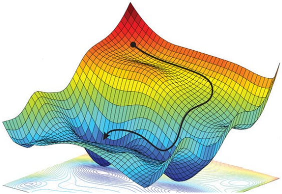
Introduction to Machine Learning

Advanced Regression Losses



Learning goals

- Know the Log-Barrier loss
- Know the Epsilon-Insensitive loss
- Know the Quantile loss

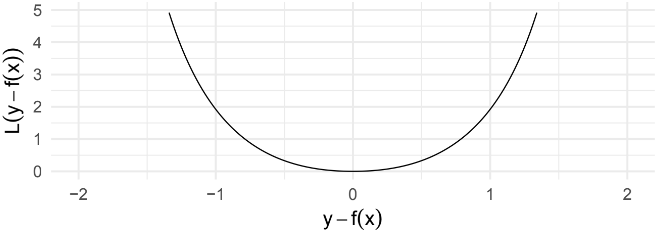
ADVANCED LOSS FUNCTIONS

- More advanced or 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
- Some learning algorithms use specific loss function, e.g., the Hinge loss for SVMs
- Sometimes a custom loss must be designed specifically for the given application.

LOG-BARRIER LOSS

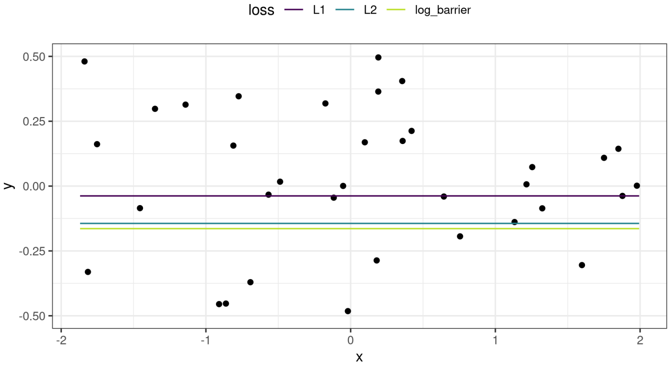
$$L\left(y, f(\mathbf{x})\right) = \begin{cases} -a^2 \cdot \log\left(1 - \left(\frac{|y - f(\mathbf{x})|}{a}\right)^2\right), & \text{if } |y - f(\mathbf{x})| \leq a \\ \infty, & \text{if } |y - f(\mathbf{x})| > a \end{cases}$$

- Behaves like 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 $a = 2$



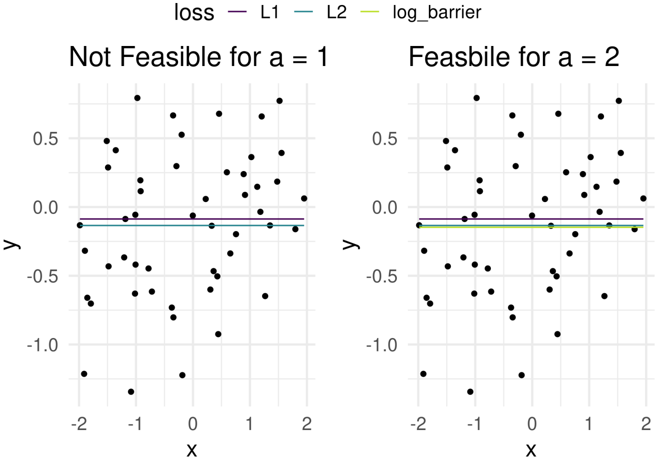
LOG BARRIER: OPTIMAL CONSTANT MODEL

- 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

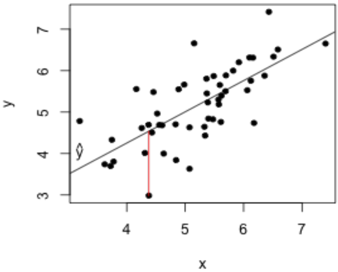
Note that the optimization problem has no (finite) solution, if there is no way to fit a constant where all residuals are smaller than *a*.

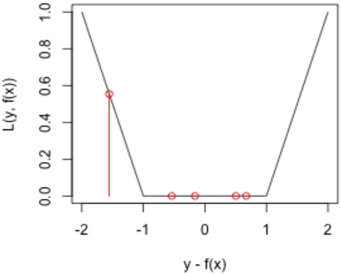


EPSILON-INSENSITIVE LOSS

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

- $\epsilon \in \mathbb{R}_+$
- Modification of L1-loss, errors below ϵ accepted without penalty
- Properties: convex and not differentiable for $y - f(\mathbf{x}) \in \{-\epsilon, \epsilon\}$
- no-closed form solution





ϵ-INSENSITIVE LOSS: OPTIMAL CONSTANT

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

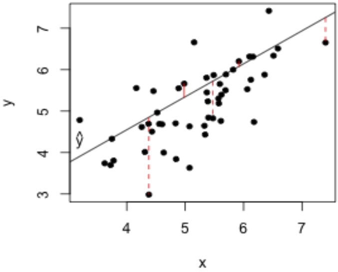
$$
\begin{aligned}
f &= \arg\min_{f \in \mathcal{H}} \mathcal{R}_{\text{emp}}(f) \\
\Leftrightarrow \quad \hat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta} \in \mathbb{R}} \sum_{i=1}^{n} L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right) \\
&= \arg\min_{\boldsymbol{\theta} \in \mathbb{R}}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

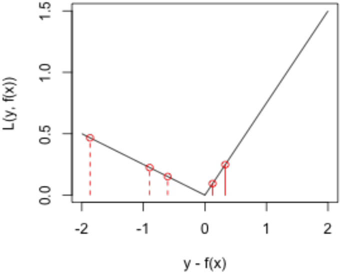
with $I_\epsilon := \left\{ i : |y^{(i)} - f\left(\mathbf{x}^{(i)}\right)| \leq \epsilon \right\}$.

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?

$$
\begin{aligned}
f &= \arg\min_{f \in \mathcal{H}} \mathcal{R}_{\text{emp}} \\
\Leftrightarrow \quad \hat{\theta} &= \arg\min_{\theta \in \mathbb{R}} \left\{ (1-\alpha) \sum_{y^{(i)} < \theta} \left| y^{(i)} - \theta \right| + \alpha \sum_{y^{(i)} \geq \theta} \left| y^{(i)} - \theta \right| \right\} \\
\Leftrightarrow \quad \hat{\theta} &= Q_\alpha(\{y^{(i)}\})
\end{aligned}
$$

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