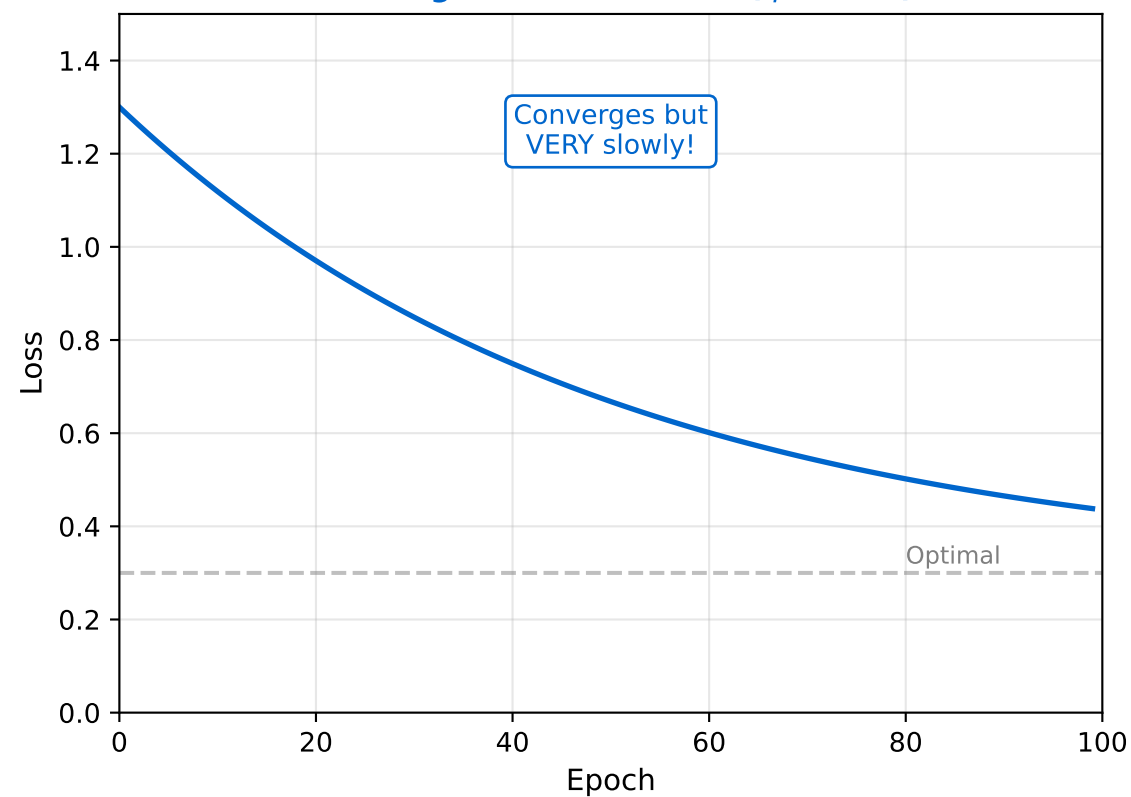
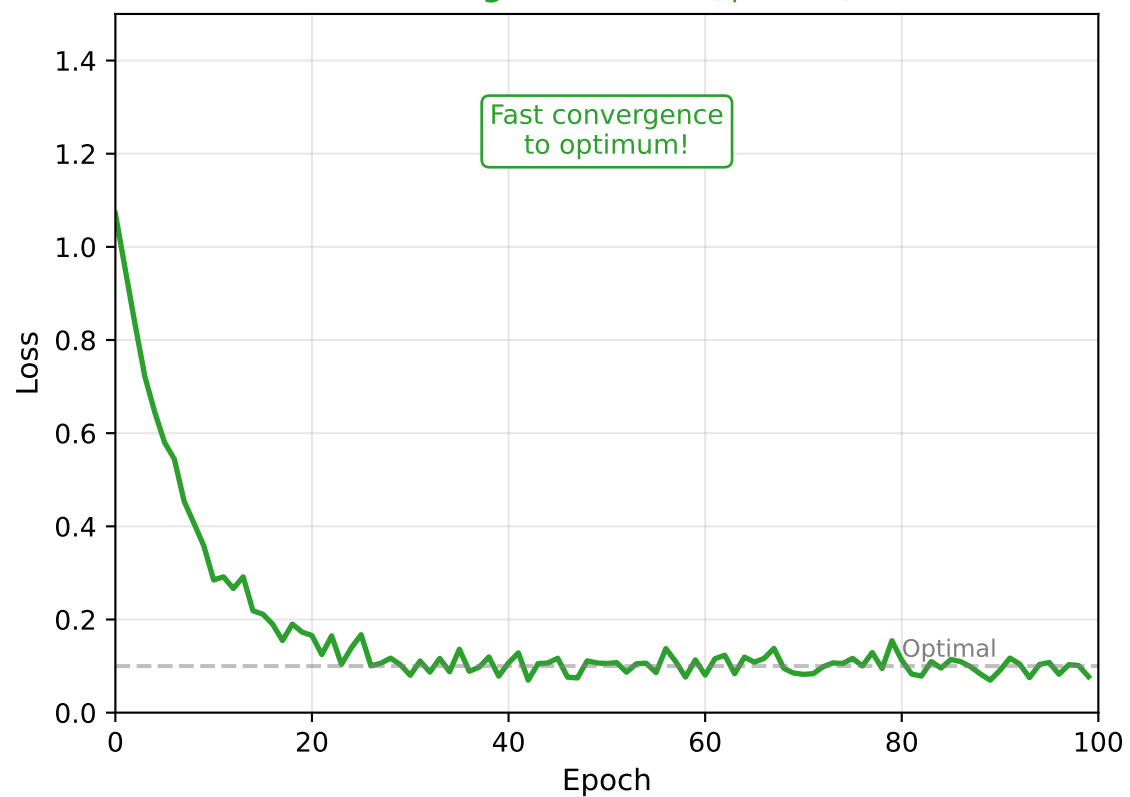


