

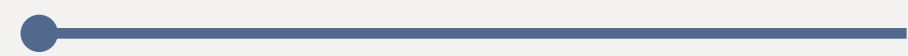


Early Stopping

Stochastic Gradient Descent



Validation





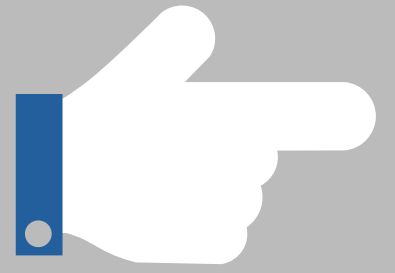
How do we combine validation with
our optimization algorithms?



Early stopping



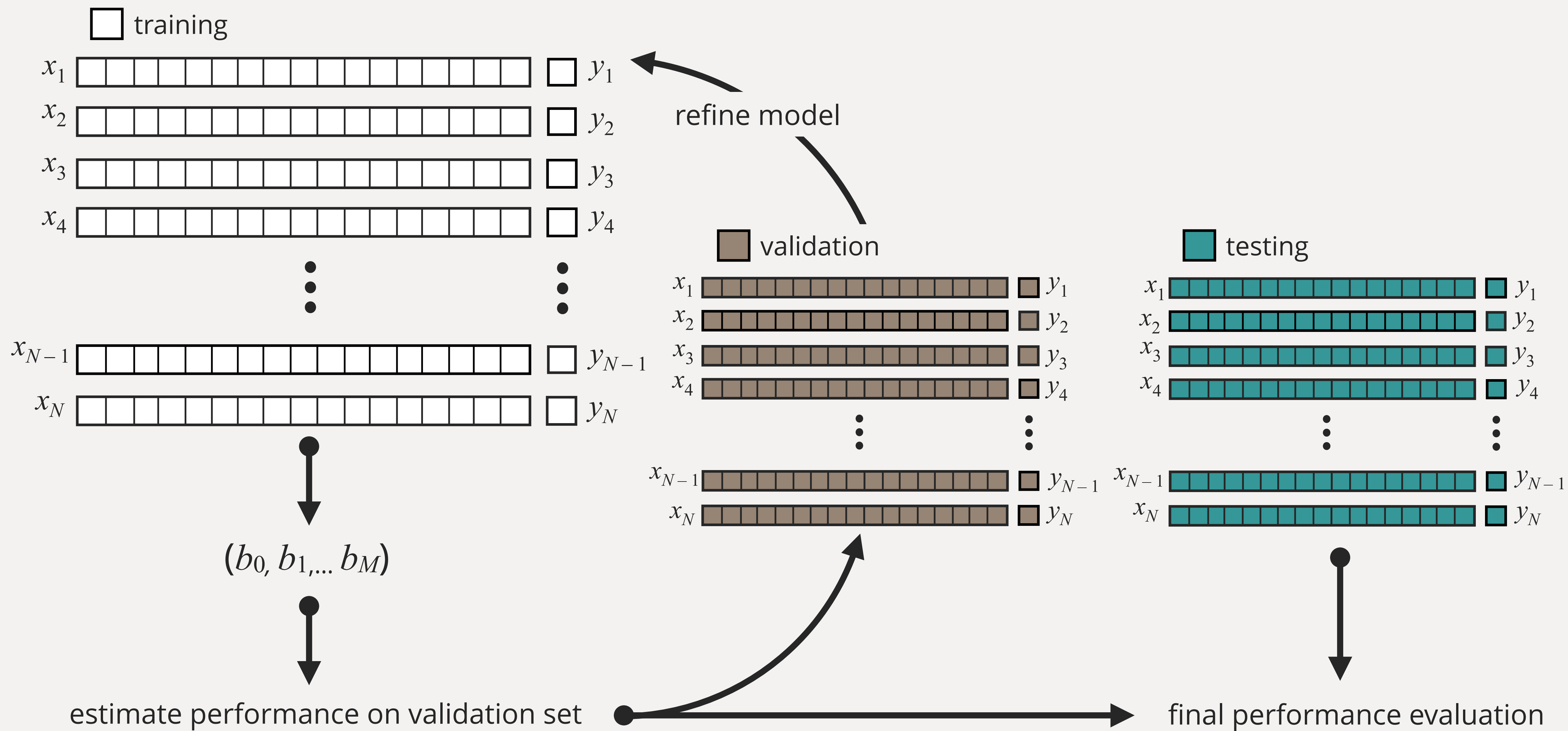
Maximizing generalization of network is mismatched with our optimization goal



