Loss functions for Classification

Machine Learning Techniques

References and Credits

- The content presented in these slides is derived from professor Arun Rajkumar's lectures and slides in the MLT course. These lectures form the "ground truth" for almost all the content in these slides.
- These slides should be viewed as a presentation of the same content in the professor's lectures using a different medium.
- These slides are not meant to be a replacement for the lectures.
- The method of incrementally displaying content on slides is borrowed from professor Mitesh Khapra.
- These slides were prepared using the tool mathcha.io.

0-1 Loss

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T \mathbf{x}\right) y$$

0-1 Loss

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T \mathbf{x}\right) y$$

$$L(\mathbf{x}, y, \mathbf{w}) =$$

0-1 Loss

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \big(\mathbf{w}^T\mathbf{x}\big)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \begin{cases} 0, & \big(\mathbf{w}^T\mathbf{x}\big)y \geqslant 0 \\ 1 & \big(\mathbf{w}^T\mathbf{x}\big)y < 0 \end{cases}$$

$$L(u) =$$

0-1 Loss

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \big(\mathbf{w}^T\mathbf{x}\big)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \begin{cases} 0, & \big(\mathbf{w}^T\mathbf{x}\big)y \geqslant 0 \\ 1 & \big(\mathbf{w}^T\mathbf{x}\big)y < 0 \end{cases}$$

$$L(u) = \begin{cases} 0, & u \geqslant 0 \\ 1 & u < 0 \end{cases}$$

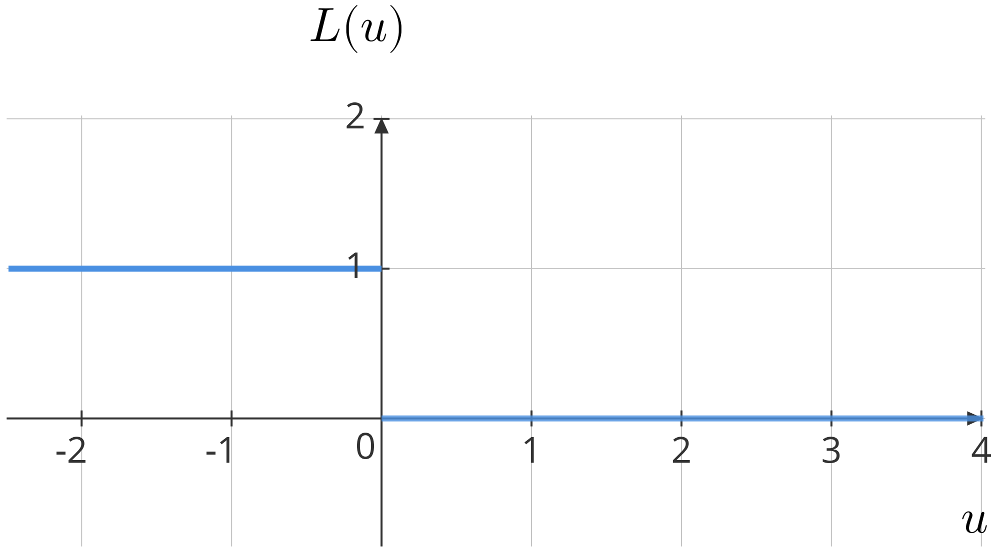
0-1 Loss

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \begin{cases} 0, & \left(\mathbf{w}^T\mathbf{x}\right)y \geqslant 0 \\ 1 & \left(\mathbf{w}^T\mathbf{x}\right)y < 0 \end{cases}$$

$$L(u) = \begin{cases} 0, & u \geqslant 0 \\ 1 & u < 0 \end{cases}$$



0-1 Loss

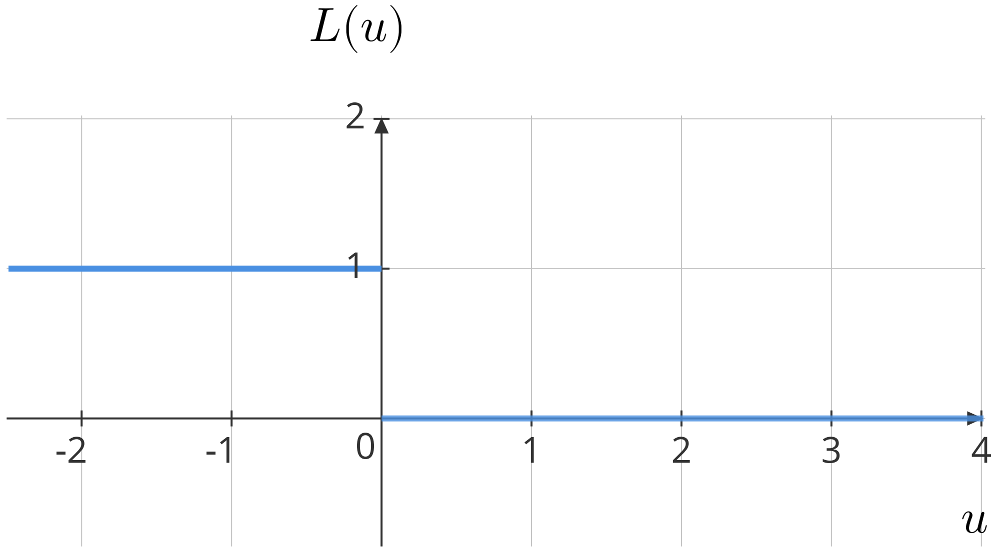
$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \begin{cases} 0, & \left(\mathbf{w}^T\mathbf{x}\right)y \geqslant 0 \\ 1 & \left(\mathbf{w}^T\mathbf{x}\right)y < 0 \end{cases}$$

$$L(u) = \begin{cases} 0, & u \geqslant 0 \\ 1 & u < 0 \end{cases}$$

- Not convex
- Optimizing it is NP-hard



Squared Loss

(Least Squares Classification)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T \mathbf{x}\right) y$$

$$L(\mathbf{x}, y, \mathbf{w}) =$$

Squared Loss

(Least Squares Classification)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \left(\mathbf{w}^T\mathbf{x} - y\right)^2$$

