Introduction to Machine Learning

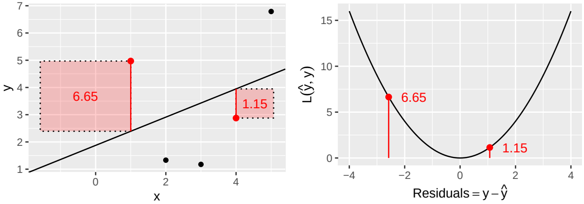
Evaluation: Measures for Regression

compstat-lmu.github.io/lecture_i2ml

MEAN SQUARED ERROR

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)})^2 \in [0;\infty] \qquad \rightarrow \text{L2 loss.}$$

Single observations with a large prediction error heavily influence the **MSE**, as they enter quadratically.

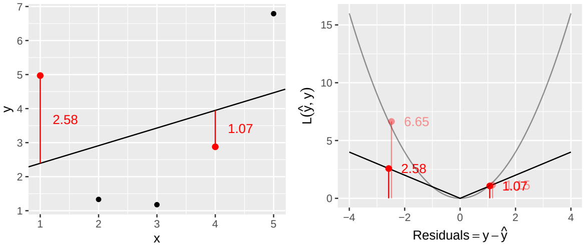


Similar measures: sum of squared errors (SSE), root mean squared error (RMSE, brings measurement back to the original scale of the outcome).

MEAN ABSOLUTE ERROR

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y^{(i)} - \hat{y}^{(i)}| \in [0; \infty] \qquad \to \text{L1 loss.}$$

Less influenced by large errors and maybe more intuitive than the MSE.



Similar measures: median absolute error (for even more robustness).

R^2

Well known measure from statistics.

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)})^2}{\sum\limits_{i=1}^{n}(y^{(i)} - \bar{y})^2} = 1 - \frac{SSE_{LinMod}}{SSE_{Intercept}}$$

- Usually introduced as *fraction of variance explained* by the model
- Simpler: Compares SSE of constant model (baseline) with complex model (LM)
- $R^2 = 1$: all residuals are 0, we predict perfectly,
  $R^2 = 0$: we predict as badly as the constant model
- If measured on the training data, $R^2 \in [0; 1]$ (LM must be at least as good as the constant)
- On other data R^2 can even be negative, as there is no guarantee that the LM generalizes better than a constant (overfitting)

GENERALIZED R^2 FOR ML

A simple generalization of R^2 for ML seems to be:

$$1 - \frac{Loss_{ComplexModel}}{Loss_{SimplerModel}}$$

- Works for arbitrary measures (not only SSE), for arbitrary models, on any data set of interest
- E.g. model vs constant, LM vs. non-linear model, tree vs. forest, model without some features vs. model with them included
- Fairly unknown; our terminology (generalized R^2) is non-standard