

# Introduction to Machine Learning

## Evaluation: Introduction and Remarks

### Learning goals

- Understand the goal of performance estimation
- Understand the difference between outer and inner loss
- Know the definition of generalization error

# PERFORMANCE EVALUATION

How well does my model perform...



... on data from the same data-generating process?

In practice:

... on current data (training data)?

... on new data (test data)?

... based on a certain measure/metric?

...

# PERFORMANCE EVALUATION

ML performance evaluation provides clear and simple protocols for reliable model validation.

- Often simpler than classical statistical model diagnosis
- Relies only on few assumptions
- Still hard enough and offers **lots** of options to cheat / make mistakes

# PERFORMANCE MEASURES

We measure performance using a statistical estimator for the **generalization error** (GE).

GE = expected loss of a fixed model

$\hat{GE}$  = average loss

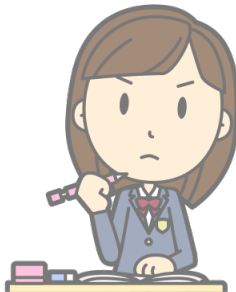
Example: Mean squared error (L2 loss)

$$\hat{GE} = MSE = \frac{1}{n} \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2$$

# MEASURES: INNER VS. OUTER LOSS

Inner loss = loss used in learning

Outer loss = loss used in evaluation  
= evaluation measure



# MEASURES: INNER VS. OUTER LOSS

Optimally: inner loss = outer loss

Not always possible:

some losses are hard to optimize / no loss is specified directly

Example:

Logistic Regression → minimize binomial loss

kNN → no explicit loss minimization

- When evaluating the models we might be interested in (cost-weighted) classification error
- Or some of the more advanced measures from ROC analysis like AUC