Squared Loss

(Least Squares Classification)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$\begin{aligned} L(\mathbf{x}, y, \mathbf{w}) &= \left(\mathbf{w}^T\mathbf{x} - y\right)^2 \\ &= \left(\mathbf{w}^T\mathbf{x}\right)^2 + y^2 - 2\left(\mathbf{w}^T\mathbf{x}\right)y \end{aligned}$$

Squared Loss

(Least Squares Classification)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$\begin{aligned} L(\mathbf{x}, y, \mathbf{w}) &= \left(\mathbf{w}^T\mathbf{x} - y\right)^2 \\ &= \left(\mathbf{w}^T\mathbf{x}\right)^2 + y^2 - 2\left(\mathbf{w}^T\mathbf{x}\right)y \\ &= \left(\mathbf{w}^T\mathbf{x}\right)^2 y^2 + 1 - 2\left(\mathbf{w}^T\mathbf{x}\right)y \end{aligned}$$

Squared Loss

(Least Squares Classification)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$\begin{aligned} L(\mathbf{x}, y, \mathbf{w}) &= \left(\mathbf{w}^T\mathbf{x} - y\right)^2 \\ &= \left(\mathbf{w}^T\mathbf{x}\right)^2 + y^2 - 2\left(\mathbf{w}^T\mathbf{x}\right)y \\ &= \left(\mathbf{w}^T\mathbf{x}\right)^2 y^2 + 1 - 2\left(\mathbf{w}^T\mathbf{x}\right)y \end{aligned}$$

$$\begin{aligned} L(u) &= u^2 - 2u + 1 \\ &= (u-1)^2 \end{aligned}$$

Squared Loss

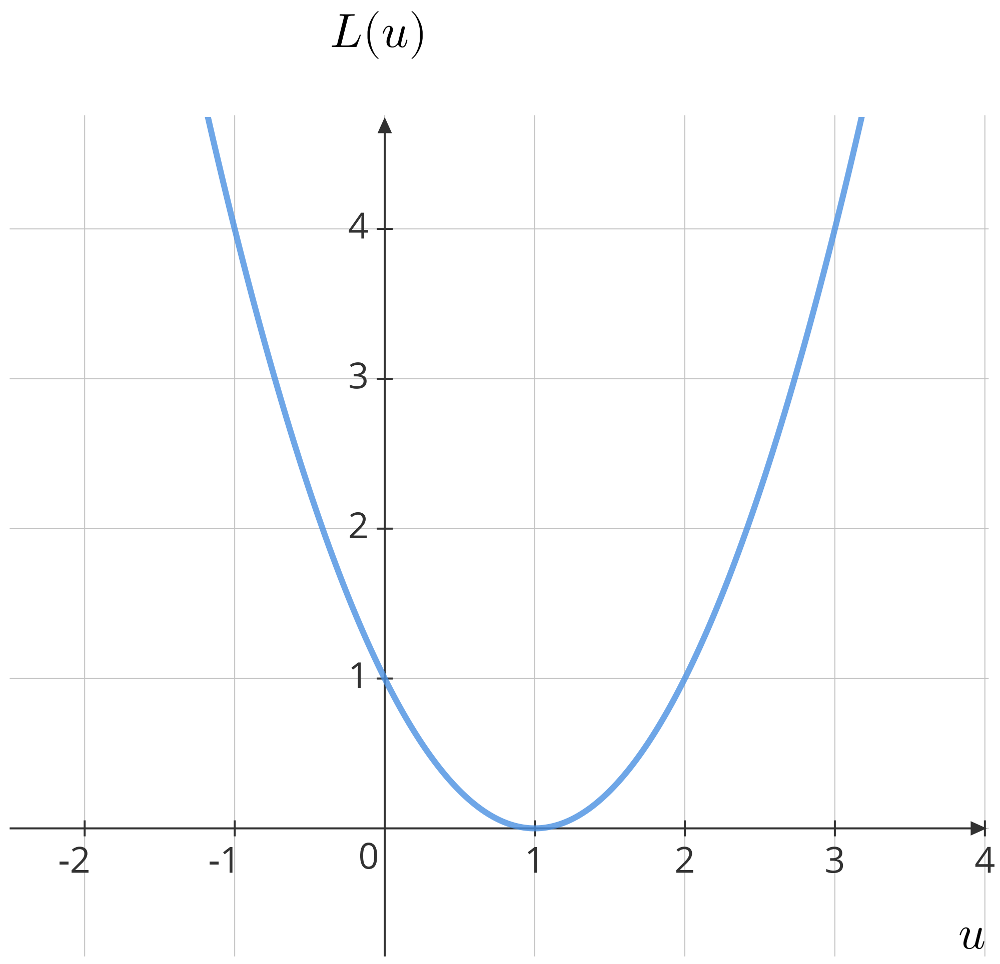
(Least Squares Classification)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ \ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$\begin{aligned} L(\mathbf{x}, y, \mathbf{w}) &= \left(\mathbf{w}^T\mathbf{x} - y\right)^2 \\ &= \left(\mathbf{w}^T\mathbf{x}\right)^2 + y^2 - 2\left(\mathbf{w}^T\mathbf{x}\right)y \\ &= \left(\mathbf{w}^T\mathbf{x}\right)^2 y^2 + 1 - 2\left(\mathbf{w}^T\mathbf{x}\right)y \end{aligned}$$

