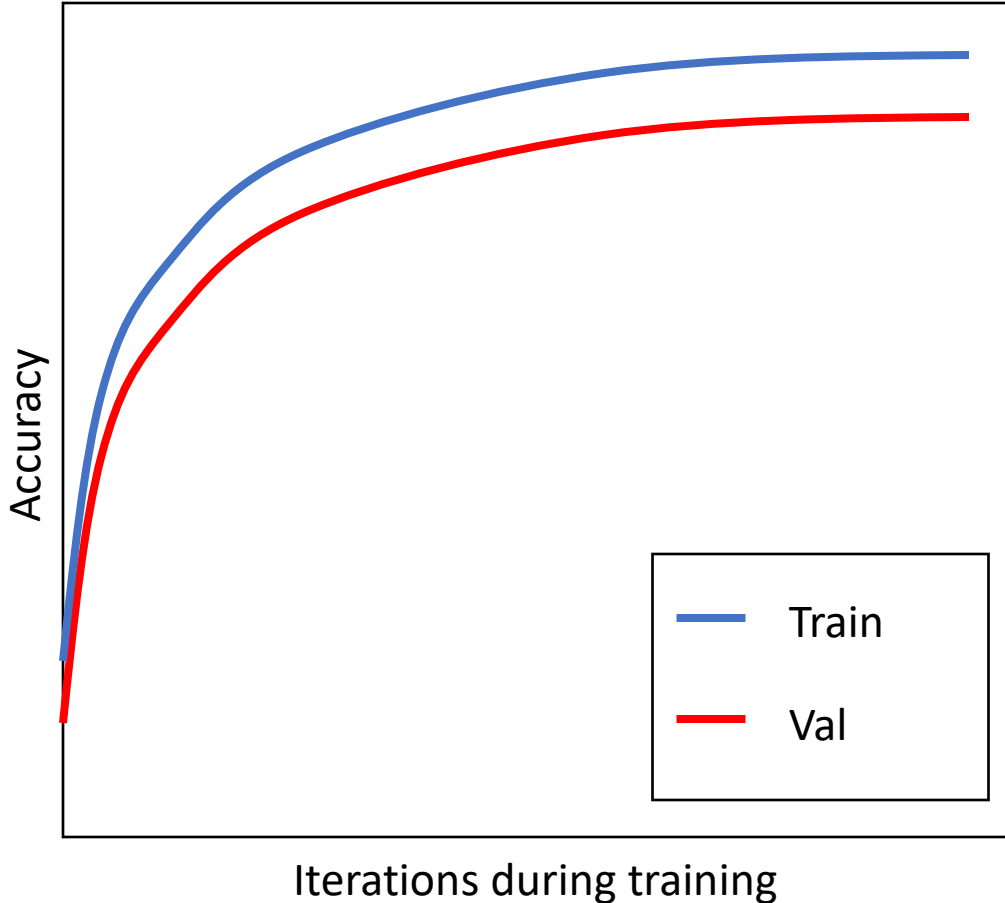
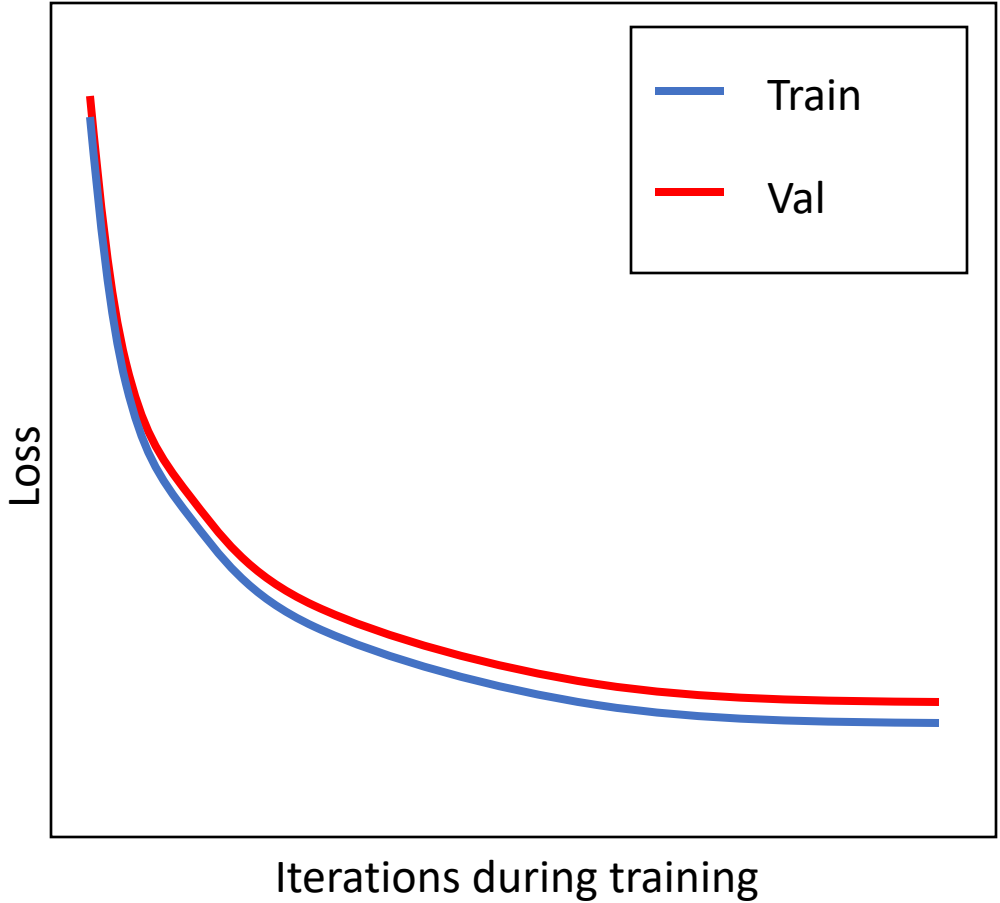


