Introduction to Machine Learning

Advanced Risk Minimization
Advanced Regression Losses
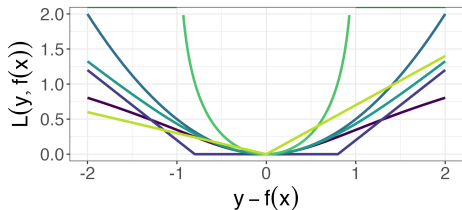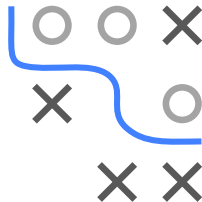

Cauchy
Epsilon-ins (eps = 0.8)
Huber (c = 1)
Log-cosine
Log-barrier (eps = 1)
Quant (alpha = 0.7)
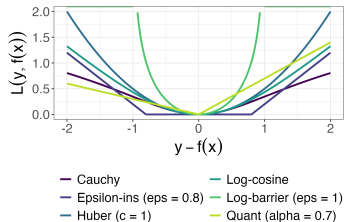
Learning goals

- Huber loss
- Log-Cosh loss
- Cauchy loss
- ϵ-Insensitive loss
- Quantile loss

ADVANCED LOSS FUNCTIONS ▸ Wang et al. 2020

- Handle errors in custom fashion
- Model other error distributions
  (see section on max. likelihood)
- Induce properties like robustness
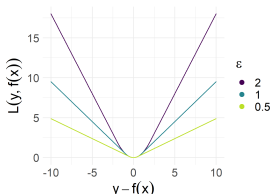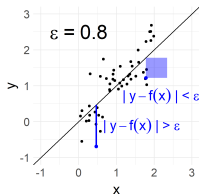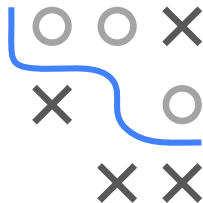- Handle other predictive tasks

HUBER LOSS

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

- Piece-wise combination of *L*1/*L*2 to have robustness/smoothness
- Analytic properties: convex, differentiable (once)

- No closed-form solution even for constant or linear model
- Solution behaves like **trimmed mean**:
  a (conditional) mean of two (conditional) quantiles
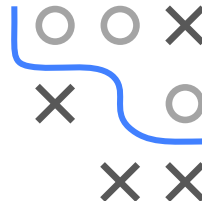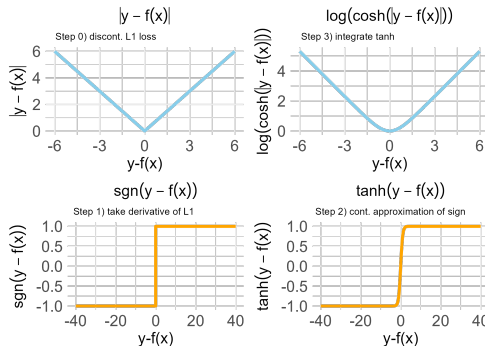
LOG-COSH LOSS ▸ R. A. Saleh and A. Saleh 2022

$$L(y, f(\mathbf{x})) = \log(\cosh(|y - f(\mathbf{x})|)) \qquad \cosh(x) = \frac{e^x + e^{-x}}{2}$$

- Approx. $0.5(|y - f(\mathbf{x})|)^2$ for small residuals;
  $|y - f(\mathbf{x})| - \log 2$ for large residuals
- Smoothed combo of $L1$ / $L2$ loss
- Similar to Huber, but twice differentiable

LOG-COSH LOSS ▸ R. A. Saleh and A. Saleh 2022
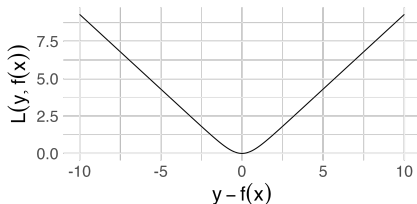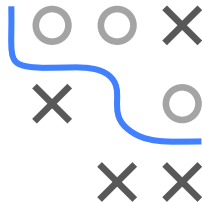
Essential idea:

1 Derivative of $L1$ w.r.t. residual

2 Approx. sign with tanh

3 Integrate "up again"

Same trick can be used to get differentiable pinball losses

LOG-COSH LOSS   ▸ R. A. Saleh and A. Saleh 2022

cosh(θ, σ) distribution:

- Normalized reciprocal cosh(x) is pdf: positive and
  $\int_{-\infty}^{\infty} \frac{1}{\pi \cosh(x)} dx = 1$

- Location-scale type (θ, σ) resembling Gaussian with heavy tails

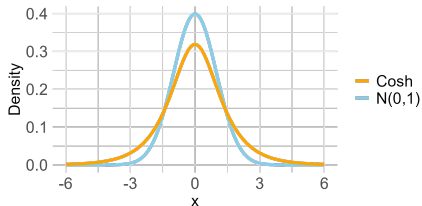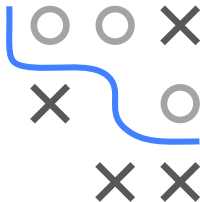- ERM using log-cosh is equivalent to MLE of cosh($\theta, 1$) distribution