Goal of optimization is to do as well as possible on our training set

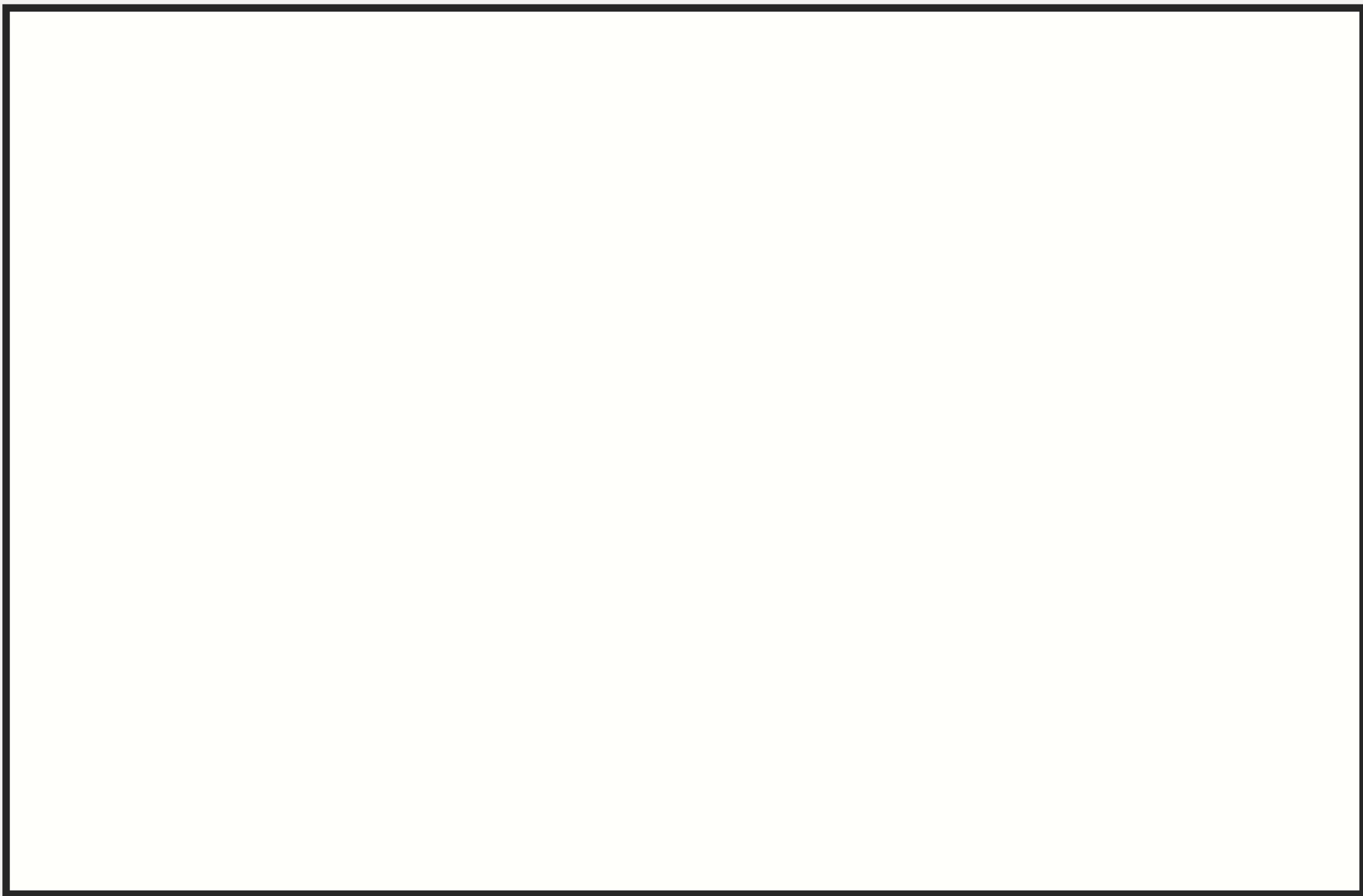


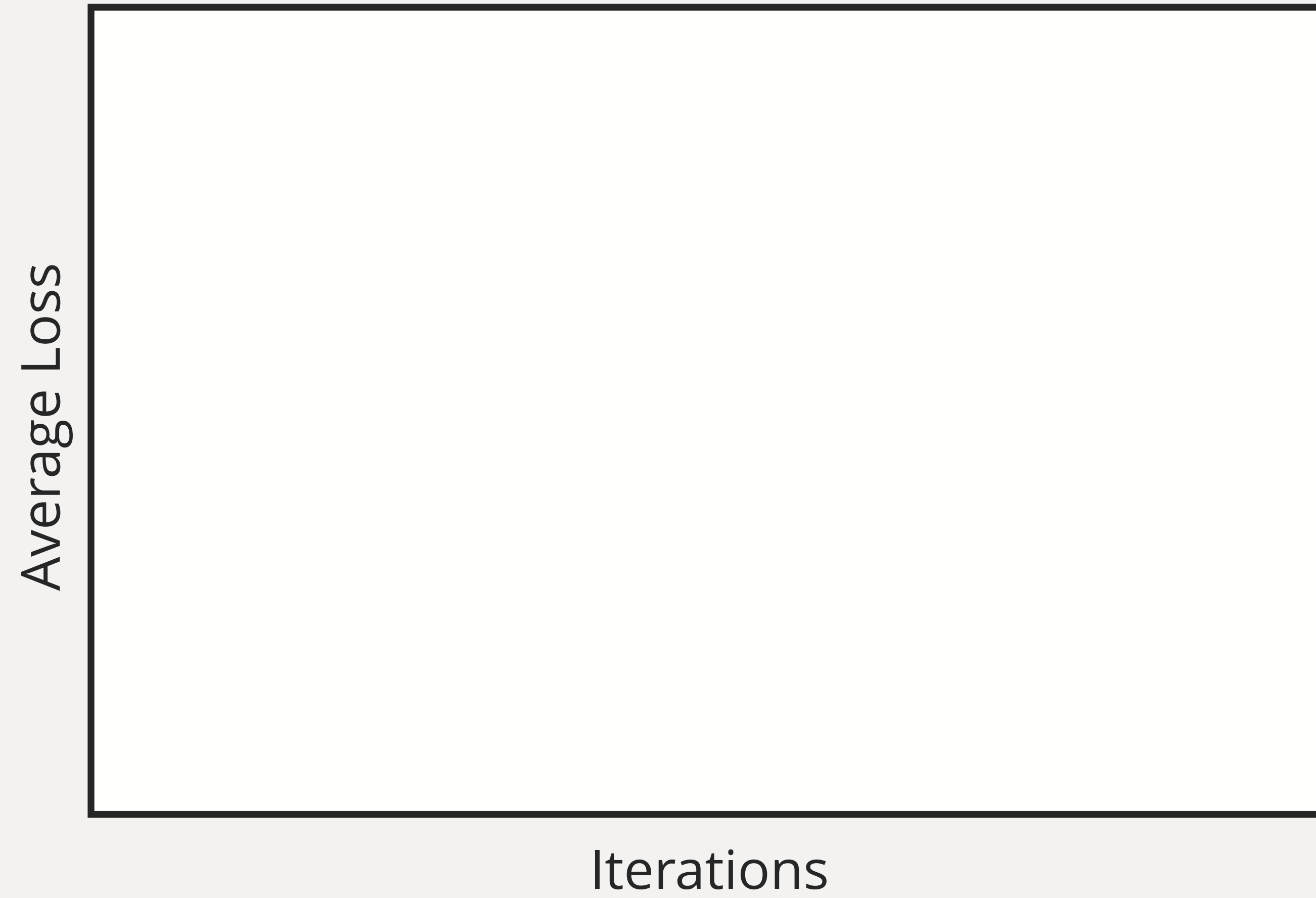
Stochastic Gradient Decent is optimizing a mathematical goal which is mismatched with our true goal



