CSC 461: Machine Learning

Fall 2024

Logistic regression (part I)

Prof. Marco Alvarez, Computer Science
University of Rhode Island

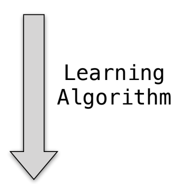
Linear classifiers

Binary classification

- Goal
  - learn a decision boundary between two classes

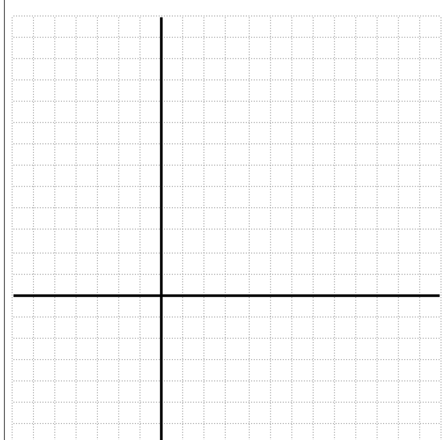
X_1	…	X_d	Y
0.5	…	0.1	+1
0.3	…	0.9	−1
0.3	…	0.875	−1
0.45	…	0.15	+1
…	…	…	…

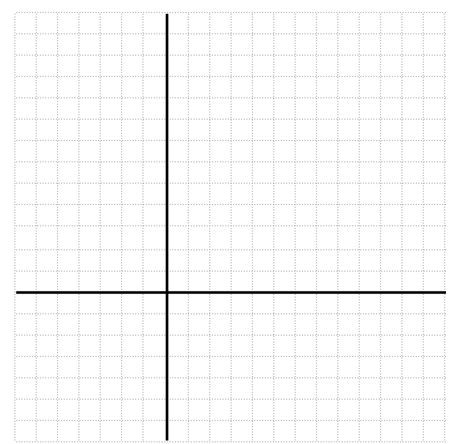
$X, \mathbf{y}$



$h(\mathbf{x})$

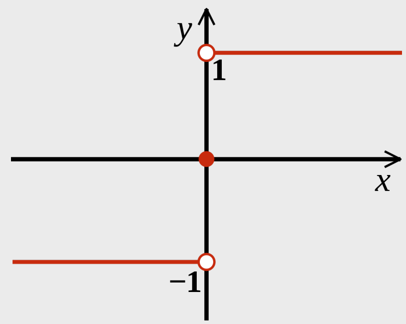
Plots (regression x classification)





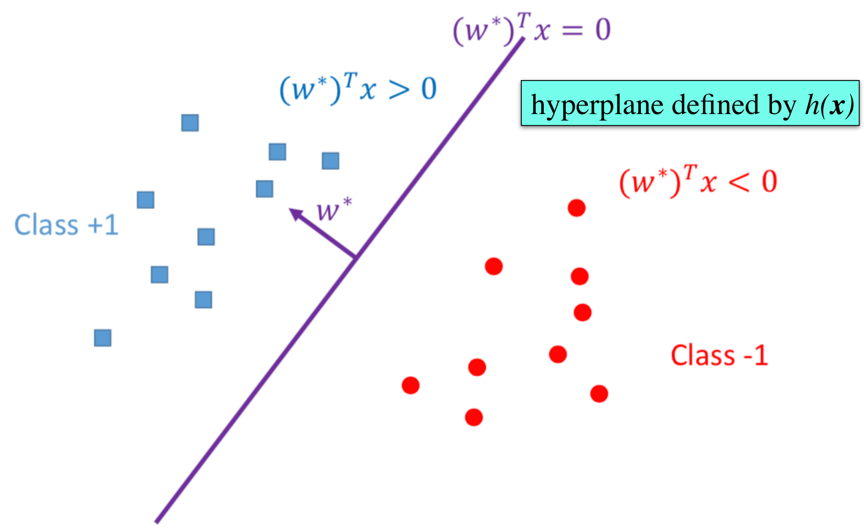
The *sign* function

$$sign(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ +1 & \text{if } x > 0 \end{cases}$$