Learning Rate: The Most Important Hyperparameter

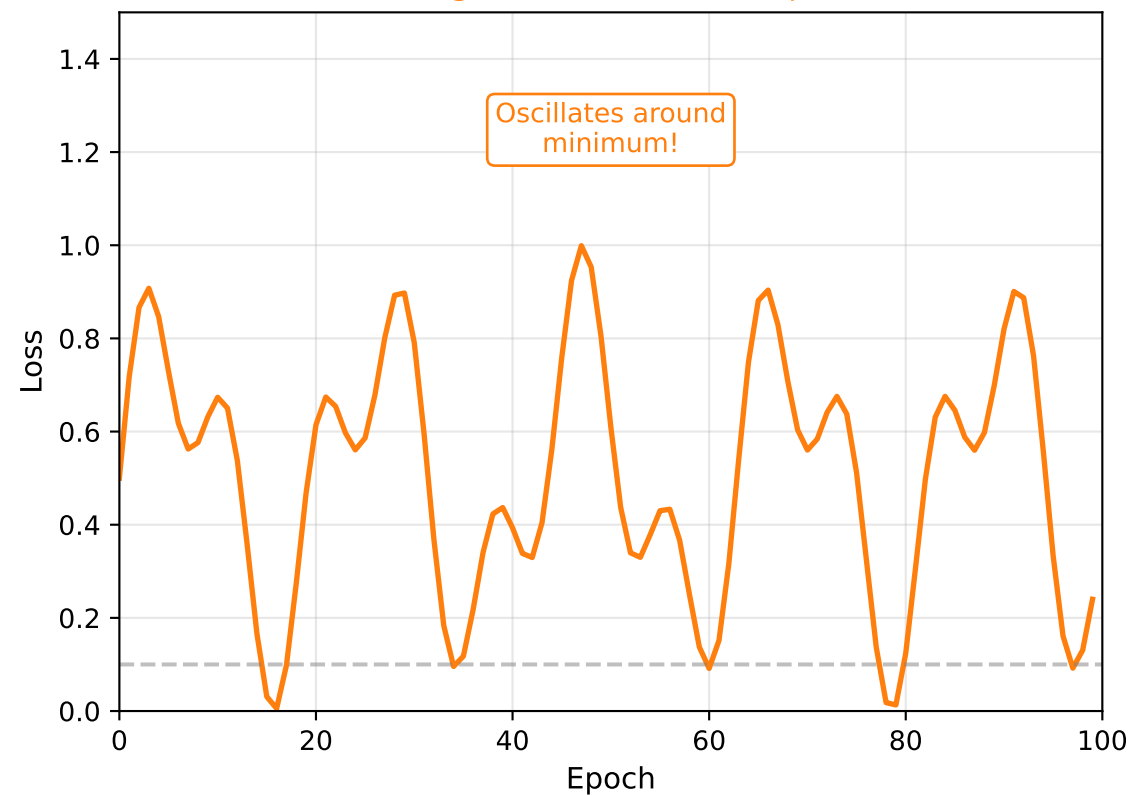
Learning Rate TOO SMALL ($\eta = 0.001$)



