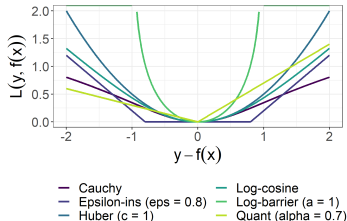


# Introduction to Machine Learning

## 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

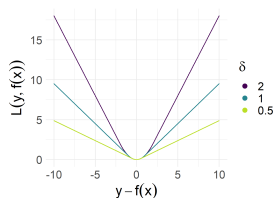
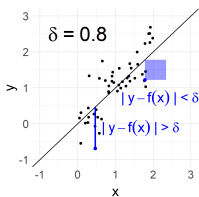
# ADVANCED LOSS FUNCTIONS

- 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 put company profit at risk.
  - $\epsilon$ -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 functions, e.g., the hinge loss for SVMs.

# HUBER LOSS

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

- Piece-wise combination of  $L1/L2$  to have robustness/smoothness
- Analytic properties: convex, differentiable, robust

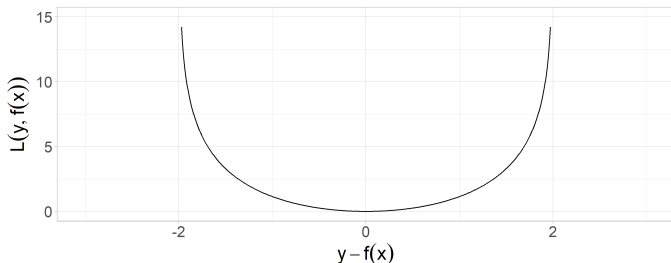


- Risk minimizer and optimal constant model do not have a closed-form solution. @BB: pls insert text / Okay like this?
- The solution is a **trimmed mean**: a (conditional) mean between two (conditional) quantiles.
- To fit a model numerical optimization is necessary.

# LOG-BARRIER LOSS

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

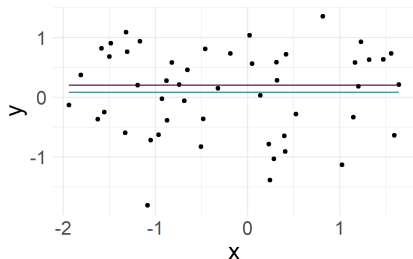
- Behaves like  $L2$  loss for small residuals.
- We use this if we don't want residuals larger than  $a$  at all.
- No guarantee that the risk minimization problem has a solution.
- Plot shows log-barrier loss for  $a = 2$ :



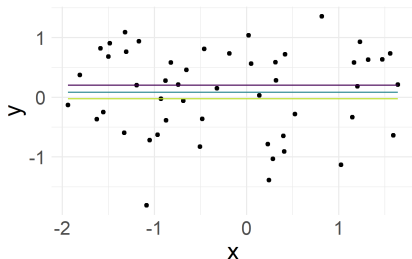
# LOG-BARRIER: OPTIMAL CONSTANT MODEL

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

Not feasible for  $a = 1$



Feasible for  $a = 2$

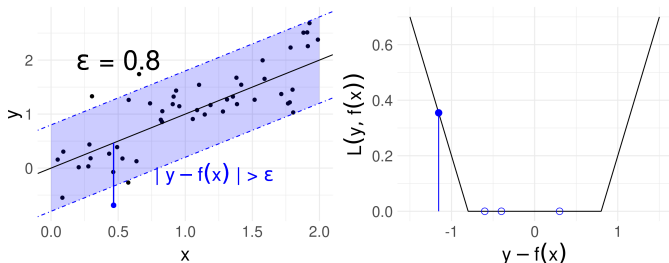


Loss — L1 — L2 — log-barrier

# $\epsilon$ -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  $L(y, f(\mathbf{x})) = |y - f(\mathbf{x})| \mathbb{1}_{\{|y - f(\mathbf{x})| > \epsilon\}}$ ?

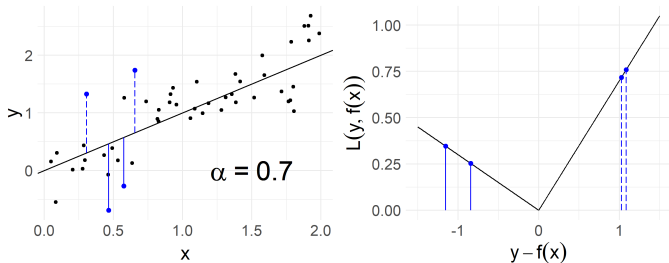
$$\begin{aligned}\hat{\theta} &= \arg \min_{\theta \in \mathbb{R}} \sum_{i=1}^n L(y^{(i)}, f(\mathbf{x}^{(i)})) \\ &= \arg \min_{\theta \in \mathbb{R}} \sum_{i \in I_\epsilon} |y^{(i)} - \theta| - \epsilon \\ &= \arg \min_{\theta \in \mathbb{R}} \sum_{i \in I_\epsilon} |y^{(i)} - \theta| - \sum_{i \in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

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

# QUANTILE LOSS

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

- Extension of  $L1$  loss (equal to  $L1$  for  $\alpha = 0.5$ ).
- Weights either positive or negative residuals more strongly.
- $\alpha < 0.5$  ( $\alpha > 0.5$ ) penalty to over-estimation (under-estimation)
- Also known as **pinball loss**.





# QUANTILE LOSS

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

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

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

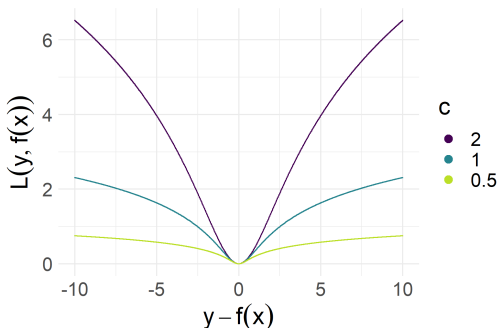
$$\Leftrightarrow \hat{\theta} = Q_{\alpha}(\{y^{(i)}\})$$

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

# CAUCHY LOSS

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

- Particularly robust toward outliers (controllable via  $c$ ).
- Analytic properties: differentiable, robust, but not convex!



# LOG-COSH LOSS

$$L(y, f(\mathbf{x})) = \log(\cosh(|y - f(\mathbf{x})|))$$

- Logarithm of the hyperbolic cosine of the residual.
- Approximately equal to  $0.5(|y - f(\mathbf{x})|)^2$  for small  $\mathbf{x}$  and to  $|y - f(\mathbf{x})| - \log 2$  for large  $\mathbf{x}$ , meaning it works mostly like  $L_2$  loss but is less outlier-sensitive.
- Has all the advantages of Huber loss and is, moreover, twice differentiable everywhere.