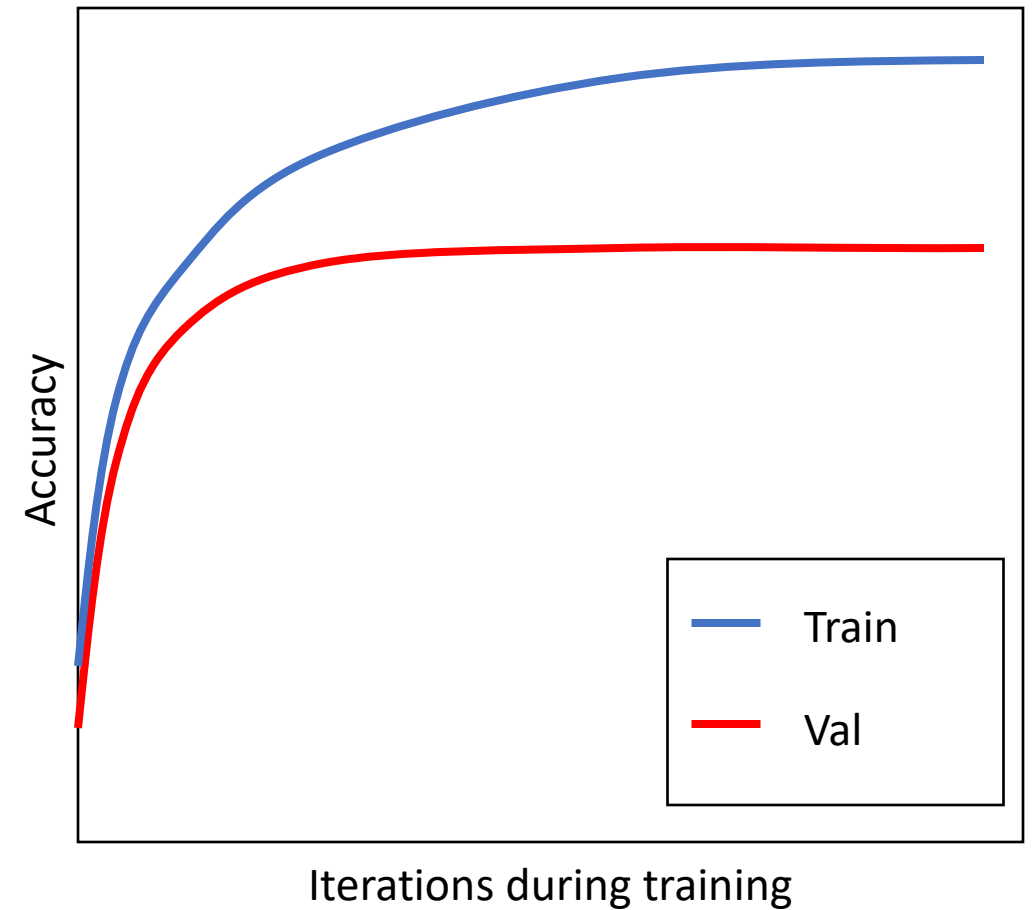
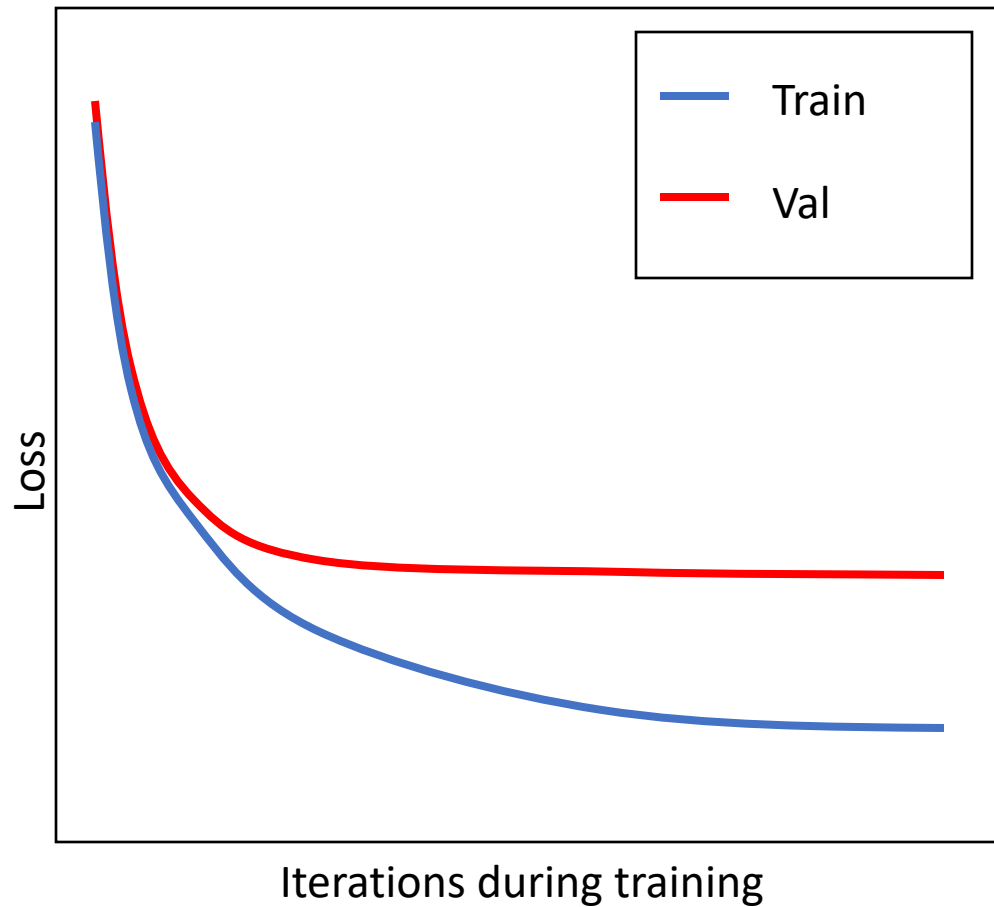
# Debugging ML