$$\begin{aligned} L(u) &= u^2 - 2u + 1 \\ &= (u-1)^2 \end{aligned}$$



Hinge Loss

(SVM)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T \mathbf{x}\right) y$$

$$L(\mathbf{x}, y, \mathbf{w}) =$$

Hinge Loss

(SVM)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \max\left(0, 1 - \left(\mathbf{w}^T\mathbf{x}\right)y\right)$$

Hinge Loss

(SVM)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ \ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \max\left(0, 1 - \left(\mathbf{w}^T\mathbf{x}\right)y\right)$$

$$\begin{aligned} L(u) &= \max(0, 1-u) \\ &= \begin{cases} 0, & u \geqslant 1 \\ 1-u, & u < 1 \end{cases} \end{aligned}$$

Hinge Loss

(SVM)

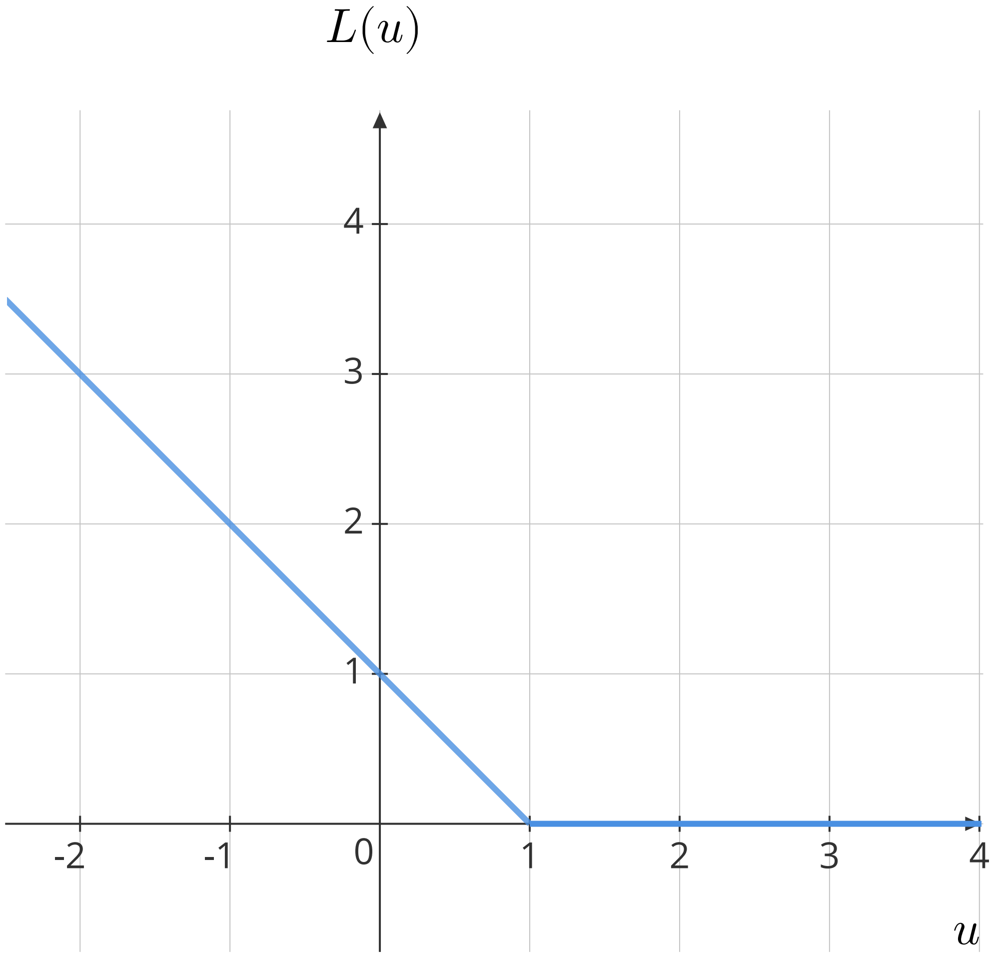
$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ \ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \max\Big(0, 1 - \left(\mathbf{w}^T\mathbf{x}\right)y\Big)$$

$$L(u) = \max(0, 1-u)$$

$$= \begin{cases} 0, & u \geqslant 1 \\ 1-u, & u < 1 \end{cases}$$



Logistic Loss

(Logistic Regression)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T \mathbf{x}\right) y$$

$$L(\mathbf{x}, z, \mathbf{w}) =$$

Logistic Loss

(Logistic Regression)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, z, \mathbf{w}) = \begin{cases} -\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 1 \\ -\log\left[1 - \sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 0 \end{cases}$$

Logistic Loss

(Logistic Regression)

$$\begin{aligned}\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] &= \log\left[\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right] \\ &= -\log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right) \\ \log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] &= \log\left[1-\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right] \\ &= -\log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right)\end{aligned}$$

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, z, \mathbf{w}) = \begin{cases} -\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 1 \\ -\log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 0 \end{cases}$$

