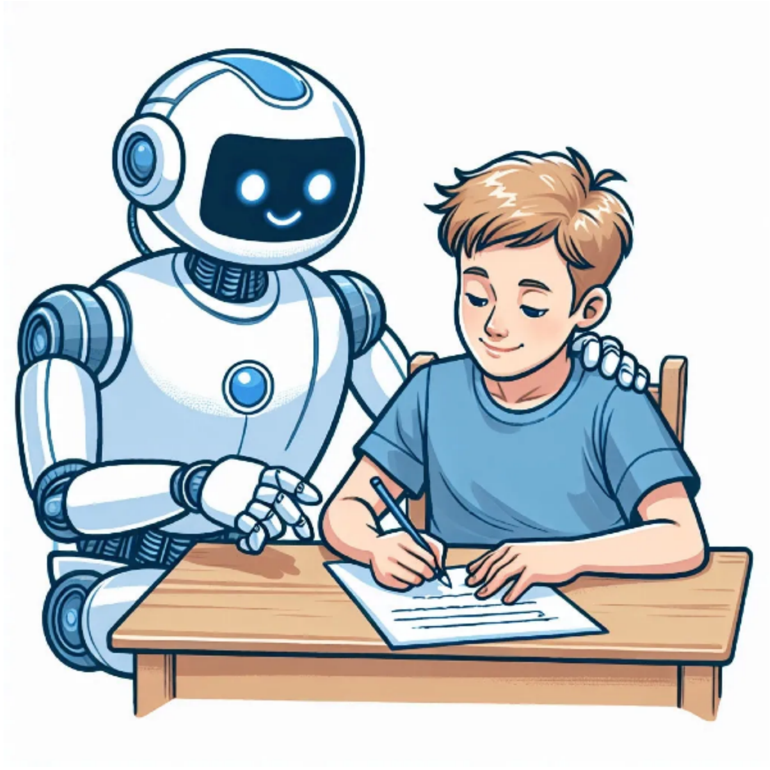
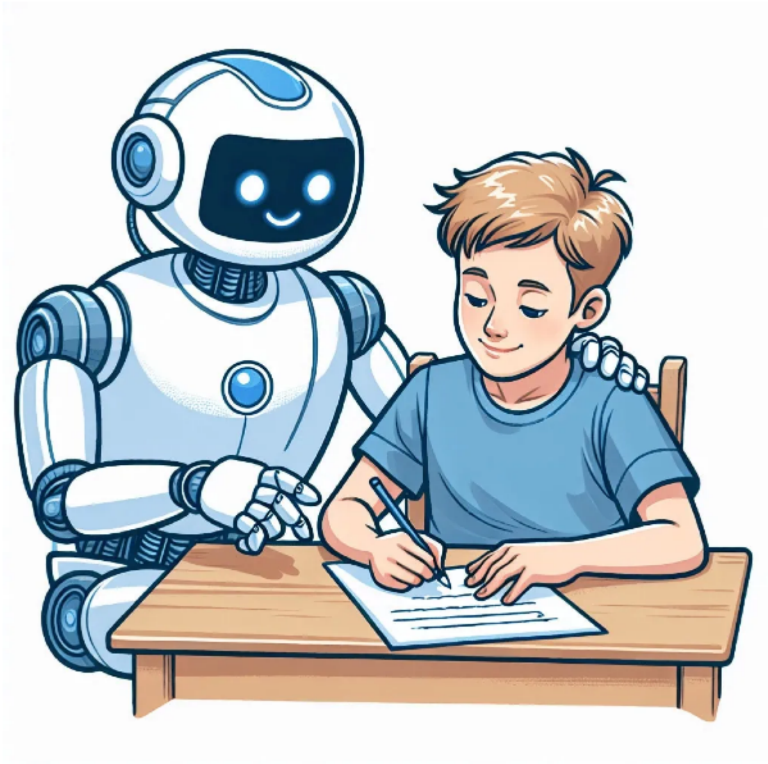
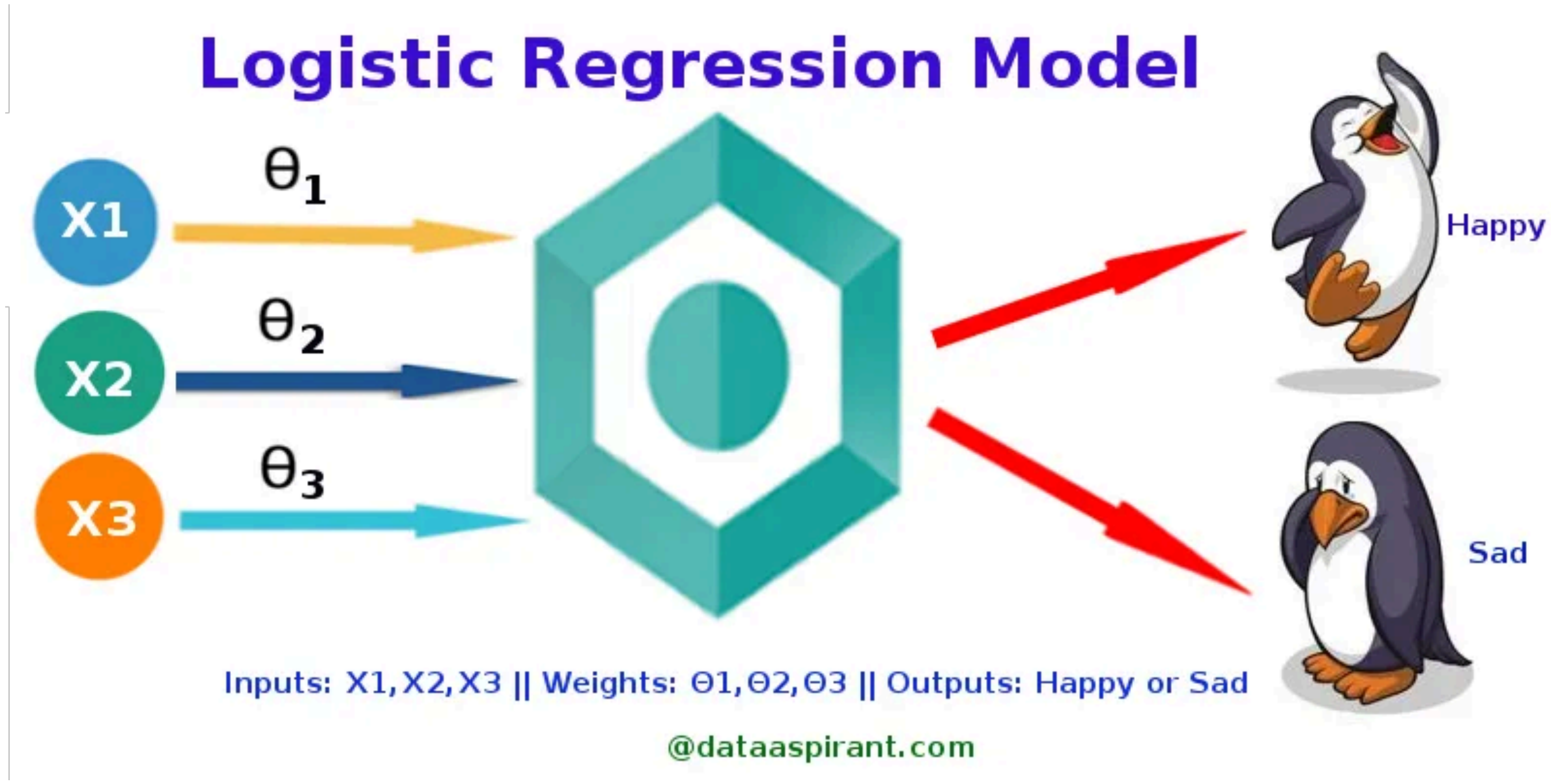
Content



