Classification

0 = not spam

1 = spam

$\xrightarrow{h}$ $\boxed{\text{model}}$ $\longrightarrow$ 0 or 1

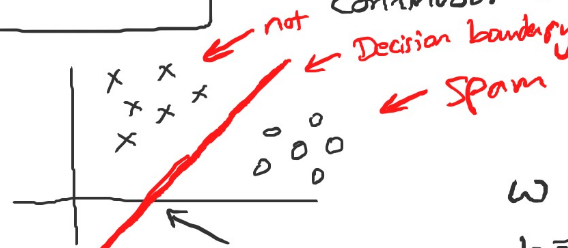
↑

email

W/ logistic regression, we are trying to predict discrete values and not continuous ones.

Logistic Regression

- Binary Classification



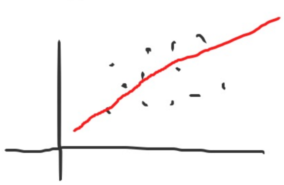
ω = parameter

b = parameter

In Linear Regression

$$h(x) = mx + b$$

minimize $J(m, b)$

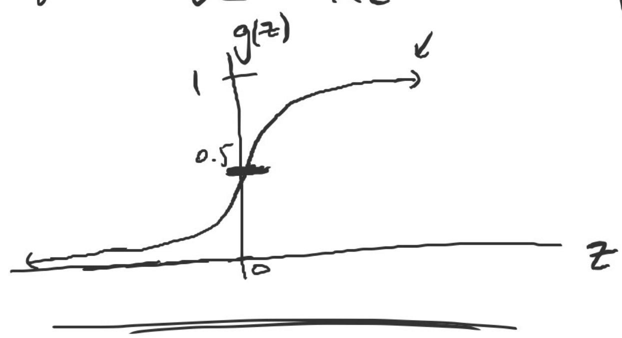


$$0 \le h(x) \le 1$$

$$h(x) = g(\underbrace{\omega x + b})$$

$$g(z) = \frac{1}{1+e^{-z}}$$ ← Sigmoid Function

Sigmoid : $g(z) = \frac{1}{1+e^{-z}}$



x	y
email 1	0
email 2	1
⋮	⋮
	0
	0
email 50	1

$g(z) \geq 0.5$

$z \geq 0$

$h(x) = g(wx+b)$; $z = wx+b$

$h(x) = \frac{1}{1+e^{-(wx+b)}}$

between 0 and 1

$0 \leq h(x) \leq 1$

$h(x) = 0.7 = P(y=1)$

70% that our output is 1

30% that our output is 0

Interpretation

$$\begin{cases} h(x) \geq 0.5 \longrightarrow y=1 \\ h(x) < 0.5 \longrightarrow y=0 \end{cases}$$

$h(x) \geq 0.5 \longrightarrow y=1$

$h(x) < 0.5 \longrightarrow y=0$

$x \geq 0$

$x < 0$

$h(x) = g(wx+b)$

$$\begin{cases} \text{if } g(wx+b) \geq 0.5, \text{ then } wx+b \geq 0 \\ \text{if } g(wx+b) < 0.5, \text{ then } wx+b < 0 \end{cases}$$

Decision Boundary. The line that separates the area where $y=1$ from where $y=0$.

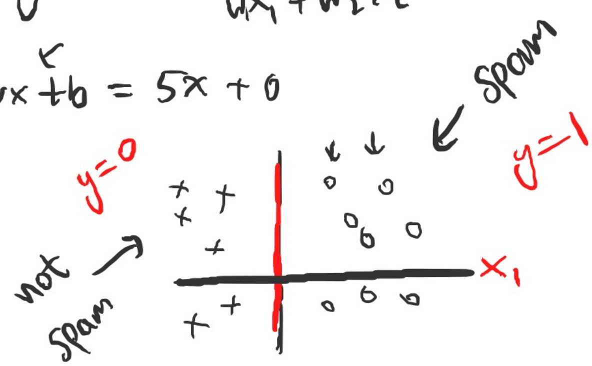
$w_1x_1 + w_2x_2 + b$

Example. $w = 5$, $b = 0$

$h(x) = wx + b = 5x + 0$

$y=1$ if $5x \geq 0 \Rightarrow x \geq 0$

$y=0$ if $5x < 0 \Rightarrow x < 0$



Cost Function

$$J(m,b) = \frac{1}{2n} \sum_{i=0}^{n} (\hat{y}^{(i)} - y^{(i)})^2$$ ← linear regression cost

$\boxed{n = \#\text{ samples}}$

$$J(w,b) = \frac{1}{n} \sum_{i=0}^{n} \text{cost}(h(x^{(i)}), y^{(i)})$$

→ $\text{cost}(h(x), y) = -\log(h(x))$ if $y = 1$

$\text{cost}(h(x), y) = -\log(1 - h(x))$ if $y = 0$

$$\boxed{J(w,b) = \frac{1}{n} \sum_{i=0}^{n} -\left(\log(h(x)^{(i)})\, y^{(i)} + \log(1 - h(x)^{(i)})(1 - y^{(i)}) \right)}$$

$g(wx+b)$ $g(wx+b)$

Goal of Logistic Regression:

$$\min_{w,b} J(w,b)$$

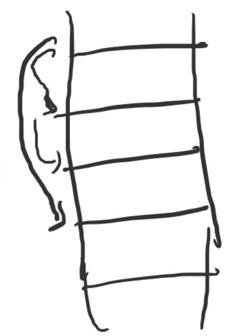
Optimize using <u>Gradient Descent</u> to get best parameter values to graph Decision Boundary (w,b).

learning rate

repeat:

$$w = w - \alpha \boxed{\frac{\partial J(w,b)}{\partial w}}$$

$$b = b - \alpha \boxed{\frac{\partial J(w,b)}{\partial b}}$$



1. Get prediction from model $h(x) = g(wx+b)$
2. Evaluate cost $J(w,b)$
3. Update w and b using Gradient Descent

$\overline{w}, \overline{b}$