Machine Learning

Prof. Sergei Gleyzer

Week 2

PH451 PH551

January 21, 2025

Announcements

- **This week we will start HS#1:**
  - **Classification, metrics**
  - **Group/Team activity**
  - **Due next Tuesday, Jan 28 at 1pm**

- **Quiz date: Thu, Feb. 13**

Recap: Machine Learning

General Approach:

Given **training** data $T_D = \{y, \mathbf{x}\} = (y,x)_1 \ldots (y,x)_N$,

function space $\{f\}$ and a
constraint on these functions

Teach a machine to learn the **mapping** $y = f(x)$

Machine Learning

Find hypothesis that minimizes the error

Function Space $\boldsymbol{F} = \{ f(x, \mathbf{w}) \}$
possibly constrained
Loss Function L

Learn y = f(x)

- By minimizing empirical risk, computed from the loss function on a known set of training data
  - **How far are estimated values from true values?**

Optimization

Given real-valued function f: $R^p \rightarrow R$

$$\underset{\boldsymbol{x} \in \mathbb{R}^p}{\text{minimize}}\ f(\boldsymbol{x})$$

f is the objective function

- also loss function or cost function

Loss Functions

$\vec{x}_i = \{x_1, x_2 \dots x_m\}_i$ | y_i

Input | Output

GOAL: Evaluate hypothesis on $\boxed{\text{training}}$ data (how bad?)

$\uparrow$ Loss (worse) $\quad \downarrow$ Loss (better)

Loss = 0 $\longrightarrow$ Perfect

Examples

0/1 Loss

Count the Mistakes

Usually use normalized 0/1 Loss

$\Longrightarrow$ fraction of missclassified samples ("training error")

$$\boxed{L_{0/1}(f) = \frac{1}{n}\sum_{i=1}^{n} \delta f(x_i)}$$

where

$$\delta_{f(x_i)} = \overbrace{\begin{cases} 1 & \underbrace{f(\vec{x}_i) \neq 1}_{\text{miss-classified}} \\ 0 \end{cases}}^{\text{non-continuos "impractical"}}$$

Absolute Loss

$$\boxed{L_{Abs}(f) = \frac{1}{n}\sum_{i=1}^{n} |f(x_i) - y_i|}$$

$\boxed{\boxed{\text{MAE, L1}}}$

"Manhattan" Norm

$\longrightarrow$ non-negative

$\longrightarrow$ grows linearly w. missclassification

$\longrightarrow$ typically useful for (noisy) regression $\leftarrow$ more robust to outliers

Squared Loss

$$\boxed{L_{SQ}(f) = \frac{1}{n}\sum_{i=1}^{n} \left(f(\vec{x}_i) - y_i \right)^2}$$

$\boxed{\boxed{\text{MSE, L2}}}$

RMSE: "Euclidean" Norm

$\longrightarrow$ non-negative

$\longrightarrow$ grows quadratically w. missed predictions

$\longrightarrow$ useful for regression [Ordinary Least Squares]

$\longrightarrow$ estimates mean given x_i

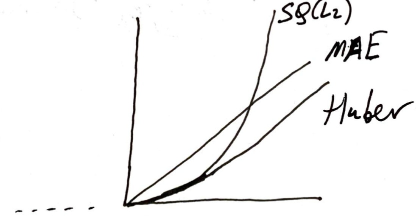
Huber

$$\boxed{L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & |a| \leq \delta \\ \delta(|a| - \frac{1}{2}\delta) \end{cases}}$$

$a = y - f(x)$ "residual"