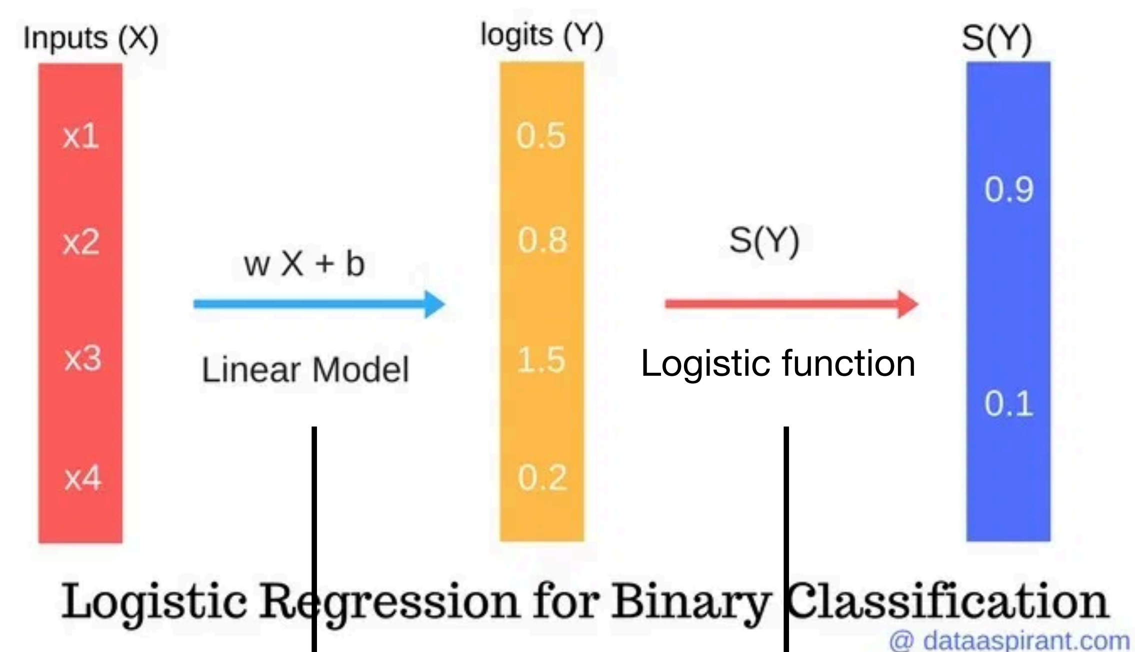
Content





Used when the output is categorical;
Classification;
Bounded output;
……

N	Penguin Activity	Penguin Activity Description	How Penguin felt (Target)
1	X1	Eating squids	Happy
2	X2	Eating small Fishes	Happy
3	X3	Hit by other Penguin	Sad
4	X4	Eating Crabs	Sad



$$\theta^T \cdot x$$

logit

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T \cdot x}}$$

hypothesis/activation/sigmoid

Logistic Regression — Detailed Overview

https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc

Logistic Regression

$$\{(x_j^{(i)}, y^{(i)}), \theta_j\};\ j = 1, 2, \cdots, N;\ i = 1, 2, \cdots, M; N < M$$

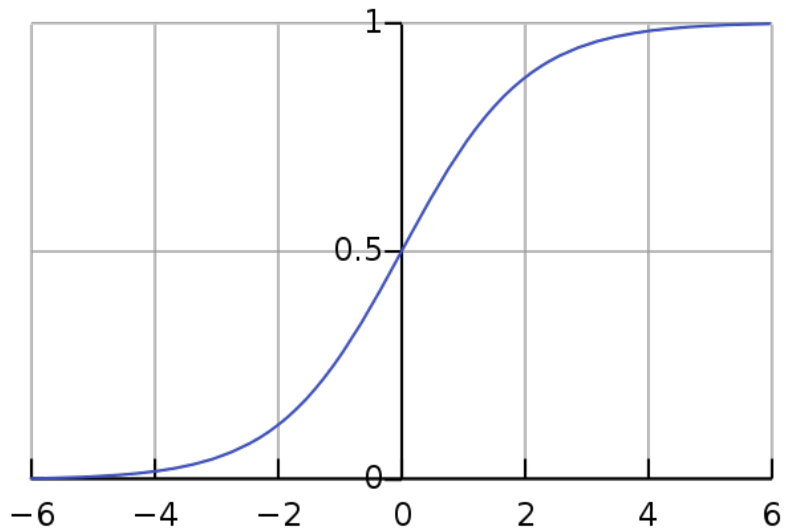
$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(M)} \end{bmatrix}$$

$$y = h_\theta(x) = \frac{1}{1 + e^{-\theta^T \cdot x}}$$

$$\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_N^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_N^{(2)} \\ & \vdots & & \vdots & \\ 1 & x_1^{(M)} & x_2^{(M)} & \cdots & x_N^{(M)} \end{bmatrix} \cdot \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_N \end{bmatrix}$$