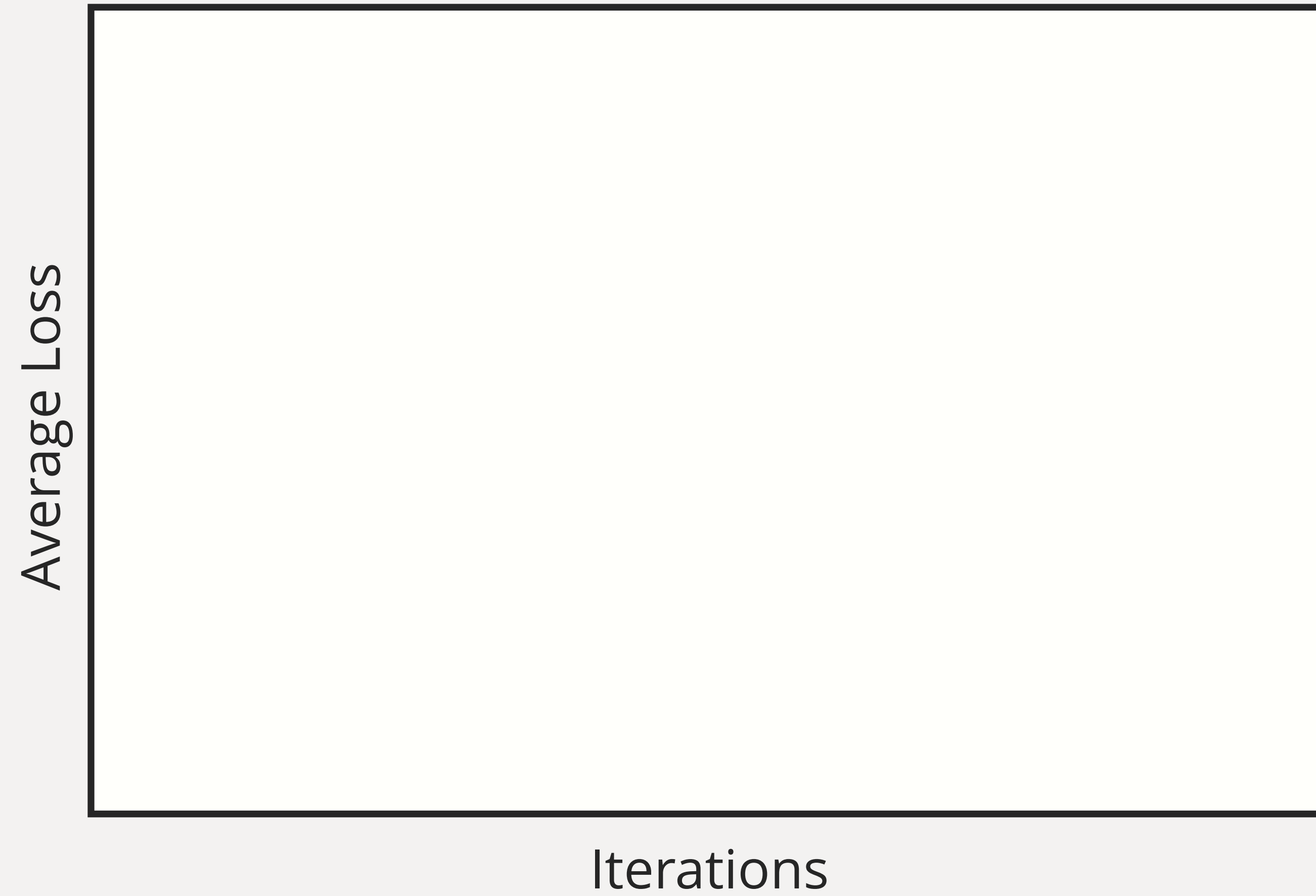
Average Loss

Iterations



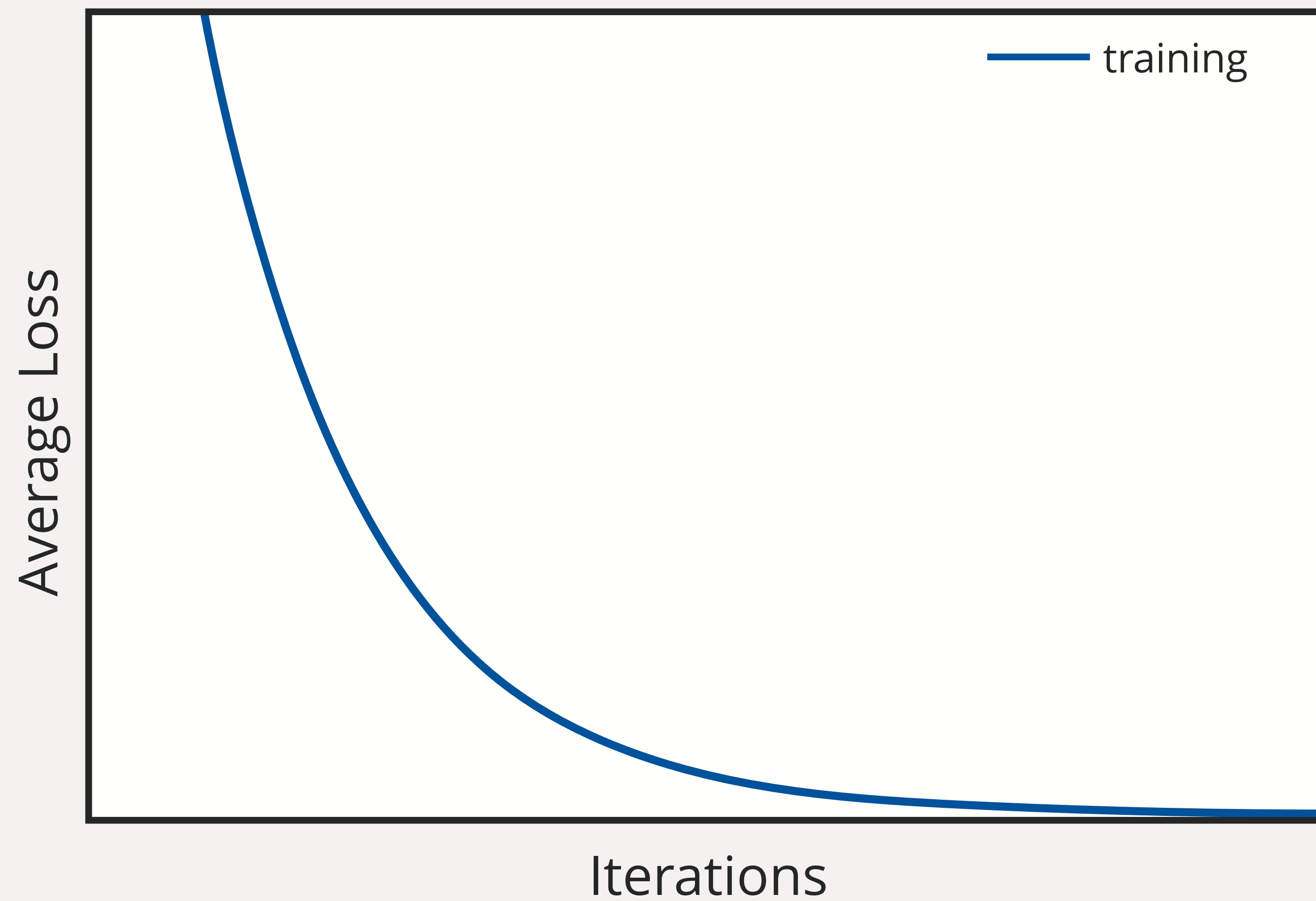


Early Stopping



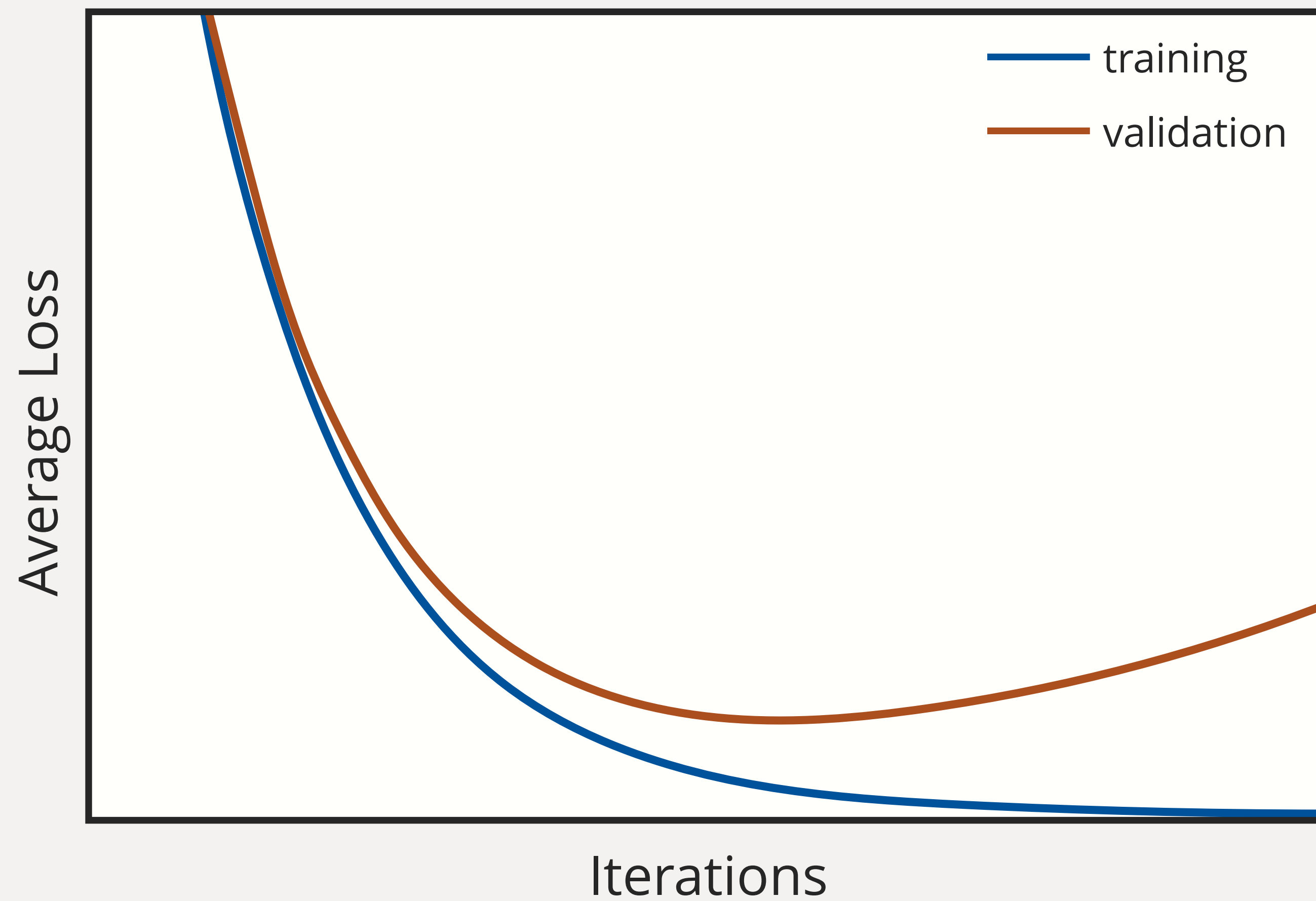
