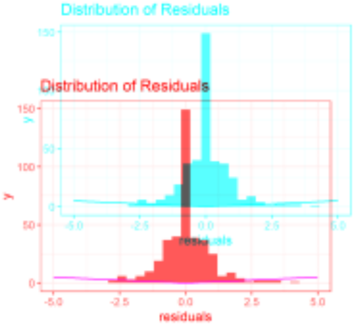
Introduction to Machine Learning

Advanced Risk Minimization

Maximum Likelihood Estimation vs. Empirical Risk Minimization



Learning goals

- Correspondence between Laplace errors and L1 loss
- Correspondence between Bernoulli targets and the Bernoulli / log loss

Learning goals
- Correspondence between Laplace errors and L1 loss
- Correspondence between Bernoulli targets and the Bernoulli / log loss

LAPLACE ERRORS - L1-LOSS / 2

The likelihood is then

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \prod_{i=1}^{n} p\left(y^{(i)} \,\middle|\, f\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right), \sigma\right) \\ &\propto \exp\left(-\frac{1}{\sigma}\sum_{i=1}^{n}\left|y^{(i)} - f\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)\right|\right).\end{aligned}$$

The negative log-likelihood is

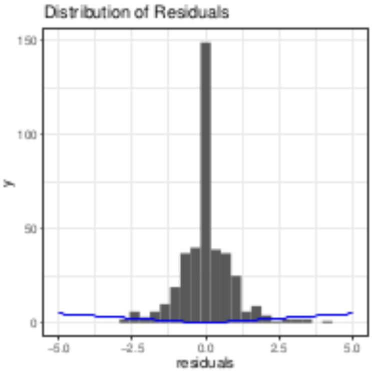
$$-\ell(\boldsymbol{\theta}) \propto \sum_{i=1}^{n}\left|y^{(i)} - f\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)\right|.$$

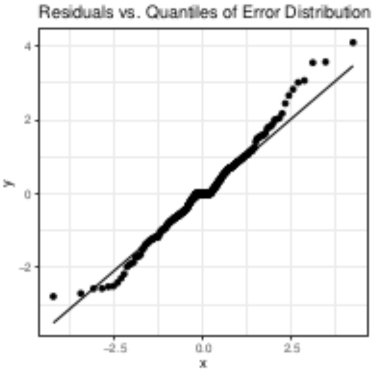
MLE for Laplacian errors = ERM with L1-loss.

- Some losses correspond to more complex or less known error densities, like the Huber loss ▸ Meyer 2021
- Huber density is (unsurprisingly) a hybrid of Gaussian and Laplace

LAPLACE ERRORS - L1-LOSS / 3

- We simulate data $y \mid \mathbf{x} \sim \text{Laplacian}(f_{\text{true}}(\mathbf{x}), 1)$ with $f_{\text{true}} = 0.2 \cdot \mathbf{x}$.
- We can plot the empirical error distribution, i.e. the distribution of the residuals after fitting a regression model w.r.t. $L1$-loss.
- With the help of a Q-Q-plot we can compare the empirical residuals vs. the theoretical quantiles of a Laplacian distribution.





MAXIMUM LIKELIHOOD IN CLASSIFICATION / 2

This gives rise to the following loss function

$$L(y, \pi(\mathbf{x})) = -y \log(\pi(\mathbf{x})) - (1 - y) \log(1 - \pi(\mathbf{x})), \quad y \in \{0, 1\}$$

which we introduced as **Bernoulli** loss.