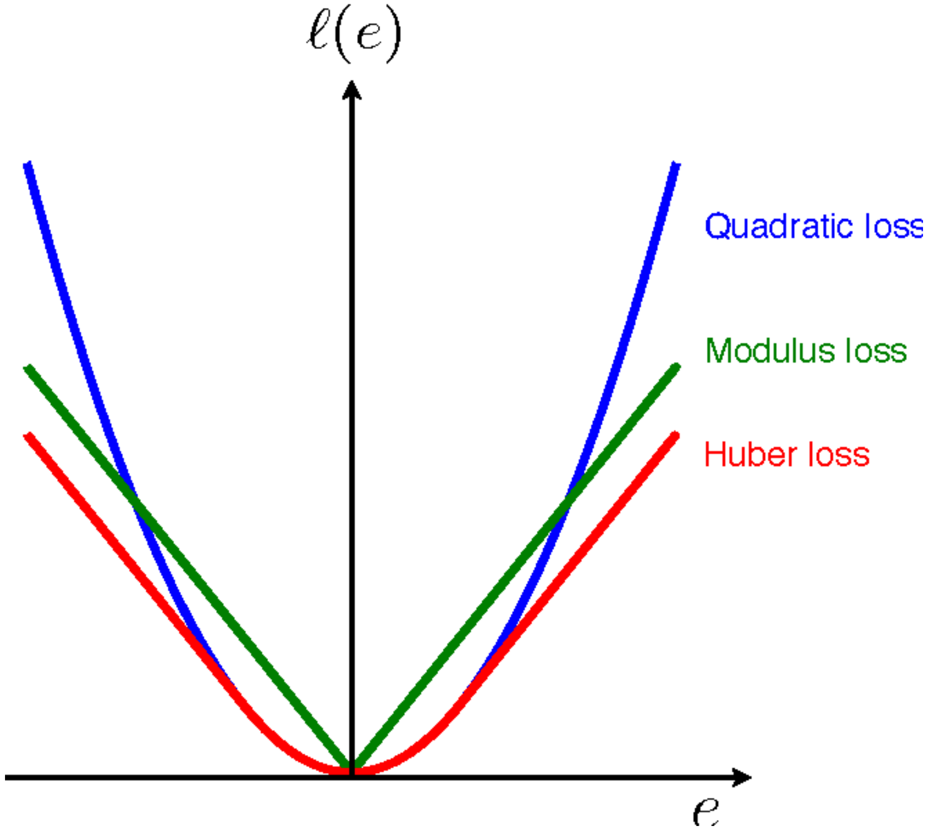
$\longrightarrow$ quadratic for small x

$\longrightarrow$ linear for large x

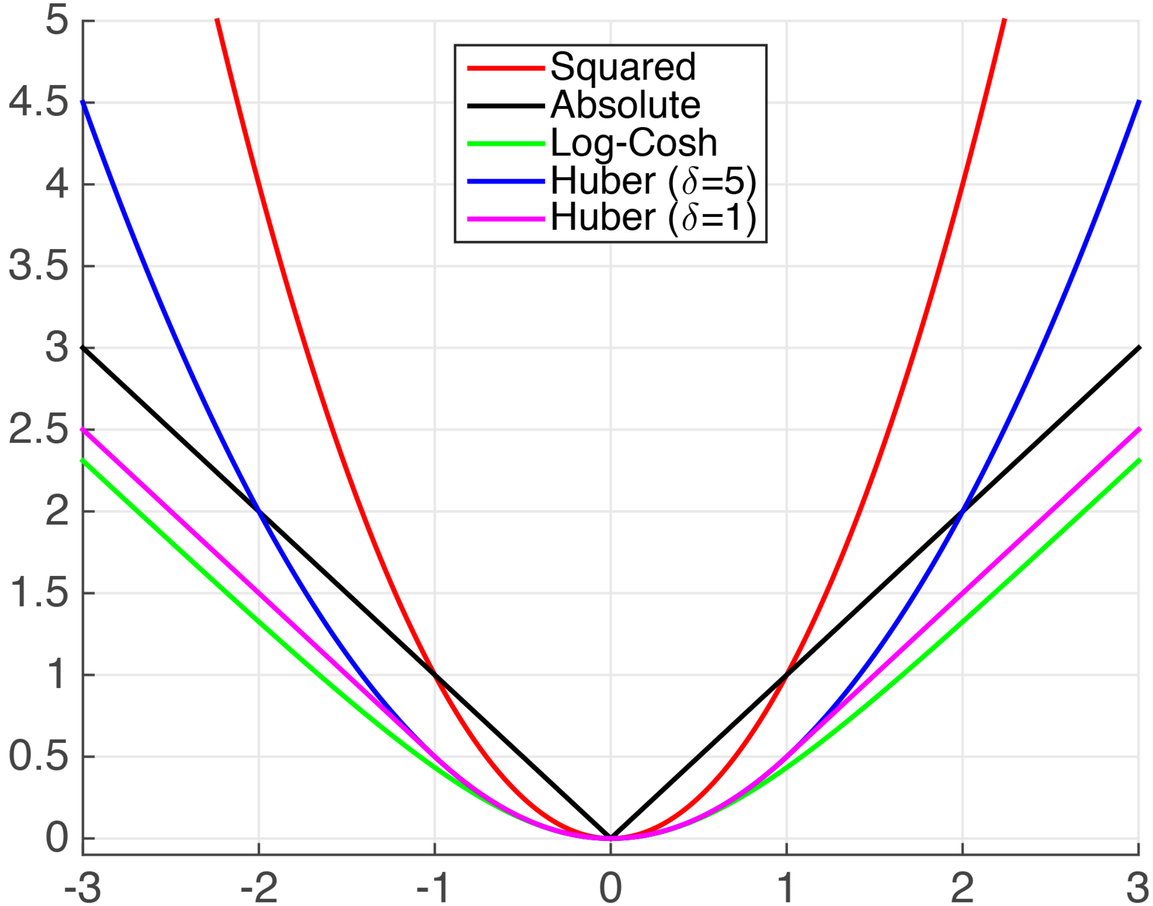


$\longrightarrow$ "best of both worlds"

Loss Functions



Loss Functions



Cross Entropy

$$L_{CE} = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$$

Log loss

binary class.

$f(x)$

