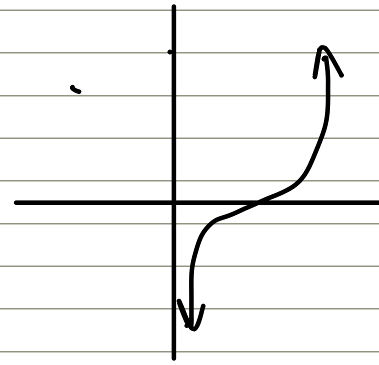
Logistic Regression:

Output is between 0 → 1 (binary classifier)

Logit Function: $f(p) = \log\left(\frac{p}{1-p}\right)$

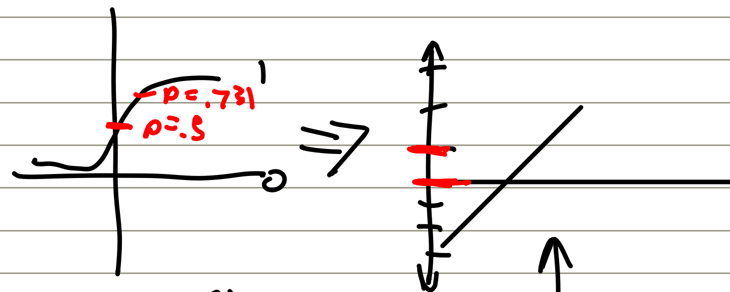
↳ log odds

↳ Takes in (0,1) and makes it $(-\infty, \infty)$



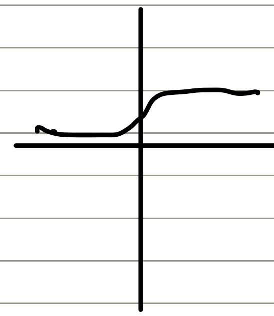
$f(.5) = \log 1 = 0$

$f(.731) = \log(2.712) = 1$



Stretches out into line

Logit inverse = sigmoid $\sigma = \frac{1}{1+e^{-x}}$



↳ Takes $(-\infty, \infty)$ and makes it (0,1)

Linear Regression = $h(x) = \theta^T x + b$

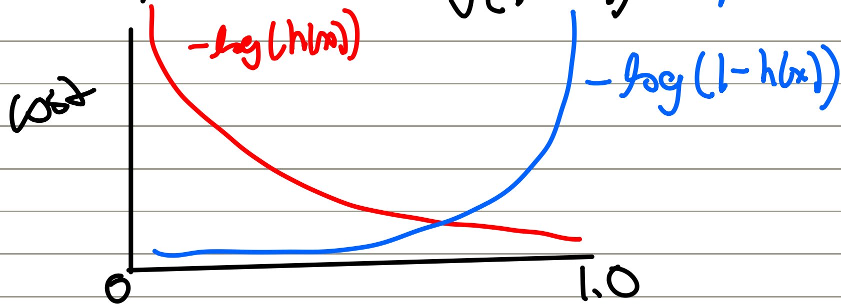
⇓

Logistic $h(x) = \sigma(h(x)) = \frac{1}{1+e^{-\theta^T x}} \Rightarrow h(x) = \begin{cases} > 0.5, & \text{if } \theta^T x > 0 \\ < 0.5, & \text{if } \theta^T x < 0 \end{cases}$

Cost Function

if $y=1$: $\text{cost} = -\log(h(x))$ ⇒ penalize for close to 0

if $y=0$: $\text{cost} = -\log(1-h(x))$ ⇒ penalize for close to 1



$\text{cost}(h(x), y) =$

$-y\log(h(x)) - (1-y)\log(1-h(x))$

Putting all errors in sample together:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} y^i \log(h(x^i)) + (1-y^i)\log(1-h(x^i))$$

Gradient Descent:

$$J' = \frac{1}{m} \sum_{i=1}^{m} (h(x^i) - y^i) \cdot x^i$$

Let p = success probability

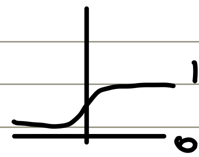
odds $= \frac{p}{1-p}$ → continuous response variable

⇓

we need discrete response variable (0 or 1) ⇒ apply the sigmoid function

⇓

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



⇓

Take the linear regression as input:

$\sigma(\beta_0 + \beta_1 x) = \frac{1}{1+e^{-\beta_0 - \beta_1 x}}$ = Probability between $(0,1)$

Gradient Descent Part

$$\theta_{new} = \theta_{curr} - \underbrace{L}_{\text{learning rate}} \nabla \underbrace{C}_{\text{cost function}}(\overbrace{f(x,\theta)}^{\hat{y}}, y), \text{ where } f(x,\theta) = \sigma(wx+b)$$

$$C(\hat{y}, y) = -\frac{1}{m} \sum_{i=1}^{m} y \log(\hat{y}) + (1-y)\log(1-\hat{y})$$

⇓

$$C(f(x,\theta), y) = -\frac{1}{m} \sum_{i=1}^{m} y \log(f(x,\theta)) + (1-y)\log(1-f(x,\theta))$$

$$\frac{\partial C}{\partial x} = -\frac{1}{m} \sum_{i=1}^{m} \left[y \log(f(x,\theta)) + (1-y)\log(1-f(x,\theta)) \right] X$$

As matrices:

$X_{m \times f}$ → m inputs, Y → m answers

↓

$$\frac{\partial C}{\partial w} = \frac{1}{m}(\hat{y} - y)^T X$$

$$\frac{\partial C}{\partial b} = \frac{1}{m}(\hat{y} - y)$$