Sigmoid function

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

Graph of sigmoid function

Looking at the graph, we can see that the value of sigmoid functi
is in the range of [0; 1]

+) As the value of x get larger, the value of sigmoid function get closer to 1 and vice versa

Derivative of sigmoid function

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

$$\Leftrightarrow \quad = \frac{d}{dx}(1 + e^{-x})^{-1}$$

Applying ~~rep~~ reciprocal rule (1)

$$\Leftrightarrow \quad = -(-1 + e^{-x})^{-2} \cdot \frac{d}{dx}(1 + e^{-x})$$

Using rule of linearity (2)

$$\Leftrightarrow \quad = -(1 + e^{-x})^{-2} \cdot \left(\frac{d}{dx}(1) + \frac{d}{dx}e^{-x}\right)$$

$$\Leftrightarrow \quad = -(1+e^{-x})^{-2} \cdot \left(0 + \frac{d}{dx} e^{-x}\right)$$

Using the exponential rule (3)

$$\Leftrightarrow \quad = -(1+e^{-x})^{-2} \cdot \left[e^{-x} \cdot \frac{d}{dx}(-x)\right]$$

$$\Leftrightarrow \quad = -(1+e^{-x})^{-2} \cdot \left[e^{-x} \cdot -\frac{d}{dx}(x)\right] \quad (2)$$

$$\Leftrightarrow \quad = -(1+e^{-x})^{-2} \cdot \left[e^{-x} \cdot (-1)\right]$$

$$\Leftrightarrow \quad = (1+e^{-x})^{-2} \cdot e^{-x}$$

Which this can be written as

$$\Leftrightarrow \quad = \frac{e^{-x}}{(1+e^{x})^2}$$

This is the derivative of sigmoid function, but we can extend the formula as bellow:

$$\Leftrightarrow \quad = \frac{1 \cdot e^{-x}}{(1+e^{-x}) \cdot (1+e^{-x})}$$

$$\Leftrightarrow \quad = \frac{1}{(1+e^{-x})} \cdot \frac{e^{-x}}{1+e^{-x}}$$

$$\Leftrightarrow \quad = \frac{1}{1+e^{-x}} \cdot \left[\frac{e^{-x}+1}{1+e^{-x}} + \frac{-1}{1+e^{-x}}\right]$$

$$\Leftrightarrow \quad = \frac{1}{1+e^{-x}} \cdot \left(1 - \frac{1}{1+e^{-x}}\right)$$

$$= \sigma(x) \cdot (1 - \sigma(x))$$ [This can be used as a shaten]

So derivative of sigmod function

$$\sigma'(x) = \sigma(x) \cdot [1 - \sigma(x)]$$

Explain the rule

(1). Reciprocal rule. $\left[\frac{1}{u(x)}\right]' = [u(x)^{-1}]'$

$$= \left[\frac{u'(x)}{u(x)^2}\right] = -u(x)^{-2} \cdot u'(x)$$

(2) Linearity rule $[a \cdot u(x) + b \cdot v(x)]' = a u'(x) + b v'(x)$

(3) Exponential rule $[e^{u(x)}]' = e^{u(x)} \cdot u'(x)$

b) Given the formula of loss function in logistic regression

The loss function using in logistic regression is called "cross-entropy loss". (or log-loss). The function is

$$\mathcal{L}(y, \hat{y}) = -(y \log(\hat{y}) + (1-y) \log(\log 1 - \hat{y}))$$

Where

y is the true label (either 0 and 1)

$\hat{y}$ is the predicted probability of the label being 1

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

The hypothesis returns in the probability that $y = 1$, given x, parameterized by θ.