- $p(x|\theta, \sigma) = \frac{1}{\pi \sigma \cosh(\frac{x-\theta}{\sigma})}$

- $\mathbb{E}_{x \sim p}[x] = \theta$

- $\text{Var}_{x \sim p}[x] = \frac{1}{4}\pi^2 \sigma^2$

- $\hat{\theta}^{MLE} = \arg\max_{\theta} \prod_{i=1}^{n} \frac{1}{\pi \cosh(y^{(i)} - \theta)} =$
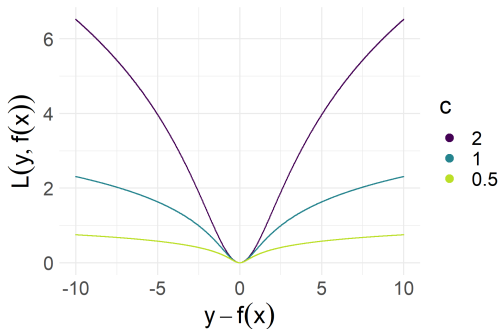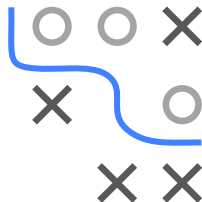 $\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \log(\cosh(y^{(i)} - \theta))$
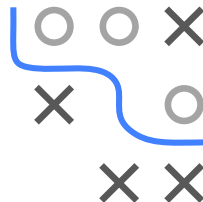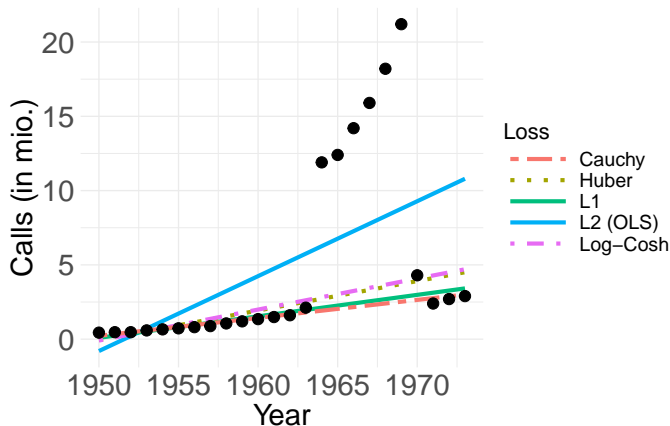
CAUCHY LOSS

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

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

TELEPHONE DATA

- Illustrate the effect of robust losses on telephone data set
- Nr. of calls (in 10mio units) in Belgium 1950-1973
- Outliers due to a change in measurement without re-calibration

ϵ-**INSENSITIVE LOSS**

$$L\left(y, f(\mathbf{x})\right) = \begin{cases} 0 & \text{if } |y - f(\mathbf{x})| \le \epsilon \\ |y - f(\mathbf{x})| - \epsilon & \text{otherwise} \end{cases}, \quad \epsilon \in \mathbb{R}_+$$
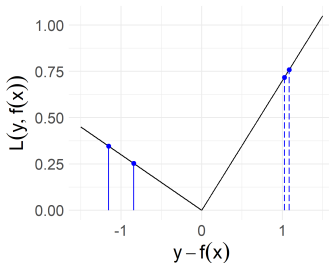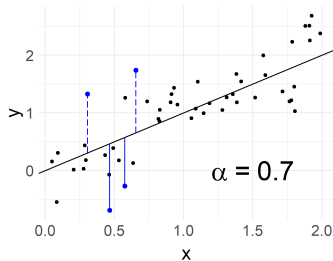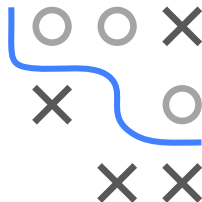
- Modification of $L1$, errors below ϵ get no penalty
- Used in SVM regression
- Properties: convex, not differentiable for $y - f(\mathbf{x}) \in \{-\epsilon, \epsilon\}$
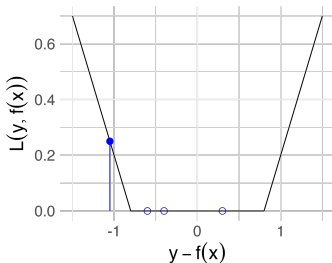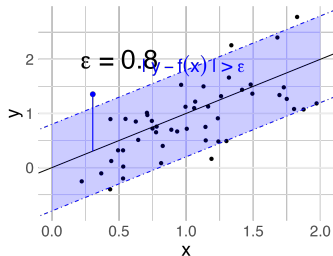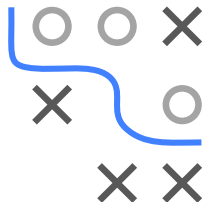
QUANTILE LOSS / PINBALL LOSS

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

- Extension of $L1$ loss (equal to $L1$ for $\alpha = 0.5$).
- Penalizes either over- or under-estimation more
- Risk minimizer is (conditional) α-quantile (median for $\alpha = 0.5$)

QUANTILE LOSS / PINBALL LOSS

- Simulate $n = 200$ samples from heteroskedastic LM
- $y = 1 + 0.2x + \varepsilon; \quad \varepsilon \sim \mathcal{N}(0, 0.5 + 0.5x); \quad x \sim \mathcal{U}[0, 10]$
- Fit LM with pinball losses to estimate α-quantiles