Early Stopping

- Can check validation loss as we go



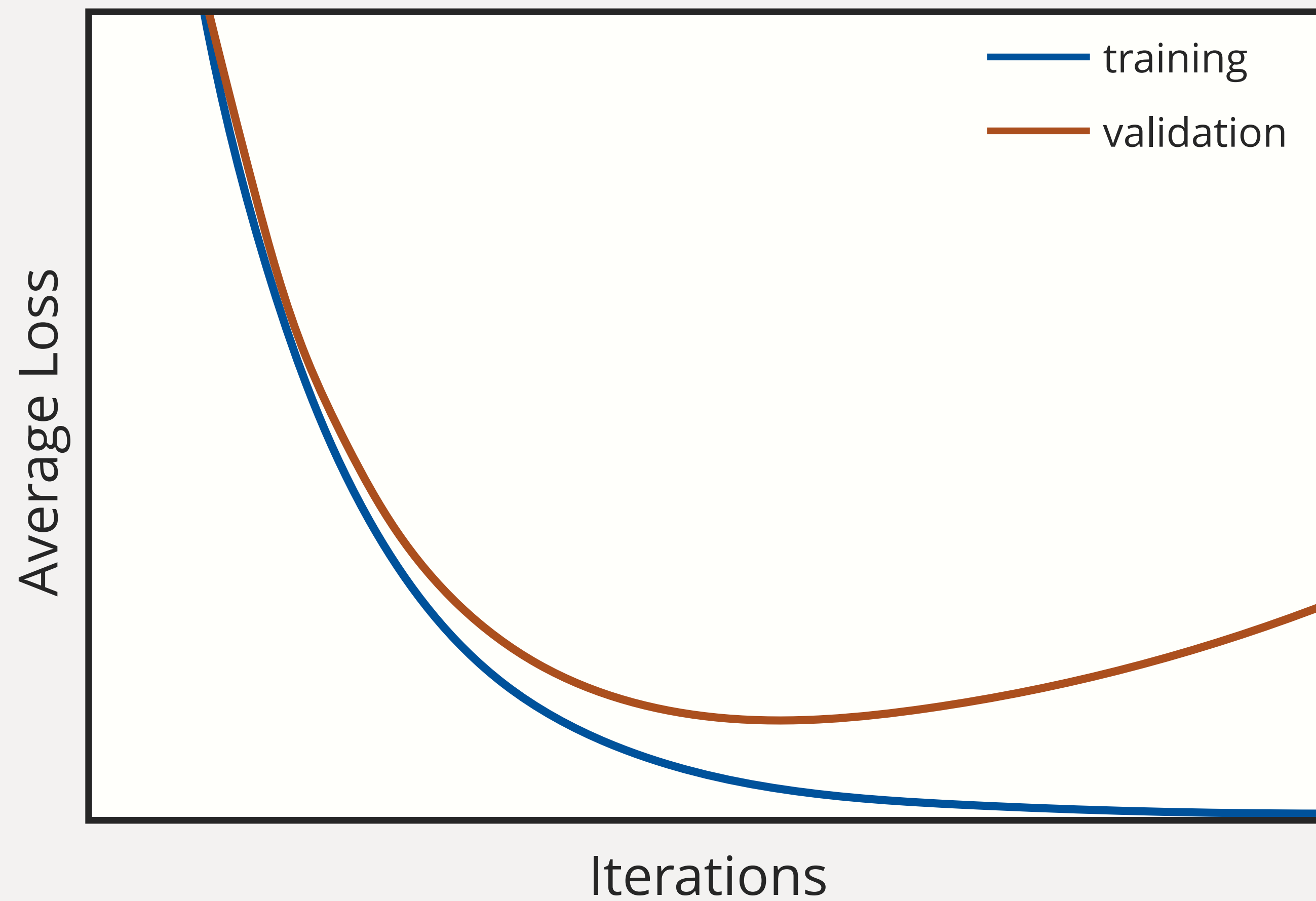
Early Stopping

