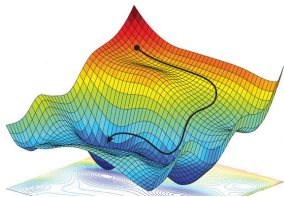


# Introduction to Machine Learning

## Advanced Regression Losses



### Learning goals

- Know the Huber loss
- Know the Log-Barrier loss
- Know the  $\epsilon$ -Insensitive loss
- Know the Quantile loss
- Know the Cauchy loss

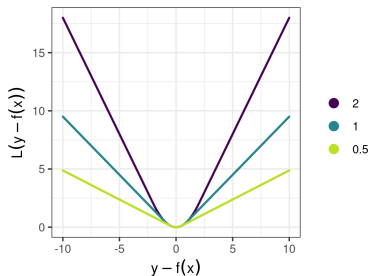
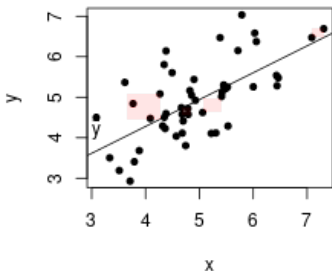
# ADVANCED LOSS FUNCTIONS

- Advanced loss functions are designed to achieve special properties (e.g. robustness and smoothness for the Huber or Cauchy loss)
- Furthermore, special loss functions are necessary in certain applications
- Examples:
  - Quantile loss: Overestimating a clinical parameter might not be as bad as underestimating it
  - Log-Barrier loss: Extremely under- or overestimating demand in production would yield to bankruptcy
  - $\epsilon$ -Insensitive loss: A certain amount of deviation in production does no harm, larger deviations do
- Sometimes a custom loss must be designed specifically for the given application.
- Some learning algorithms use specific loss function, e.g., the Hinge loss for SVMs

# HUBER LOSS

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

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



# HUBER LOSS

## Risk minimizer:

- There is no closed-form solution for the risk minimizer.
- 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  $\delta$ .

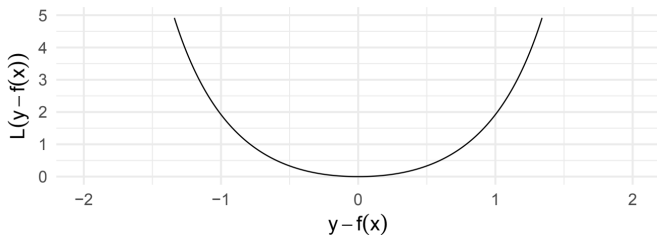
## Optimal constant model:

- Similarly, there is no closed-form solution for the optimal constant model.
- Numerical optimization methods are necessary.
- The “optimal” solution can only be approached to a certain degree of accuracy via iterative optimization.

# LOG-BARRIER LOSS

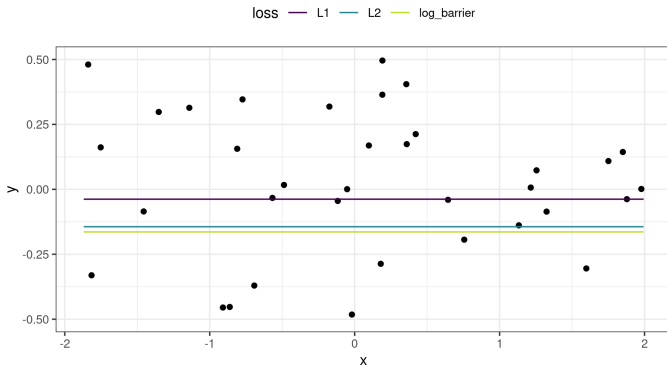
$$L(y, f(\mathbf{x})) = \begin{cases} -a^2 \cdot \log\left(1 - \left(\frac{|y-f(\mathbf{x})|}{a}\right)^2\right), & \text{if } |y - f(\mathbf{x})| \leq a \\ \infty, & \text{if } |y - f(\mathbf{x})| > a \end{cases}$$

- Behaves like L2 loss for small residuals
- We use this, if we don't want residuals larger than  $a$  at all
- No guarantee that the risk minimization problem has a solution
- Plot shows Log-Barrier Loss for  $a = 2$



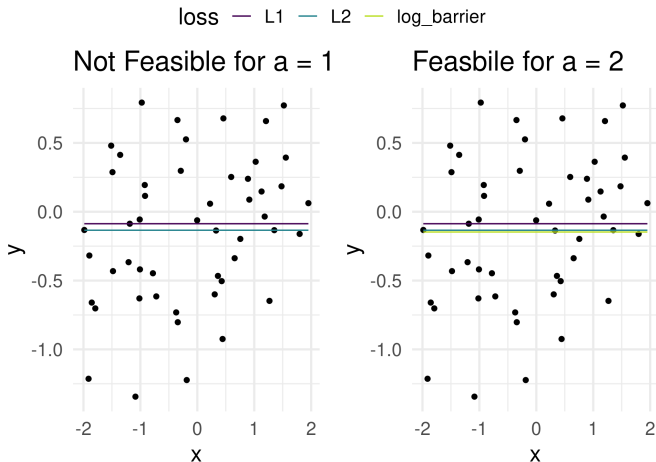
# LOG BARRIER: OPTIMAL CONSTANT MODEL

- Similarly to the Huber loss, there is no closed-form solution for the optimal constant model  $f(\mathbf{x}) = \theta$  w.r.t. the Log Barrier loss.
- Again, numerical optimization methods are necessary.



