CMSE 820: Mathematical Foundations of Data Science

Lecture 09

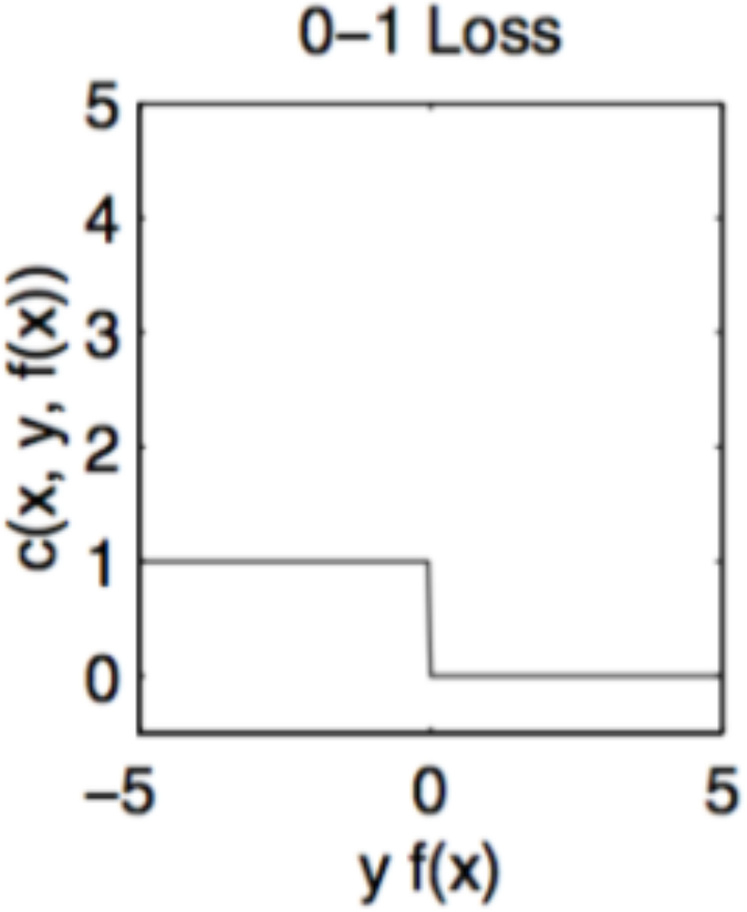
Loss functions for binary classification

- $\mathcal{Y} = \{+1, -1\}$
- Assume that our machine learning algorithm constructs a function

$$f : \mathcal{X} \to \mathcal{Y}$$

0-1 loss

$$c(x, y, f(x)) = \begin{cases} 0 & \text{if } y = f(x), \\ 1 & \text{if } y \neq f(x). \end{cases}$$



Loss functions for binary classification

- Now suppose instead our machine learned algorithm outputs a function

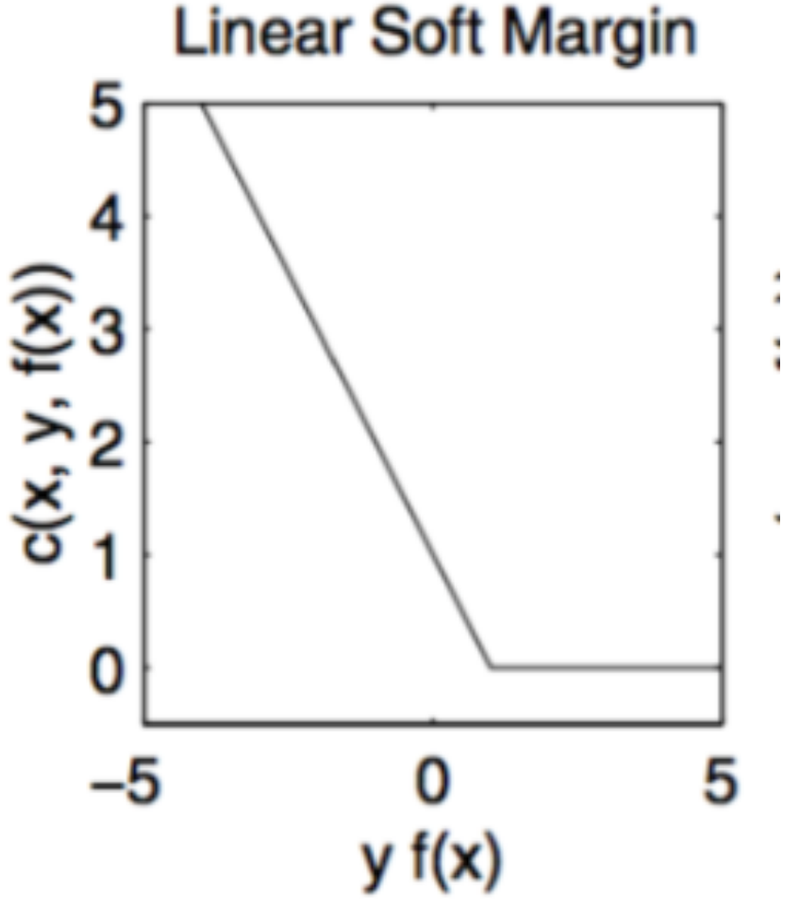
$$f : \mathcal{X} \to \mathbb{R}$$

  which gives the confidence of our prediction.