■ Can check validation loss as we go



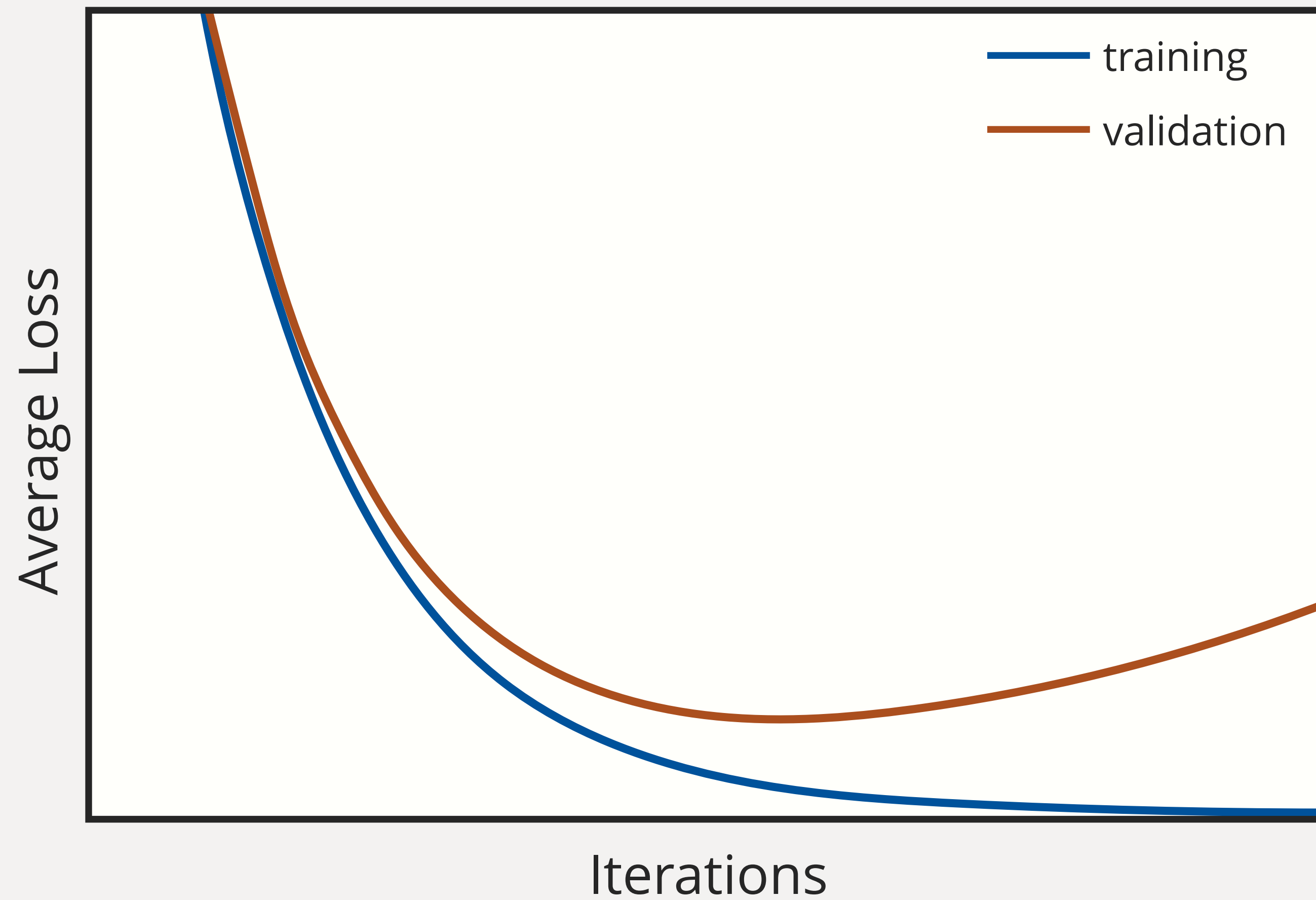
Early Stopping

Can check validation loss as we go



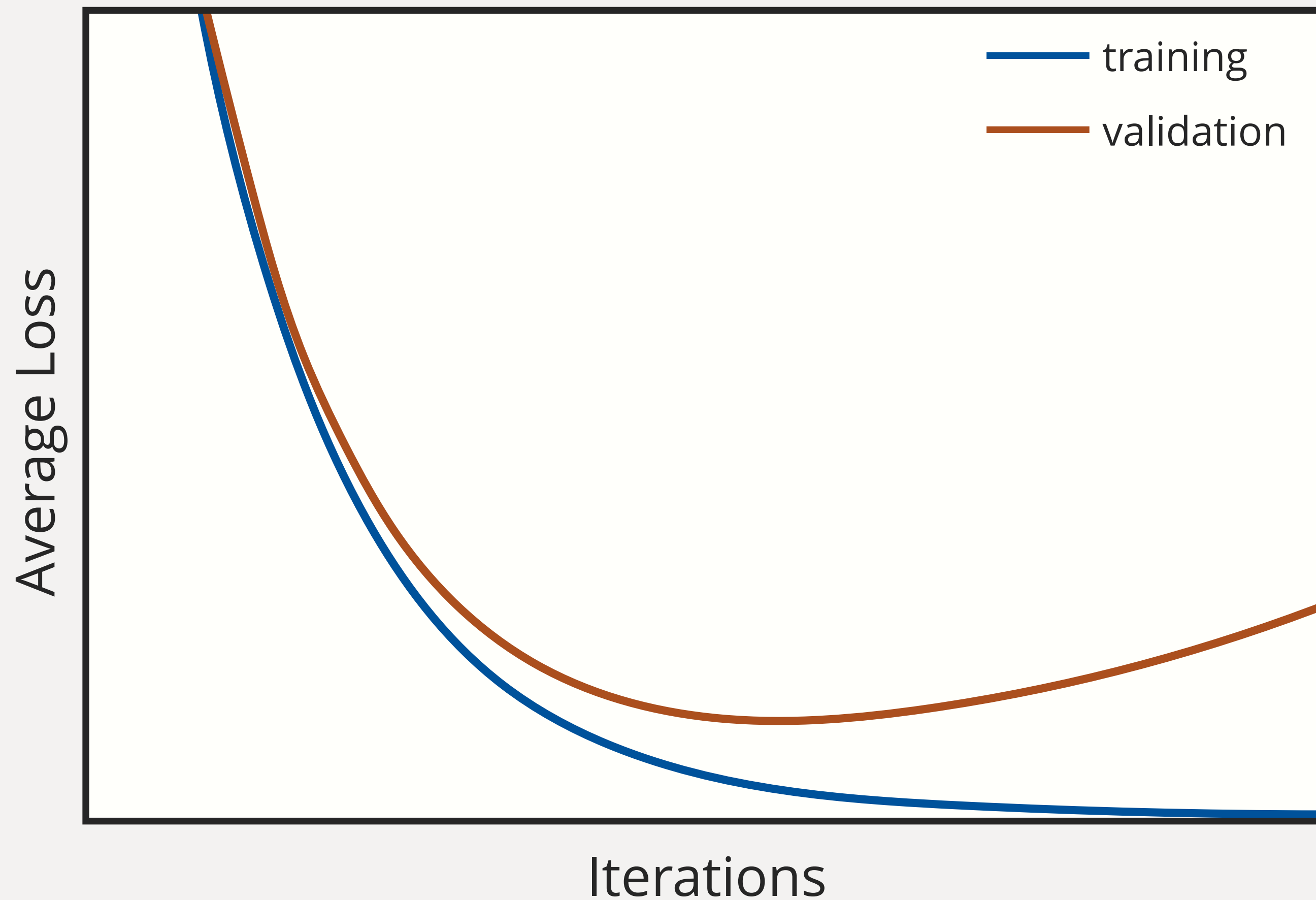
Early Stopping

- Can check validation loss as we go
- Instead of optimizing to convergence, optimize until validation loss stops improving



