

Machine Learning

Over/Under Fit

Curtis Larsen

Utah Tech University—Computing

Spring 2025

Objectives

Objectives:

- ▶ Recognize over fitting and under fitting.
- ▶ Demonstrate symptoms of over fitting and under fitting.
- ▶ Build training pipelines to detect over fitting and under fitting.
- ▶ Regularize models to reduce over fitting.
- ▶ Select more powerful model to reduce under fitting.

Recognize Over/Under Fit

- ▶ Show some data
- ▶ Show an over fit model
- ▶ Show an under fit model

Demonstrate Symptoms

- ▶ Show training error for over/under fit
- ▶ Show testing error for over/under fit
- ▶ Show learning curve?
- ▶ Introduce validation data (dev set?).

Detect Over/Under Fit

- ▶ Use cross validation.
- ▶ Select model with lowest validation error.
- ▶ Allows us to select the best model (model variety, and model training hyperparameters)
- ▶ if training error \ll validation error; then we are over fit
- ▶ if training error and validation error are high; then we are under fit
- ▶ if training error and validation error are low; then we are goldilocks fit

Regularization

- ▶ Simpler model.
- ▶ Smaller parameters in model.
- ▶ Hypothesis space reduction.

Summary

- ▶ Right-size model's complexity to match the data.
- ▶ Use Cross Validation to find best model.
- ▶ Does not violate test data set.