# The (ideal) plots



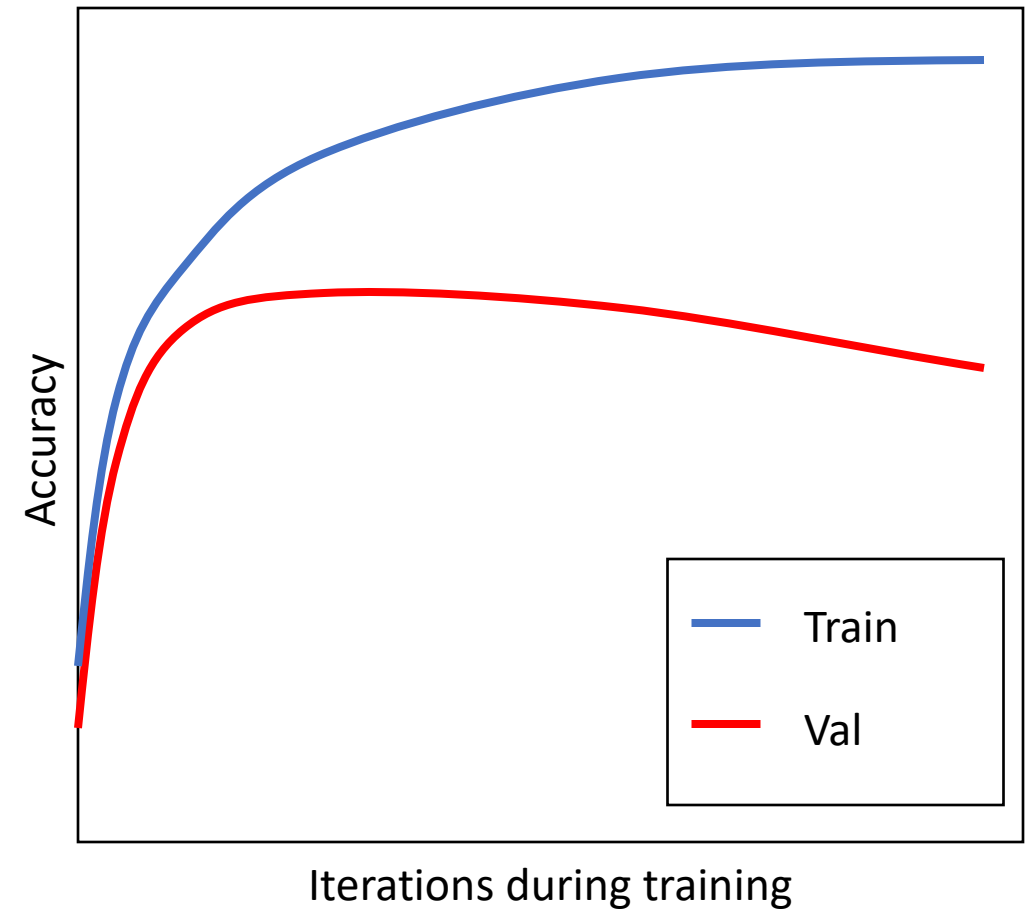