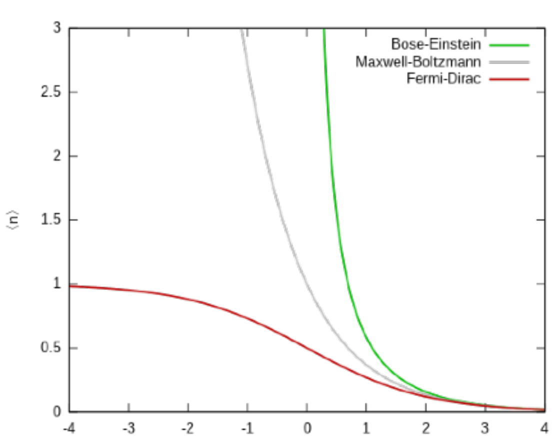
Discrete/binary Output

$$y^{(i)} = 0 \text{ or } 1$$



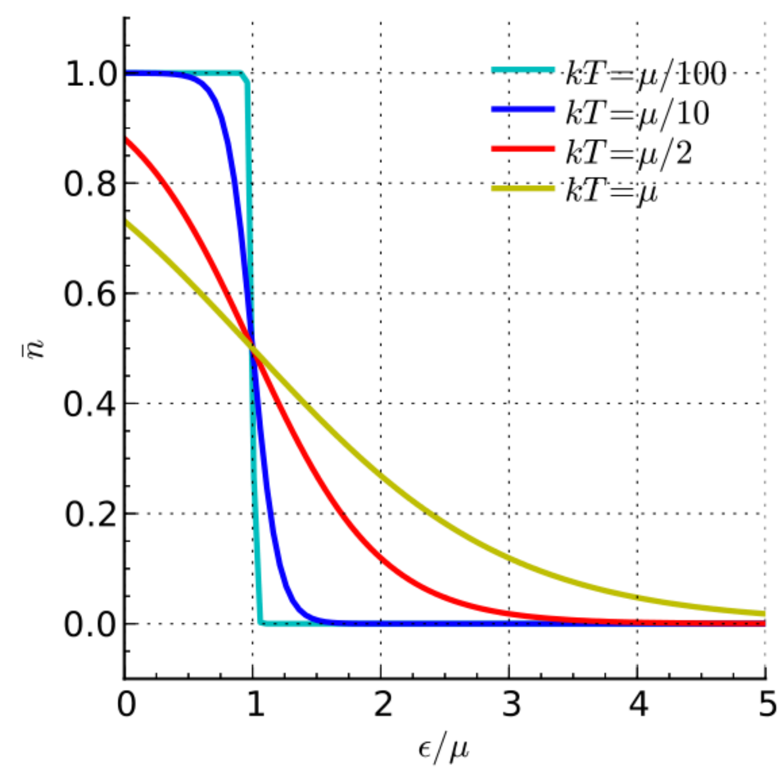
logistic unit (logit): $\theta^T \cdot x$

hypothesis/activation/sigmoid: $h_\theta(x) = \dfrac{1}{1 + e^{-\theta^T \cdot x}}$



Fermi-Dirac distribution

$$\bar{n}_i = \frac{1}{1 + e^{(\epsilon_i - \mu)/k_B T}}$$



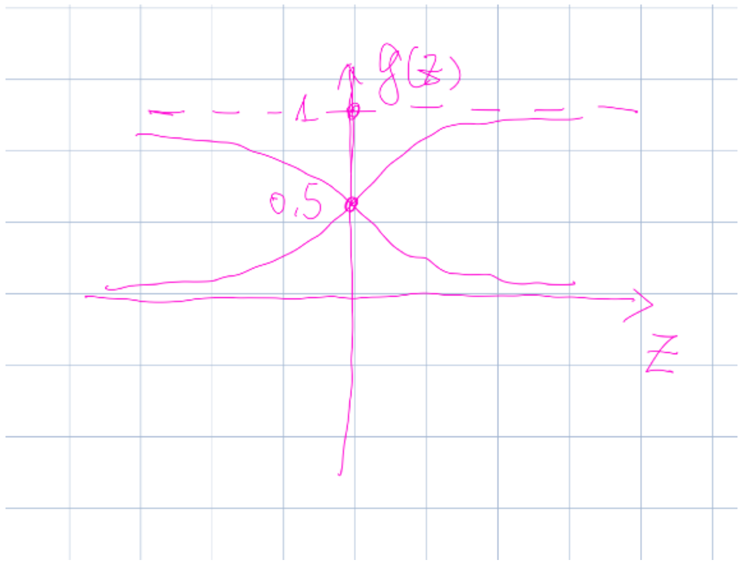
$$\bar{n}_i = \frac{g_i}{e^{(\epsilon_i - \mu)/k_B T} - 1}$$

