Loss Functions for Classification

Cynthia Rudin
Duke
Loss Functions for Classification

Classification error:

Fraction of times \( \text{sign}(f(x_i)) \) is not \( y_i \)

\[
\text{\frac{1}{n} \sum_{i=1}^{n} [y_i \neq \text{sign}(f(x_i))]} \]

Minimizing this directly is usually computationally hard.
Loss Functions for Classification

Classification error:

Fraction of times $\text{sign}(f(x_i))$ is not $y_i$

$$= \frac{1}{n} \sum_{i=1}^{n} [y_i \neq \text{sign}(f(x_i))]$$

$$= \frac{1}{n} \sum_{i=1}^{n} 1_{[y_i f(x_i) \leq 0]}$$
Loss Functions for Classification

Classification error:

Fraction of times \(\text{sign}(f(x_i))\) is not \(y_i\):

\[
\frac{1}{n} \sum_{i=1}^{n} [y_i \neq \text{sign}(f(x_i))]
\]
\[ yf(x) < 0 \]

\[ yf(x) > 0 \]
\[ yf(x) < 0 \]

\[ yf(x) > 0 \]

\[ 1_{[yf(x) \leq 0]} \]
\[ yf(x) < 0 \quad \text{"very wrong"} \quad yf(x) > 0 \quad \text{"very correct"} \]

\[ yf(x) < 0 \quad \text{"sort of wrong"} \quad yf(x) > 0 \quad \text{"sort of correct"} \]

\[ 1_{[yf(x) \leq 0]} \]
$yf(x) < 0$

- "very wrong"

- "sort of wrong"

+ 

$l(yf(x))$

$1_{yf(x) \leq 0}$

$yf(x) > 0$

- "sort of correct"

+ 

+ 

- 

+ 

+ 

+ 

- 

+ 

+ 

+ 

"very correct"
Loss Functions for Classification

\[ l(yf(x)) \]

\[ 1_{[yf(x) \leq 0]} \]

\[ yf(x) \]
Loss Functions for Classification

\[ l(yf(x)) \]
Loss Functions for Classification

\[ l(yf(x)) \]
Loss Functions for Classification

\[ l(\gamma f(x)) \]

\[ e^{-\gamma f(x)} \] (AdaBoost)

\[ 1_{[\gamma f(x) \leq 0]} \]
Loss Functions for Classification

\[ l(yf(x)) \]

- \( e^{-yf(x)} \) (AdaBoost)
- \( \max(0, 1 - yf(x)) \) (SVM)
- \( 1_{[yf(x) \leq 0]} \)

Graph showing different loss functions with their corresponding expressions.
Loss Functions for Classification

\[ l(y_f(x)) \]

- \( e^{-y_f(x)} \) (AdaBoost)
- \( \log(1 + e^{-y_f(x)}) \) (Logistic Regression)
- \( \max(0, 1 - y_f(x)) \) (SVM)

\( y_f(x) \)
Loss Functions for Classification

Fraction of times \( \text{sign}(f(x_i)) \) is not \( y_i \):

\[
\frac{1}{n} \sum_{i=1}^{n} 1[y_i \neq \text{sign}(f(x_i))] \\
= \frac{1}{n} \sum_{i=1}^{n} 1[y_i f(x_i) \leq 0] \\
\leq \frac{1}{n} \sum_{i=1}^{n} \ell(y_i f(x_i))
\]
To summarize

• Margin: $y_i f(x_i)$ measures how well point $i$ is classified.
• It’s computationally hard to minimize the 0-1 loss directly
• Several ML methods minimize upper bounds for the 0-1 loss that involve the margin.