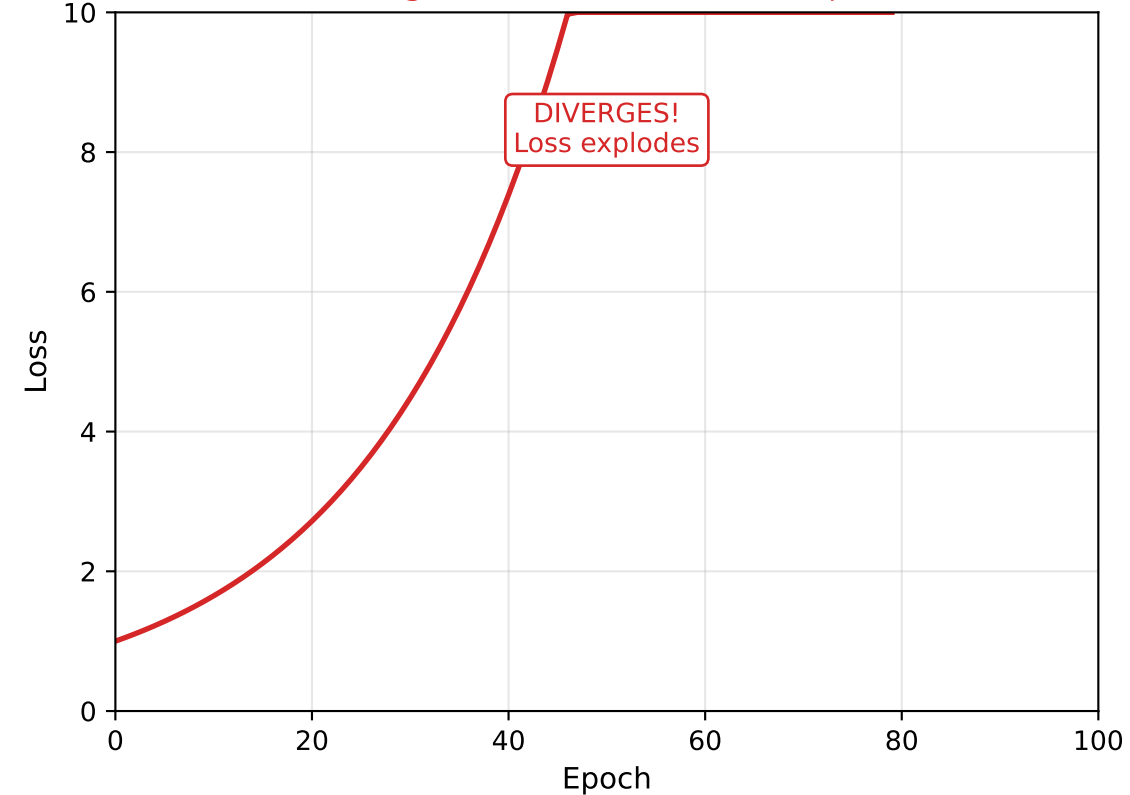
Learning Rate GOOD ($\eta = 0.01$)



Learning Rate TOO LARGE ($\eta = 0.5$)



Learning Rate WAY TOO LARGE ($\eta = 1.0$)



Rule of thumb: Start with 0.001-0.01 and adjust based on training dynamics