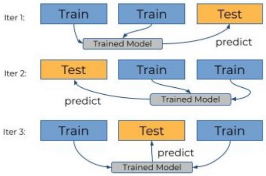
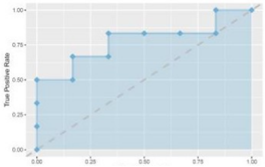
Introduction to Machine Learning

Evaluation: Introduction and Remarks





Learning goals

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

REGRESSION: DEFINING A CUSTOM LOSS

Assume a use case, where the target variable can have a wide range of values across different orders of magnitude. A possible solution would be to use a loss functions that allows for better model evaluation and comparison. The **Mean Squared Logarithmic Absolute Error** is not strongly influenced by large values due to the logarithm.

$$\frac{1}{n}\sum_{i=1}^{n}(\log(|y^{(i)} - \hat{y}^{(i)}| + 1))^2$$

LIST OF CLASSIFICATION PERFORMANCE MEASURES

Classification	Explanation
Accuracy	Fraction of correct classifications
Balanced Accuracy	Fraction of correct classifications in each class
Recall	Fraction of positives a classifier captures
Precision	Fraction of positives in instances predicted as positive
F1-Score	Tradeoff between precision and recall
AUC	Measures calibration of predicted probabilities
Brier Score	Squared difference between predicted probability and true label
LogLoss	Emphasizes errors for predicted probabilities close to 0 and 1

BIAS-VARIANCE TRADEOFF

We can decompose the generalization error for L_2-loss as follows:

$$
\begin{aligned}
GE_n\left(\hat{f}_{\mathcal{D}}\right) &= \mathbb{E}(L(y, \hat{f}_{\mathcal{D}}(x))|\mathcal{D}) \\
&= \mathbb{E}((y - \hat{f}(x))^2) \\
&\overset{Var(y)=\mathbb{E}(y^2)-\mathbb{E}(y)^2}{=} Var(y) + \mathbb{E}(y)^2 + Var(\hat{f}(x)) + \\
&\quad \mathbb{E}(\hat{f}(x))^2 - 2\mathbb{E}(y)\mathbb{E}(\hat{f}(x)) \\
&\overset{\mathbb{E}(y)=f(x)}{=} Var(y) + Var(\hat{f}(x)) + (f(x) - \mathbb{E}(\hat{f}(x)))^2 \\
&= \sigma^2 + Var(\hat{f}(x)) + Bias(\hat{f}(x))^2
\end{aligned}
$$

where

- σ^2: intrinsic variability of the data, cannot be avoided
- $Var(\hat{f}(x))$: variance of the model, the learners's tendency to learn random things irrespective of the signal (*overfitting*)
- $Bias(\hat{f}(x))^2$: systematic bias of the model (*underfitting*)

BIAS-VARIANCE TRADEOFF

- We can reduce the model's variance on the cost of its bias and vice versa by controlling the model complexity.
- We search for the perfect Bias-Variance-Tradeoff that minimizes our expected prediction error.