Logistic Regression

Part I

Step One: Function Set

Function set: **Including all different w and b**

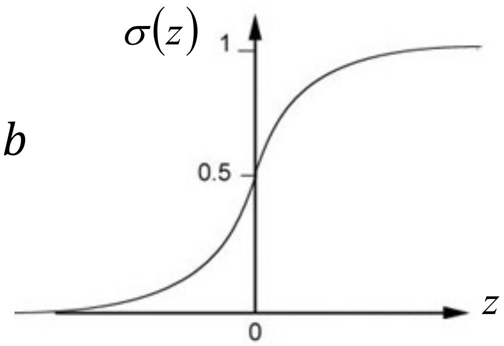
$$\begin{cases} z \geq 0 & \text{class 1} \\ z < 0 & \text{class 2} \end{cases}$$

$$P_{w,b}(C_1|x) = \sigma(z)$$

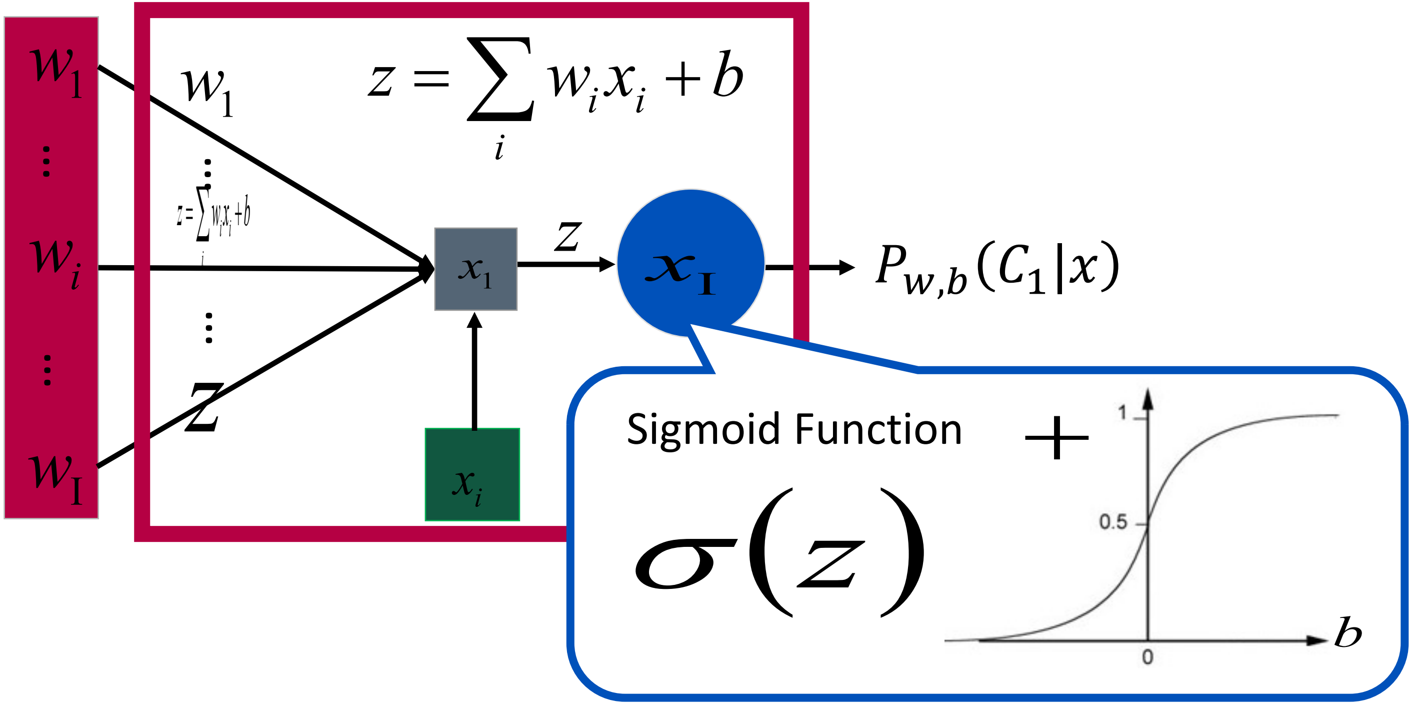
$$z = w \cdot x + b = \sum_i w_i x_i + b$$

$$\sigma(z) = \frac{1}{1 + exp(-z)}$$

Sigmoid function
Logistic function



Step One: Function Set, Continued



Step Two: Goodness of a Function: Part I

Training Data					
	x^1	x^2	x^3	……	x^N
	C_1	C_1	C_2		C_1

Assume the data is generated based on $f_{w,b}(x) = P_{w,b}(C_1|x)$

Given a set of w and b, what is its probability of generating the data?