Bose-Einstein distribution

Logistic Regression

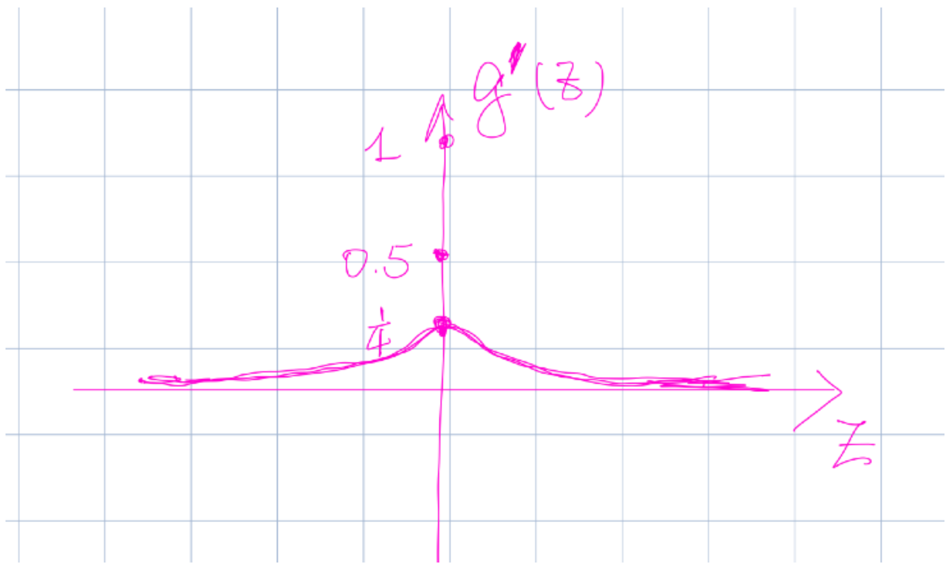
$$z = \theta^T \cdot x \qquad h_\theta(x) = \frac{1}{1+e^{-\theta^T \cdot x}} \rightarrow g(z) = \frac{1}{1+e^{-z}}$$

Several important properties of logistic function



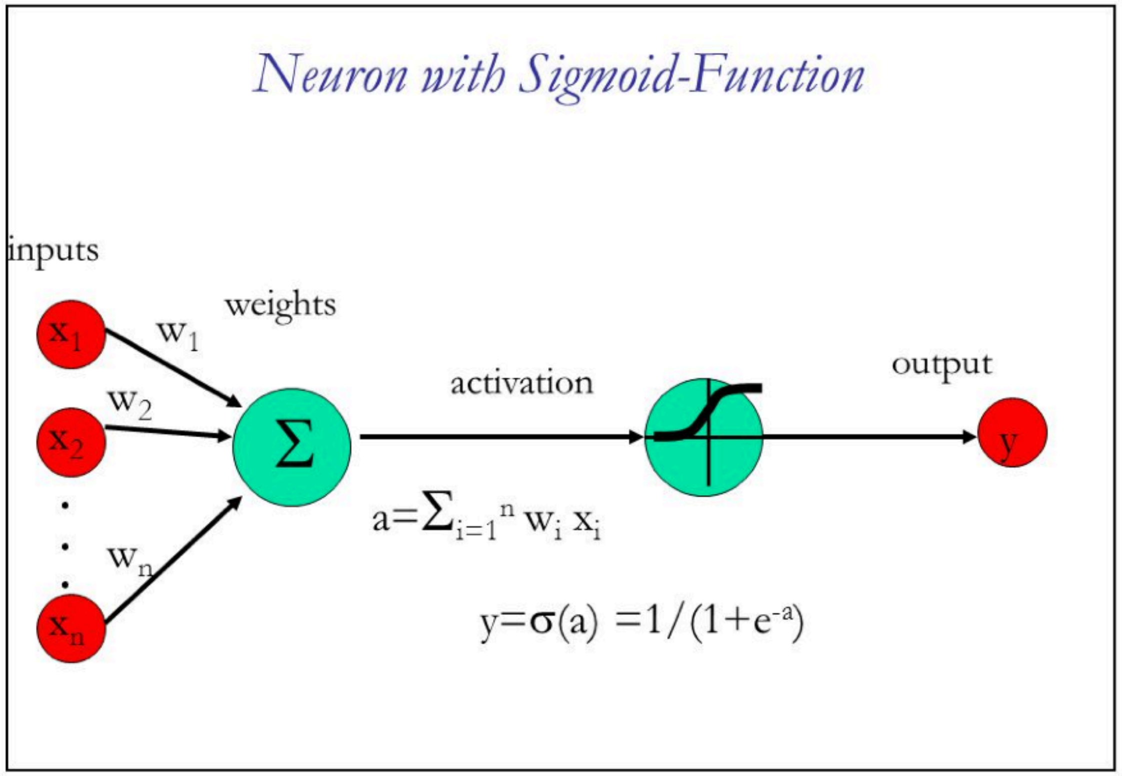
$$g(z) + g(-z) = 1$$

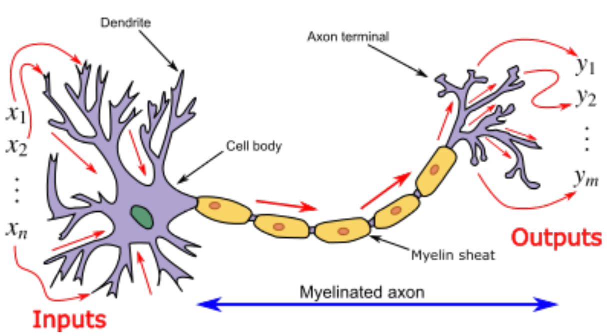
$$g'(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \cdots = g(z)g(-z)$$