Logistic Loss

(Logistic Regression)

$$\begin{aligned}\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] &= \log\left[\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right] \\ &= -\log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right) \\ \log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] &= \log\left[1-\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right] \\ &= -\log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right)\end{aligned}$$

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, z, \mathbf{w}) = \begin{cases} -\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 1 \\ -\log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 0 \end{cases}$$

$$= \begin{cases} \log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right), & z = 1 \\ \log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right), & z = 0 \end{cases}$$

Logistic Loss

(Logistic Regression)

$$\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] = \log\left[\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right] = -\log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right)$$

$$\log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] = \log\left[1-\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right] = -\log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right)$$

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, z, \mathbf{w}) = \begin{cases} -\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 1 \\ -\log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 0 \end{cases}$$

$$= \begin{cases} \log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right), & z = 1 \\ \log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right), & z = 0 \end{cases}$$

$$L(\mathbf{x}, y, \mathbf{w}) = \log\left(1+e^{-\left(\mathbf{w}^T\mathbf{x}\right)y}\right)$$

Logistic Loss

(Logistic Regression)

$$\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] = \log\left[\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right]$$
$$= -\log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right)$$
$$\log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] = \log\left[1-\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right]$$
$$= -\log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right)$$

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$
$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, z, \mathbf{w}) = \begin{cases} -\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z=1 \\ -\log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z=0 \end{cases}$$

$$= \begin{cases} \log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right), & z=1 \\ \log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right), & z=0 \end{cases}$$

$$L(\mathbf{x}, y, \mathbf{w}) = \log\left(1+e^{-\left(\mathbf{w}^T\mathbf{x}\right)y}\right)$$

$$L(u) = \log\left(1+e^{-u}\right)$$

Logistic Loss

(Logistic Regression)

$$\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] = \log\left[\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right]$$
$$= -\log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right)$$
$$\log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right] = \log\left[1-\frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}\right]$$
$$= -\log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right)$$

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$
$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

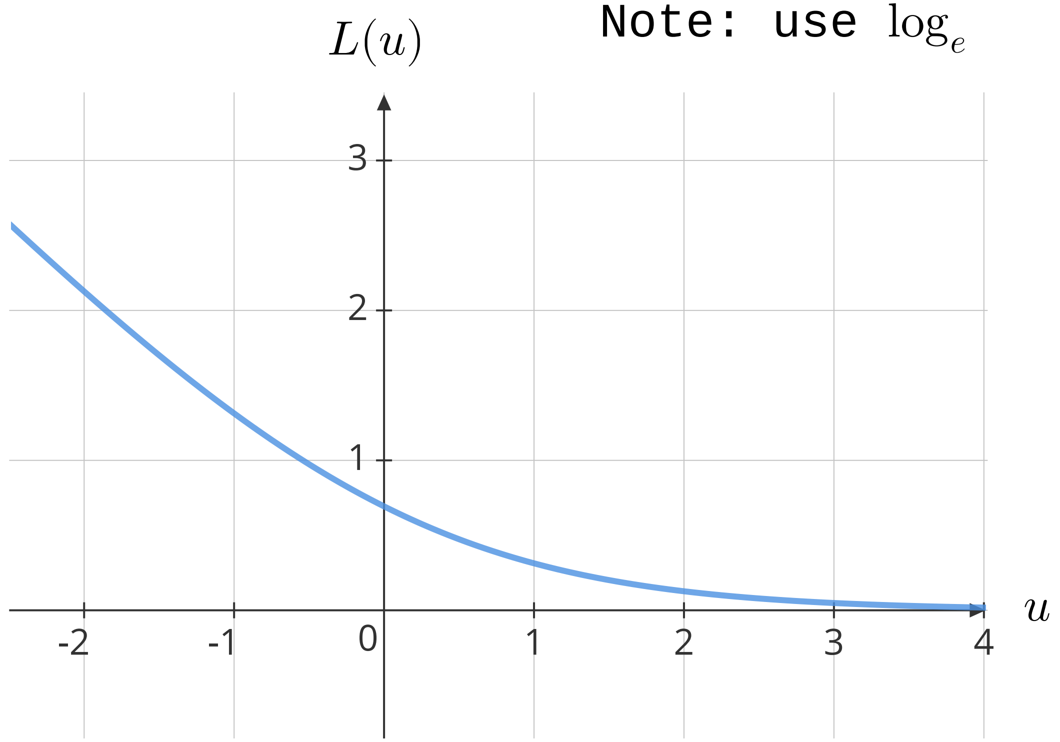
$$L(\mathbf{x}, z, \mathbf{w}) = \begin{cases} -\log\left[\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 1 \\ -\log\left[1-\sigma\left(\mathbf{w}^T\mathbf{x}\right)\right], & z = 0 \end{cases}$$

$$= \begin{cases} \log\left(1+e^{-\mathbf{w}^T\mathbf{x}}\right), & z = 1 \\ \log\left(1+e^{\mathbf{w}^T\mathbf{x}}\right), & z = 0 \end{cases}$$

$$L(\mathbf{x}, y, \mathbf{w}) = \log\left(1+e^{-\left(\mathbf{w}^T\mathbf{x}\right)y}\right)$$

$$L(u) = \log\left(1+e^{-u}\right)$$

Note: use $\log_e$



Modified Hinge Loss

(Perceptron)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T \mathbf{x}\right) y$$

$$L(\mathbf{x}, y, \mathbf{w}) =$$

Modified Hinge Loss

(Perceptron)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \begin{cases} 0, & \left(\mathbf{w}^T\mathbf{x}\right)y \geqslant 0 \\ -\left(\mathbf{w}^T\mathbf{x}\right)y, & \left(\mathbf{w}^T\mathbf{x}\right)y < 0 \end{cases}$$

