Diagnosing Model Behavior from Loss Curves

Luca Mondonico

2025-09-09

1 Underfit Learning Curves

Underfitting occurs when models cannot learn the training dataset sufficiently. Two indicators:

- Training loss remains flat regardless of training
- Training loss continues decreasing until training ends (premature halt)

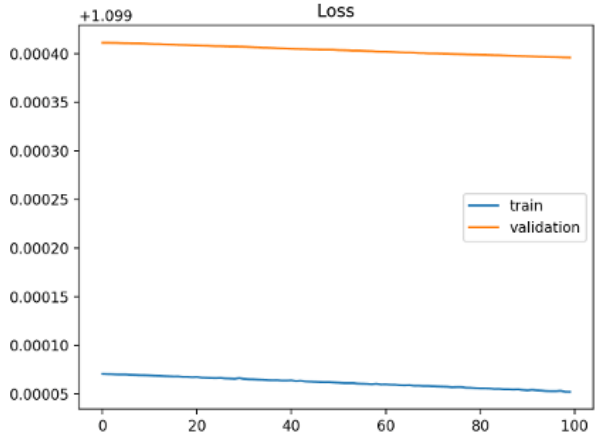

Figure 1: Underfit: Insufficient capacity

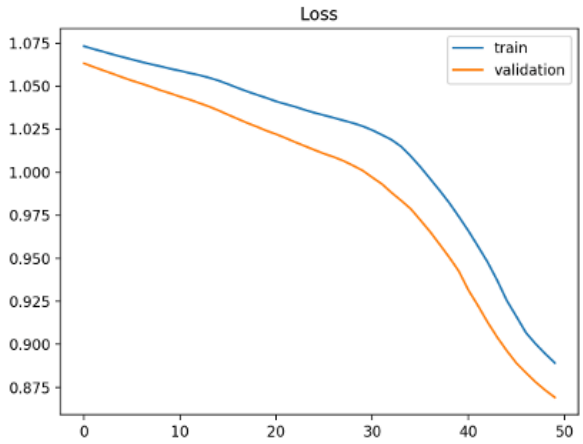

Figure 2: Underfit: Needs more training

2 Overfit Learning Curves

Overfitting occurs when models learn training data too well, including noise, reducing generalization ability. Indicators:

- Training loss continues decreasing with experience
- Validation loss decreases then increases again

The validation loss inflection point indicates optimal stopping time.

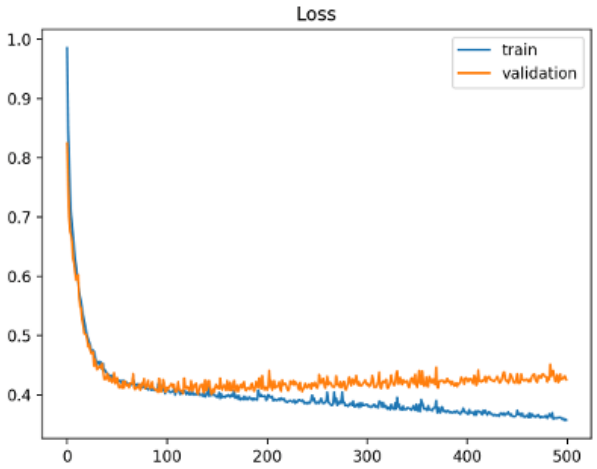

Figure 3: Overfit model showing diverging train/validation curves

3 Good Fit Learning Curves

Good fit exists between underfit and overfit. Characteristics:

- Training loss decreases to stability
- Validation loss decreases to stability with small "generalization gap"

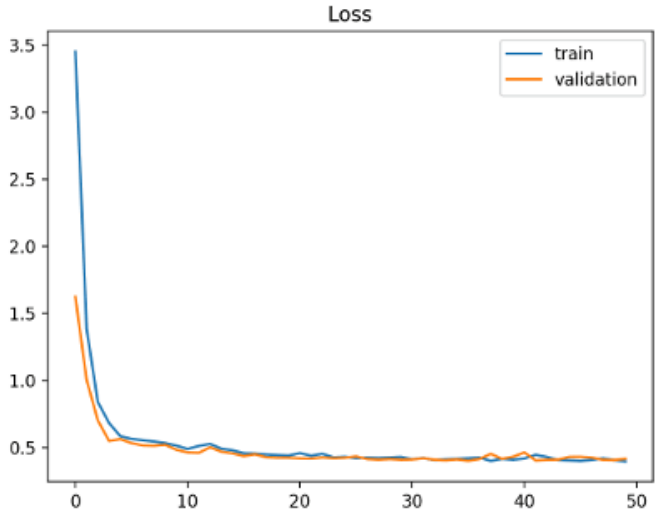

Figure 4: Good fit with stable, converging curves

4 Diagnosing Unrepresentative Datasets

Unrepresentative Training Dataset: Training loss and validation loss both improve but large gap remains between curves, suggesting insufficient training examples.

Unrepresentative Validation Dataset: Two scenarios:

- Validation loss shows noisy movements around training loss (too few validation samples)
- Validation loss lower than training loss (validation dataset easier to predict)

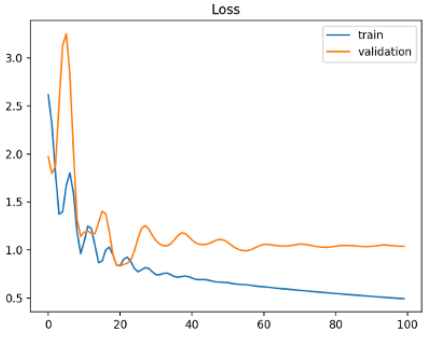

Figure 5: Small training dataset

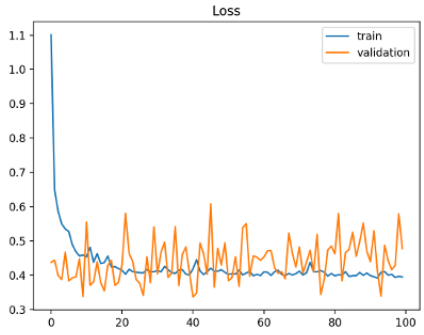

Figure 6: Small validation (noisy)

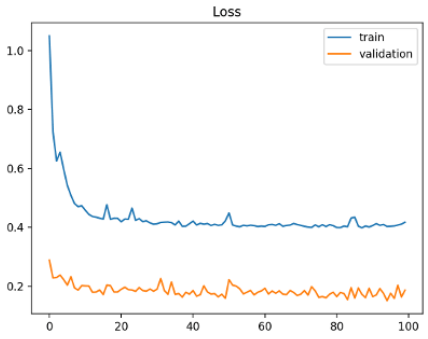

Figure 7: Easy validation dataset

Summary: Learning curves provide crucial insights into model behavior. Underfit models need more capacity or training time. Overfit models require regularization or early stopping. Good fit models balance bias-variance tradeoff. Dataset issues can be identified through gap analysis and curve stability patterns.