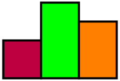
Introduction to Machine Learning

Evaluation: Introduction and Remarks

compstat-lmu.github.io/lecture_i2ml

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
- Rely 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

$\hat{\text{GE}}$ $=$ average loss

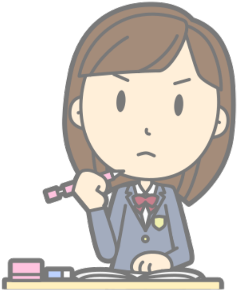
Example: Mean squared error (L2 loss)

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

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	$\rightarrow$ minimize binomial loss
kNN	$\rightarrow$ 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