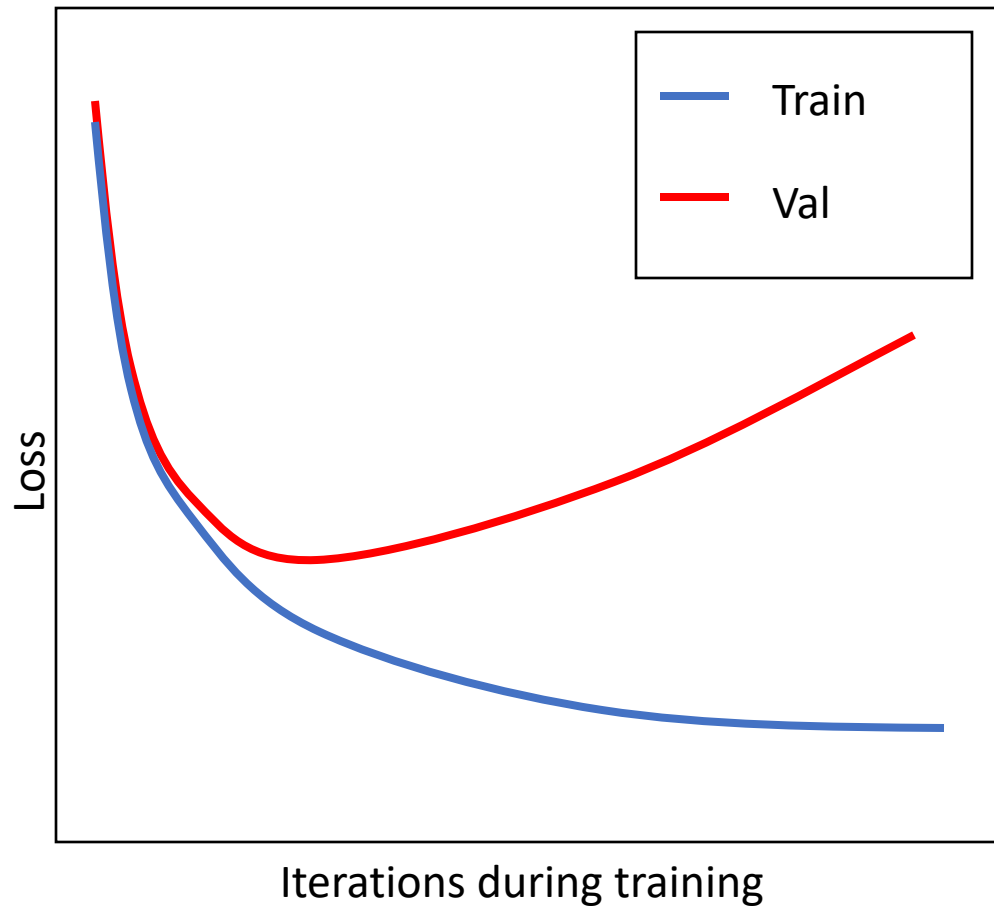
# Overfitting (mild)