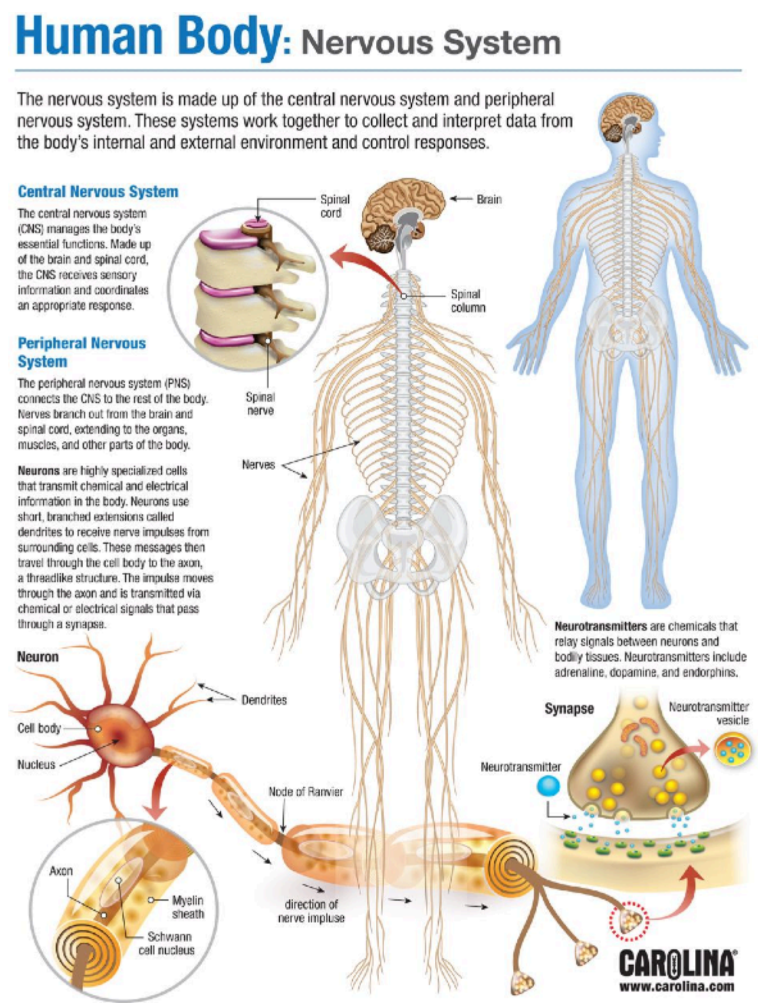
$$g'(z) = \cdots = g'(-z)$$

Perceptron

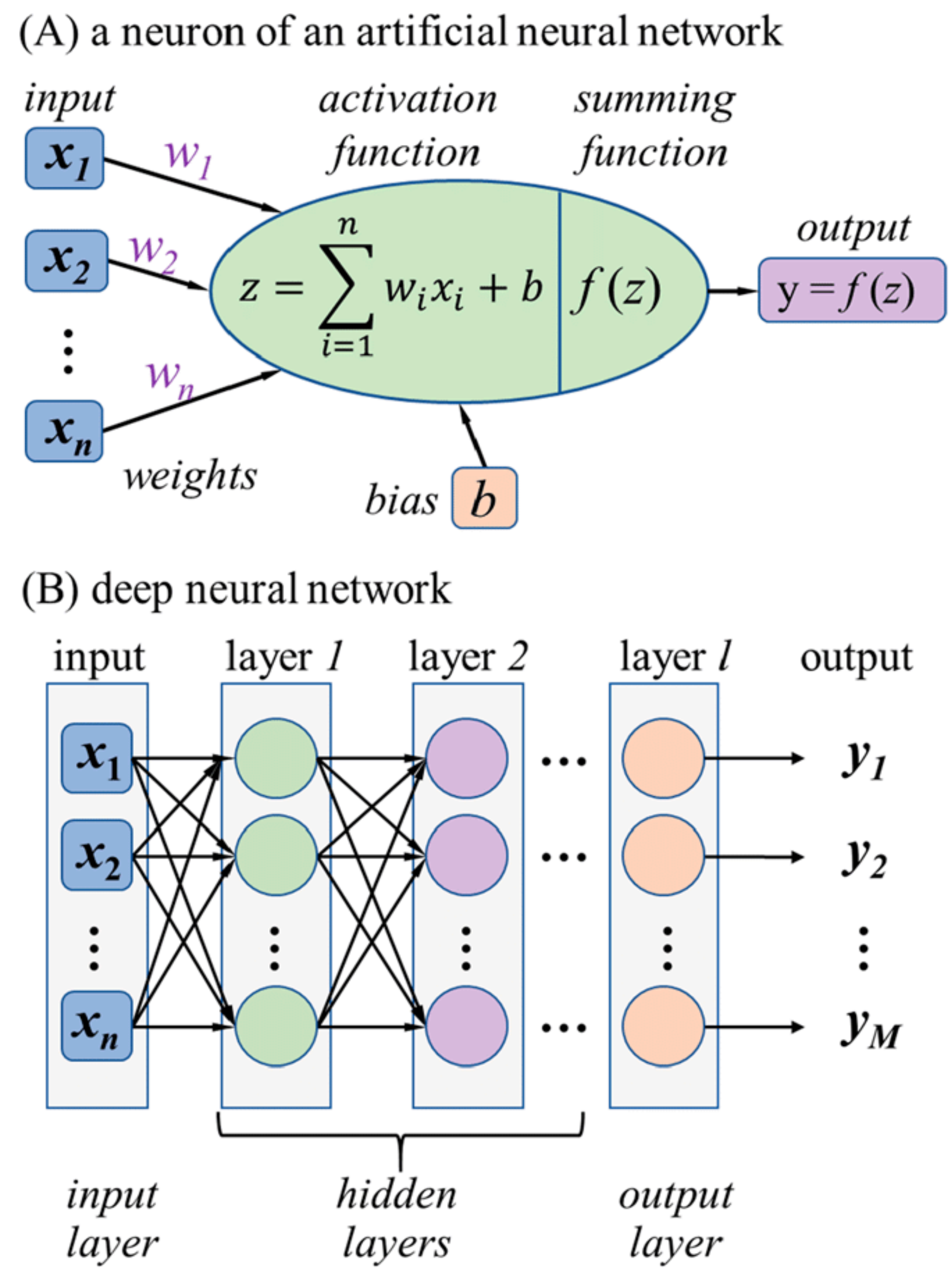




Γ



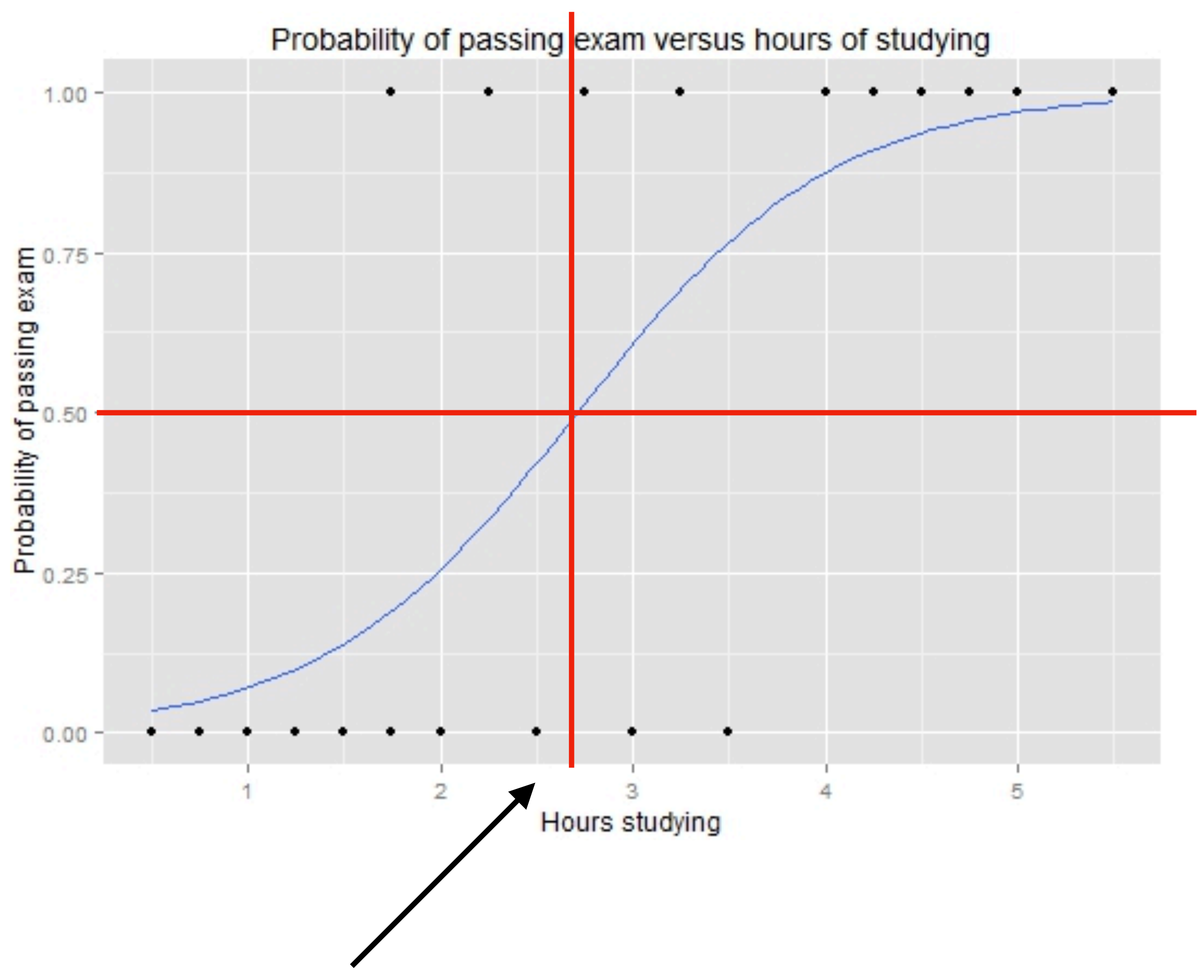
Neural network



Logistic Regression — it is very logical !

A group of students spends between 0 and 6 hours studying for an exam. How does the number of hours spent affect the probability of the student passing the exam?

Hours	0.50	0.75	1.00	1.25	1.50	1.75	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Pass	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1