$$L(w,b) = f_{w,b}(x^1) f_{w,b}(x^2) \left(1 - f_{w,b}(x^3)\right) \cdots f_{w,b}(x^N)$$

The most likely w* and b* is the one with the largest $L(w,b)$.

$$w^*, b^* = arg \max_{w,b} L(w,b)$$

Step Two: Goodness of a Function: Part II

x^1	x^2	x^3	……
C_1	C_1	C_2	

$\Rightarrow$

x^1	x^2	x^3	……
$\hat{y}^1 = 1$	$\hat{y}^2 = 1$	$\hat{y}^3 = 0$	

$\hat{y}^n$: 1 for class 1, 0 for class 2

$$L(w,b) = f_{w,b}(x^1) f_{w,b}(x^2) \left(1 - f_{w,b}(x^3)\right) \cdots$$

$$\boxed{w^*, b^* = arg \max_{w,b} L(w,b)} = \boxed{w^*, b^* = arg \min_{w,b} -lnL(w,b)}$$

$$-lnL(w,b)$$

$$= -lnf_{w,b}(x^1) \rightarrow -[1\, lnf(x^1) + \cancel{0\, ln(1 - f(x^1))}]$$

$$-lnf_{w,b}(x^2) \rightarrow -[1\, lnf(x^2) + \cancel{0\, ln(1 - f(x^2))}]$$

$$-ln\left(1 - f_{w,b}(x^3)\right) \rightarrow -[\cancel{0\, lnf(x^3)} + 1\, ln(1 - f(x^3))]$$

⋮

Step Two: Goodness of a Function: Part III

$$L(w,b) = f_{w,b}(x^1) f_{w,b}(x^2) \left(1 - f_{w,b}(x^3)\right) \cdots f_{w,b}(x^N)$$

$$-lnL(w,b) = -[ln f_{w,b}(x^1) + ln f_{w,b}(x^2) + ln\left(1 - f_{w,b}(x^3)\right) \cdots]$$

$\hat{y}^n$: 1 for class 1, 0 for class 2

$$= \sum_n \underline{-\left[\hat{y}^n ln f_{w,b}(x^n) + (1-\hat{y}^n) ln\left(1 - f_{w,b}(x^n)\right)\right]}$$

Cross entropy between two Bernoulli distribution

Distribution p:		Distribution q:
$p(x=1) = \hat{y}^n$	cross entropy	$q(x=1) = f(x^n)$
$p(x=0) = 1 - \hat{y}^n$		$q(x=0) = 1 - f(x^n)$

$$H(p,q) = -\sum_x p(x) ln(q(x))$$

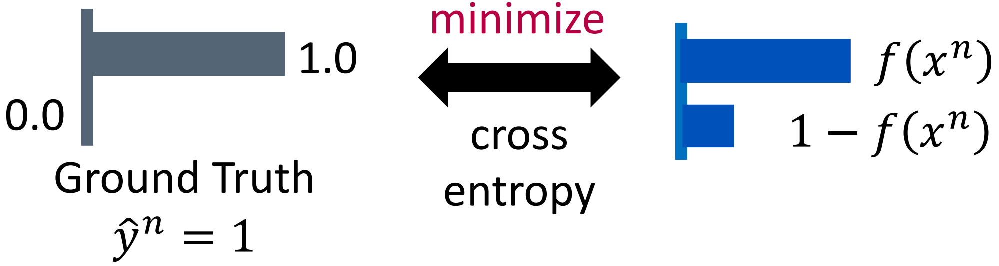
Step Two: Goodness of a Function: Part IV

$$L(w,b) = f_{w,b}(x^1) f_{w,b}(x^2)\left(1 - f_{w,b}(x^3)\right) \cdots f_{w,b}(x^N)$$

$$-lnL(w,b) = -[lnf_{w,b}(x^1) + lnf_{w,b}(x^2) + ln\left(1 - f_{w,b}(x^3)\right) \cdots]$$

$\hat{y}^n$: 1 for class 1, 0 for class 2

$$= \sum_n \underbrace{-\left[\hat{y}^n lnf_{w,b}(x^n) + (1-\hat{y}^n) ln\left(1 - f_{w,b}(x^n)\right)\right]}_{\text{Cross entropy between two Bernoulli distribution}}$$