- Validation accuracy does not improve as training loss goes down

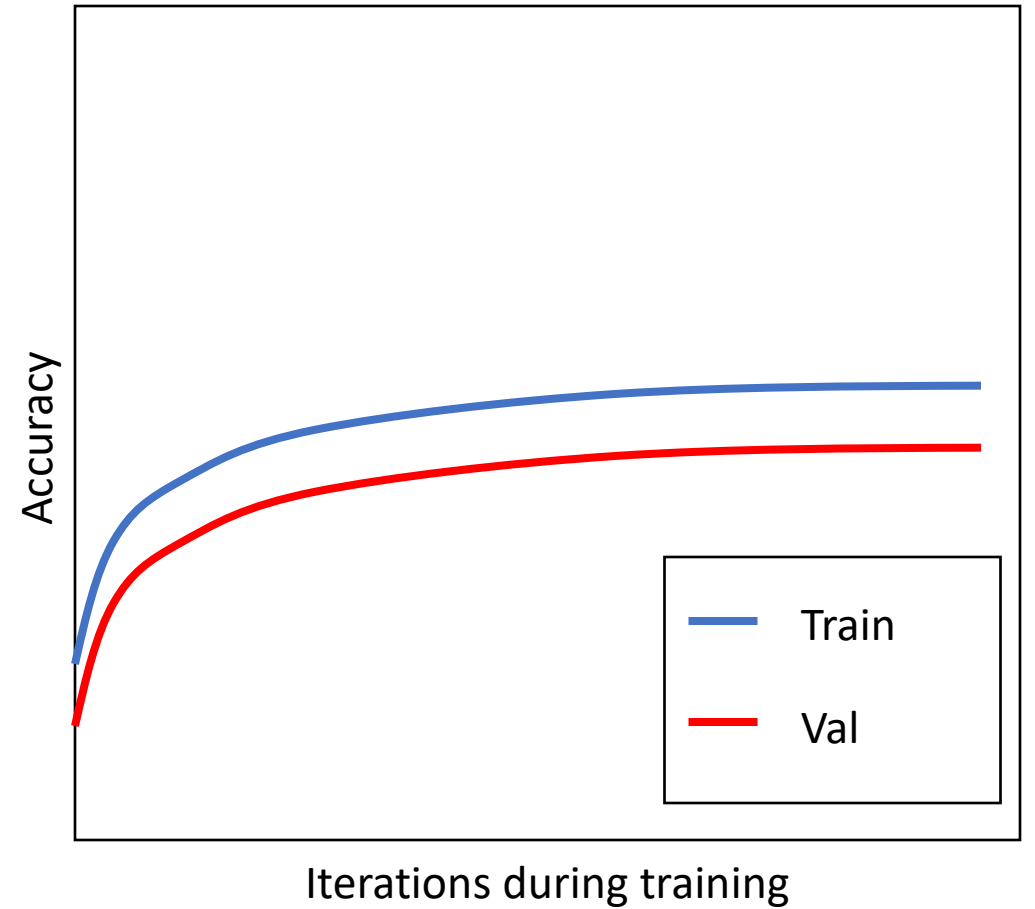
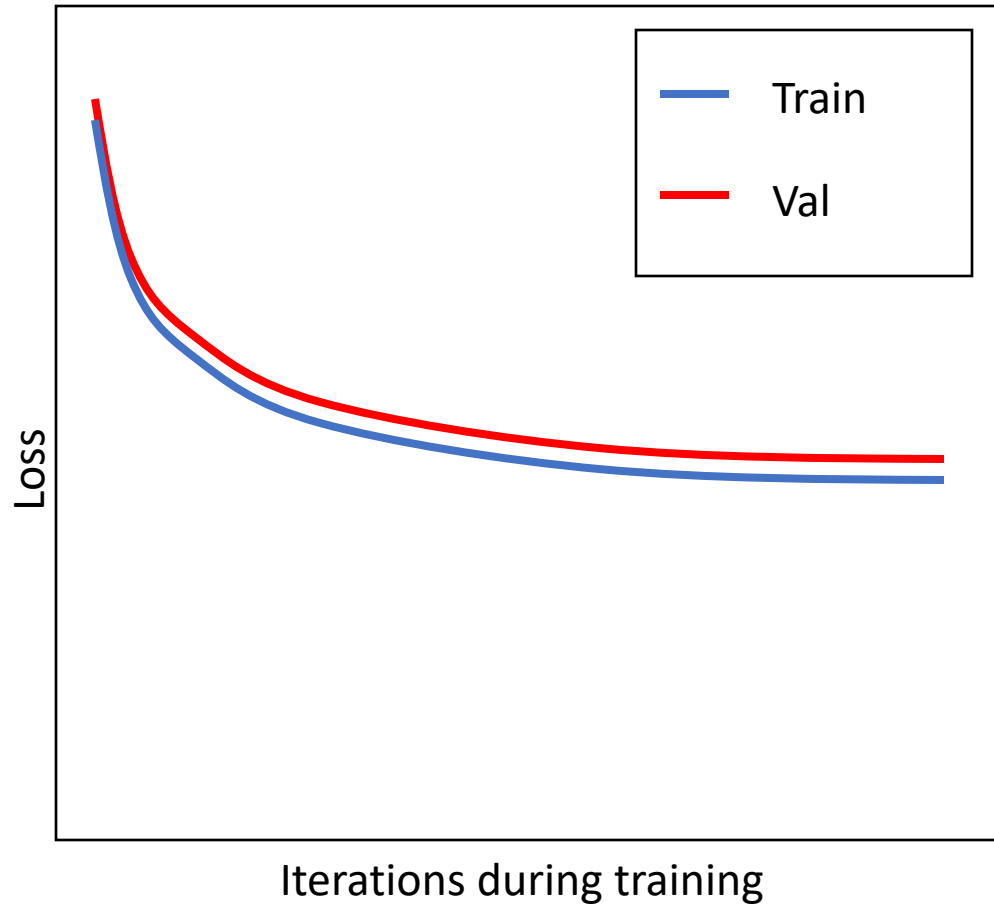