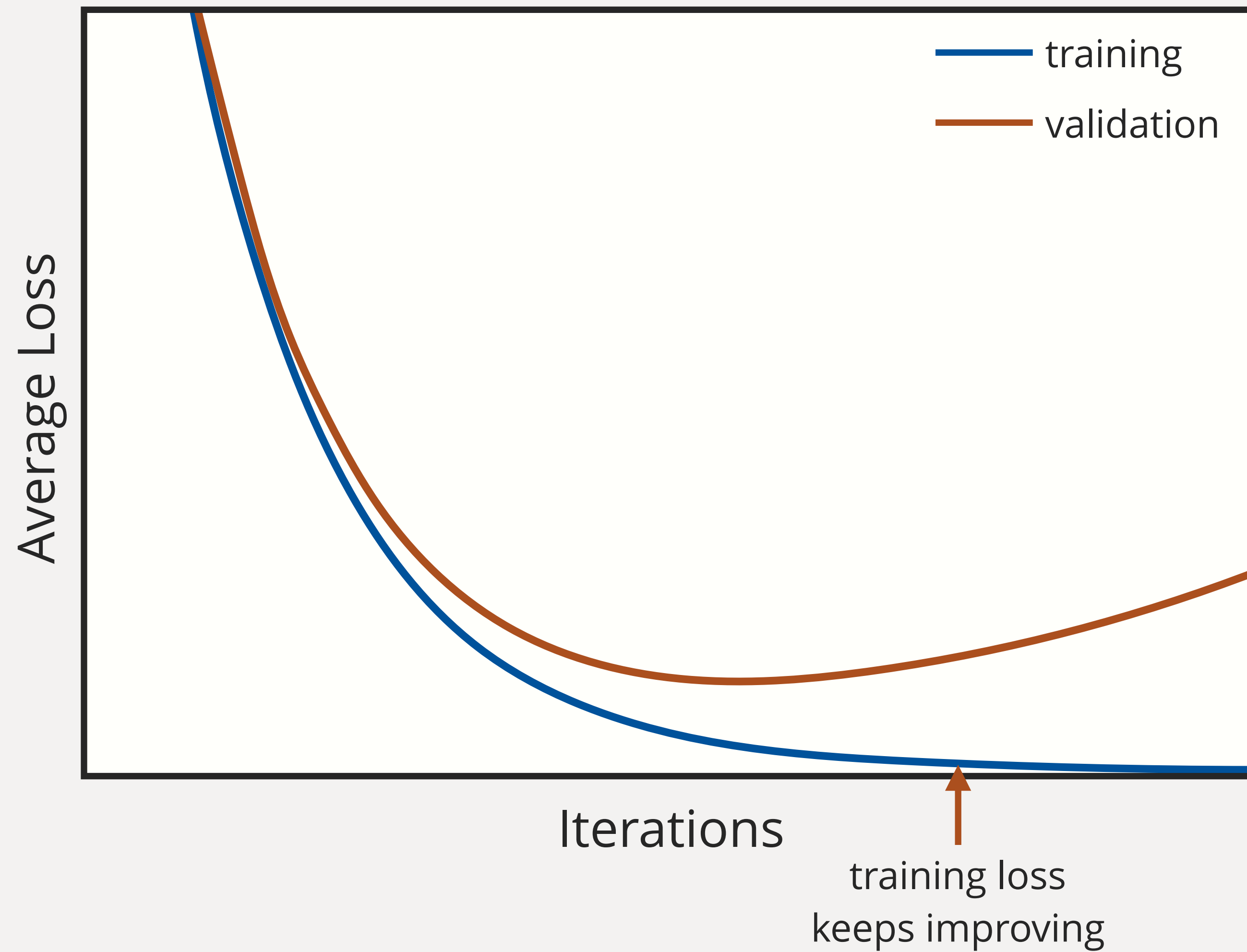
Early Stopping

- Can check validation loss as we go
- Instead of optimizing to convergence, optimize until validation loss stops improving
- Helps save computational cost



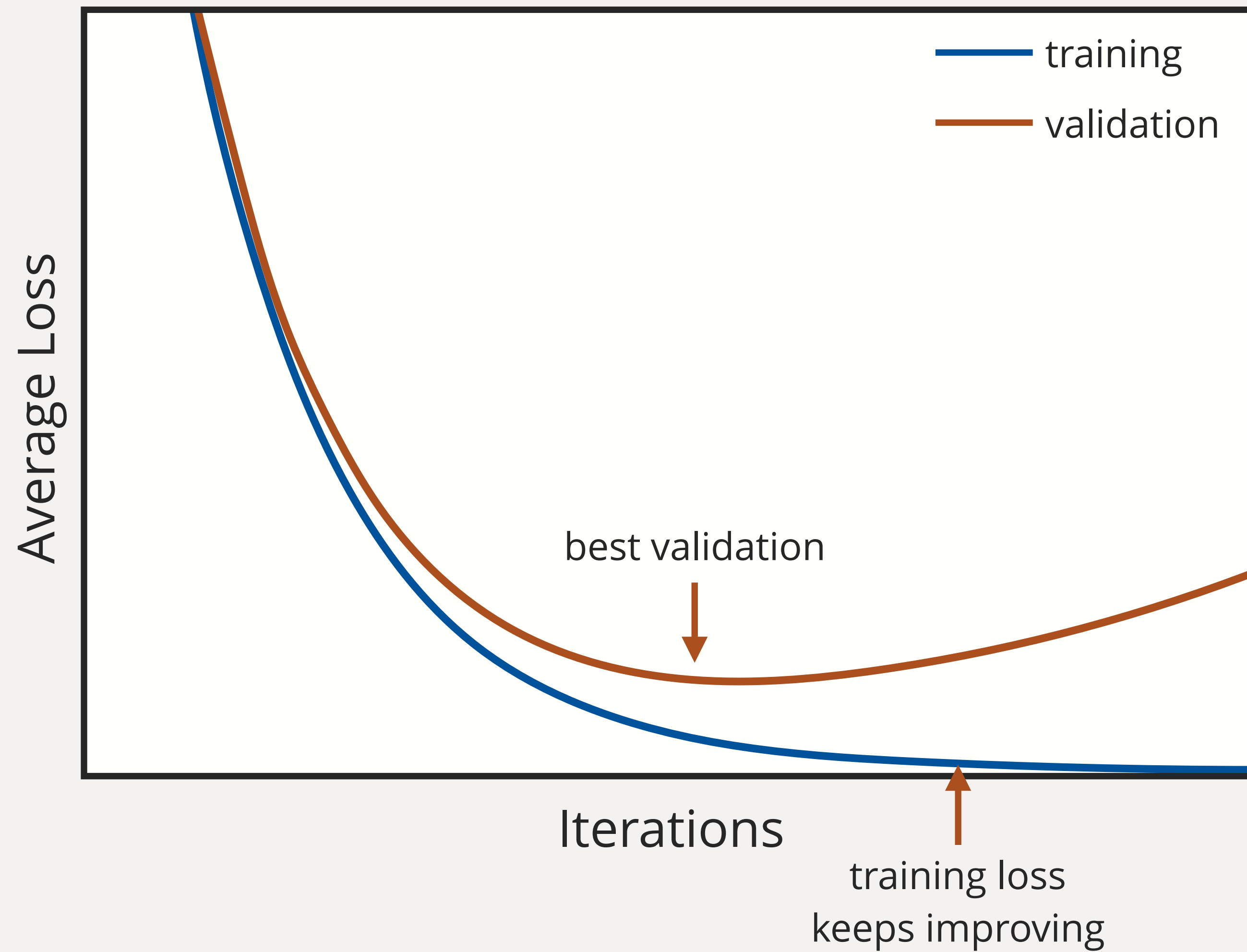
Early Stopping

- Can check validation loss as we go
- Instead of optimizing to convergence, optimize until validation loss stops improving
- Helps save computational cost
- Will perform better in the real world



Early Stopping

- Can check validation loss as we go
- Instead of optimizing to convergence, optimize until validation loss stops improving
- Helps save computational cost
- Will perform better in the real world



Early Stopping

- Can check validation loss as we go
- Instead of optimizing to convergence, optimize until validation loss stops improving
- Helps save computational cost
- Will perform better in the real world

