**Introduction to Machine Learning**

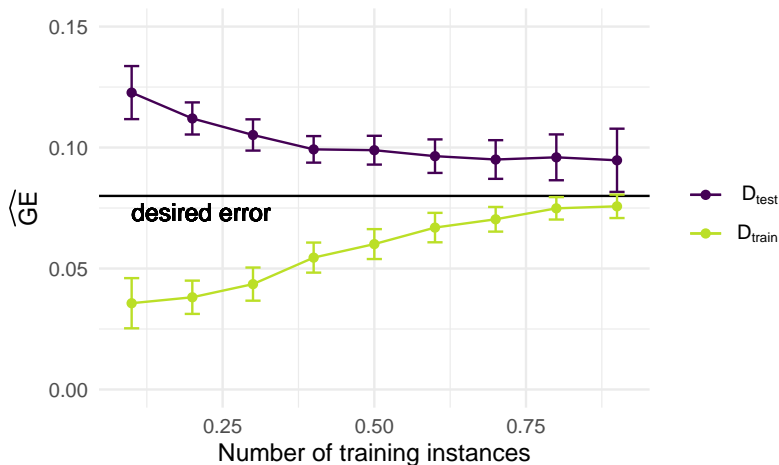**Evaluation: Learning Curves**

compstat-lmu.github.io/lecture_i2ml

LEARNING CURVES

- The *Learning Curve* compares the performance of a model on training and test data over a varying number of training instances.
  $\rightarrow$ How fast can learner learn the given relationship in the data?
- Learning usually fast in the beginning
- Visualizes when a learner has learned as much as it can:
  - when performance on training and test set reach a plateau.
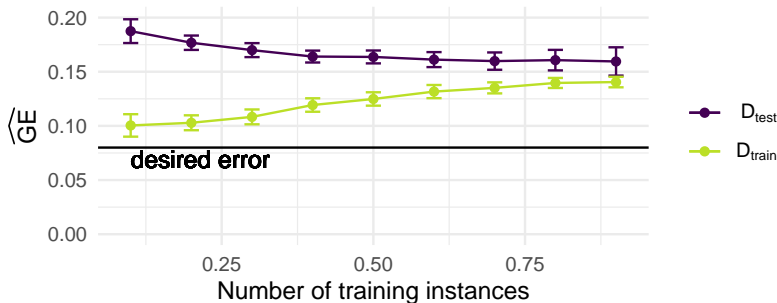 - when gap between training and test error remains the same.
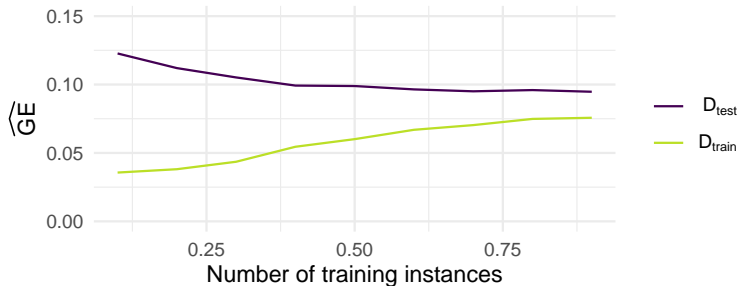
LEARNING CURVES

An ideal learning curve looks like:

LEARNING CURVES

In general, there are two reasons for a bad looking learning curve:

1 High bias in model / underfitting

- training and test errors converge at a high value.
- model can't learn underlying relationship and has high systematic errors, no matter how big the training set.
- poor fit, which also translates into high test error.

LEARNING CURVES

② High variance in model / Overfitting

- large gap between training and test errors.
- model requires more training data to improve.
- model has a poor fit and does not generalize well.