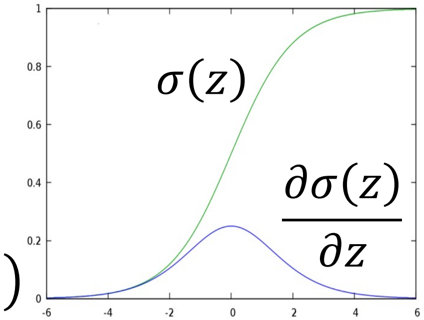


Step Three: Find the Best Function: Part I

$$\left(1 - f_{w,b}(x^n)\right) x_i^n$$

$$\frac{-lnL(w,b)}{\partial w_i} = \sum_n -\left[\hat{y}^n \boxed{\frac{lnf_{w,b}(x^n)}{\partial w_i}} + (1-\hat{y}^n)\frac{ln\left(1 - f_{w,b}(x^n)\right)}{\partial w_i}\right]$$

$$\frac{\partial lnf_{w,b}(x)}{\partial w_i} = \frac{\partial lnf_{w,b}(x)}{\partial z}\frac{\partial z}{\partial w_i} \qquad \frac{\partial z}{\partial w_i} = x_i$$



$$\frac{\partial ln\sigma(z)}{\partial z} = \frac{1}{\sigma(z)}\frac{\partial \sigma(z)}{\partial z} = \frac{1}{\cancel{\sigma(z)}}\cancel{\sigma(z)}\left(1 - \sigma(z)\right)$$

$$f_{w,b}(x) = \sigma(z)$$

$$= 1/[1 + exp(-z)]$$

$$z = w \cdot x + b = \sum_i w_i x_i + b$$

Step Three: Find the Best Function: Part II

$$\frac{-lnL(w,b)}{\partial w_i} = \sum_n -\left[\hat{y}^n \overbrace{\boxed{\frac{lnf_{w,b}(x^n)}{\partial w_i}}}^{\left(1-f_{w,b}(x^n)\right)x_i^n} + (1-\hat{y}^n)\overbrace{\boxed{\frac{ln\left(1-f_{w,b}(x^n)\right)}{\partial w_i}}}^{-f_{w,b}(x^n)x_i^n}\right]$$

$$\frac{\partial ln\left(1-f_{w,b}(x)\right)}{\partial w_i} = \frac{\partial ln\left(1-f_{w,b}(x)\right)}{\partial z}\frac{\partial z}{\partial w_i} \qquad \frac{\partial z}{\partial w_i} = x_i$$

$$\frac{\partial ln\left(1-\sigma(z)\right)}{\partial z} = -\frac{1}{1-\sigma(z)}\frac{\partial \sigma(z)}{\partial z} = -\frac{1}{\cancel{1-\sigma(z)}}\sigma(z)\cancel{\left(1-\sigma(z)\right)}$$

$$f_{w,b}(x) = \sigma(z)$$

$$= 1/[1+exp(-z)]$$

$$z = w \cdot x + b = \sum_i w_i x_i + b$$

Step Three: Find the Best Function: Part III

$$\frac{-lnL(w,b)}{\partial w_i} = \sum_n -\left[\hat{y}^n \overset{\left(1-f_{w,b}(x^n)\right)x_i^n}{\boxed{\frac{lnf_{w,b}(x^n)}{\partial w_i}}} + (1-\hat{y}^n)\overset{-f_{w,b}(x^n)x_i^n}{\boxed{\frac{ln\left(1-f_{w,b}(x^n)\right)}{\partial w_i}}}\right]$$

$$= \sum_n -\left[\hat{y}^n \left(1-f_{w,b}(x^n)\right)x_i^n - (1-\hat{y}^n)f_{w,b}(x^n)x_i^n\right]$$

$$= \sum_n -\left[\hat{y}^n \cancel{-\hat{y}^n f_{w,b}(x^n)} - f_{w,b}(x^n) \cancel{+\hat{y}^n f_{w,b}(x^n)}\right]x_i^n$$

$$= \sum_n -\left(\hat{y}^n - f_{w,b}(x^n)\right)x_i^n$$

Larger difference, larger update

$$w_i \leftarrow w_i - \eta \sum_n -\left(\hat{y}^n - f_{w,b}(x^n)\right)x_i^n$$