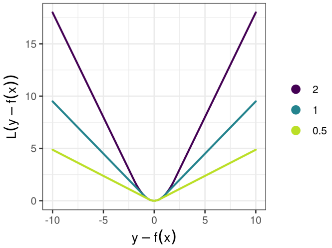
Introduction to Machine Learning

Regression Losses: Huber loss



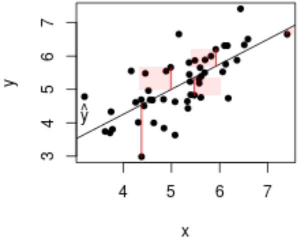
Learning goals

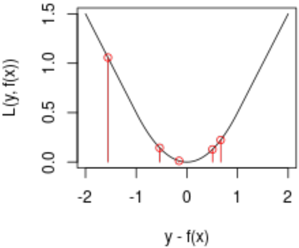
- Know the Huber loss
- Understand that there is no closed-form risk minimizer to the Huber loss
- Find the optimal constant model via iterative optimization

HUBER-LOSS

$$L\left(y, f(\mathbf{x})\right) = \begin{cases} \frac{1}{2}(y - f(\mathbf{x}))^2 & \text{if } |y - f(\mathbf{x})| \leq \delta \\ \delta|y - f(\mathbf{x})| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}, \delta > 0$$

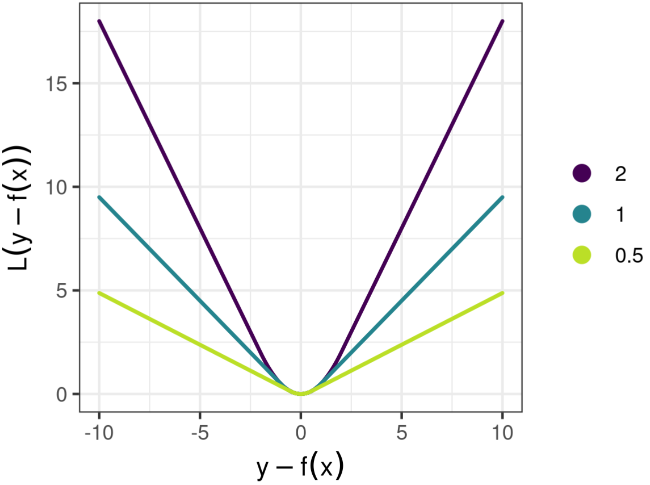
- Piece-wise combination of $L1$ and $L2$ loss
- Analytic properties: Convex, differentiable, robust
- Combines advantages of $L1$ and $L2$ loss: differentiable + robust





HUBER-LOSS

The following plot shows the Huber loss for different values of δ.



L2-LOSS: RISK MINIMIZER

Let us consider the (true) risk for $\mathcal{Y} = \mathbb{R}$ and the Huber loss

$$L\left(y, f(\mathbf{x})\right) = \begin{cases} \frac{1}{2}(y - f(\mathbf{x}))^2 & \text{if } |y - f(\mathbf{x})| \leq \delta \\ \delta|y - f(\mathbf{x})| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}, \delta > 0$$

with unrestricted $\mathcal{H} = \{f : \mathcal{X} \to \mathbb{R}^g\}$.

There is no closed-form solution for the risk minimizer of the Huber loss.

However, the risk minimizer for the Huber loss is a **trimmed mean**: The risk minimizer is the (conditional) mean of values between two (conditional) quantiles. The location of the quantiles depends on the distribution as well as the value of δ.

HUBER LOSS: OPTIMAL CONSTANT MODEL

What is the optimal constant model $f(\mathbf{x}) = \boldsymbol{\theta}$ w.r.t. the Huber loss?

$$
\begin{aligned}
f &= \operatorname*{arg\,min}_{f \in \mathcal{H}} \mathcal{R}_{\text{emp}}(f) \\
\Leftrightarrow \hat{\boldsymbol{\theta}} &= \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \mathbb{R}} \sum_{i=1}^{n} L(y, \boldsymbol{\theta})
\end{aligned}
$$

with $L(y, \boldsymbol{\theta}) = \begin{cases} \frac{1}{2}(y - \boldsymbol{\theta})^2 & \text{if } |y - \boldsymbol{\theta}| \leq \delta \\ \delta|y - \boldsymbol{\theta}| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$.

- There is no closed-form solution.
- Numerical optimization methods are necessary.
- $\rightarrow$ the “optimal” solution can only be approached to a certain degree of accuracy via iterative optimization.

$L1$- VS. $L2$- VS. HUBER LOSS

- **Optimization**: $L2$ loss can be differentiated and the empirical risk minimization problem has a closed-form solution; $L1$ is not differentiable and has no closed-form solution.
- **Robustness**: $L1$ loss penalizes large residuals less than $L2$ loss, thus, $L1$ loss is more robust to outliers.
- Huber loss has the robustness of $L1$ loss where residuals are large and flexibility of $L2$ loss where residuals are small.