Modified Hinge Loss

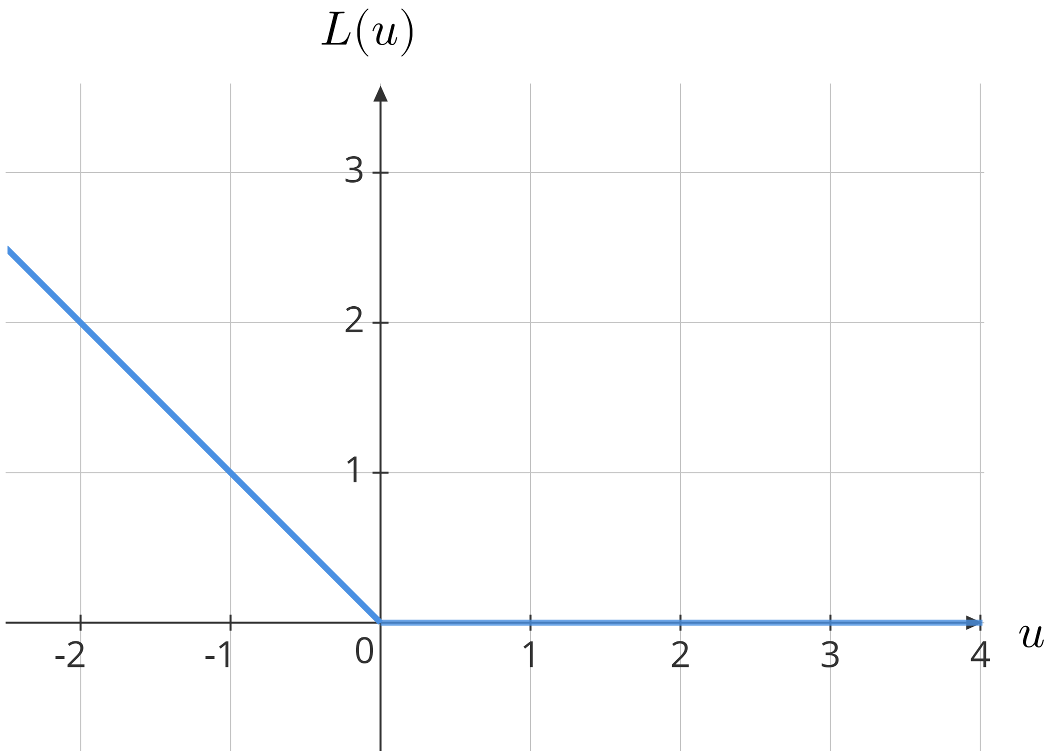
(Perceptron)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \begin{cases} 0, & \left(\mathbf{w}^T\mathbf{x}\right)y \geqslant 0 \\ -\left(\mathbf{w}^T\mathbf{x}\right)y, & \left(\mathbf{w}^T\mathbf{x}\right)y < 0 \end{cases}$$

$$L(u) = \max(0, -u)$$



Modified Hinge Loss

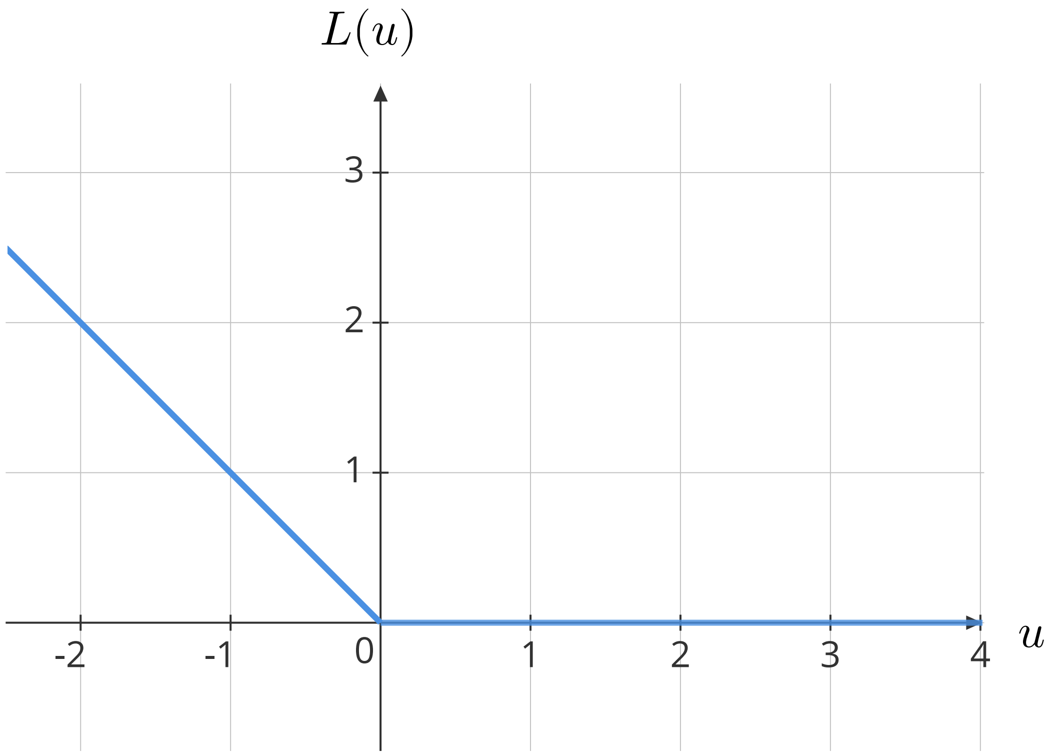
(Perceptron)

