Họ và tên : Trần Hải Nam

ID: 11219279

Class: DSEB 63

Exercise 1:

a) Sigmoid function:

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

Đạo hàm:

$$\frac{\partial \sigma(z)}{\partial z} = \frac{(1+e^{-z})'}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2} = \sigma(z).(1-\sigma(z))$$

b) Loss function:

$$L = -\log p(y|w) = -\sum_{i=1}^{N} \left(y_i \log \hat{y_i} + (1-y_i) \log(1-\hat{y_i}) \right)$$

with $\hat{y_i} = \sigma(Xw)$ and $y_i = \{0, 1\}$

→ This is ~~entropy~~ cross-entropy error function

c)

$$L = -\sum_{i=1}^{N} \left(y_i \log \hat{y_i} + (1-y_i) \log(1-\hat{y_i}) \right)$$

$\nabla(L, w) = \frac{\partial L}{\partial w}$

According the Chain Rule:

$$\frac{\partial L}{\partial \vec{w}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial \vec{w}}, \quad \text{with } z = X^T w$$

We can calculate ~~each~~ more details for each component in the right side!

$$\frac{\partial L_i}{\partial \hat{y}} = -\left(\frac{y_i}{\hat{y_i}} - \frac{(1-y_i)}{(1-\hat{y_i})} \right)$$

$$\frac{\partial \hat{y_i}}{\partial z_i} = \frac{\partial \sigma(z_i)}{\partial z_i} = \frac{e^{-z_i}}{(1+e^{-z_i})^2} = \sigma(z_i).(1-\sigma(z_i))$$

$$= \hat{y_i}.(1-\hat{y_i})$$

and $\frac{\partial z_i}{\partial \vec{w_i}} = x_i^T = \vec{x_i}$

$$\Rightarrow \frac{\partial L_i}{\partial \vec{w}_i} = \frac{\partial L}{\partial} - \left(\frac{y}{\hat{y}} - \frac{(1-y)}{(1-\hat{y})}\right) . \hat{y}(1-\hat{y}) . x_i^T$$

$$= -(y-\hat{y})\vec{x}^T$$

$$\Rightarrow \frac{\partial L}{\partial \vec{w}} = -\sum_{i=1}^{N}(y_i - \hat{y}_i)\vec{x}_i$$

$$= X^T(\hat{y} - y)$$

Exercise 3 -

Assume that we can apply MSE as the loss function of logistics model:

we have:

$$\hat{y} = \frac{1}{1+e^{-wx}}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

We know that a function $f(x)$ is convex if $f''(x) > 0$

$$f(x) = (y - \hat{y})^2$$

$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w}$$

$$= -2(y-\hat{y}) . \hat{y}(1-\hat{y}) . x$$

$$= -2(y\hat{y} - y\hat{y}^2 - \hat{y}^2 + \hat{y}^3)x$$

$$\frac{\partial^2 f}{\partial w^2} = \frac{\partial}{\partial w}\left(\frac{\partial f}{\partial w}\right) = \frac{\partial}{\partial w}\left(\frac{\partial f}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w}\right)$$

$$= -2(y - 2y\hat{y} - 2\hat{y} + 3\hat{y}^2)x . x\hat{y}(1-\hat{y})$$

$$= -2(3\hat{y}^2 - 2\hat{y}(y+1) + y) \quad + y) . \underbrace{x^2 . \hat{y}(1-\hat{y})}_{>0}$$

Consider the sign of:

$$H(\hat{y}) = -2(3\hat{y}^2 - 2\hat{y}(y+1) + y)$$

When $y = 0$:

$$H(\hat{y}) = -2(3\hat{y}^2 - 2\hat{y}) = -2\left[3\hat{y}\left(\hat{y} - \frac{2}{3}\right)\right]$$

negative when $\hat{y} \in \left[\frac{2}{3}, 1\right]$

when $y = 1$.

$$H(\hat{y}) = -2(3\hat{y}^2 - 4\hat{y} + 1)$$
$$= -6(\hat{y} - \frac{1}{3})(\hat{y} - 1)$$

Negative when: $\hat{y} \in [0, \frac{1}{3}]$

→ This show that when we apply MSE, loss function can not be convex

→ Hard to find local minium.