Decision boundary

$$h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}} = \frac{1}{1+e^{-(-4.0777+1.5046\cdot \text{Hours})}}$$

$$h_\theta(x) = 0.5 \quad \text{happens at} \quad x = 2.71$$

$$\text{Probability of passing exam} = \frac{1}{1+\exp(-(1.5046\cdot 2 - 4.0777))} = 0.26$$

$$\text{Probability of passing exam} = \frac{1}{1+\exp(-(1.5046\cdot 4 - 4.0777))} = 0.87$$

https://en.wikipedia.org/wiki/Logistic_regression

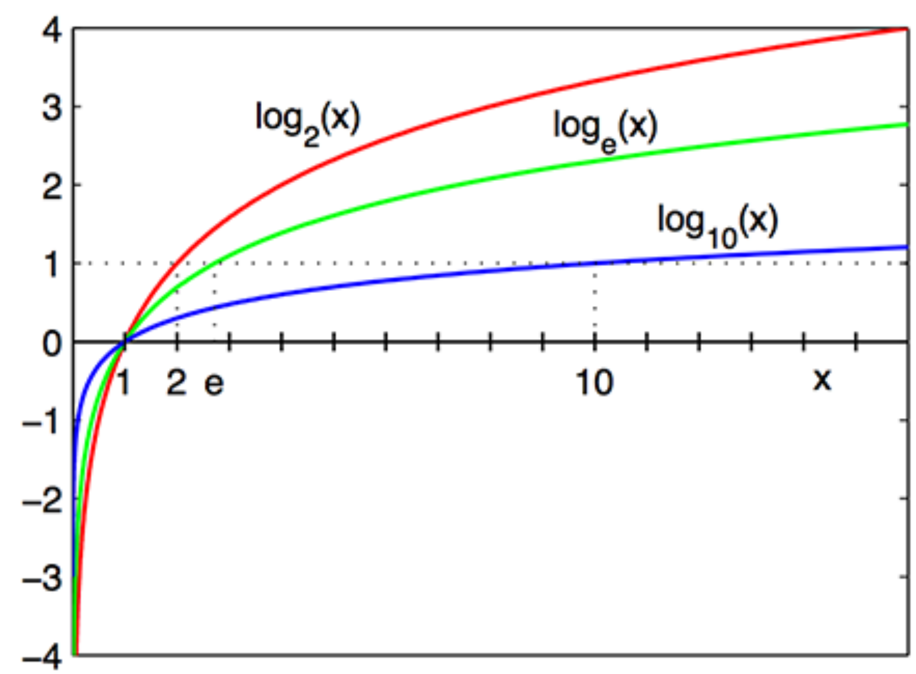
Logistic Regression — it is very logical !

