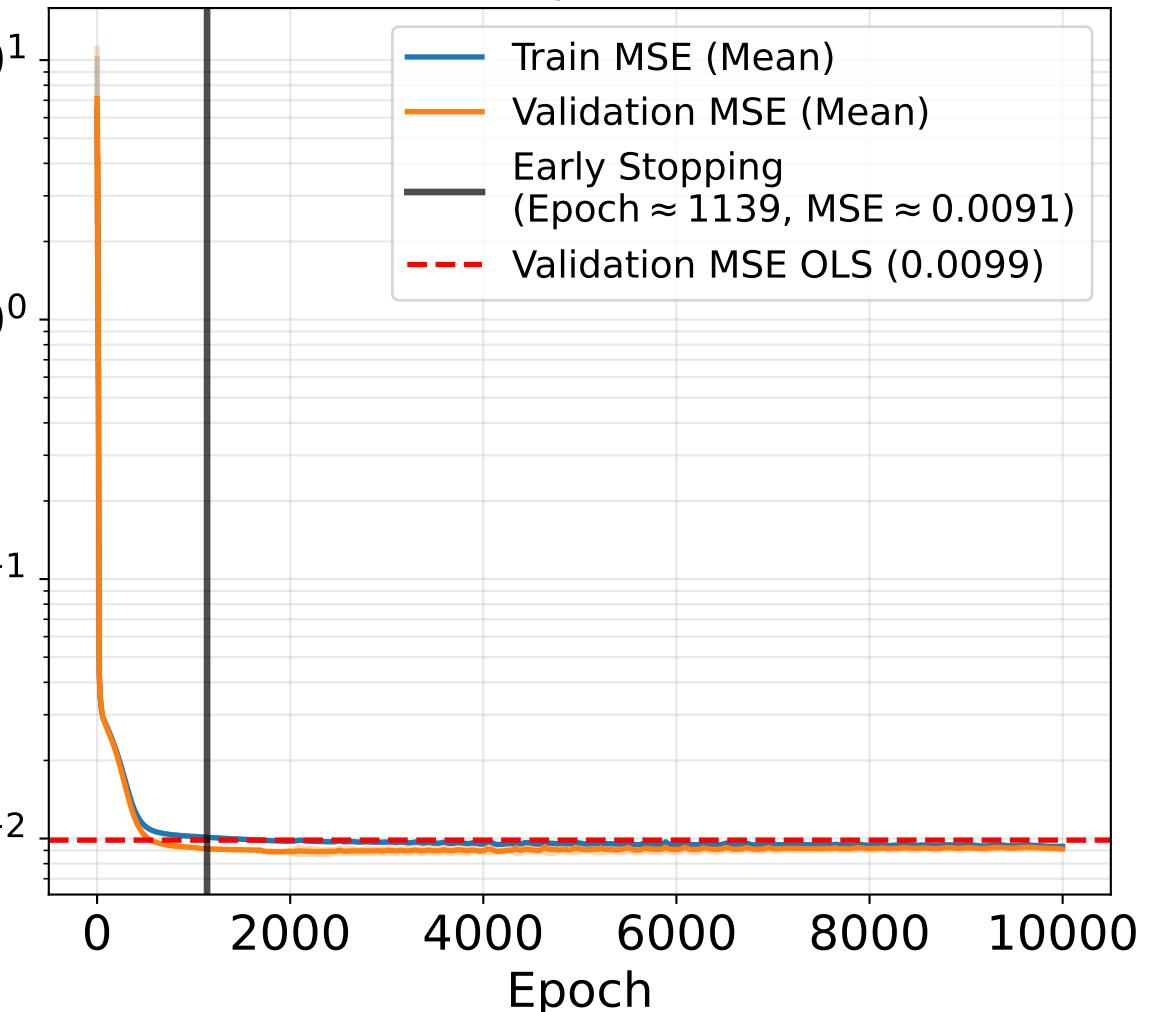
Learning Curves with GD ($N=300$, $\sigma=0.1$) - Averaged Over 3 Runs

2 Hidden Layers & 100 Nodes Each

Sigmoid: $\eta=0.0100$



ReLU: $\eta=0.1000$



LeakyReLU: $\eta=0.1000$