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

$$L(\mathbf{x}, y, \mathbf{w}) = \begin{cases} 0, & \left(\mathbf{w}^T\mathbf{x}\right)y \geqslant 0 \\ -\left(\mathbf{w}^T\mathbf{x}\right)y, & \left(\mathbf{w}^T\mathbf{x}\right)y < 0 \end{cases}$$

$$L(u) = \max(0, -u)$$



Perceptron update rule is equivalent to SGD:

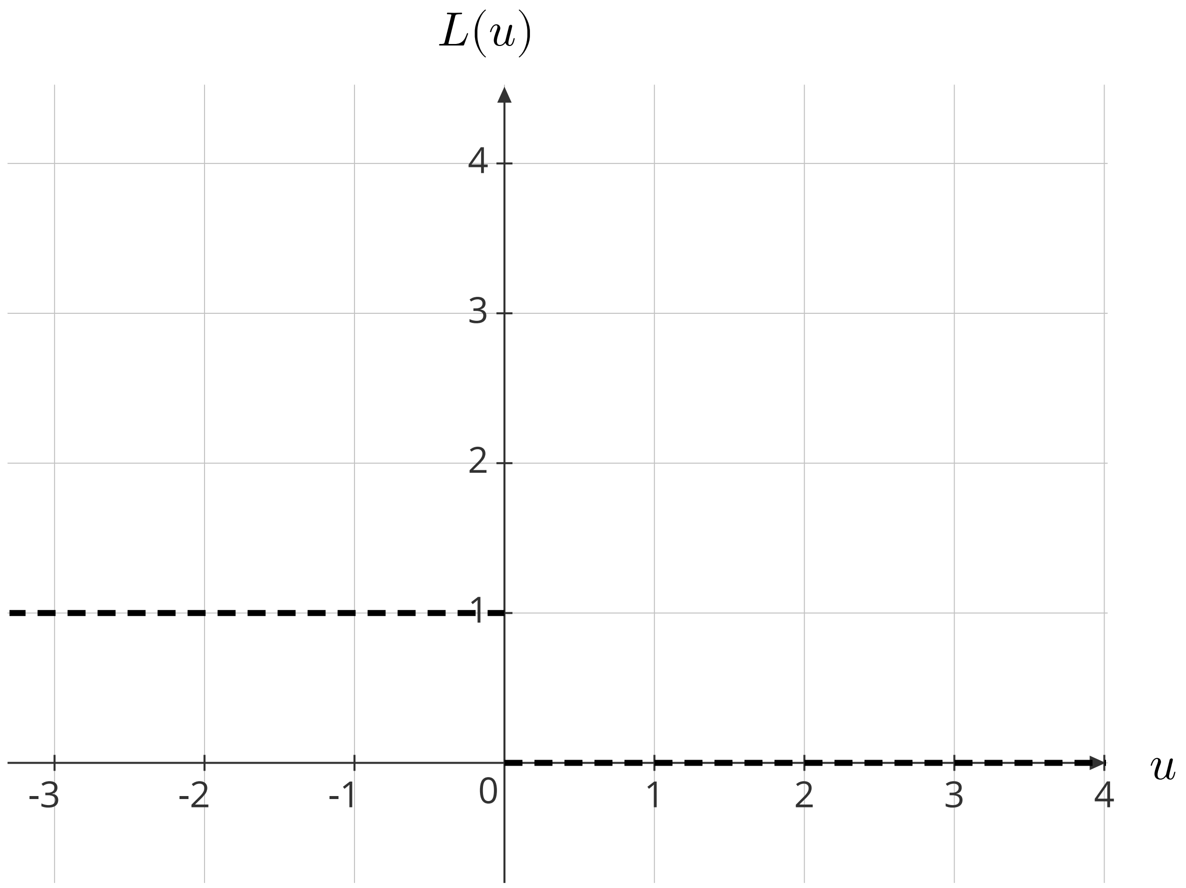
- On modified hinge loss
- with unit batch size
- and unit learning rate

Convex Surrogates

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = \left(\mathbf{w}^T\mathbf{x}\right)y$$