Model in learning $$h_\theta(x) = \frac{1}{1 + e^{-\theta^T \cdot x}}$$

Likelihood function $$L(\theta | \cdots (x^{(i)}, y^{(i)}) \cdots) \stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{M} h_\theta(x^{(i)})^{y^{(i)}} (1 - h_\theta(x^{(i)}))^{1-y^{(i)}}$$

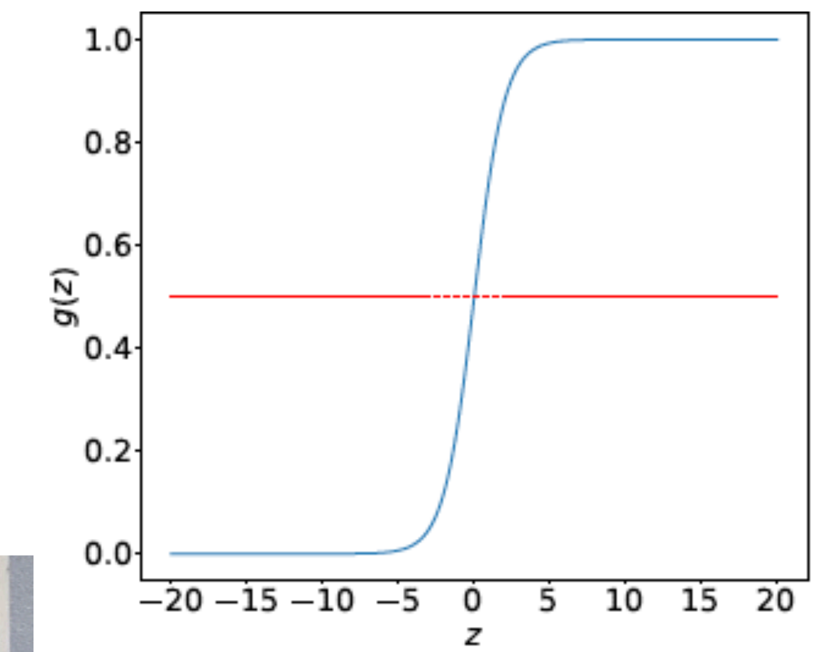
Cost function $$J(\theta) = -\frac{1}{M} \ln(L(\theta)) = -\frac{1}{M} \sum_{i=1}^{M} [y^{(i)} \ln(h_\theta(x^{(i)}) + (1 - y^{(i)}) \ln(1 - h_\theta(x^{(i)}))]$$

Optimization procedure $$\theta^* = \arg \min_\theta J(\theta)$$



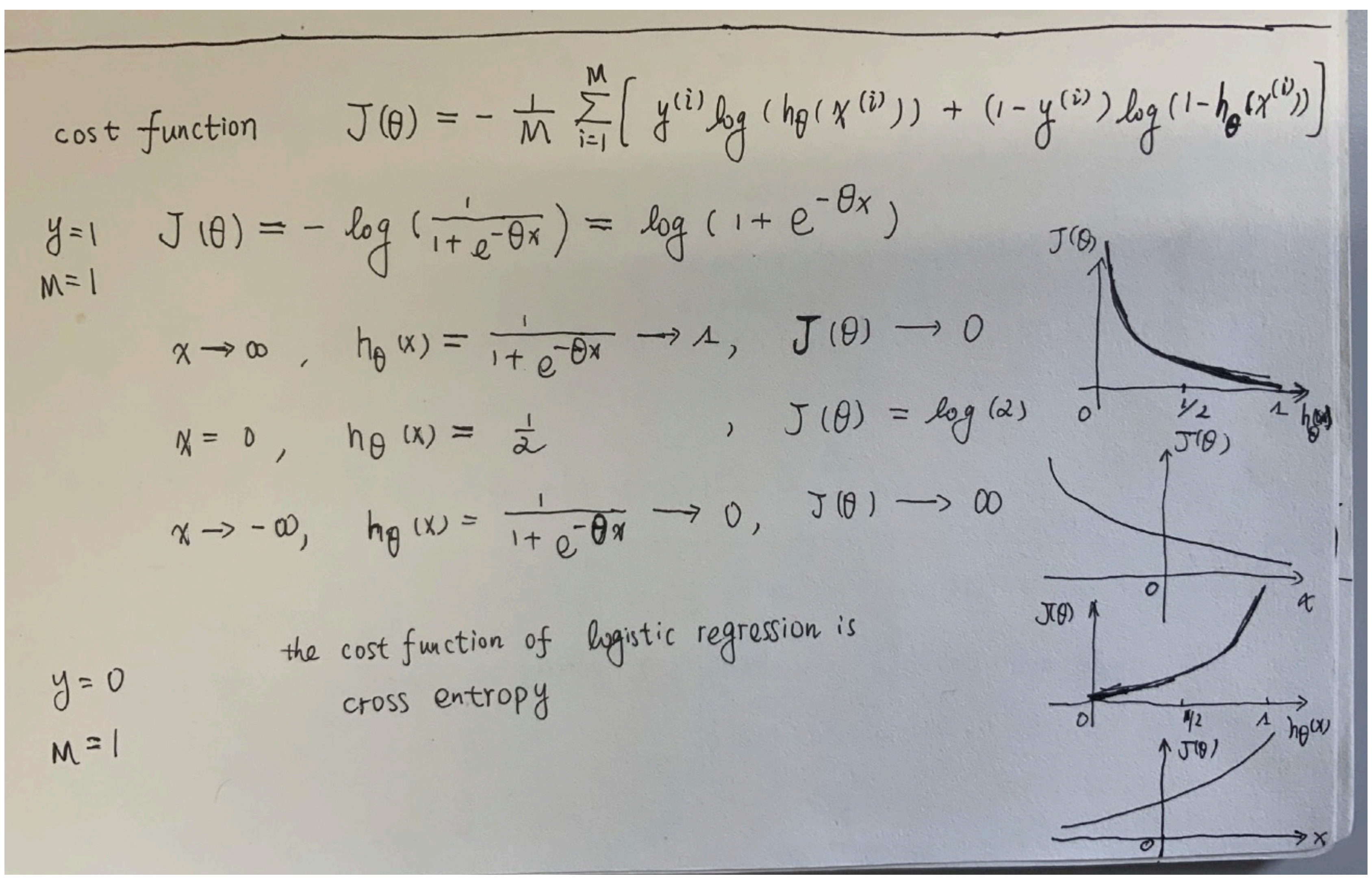
Logistic Regression — it is very logical !

