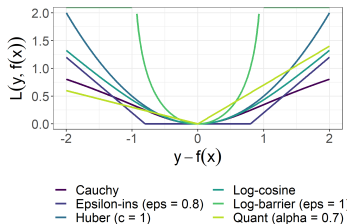


# Introduction to Machine Learning

## Advanced Regression Losses



### Learning goals

- Know the Huber loss
- Know the log-cosh loss
- Know the Cauchy loss
- Know the log-barrier loss
- Know the  $\epsilon$ -insensitive loss
- Know the quantile loss

# ADVANCED LOSS FUNCTIONS

Special loss functions can be used to estimate non-standard posterior components, to measure errors in a custom way or are designed to have special properties like robustness.

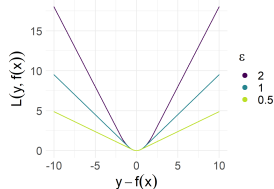
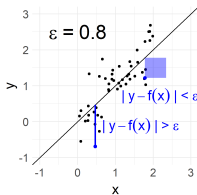
Examples:

- 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.

# HUBER LOSS

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

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

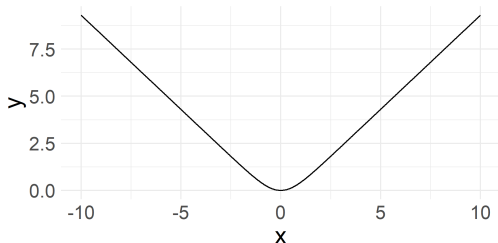


- Risk minimizer and optimal constant do not have a closed-form solution. To fit a model numerical optimization is necessary.
- Solution behaves like **trimmed mean**: a (conditional) mean of two (conditional) quantiles.

# LOG-COSH LOSS

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

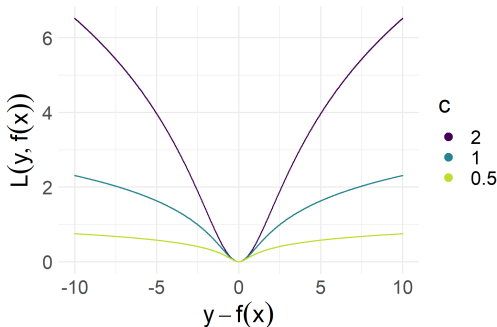
- Logarithm of the hyperbolic cosine of the residual.
- Approximately equal to  $0.5(|y - f(\mathbf{x})|)^2$  for small  $\mathbf{x}$  and to  $|y - f(\mathbf{x})| - \log 2$  for large  $\mathbf{x}$ , meaning it works mostly like  $L_2$  loss but is less outlier-sensitive.
- Has all the advantages of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



# CAUCHY LOSS

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

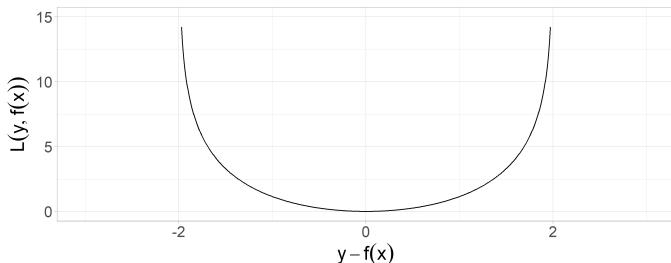
- Particularly robust toward outliers (controllable via  $c$ ).
- Analytic properties: differentiable, but not convex!



# LOG-BARRIER LOSS

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

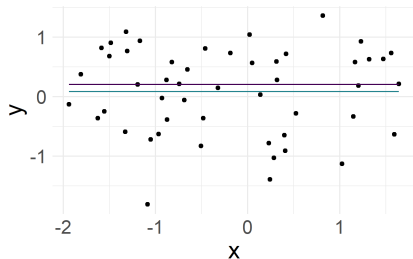
- Behaves like  $L2$  loss for small residuals.
- We use this if we don't want residuals larger than  $\epsilon$  at all.
- No guarantee that the risk minimization problem has a solution.
- Plot shows log-barrier loss for  $\epsilon = 2$ :



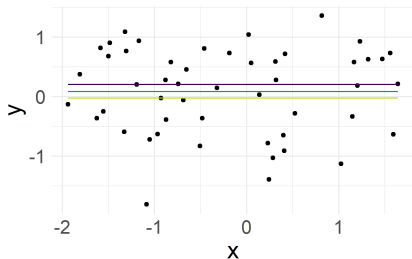
# LOG-BARRIER LOSS

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

Not feasible for  $\epsilon = 1$



Feasible for  $\epsilon = 2$

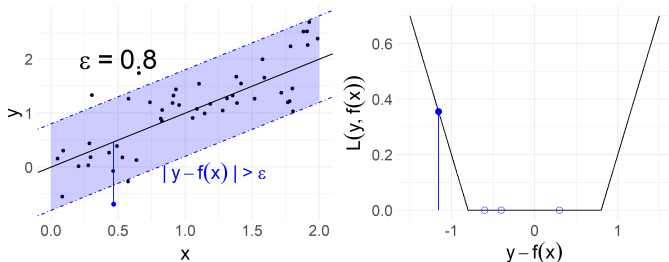


Loss — L1 — L2 — log-barrier

# $\epsilon$ -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.
- Used in SVM regression.
- Properties: convex and not differentiable for  $y - f(\mathbf{x}) \in \{-\epsilon, \epsilon\}$ .





# QUANTILE LOSS / PINBALL 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)
- Risk minimizer is (conditional)  $\alpha$ -quantile (median for  $\alpha = 0.5$ ).