- $\operatorname{sgn} f(x)$ is the class label
- $|f(x)|$ is our confidence in the prediction

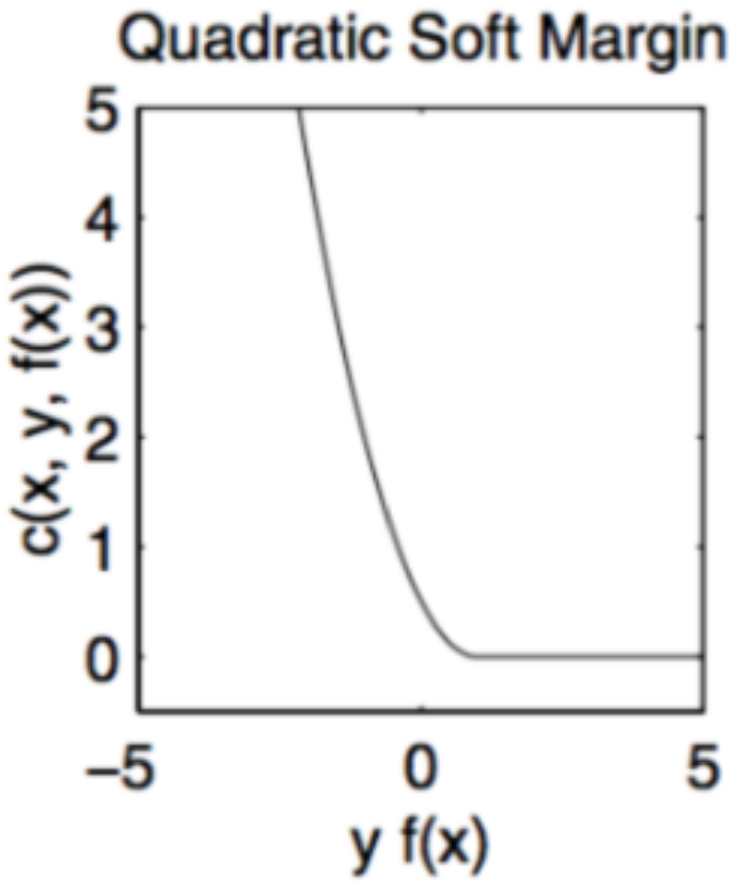
Soft margin loss

$$c(x, y, f(x)) = \max(0, 1 - yf(x)) = \begin{cases} 0 & \text{if } yf(x) \geq 1, \\ 1 - yf(x) & \text{if } yf(x) < 1. \end{cases}$$