Loss	$L(u)$	Classifier	Label

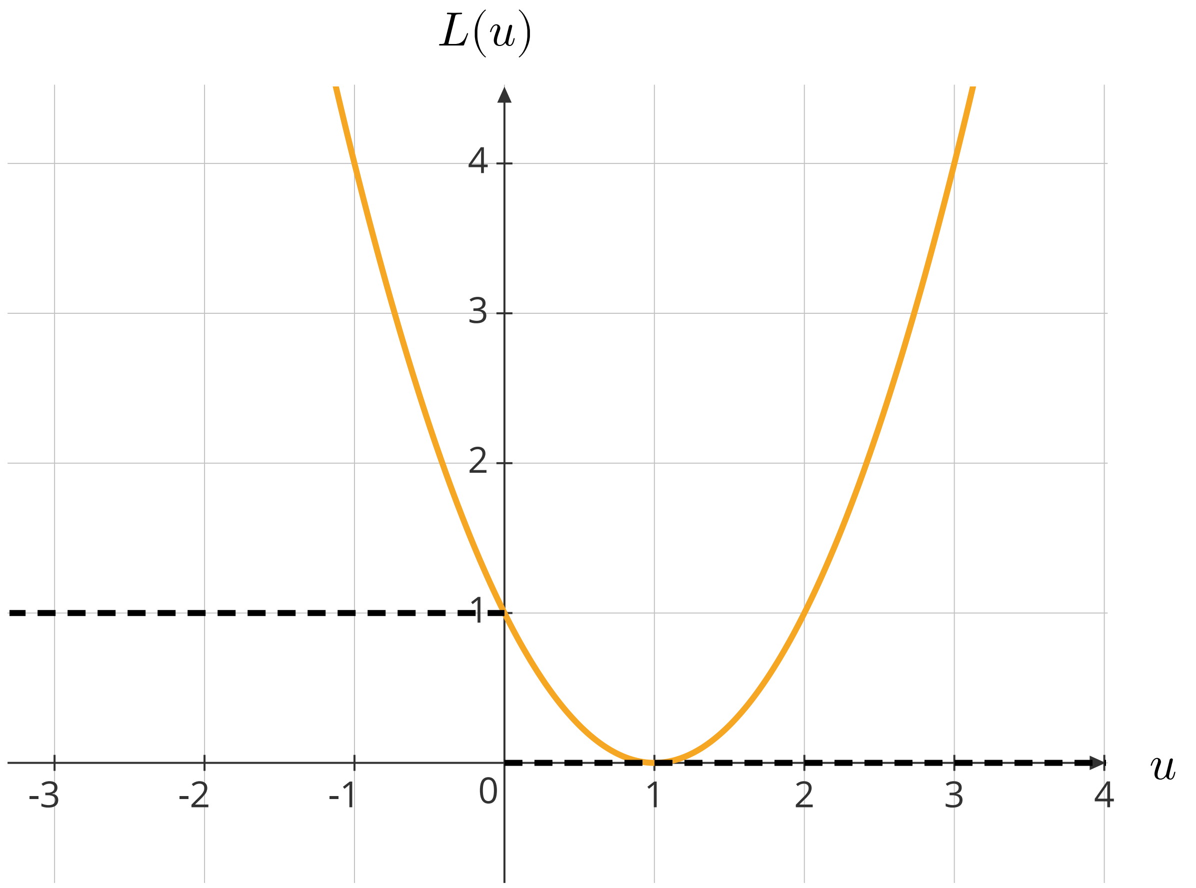


Convex Surrogates

$$D = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}, \; y_i \in \{-1, \; 1\}$$