# LOG BARRIER: OPTIMAL CONSTANT MODEL

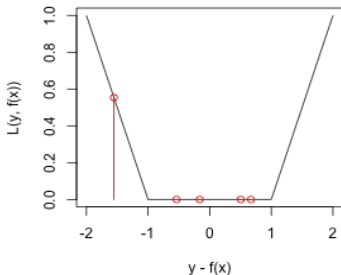
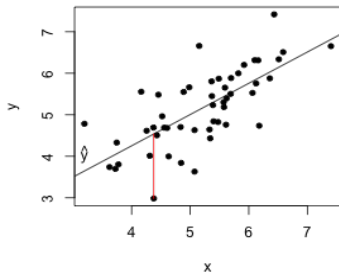
Note that the optimization problem has no (finite) solution, if there is no way to fit a constant where all residuals are smaller than  $a$ .



# $\epsilon$ -INSENSITIVE LOSS

$$L(y, f(\mathbf{x})) = \begin{cases} 0, & \text{if } |y - f(\mathbf{x})| \leq \epsilon \\ |y - f(\mathbf{x})| - \epsilon, & \text{otherwise} \end{cases}, \quad \epsilon \in \mathbb{R}_+$$

- Modification of L1-loss, errors below  $\epsilon$  accepted without penalty
- Properties: convex and not differentiable for  $y - f(\mathbf{x}) \in \{-\epsilon, \epsilon\}$





# $\epsilon$ -INSENSITIVE LOSS: OPTIMAL CONSTANT

What is the optimal constant model  $f(\mathbf{x}) = \theta$  w.r.t. the  $\epsilon$ -insensitive loss  $L(y, f(\mathbf{x})) = |y - f(\mathbf{x})| \mathbf{1}_{\{|y - f(\mathbf{x})| > \epsilon\}}$ ?

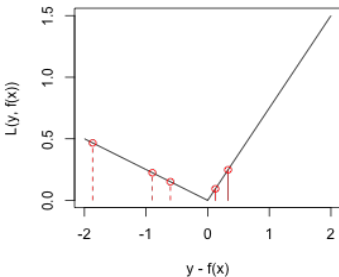
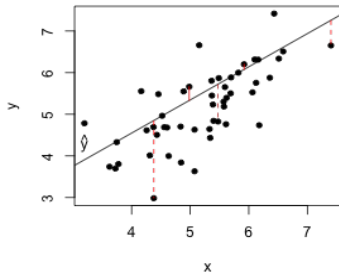
$$\begin{aligned}\hat{\theta} &= \arg \min_{\theta \in \mathbb{R}} \sum_{i=1}^n L(y^{(i)}, f(\mathbf{x}^{(i)})) \\ &= \arg \min_{\theta \in \mathbb{R}} \sum_{i \in I_\epsilon} |y^{(i)} - \theta| - \epsilon \\ &= \arg \min_{\theta \in \mathbb{R}} \sum_{i \in I_\epsilon} |y^{(i)} - \theta| - \sum_{i \in I_\epsilon} \epsilon \\ &= \text{median} \left( \left\{ y^{(i)} \mid i \in I_\epsilon \right\} \right) - |I_\epsilon| \cdot \epsilon\end{aligned}$$

with  $I_\epsilon := \{i : |y^{(i)} - f(\mathbf{x}^{(i)})| \leq \epsilon\}$ .

# REGRESSION LOSSES: QUANTILE LOSS

$$L(y, f(\mathbf{x})) = \begin{cases} (1 - \alpha)(f(\mathbf{x}) - y), & \text{if } y < f(\mathbf{x}) \\ \alpha(y - f(\mathbf{x})) & \text{if } y \geq f(\mathbf{x}) \end{cases}, \quad \alpha \in (0, 1)$$

- Is an extension of L1 loss (with  $\alpha = 0.5$  equals L1 loss)
- Weights positive / negative residuals more
- $\alpha < 0.5$  ( $\alpha > 0.5$ ) penalty to over-estimation (under-estimation)
- Also known as **pinball loss**



# REGRESSION LOSSES: QUANTILE LOSS

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

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}} \sum_{i=1}^n L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right)$$

$$\Leftrightarrow \hat{\theta} = \arg \min_{\theta \in \mathbb{R}} \left\{ (1 - \alpha) \sum_{y^{(i)} < \theta} |y^{(i)} - \theta| + \alpha \sum_{y^{(i)} \geq \theta} |y^{(i)} - \theta| \right\}$$

$$\Leftrightarrow \hat{\theta} = Q_{\alpha}(\{y^{(i)}\})$$

where  $Q_{\alpha}(\cdot)$  computes the empirical  $\alpha$ -quantile of  $\{y^{(i)}\}, i = 1, \dots, n$ .

# CAUCHY LOSS

$$L(y, f(\mathbf{x})) = \log \left( \frac{1}{2} (x/c)^2 + 1 \right), \quad c \in \mathbb{R}$$

- Particularly robust towards outliers (controllable via  $c$ )
- Analytic properties: Differentiable, robust, but not convex!