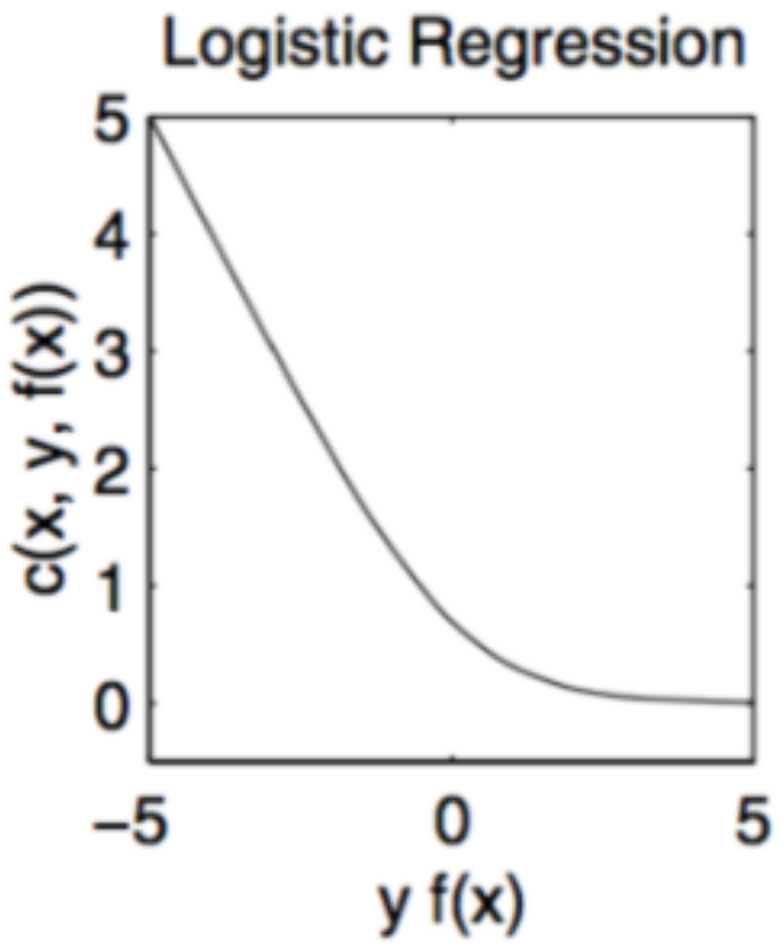
Quadratic soft margin loss

$$c(x, y, f(x)) = \max(0, 1 - yf(x))^2$$



Logistic loss

$$c(x, y, f(x)) = \log(1 + \exp(-yf(x))).$$



Loss functions for regression

- Now consider regression in which $\mathcal{Y} \subseteq \mathbb{R}$
- Our machine learned algorithm outputs a function

$$f : \mathcal{X} \to \mathbb{R}$$

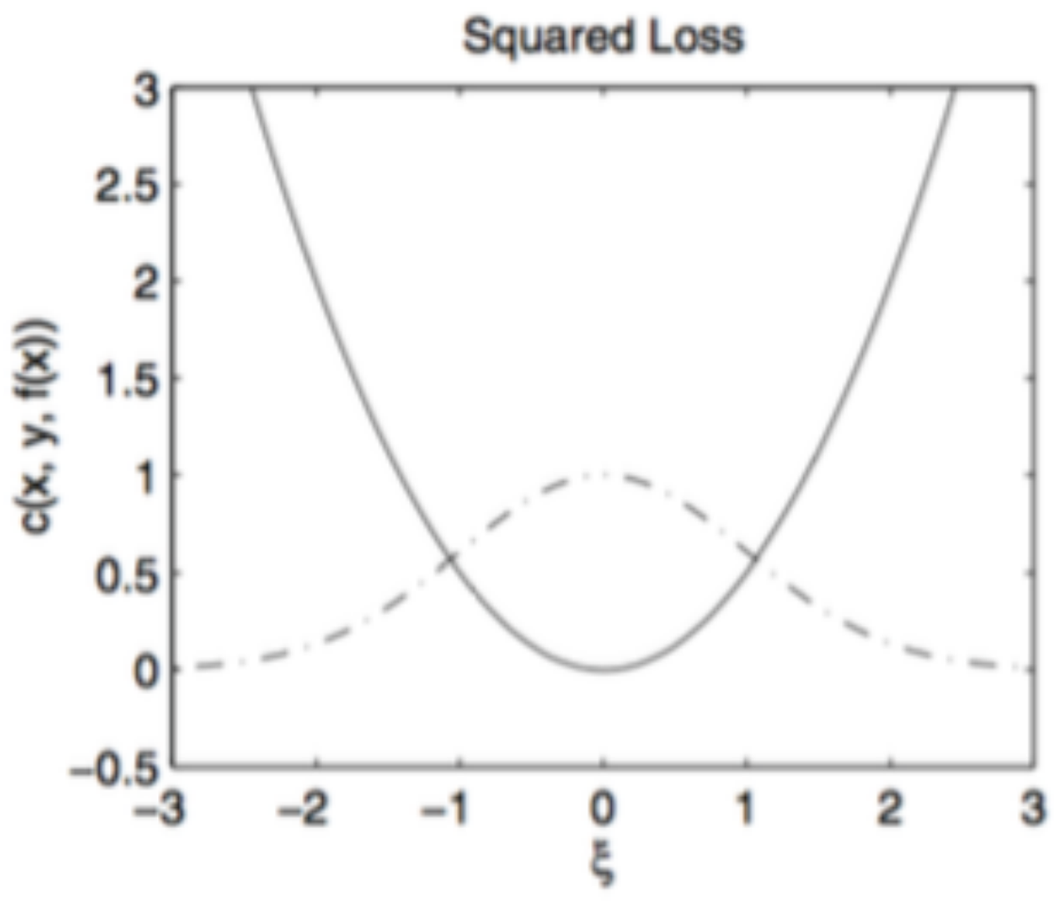
  which estimates the label y.