$$u = (\mathbf{w}^T \mathbf{x}) y$$

Loss	$L(u)$	Classifier	Label
Squared loss	$(u-1)^2$	Least squares classifer	

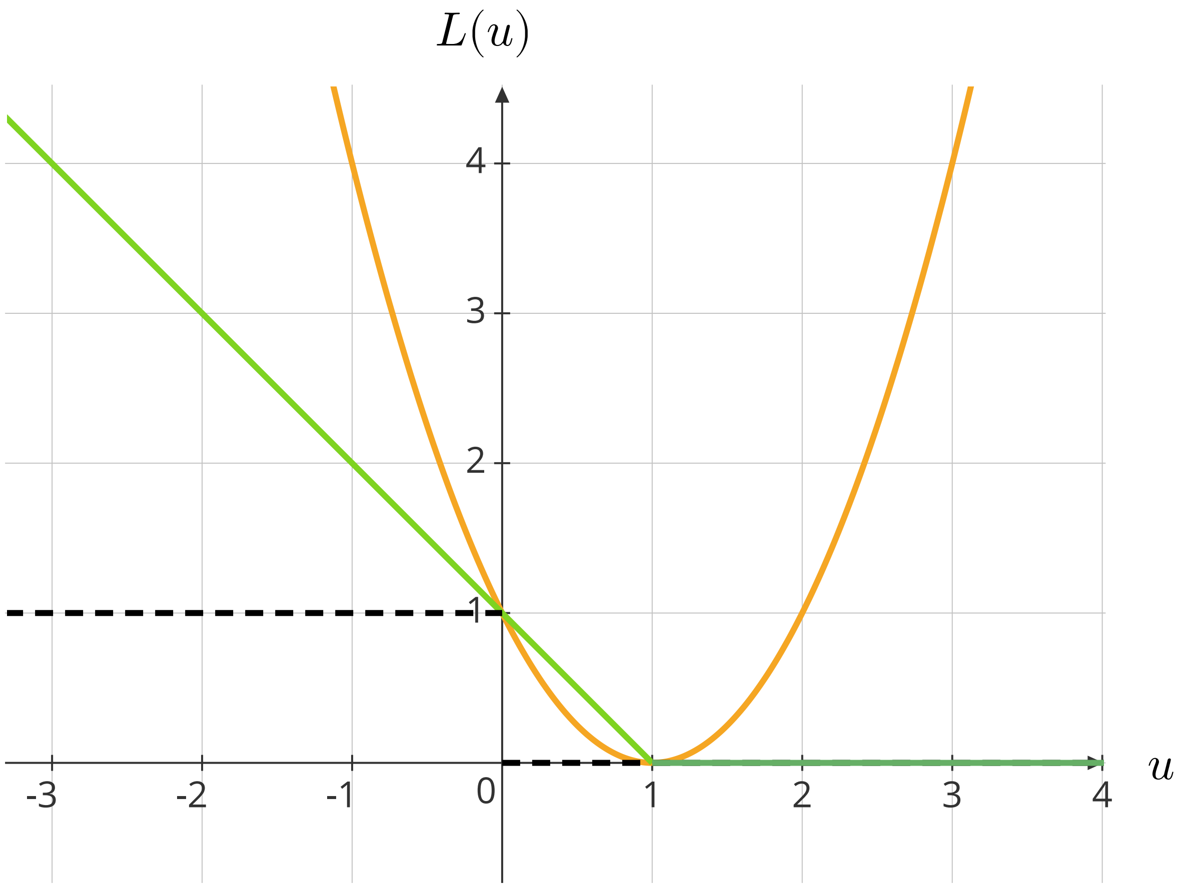


Convex Surrogates

$$D = \{(\mathbf{x}_1, y_1),\ \cdots, (\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = (\mathbf{w}^T \mathbf{x})y$$

Loss	$L(u)$	Classifier	Label
Squared loss	$(u-1)^2$	Least squares classifer	
Hinge loss	$\max(0, 1-u)$	SVM	

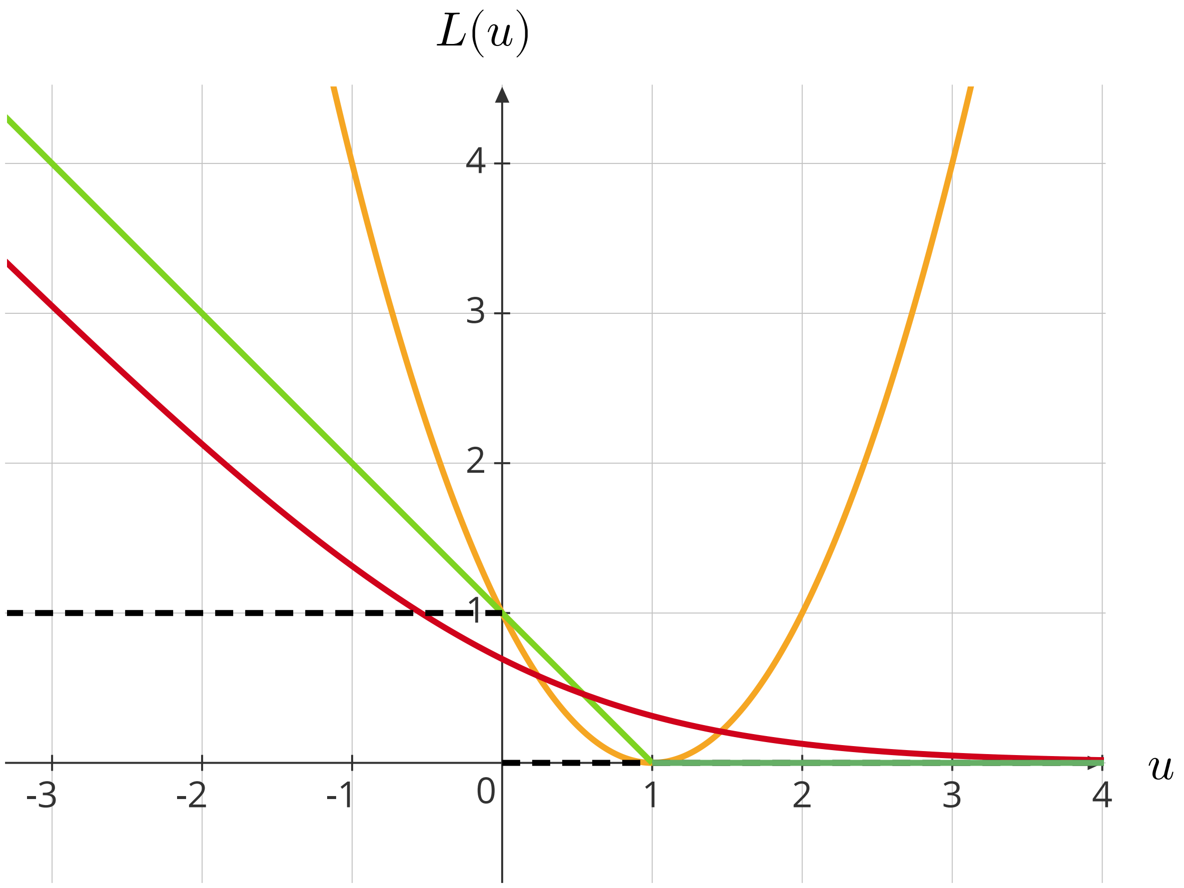


Convex Surrogates

$$D = \{(\mathbf{x}_1, y_1),\ \cdots,(\mathbf{x}_n, y_n)\},\ y_i \in \{-1,\ 1\}$$

$$u = (\mathbf{w}^T\mathbf{x})y$$

Loss	$L(u)$	Classifier	Label
Squared loss	$(u-1)^2$	Least squares classifer	
Hinge loss	$\max(0, 1-u)$	SVM	
Logistic loss	$\log(1+e^{-u})$	Logistic Regression	



Convex Surrogates

$$D = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}, \; y_i \in \{-1, \; 1\}$$

$$u = \left(\mathbf{w}^T \mathbf{x}\right) y$$

Loss	$L(u)$	Classifier	Label
Squared loss	$(u-1)^2$	Least squares classifer	
Hinge loss	$\max(0, 1-u)$	SVM	
Logistic loss	$\log\left(1+e^{-u}\right)$	Logistic Regression	
Modified hinge loss	$\max(0, -u)$	Perceptron