# Overfitting (bad)

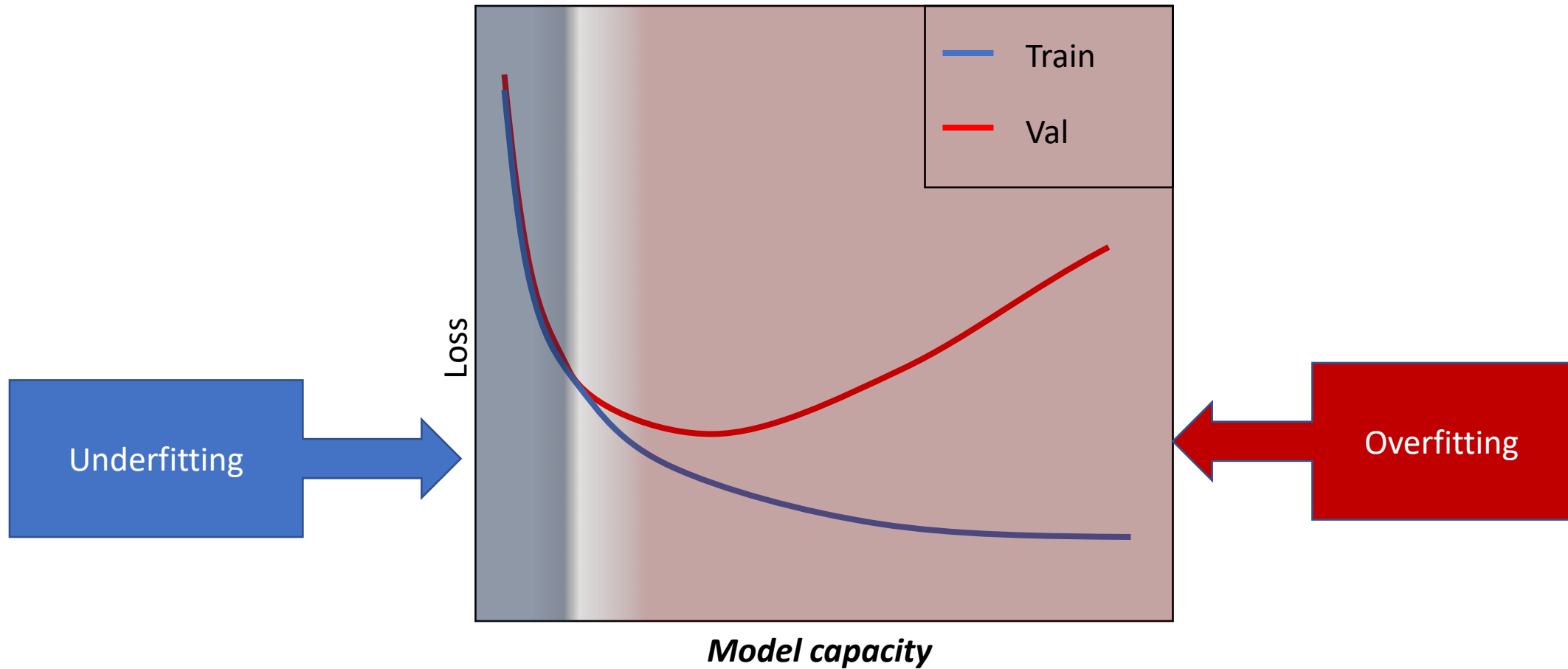
- Validation accuracy *decreases* as training loss goes down