- In most cases we can't get the label exactly right
- So our loss functions generally will measure the error as a function of

$$y - f(x)$$

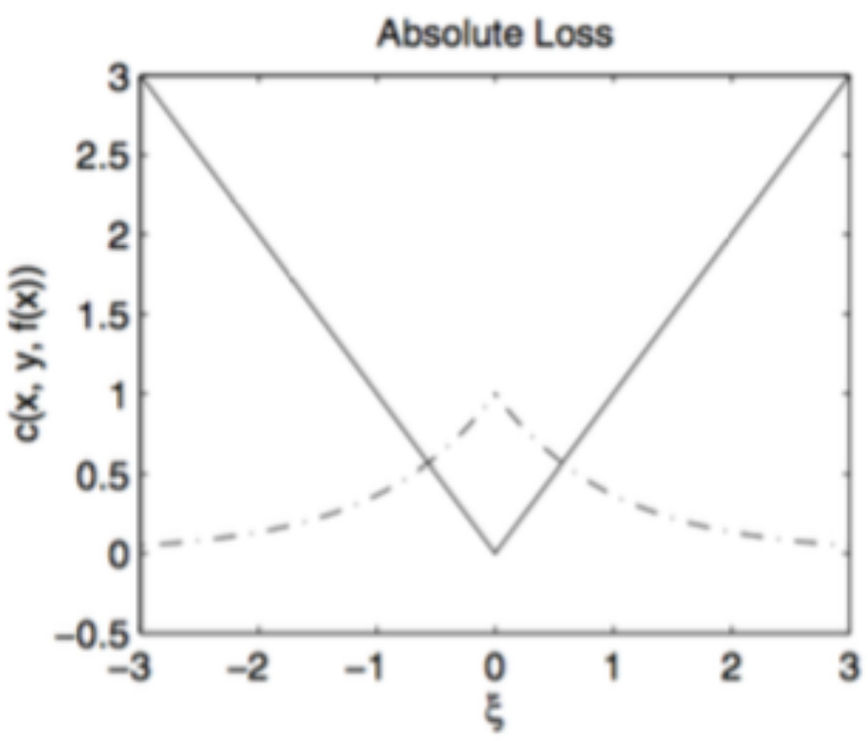
Squared loss

$$c(x, y, f(x)) = (y - f(x))^2$$

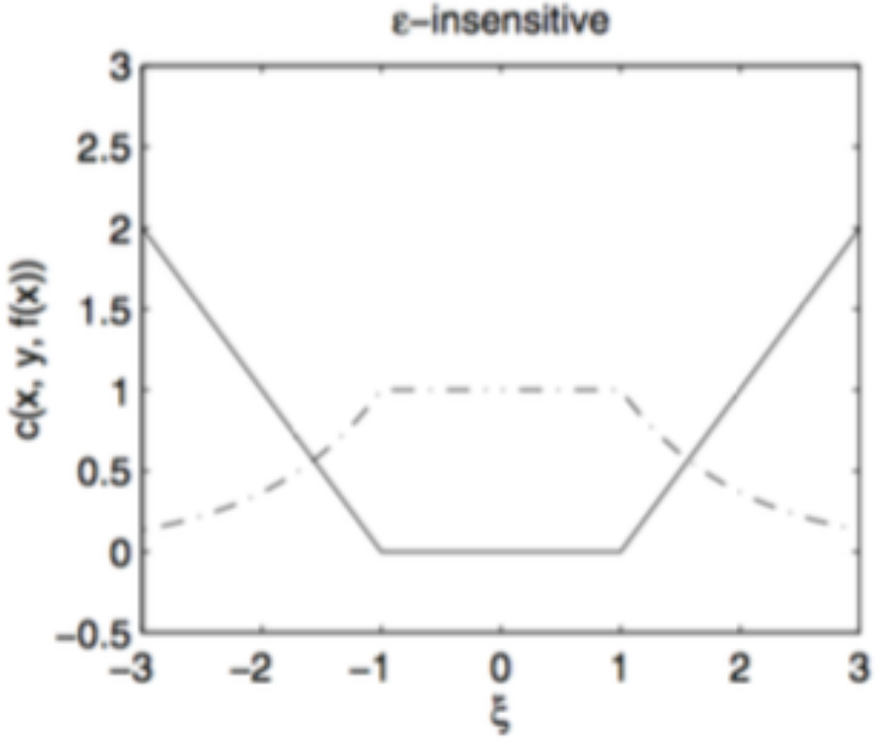


Epsilon sensitive loss

$$c(x, y, f(x)) = \max(|y - f(x)| - \epsilon, 0)$$



$\varepsilon = 0$



$\varepsilon > 0$