$$h(\mathbf{x}) = sign\left(\mathbf{w}^T\mathbf{x}\right)$$

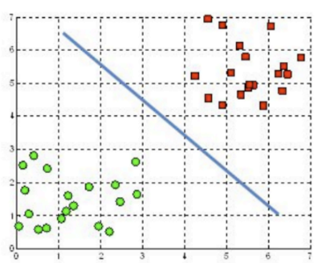
Decision boundary



credit: yingyu liang, cos 495, princeton

Decision boundary

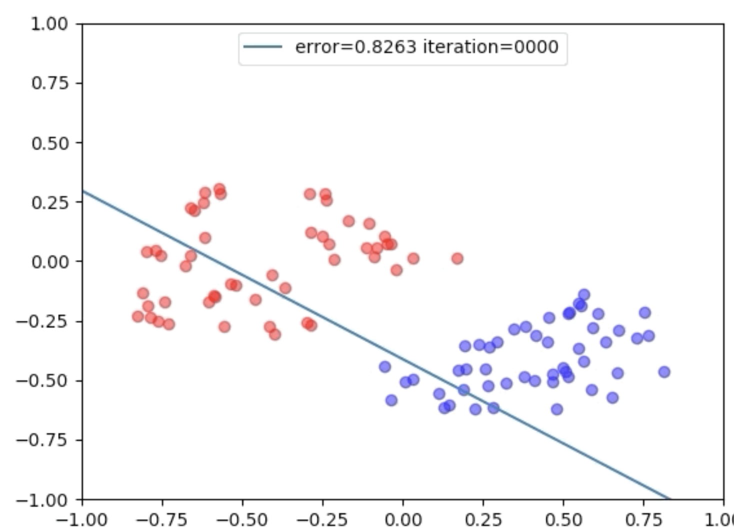
A hyperplane in $\mathbb{R}^2$ is a line



$$0 = b + w_1 x_1 + w_2 x_2$$

Image credit: https://mc.ai/why-activation-function-is-used-in-neural-network/

Learning



Example

- Provide a solution (weight vector)

x_0	x_1	x_2	y
1	0	0	-1
1	0	1	-1
1	1	0	-1
1	1	1	$+1$

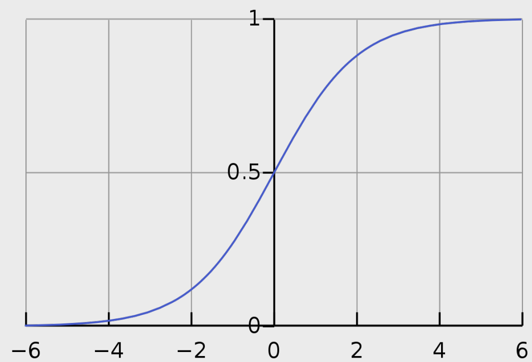
Logistic regression

Logistic regression

- Binary classifier
  - $\mathscr{D} = \left\{(\mathbf{x}^{(1)}, y^{(1)}), \ldots, (\mathbf{x}^{(n)}, y^{(n)})\right\}$
  - $\mathbf{x}^{(i)} \in \mathbb{R}^d, \quad y^{(i)} \in \{-1, +1\}$
  - uses a **logistic function**
  - models **probability** of output given input
- It is considered a **linear classifier**
  - with a non-linear *activation function*

Logistic function

$$\sigma(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$



mapping $\mathbb{R}$ to [0,1]

continuous and differentiable

Probabilistic interpretation

$$h(\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x}) = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}$$

Probability of class +1 $\quad P(y=+1 \mid \mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x})$

Probability of class -1 $\quad P(y=-1 \mid \mathbf{x}) = 1 - P(y=+1 \mid \mathbf{x})$

$P(y=-1 \mid \mathbf{x}) = \sigma(-\mathbf{w}^T\mathbf{x})$ (show)

$$P(y \mid \mathbf{x}) = \frac{1}{1+e^{-y\mathbf{w}^T\mathbf{x}}} = \sigma(y\mathbf{w}^T\mathbf{x})$$

Another look