# Underfitting

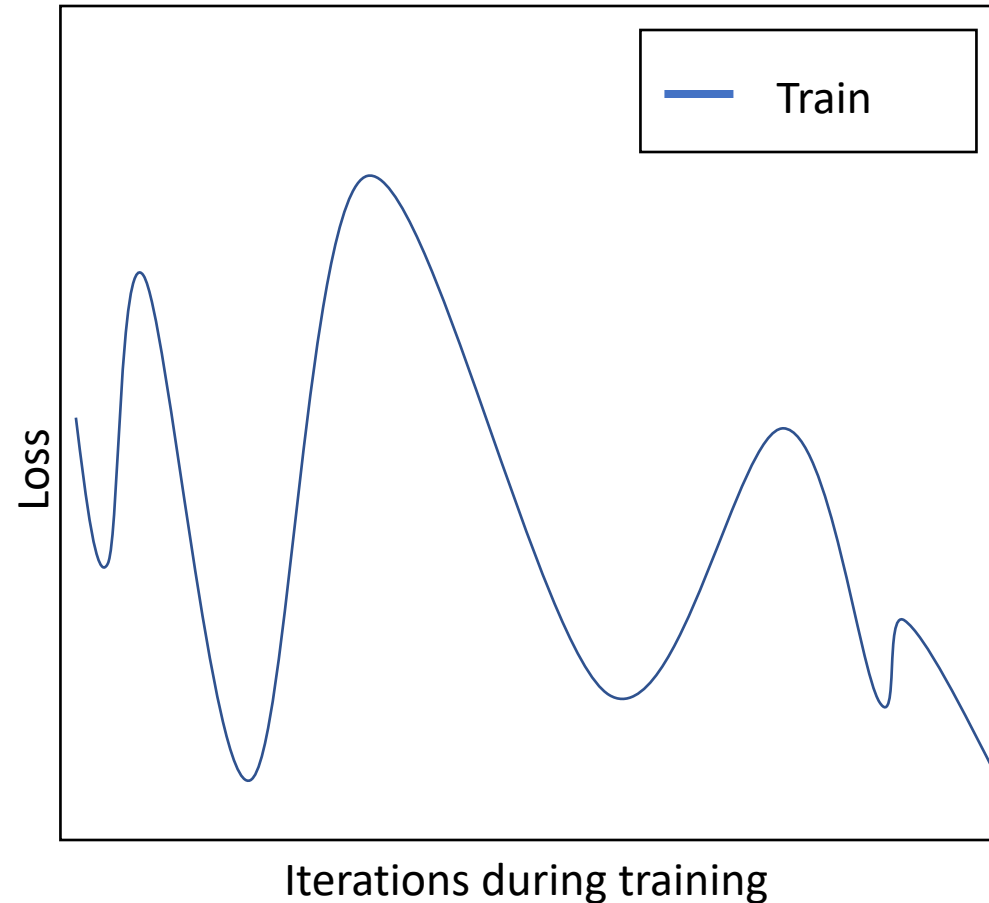


# Underfitting vs overfitting



Optimization issues

# Optimization issues: Large step size



Small step sizes  
cause slow learning,  
Large step sizes  
cause divergence



# SGD with momentum

- Stochastic gradient is stochastic
- Can reduce variance using *momentum*
- *Standard update:*

$$\theta_{t+1} = \theta_t - \lambda \nabla_{\theta} F(\theta_t)$$

- *With momentum:*

$$\begin{aligned}\theta_{t+1} &= \theta_t - \lambda p_{t+1} \\ p_{t+1} &= \gamma p_t + \nabla_{\theta} F(\theta_t)\end{aligned}$$

# Optimization issues: convergence

