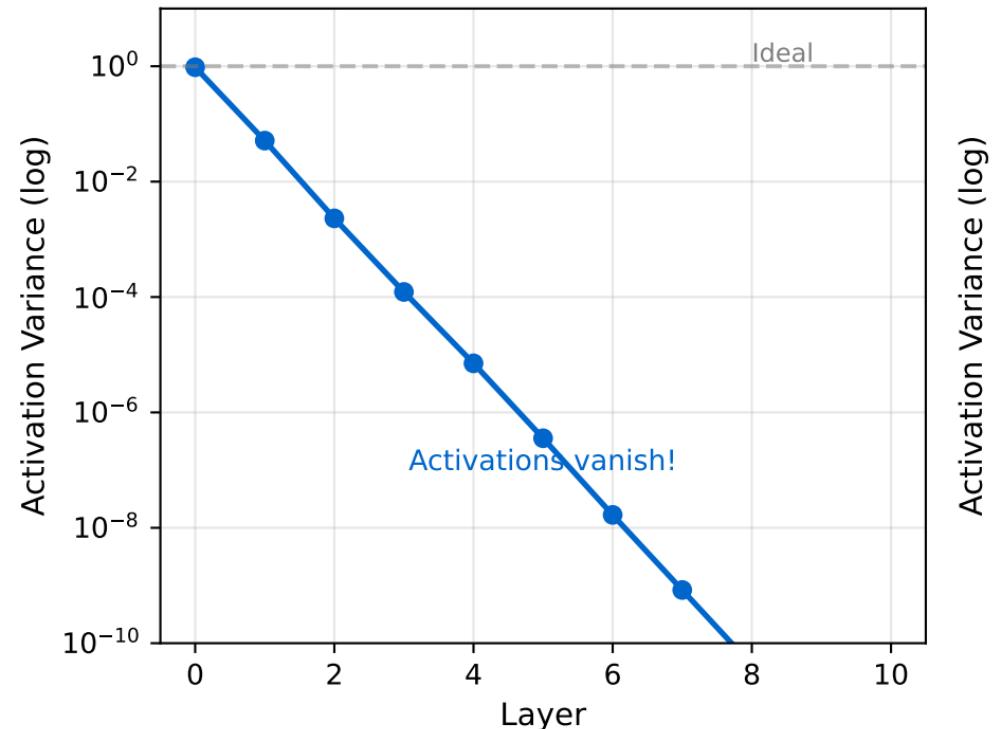
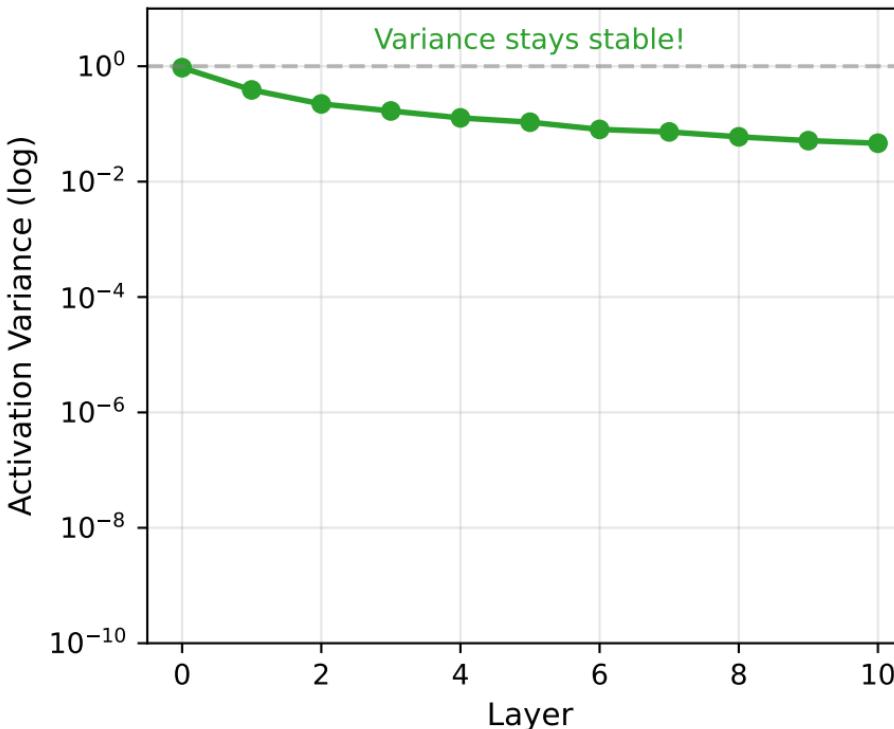


Weight Initialization: Getting It Right Matters

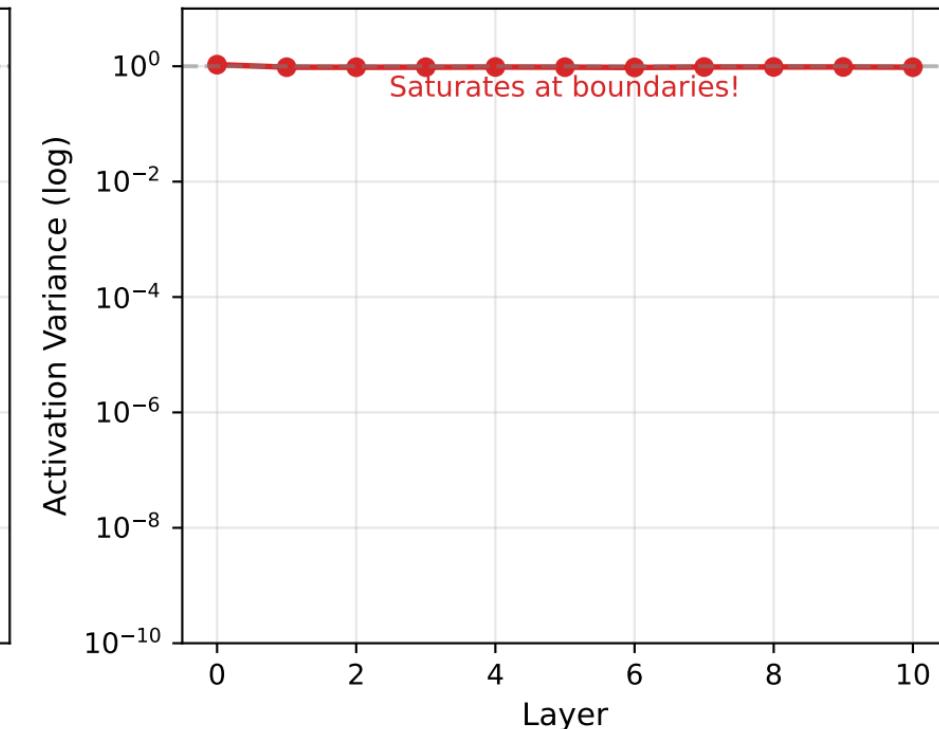
Init TOO SMALL (std=0.01)



Xavier Init (std= $1/\sqrt{n}$)



Init TOO LARGE (std=1.0)



Xavier (Glorot): $\sigma = \sqrt{1/n_{in}}$ for tanh | He: $\sigma = \sqrt{2/n_{in}}$ for ReLU