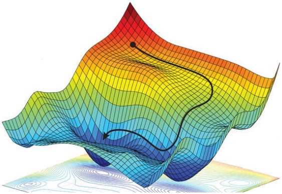
Introduction to Machine Learning

Advanced Regression Losses



Learning goals

- Know the Huber loss
- Know the log-barrier loss
- Know the ϵ-insensitive loss
- Know the quantile loss
- Know the Cauchy loss

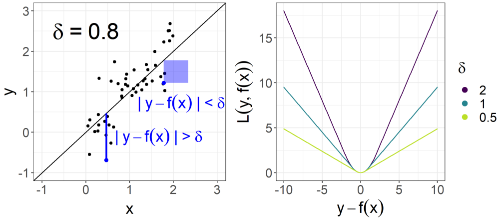
ADVANCED LOSS FUNCTIONS

- 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.
  - ϵ-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 $L1$ and $L2$ loss
- Analytic properties: convex, differentiable, robust
- Combines advantages of $L1$ and $L2$ loss: differentiable + robust



- xxx
- xxx
- xxx

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 δ.

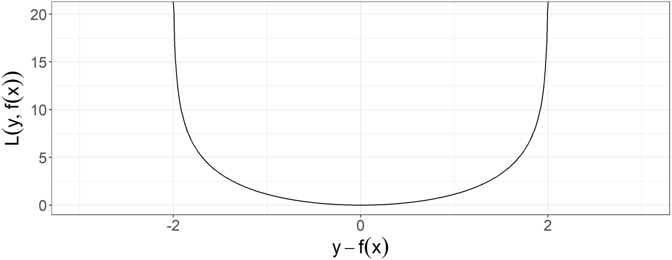
Optimal constant model:

- 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

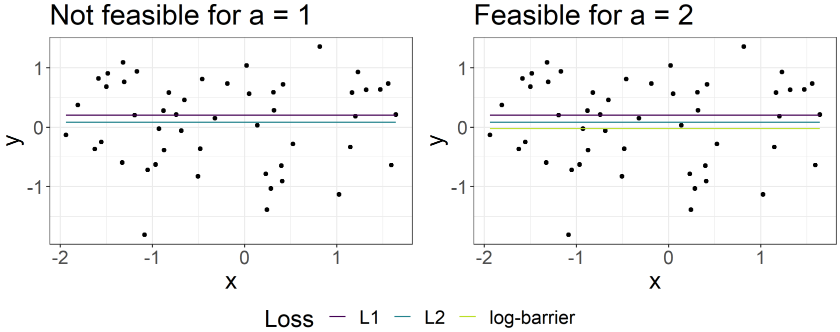
$$L\left(y, f(\mathbf{x})\right) = \begin{cases} -a^2 \cdot \log\left(1 - \left(\frac{|y - f(\mathbf{x})|}{a}\right)^2\right) & \text{if } |y - f(\mathbf{x})| \leq a \\ \infty & \text{if } |y - f(\mathbf{x})| > a \end{cases}$$

- Behaves like *L*2 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

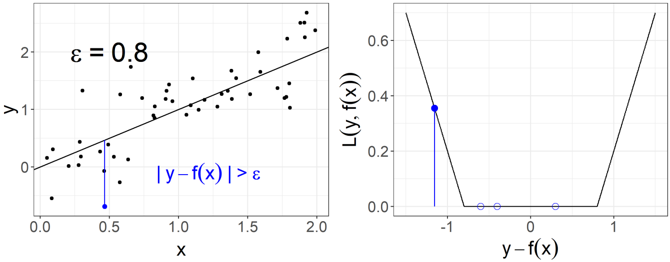
- Similarly to the Huber loss, there is no closed-form solution for the optimal constant model $f(\mathbf{x}) = \theta$ w.r.t. the log-barrier loss.
- Again, numerical optimization methods are necessary.
- Note that the optimization problem has no (finite) solution if there is no way to fit a constant where all residuals are smaller than a.



ϵ-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 ϵ accepted without penalty.
- Properties: convex and not differentiable for $y - f(\mathbf{x}) \in \{-\epsilon, \epsilon\}$.



ϵ-INSENSITIVE LOSS: OPTIMAL CONSTANT

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

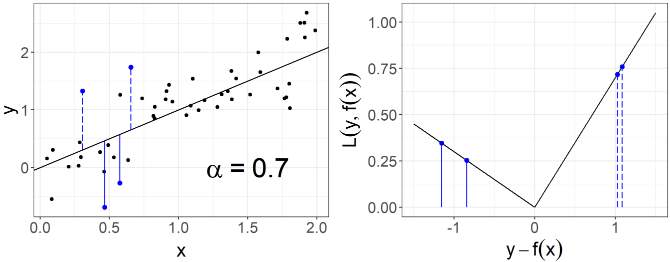
$$
\begin{aligned}
\hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta} \in \mathbb{R}}{\arg\min} \sum_{i=1}^{n} L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right) \\
&= \underset{\boldsymbol{\theta} \in \mathbb{R}}{\arg\min} \sum_{i \in I_\epsilon} \left| y^{(i)} - \boldsymbol{\theta} \right|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

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

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?

$$\begin{aligned}
& \hat{\theta} &=& \arg\min_{\theta \in \mathbb{R}} \sum_{i=1}^{n} L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right) \\
\Leftrightarrow\ & \hat{\theta} &=& \arg\min_{\theta \in \mathbb{R}} \left\{ (1-\alpha) \sum_{y^{(i)} < \theta} \left| y^{(i)} - \theta \right| + \alpha \sum_{y^{(i)} \geq \theta} \left| y^{(i)} - \theta \right| \right\} \\
\Leftrightarrow\ & \hat{\theta} &=& Q_\alpha(\{y^{(i)}\})
\end{aligned}$$

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

CAUCHY LOSS

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

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