Written as:

$$h(x) = P(y = 1 | x; \theta)$$

Decision boundary can be described as

1: if $\theta^T . x \geq 0 \rightarrow h(x) \geq 0.5$

0: if $\theta^T . x < 0 \rightarrow h_x < 0.5$

We have cost function in this hypothesis

$$Cost(\cancel{h_\theta(\theta)}\ h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

$$(=)\quad Cost(h_\theta(x), y) = -y \log(\cancel{\theta}\ h_\theta(x)) - (1-y)\log(h_\theta(x))$$

So the cost function is the summation of from all data

$$J(\theta) = \frac{-1}{m}\left[\sum_{i=1}^{m} -y^{(i)} \log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1 - h_\theta(x^{(i)}))\right]$$

Or.

$$J(\theta) = \frac{-1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log(\hat{y}^{(i)}) + (1-y^{(i)}).\log(1-\hat{y}^{(i)})\right]$$

While m is numble of samples

Types of loss function:

The cross-entropy loss is a convex loss function

"convex loss function": A convex loss function if for every pair of ~~par~~ points within its domain, the line segment connecting these points lies above the graph function.

To proof the cross-entropy loss is a convex: The target ~~of~~ is the 2nd derivative ~~the~~ with respect to z is non negative with z is linear combination of features and weights.

Date . . . . No.

$$\hat{y} = 6(x) = \frac{1}{1+e^{-z}}$$

$$L(y,\hat{y}) = -(y+\log(\hat{y}) + (1-y)\log(1-\hat{y}))$$

In the (a), we have proof that

$$6'(x) = 6(x)(1-6(x))$$

1, First derivative of loss function with respect to z

$$\frac{dL}{dz} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dz}$$

While
$$\begin{cases} \dfrac{dL}{d\hat{y}} = -\dfrac{y}{\hat{y}} + \dfrac{1-y}{1-\hat{y}} \\ \dfrac{d\hat{y}}{dz} = 6(x).(1-6(x)) \end{cases}$$

2, Second derivative of loss function

$$\frac{d^2L}{d^2z^2} = \frac{d^2L}{d\hat{y}^2} \cdot \left(\frac{d\hat{y}}{dz}\right)^2 + \frac{dL}{d\hat{y}} \cdot \frac{d^2\hat{y}}{dz^2}$$

Given that $\hat{y}$ is the probability is always between 0 and 1 and y is either 0 and 1

$$\Rightarrow \frac{d^2L}{d_z^2} \text{ is always non-negative}$$

HONG HA

Date . . . No.

c, Calculate ~~the~~ the gradient vector ~~of~~ for loss function in logistic regression

We have $\mathcal{L}(y, \hat{y}) = -(y \log(\hat{y}) + (1-y) \cdot \log(1-\hat{y}))$

Where $\hat{y}$ is the probability predicted given by $\sigma(z)$

$$\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}}$$

Where z is linear combination of features & weights

$$z = w^T x + b$$

To calulate the gradient vector for loss function with respect to the weights w and bias b

$$\frac{\partial L}{\partial w} = (y - \hat{y}) x$$

$$\frac{\partial L}{\partial b} = y - \hat{y}$$

To prove the above gradient, using the derivation

Derivative of the loss function with respect to z

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z}$$

In the question (b), we have

$$\frac{\partial L}{\partial y} = \frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}} \quad \& \quad \frac{\partial \hat{y}}{\partial z} = \hat{y} \cdot (1-\hat{y})$$

Multiplying above given

$$\frac{\partial L}{\partial z} = \hat{y} - y$$

Derivative with respect to w

$$\frac{\partial \mathcal{L}}{\partial w} = x . \frac{\partial L}{\partial z} = x.(\hat{y} - y) \quad \text{(Using chain rule)}$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial z} . \frac{\partial z}{\partial w}$$ (Since $z = w^T x + b$, the term $\frac{\partial z}{\partial w}$ is simply input vector x.

$$\Rightarrow \quad x(\hat{y} - y)$$

Derivative with respect to b

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial z} . \frac{\partial z}{\partial b}$$ Since $z = w^T x + b$, the term $\frac{\partial z}{\partial b}$ is simply 1

$$\Rightarrow \frac{\partial L}{\partial z} . 1 = \hat{y} - y$$