Thứ

Ngày . . No.

1)

a. We have,

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

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

$$L(\theta) = \log L(\theta) = \sum_{i=1}^{N} y^{[i]} \log h(x^{[i]}) + (1 - y^{[i]}) \log(1 - h(x^{[i]}))$$

$$\frac{\partial}{\partial \theta_J} L(\theta) = \left(y \frac{1}{g(\theta^T x)} - (1-y) \frac{1}{1 - g(\theta^T x)} \right) \frac{\partial}{\partial \theta_i} g(\theta^T x)$$

$$= \left(y \frac{1}{g(\theta^T x)} - (1-y) \frac{1}{1 - g(\theta^T x)} \right) g(\theta^T x)(1 - g(\theta^T x)) \theta^T x$$

$$= \left(y(1 - g(\theta^T x)) - (1-y) g(\theta^T x) \right) x_J$$

$$= (y - h_\theta(x)) x_J.$$

2)

$$w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ \vdots \\ w_d \end{bmatrix} ; x = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nd} \end{bmatrix} ; y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} ; \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

Using Chain Rule:

$$\frac{dL}{dw} = \frac{dL}{d\hat{y}} \times \frac{d\hat{y}}{dw}$$

$$\Leftrightarrow \frac{dL}{dw} = -\left[y_i \frac{d\log(\hat{y}_i)}{dw} + (1-y_i)\frac{d\log(1-\hat{y}_i)}{dw} \right]$$

$$= -\left[y_i \frac{d\log(\hat{y}_i)}{d(\hat{y}_i)} \cdot \frac{d\hat{y}_i}{dw} + (1-y_i)\frac{d\log(1-\hat{y}_i)}{d\hat{y}_i} \cdot \frac{d\hat{y}_i}{dw} \right]$$

$$= -\left[y_i \frac{1}{\hat{y}_i} + (1-y_i)\frac{1}{1-\hat{y}_i} \right] \frac{d\hat{y}_i}{dw}$$

$$= -\left[\frac{y_i - \hat{y}_i}{\hat{y}_i(1-\hat{y}_i)} \right] \frac{d\hat{y}_i}{dw} \quad (1)$$

Let $z = e^{-w^t x}$, we have:

$$\frac{d\hat{y}_i}{dw} = \frac{\frac{1}{1+z_i}}{dw} = \frac{\frac{1}{1+z_i} dz_i}{dz_i \cdot dw} = \frac{-1}{(1+z_i)^2}(z_i x_i)$$

$$= -x \cdot \frac{z_i}{(1+z_i)^2} = x_i \hat{y}_i (1-\hat{y}_i) \quad (2)$$

$$(1)(2) \Rightarrow \frac{dL}{dw} = -\left[\frac{y_i - \hat{y}_i}{\hat{y}_i(1-\hat{y}_i)} \right] x_i \hat{y}_i (1-\hat{y}_1)$$

$$= \sum_{i=1}^{N} x_i(\hat{y}_i - y_i) = x^T(\hat{y} - y)$$

By using Gradient descent, we can find w_i.

With every x, we have $\hat{y} = \delta(x^T w)$, compare with the thresh ($t \geq 0,5$):

$$\begin{cases} \hat{y} = \delta(x^T w) \geq 0,5 \to 1 \\ \hat{y} = \delta(x^T w) < 0,5 \to 0. \end{cases}$$

5)

⊕ We have:

$$\frac{\partial L}{\partial w} = x_i(\hat{y}_i - y_i) \quad (1)$$

Since: $$\frac{\partial \hat{y}_i}{\partial w} = x_i \hat{y}_i (1 - \hat{y}_i) \quad (2)$$

$$\Rightarrow \frac{\partial^2 L}{\partial w^2} = x_i \cdot \frac{\partial \hat{y}_i}{\partial w} = x_i^2 \cdot \hat{y}_i (1 - \hat{y}_i) \geq 0$$

$\Rightarrow$ The Loss binary - crossentropy with logistic model is convex

⊕ $$L = \frac{1}{N} \cdot \sum_{n=1}^{N} (\hat{y}_i - y_i)^2$$

$$L(MSE) = (y - \hat{y})^2$$

$$(1) \Rightarrow \frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w}$$

Thứ
Ngày . . No.

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

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

$$\frac{\partial^2 L}{\partial w^2} = -2x\left(y.\frac{\partial \hat{y}}{\partial w} - y\frac{\partial \hat{y}^2}{\partial \hat{y}}.\frac{\partial \hat{y}}{\partial w} - \frac{\partial \hat{y}^2}{\partial \hat{y}}.\frac{\partial \hat{y}}{\partial w} + \frac{\partial \hat{y}^3}{\partial \hat{y}}.\frac{\partial \hat{y}}{\partial w}\right)$$

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

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

$x^2\hat{y}(1-\hat{y}) \geq 0 \Rightarrow$ we need to consider only:

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

$$f(\hat{y}) = \begin{cases} 4\hat{y} - 6\hat{y}^2 & \text{when } y=0 \quad (3) \\ -2 + 4\hat{y} + 4\hat{y} - 6\hat{y}^2 = -2 + 8\hat{y} - 6\hat{y}^2 & \text{when } y=1 \quad (4) \end{cases}$$

for (3): $f(\hat{y}) \leq 0$ when $\hat{y} \leq 0$; $\hat{y} \geq \frac{2}{3}$

(4): $f(\hat{y}) \leq 0$ when $\hat{y} \leq \frac{1}{3}$, $\hat{y} \geq 1$

$\Rightarrow$ The Loss Square Error with Logistic Model is NOT convex.