probability 1

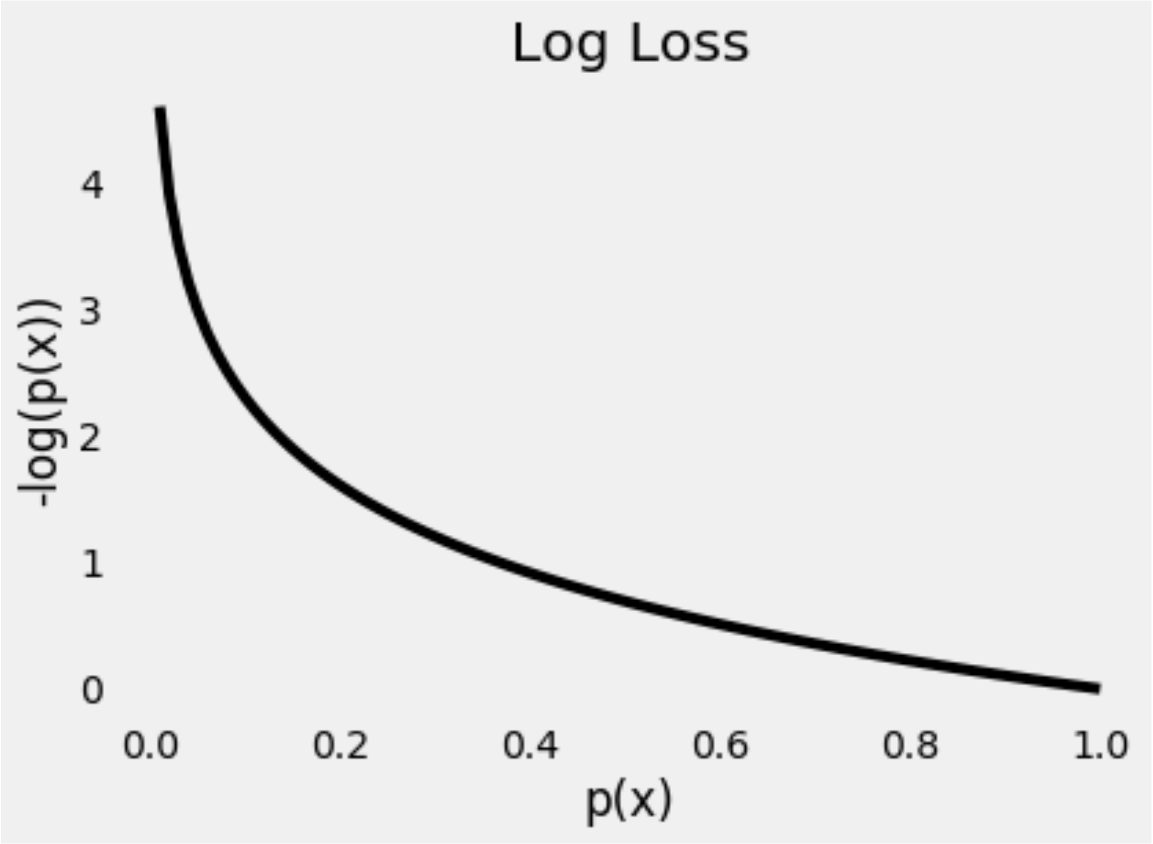
convex

$\uparrow$ loss

$y_i = 0 \quad \log(1 - f(\vec{x}))$

$y_i = 1 \quad \log(f(x))$

Cross Entropy



Loss Functions

Name	Known as	Typical Use	Estimator
Mean Absolute Error	L1	Regression	Median
Mean Squared Error	L2	Regression	Mean
Cross-Entropy		Classification	Maximum Likelihood

Hands-on Activity