6/9/24

Home Work - 1

[ML]

3. $g(z) = \frac{1}{1 + e^{-z}}$

S.T : $g(-z) = 1 - g(z)$

Also. S.T. : $g^{-1}(y) = \ln\left(\frac{y}{1-y}\right)$

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

② $g(z) \Leftarrow 1 - \frac{1}{1 + e^{-z}}$

$= \frac{\cancel{1} + e^{-z} - \cancel{1}}{1 + e^{-z}}$

$= \boxed{\frac{e^{-z}}{1+e^{-z}}}$

$= \cancel{e^{-z}\left(\right.}$

$\Rightarrow \frac{1}{1 + e^{z}} \cdot \frac{e^{-z}}{e^{-z}} = \frac{e^{-z}}{e^{-z} + \underbrace{e^{z-z}}_{e^{0}=1}}$ more simplified

$= \boxed{\frac{e^{-z}}{e^{-z} + 1}}$ $\Rightarrow$ Thus equal

(2) Inverse of $g(z)$ is given by

$$g^{-1}(y) = \ln\left(\frac{y}{1-y}\right)$$

$$g(z) = \frac{1}{1+e^{-z}} \cancel{\times \frac{1+e^{-z}}{1+e^{-z}}}$$

$$\cancel{y} = \cancel{\frac{1+e^{-z}}{1+e^{-}}}$$

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

$$y + ye^{-z} = 1$$

$$ye^{-z} = 1-y$$

※ **Log formula:**

$$-\ln(a) = \ln\left(\frac{1}{a}\right)$$

$$e^{-z} = \frac{1-y}{y}$$

$$-z = \ln\left(\frac{1-y}{y}\right)$$

$$z = -\ln\left(\frac{1-y}{y}\right)$$

$$\Rightarrow z = \ln\left(\frac{y}{1-y}\right)$$

$$\Rightarrow g^{-1}(\underset{z}{y}) = \ln\left(\frac{y}{1-y}\right)$$

$$\Rightarrow \boxed{g^{-1}(z) = \ln\left(\frac{y}{1-y}\right)}$$

4. Problem 4

Logistic Regression: Implement a linear logistic regression algorithm for binary classification in Matlab using gradient descent.

Data set $[x, y]$

Pick reasonable values for η & ϵ to use in dataset → " dataset4.mat ".

TRAINING
(whole dataset)

To find : ① Show with figures the resulting linear decision boundary on the 2D x data

② Binary classification error

③ Empirical risk

④ No. of iterations needed for η + ϵ

① Soln : $\boxed{f(x, \theta) = \left(1 + e^{(-\theta^T x)}\right)^{-1}}$ — logistic (sigmoid) function

input vector

parameter vector

$$\boxed{= \frac{1}{1 + e^{-\theta^T x}}}$$

Logical loss [Empirical Risk] =>

$$R_{emp}(\theta) = \frac{1}{N} \sum_{i=1}^{N} (y_i - 1) \log(1 - f(x_i;\theta) - y_i \log(f(x_i;\theta))$$

True labels (0 or 1)

Input feature vectors

Derivative of ① :

$\boxed{\theta^T x = z}$

Substitute

$$\Rightarrow \frac{\partial f(x;\theta)}{\partial \theta} = \frac{1}{1 + e^{-z}}$$

$$= (1 + e^{-z})^{-1}$$

$$= \underbrace{(1 + e}_{u}^{-z})^{-2} \cdot \frac{d}{dx}(1 + e^{-z})$$

$$= -1 (1 + e^{-z})^{-2} \cdot (0 - e^{-z})$$

$$-1 (1 + e^{-z})$$

$$= -1\left(1+e^{-z}\right)^{-2} \cdot -e^{-z}$$

$$= \left(1+e^{-z}\right)^{-2} \cdot e^{-z}$$

$$\boxed{z = \theta^T x}$$

$$= \frac{e^{-z}}{\left(1+e^{-z}\right)^{2}}$$

$$\boxed{f(x;\theta) = \frac{1}{1+e^{-z}}}$$

$$= \frac{e^{-z}}{1+e^{-z}} \cdot \frac{1}{1+e^{-z}}$$

$$= f(z) \cdot (1 - f(z))$$

Next, gradient of the empirical risk:

$$R_{emp}(\theta) = \frac{1}{N}\sum_{i=1}^{N} (y_i - 1)\log(1 - f(x_i;\theta)) - y_i \log(f(x_i;\theta))$$

$y_i = 1$

True label (0 or 1)

Rough : $\boxed{f(x_i;\theta) = f(z)}$

$$\Rightarrow \frac{1}{N}\sum_{i=1}^{N} (y_i - 1)\log(1 - f(z^i)) - y_i \log(f(z^i))$$

Derivative w.r.t θ

1st part

$$= (y_i - 1) \cdot d\left(\frac{1}{1 - f(z^i)}\right) \cdot -1 = \frac{-1}{1 - f_i}$$

$$= (y_i - 1) \frac{-1}{1 - f_i} \cdot \frac{d}{d\theta}(1 - f_i)$$

$$= (y_i - 1) \frac{1}{1 - f_i} \cdot (-1)$$

1st + 2nd part

$$= -\left(y\left(\log\left(f(x;\theta)\right)\right) + (1-y)\log\left(1-f(x,\theta)\right)\right)$$

Substitute $f(x;\theta) = \cancel{(\theta^T x)}$: f_z

$$= -\left(y \log f_z\right) + (1-y)\log\left(1-f_z\right)$$

Gradient $\sigma(z)$ is

$$\frac{d}{dz}\sigma(z) = \sigma z\left(1-\sigma(z)\right)$$

$$= f(x;\theta)\left(1-f(x;\theta)\right)x$$

Now computing together.

$$\frac{\partial \ell(\theta)}{\partial \theta} = -\left(\frac{y}{f(x;\theta)}\frac{\partial f(x;\theta)}{d\theta} - \frac{1-y}{1-f(x;\theta)}\frac{\partial f(x;\theta)}{\partial \theta}\right.$$

$$= -\left(y\left(1-f(x;\theta)\right)x - (1-y)f(x;\theta)x\right)$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = f\left((x;\theta)-y\right)x$$

$\therefore$ Gradient of logistic loss function w.r.t θ is

$$\boxed{\frac{\partial R_{emp}(\theta)}{\partial \theta} = \frac{1}{N}\sum_{i=1}^{N}\left(f(x_i;\theta)-y_i\right)x_i}$$