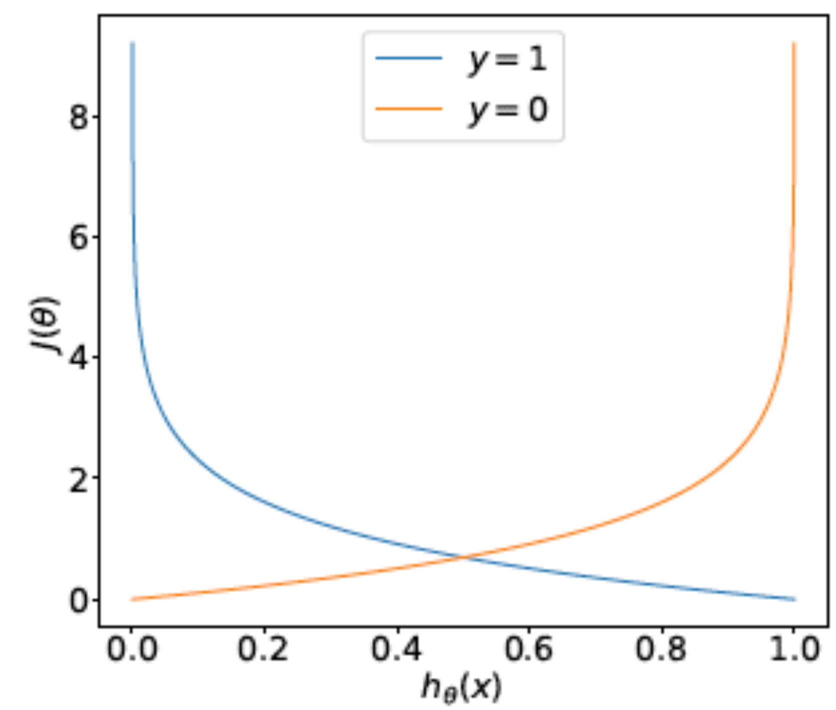
$$z = \theta^T \cdot x \qquad h_\theta(x) = \frac{1}{1+e^{-\theta^T \cdot x}} \rightarrow g(z) = \frac{1}{1+e^{-z}}$$



Several important properties of logistic function

1. $g(z) + g(-z) = 1$
2. $g'(z) = g(z)(1 - g(z)) = g(z)g(-z)$
3. $g'(-z) = g'(z)$





Logistic Regression — it is very logical !

Gradient descent
$$\theta := \theta - \alpha \nabla_\theta J(\theta)$$
$$= \theta - \alpha\big[- \langle xy \rangle + \frac{1}{M}\sum_{i=1}^{M} x^{(i)} h_\theta(x^{(i)})\big]$$

1. $g(z) + g(-z) = 1$
2. $g'(z) = g(z)(1 - g(z))$
3. $g'(-z) = g'(z)$

$$\nabla_\theta J(\theta) = -\frac{1}{M}\sum_{i=1}^{M}\Big[x^{(i)}y^{(i)}\frac{g'(\theta^T x^{(i)})}{g(\theta^T x^{(i)})} - x^{(i)}(1-y^{(i)})\frac{g'(-\theta^T x^{(i)})}{g(-\theta^T x^{(i)})}\Big]$$
$$= -\frac{1}{M}\sum_{i=1}^{M}\Big[x^{(i)}y^{(i)}g(-\theta^T x^{(i)}) - x^{(i)}(1-y^{(i)})g(\theta^T x^{(i)})\Big]$$
$$= -\frac{1}{M}\sum_{i=1}^{M}\Big[x^{(i)}y^{(i)} - x^{(i)}g(\theta^T x^{(i)})\Big]$$
$$= -\langle xy \rangle + \frac{1}{M}\sum_{i=1}^{M} x^{(i)} g(\theta^T x^{(i)})$$

the same holds for linear reg.

$$\nabla_\theta J(\theta) = \frac{1}{2M}\nabla_\theta \sum_{i=1}^{M}(\theta^T x^{(i)} - y^{(i)})^2$$
$$= \frac{1}{M}\sum_{i=1}^{M} x^{(i)}(\theta^T x^{(i)} - y^{(i)})$$
$$= -\frac{1}{M}\sum_{i=1}^{M} x^{(i)}y^{(i)} + \frac{1}{M}\sum_{i=1}^{M} x^{(i)}\theta^T x^{(i)}$$
$$= -\langle xy \rangle + \frac{1}{M}\sum_{i=1}^{M} x^{(i)} g(\theta^T x^{(i)})$$

Logistic Regression — it is very logical !

$$J(\theta) = -y \log(h_\theta(x)) - (1-y)\log(1 - h_\theta(x))$$

$$h_\theta = \frac{1}{1+e^{-z}}, \quad z = \theta_1 x_1 + \theta_2 x_2 + \theta_0$$

$$\frac{\partial J(\theta)}{\partial \theta_1} = -y \frac{1}{h_\theta(z)} h_\theta(z) h_\theta(-z) x_1 + (1-y) \frac{1}{1-h_\theta(z)} h_\theta(z) h_\theta(-z) x_1$$

$$= -y\, h_\theta(-z) x_1 + (1-y) h_\theta(z) x_1$$

$$= -x_1 y + x_1 h_\theta(z)$$

$$\nabla_\theta J(\theta) = -\langle x y \rangle + \frac{1}{M}\sum_{i=1}^{M} x^{(i)} g(\theta^T x^{(i)})$$

$$\frac{\partial J(\theta)}{\partial \theta_2} = -x_2 y + x_2 h_\theta(z)$$

$$\frac{\partial J(\theta)}{\partial \theta_0} = -y + h_\theta(z)$$