HW5

Thanh Binh

October 2022

1. Tính vector calculus $\frac{dL}{dw}$ $(x^T(y^- y))$

$$\frac{dL}{dw}\left(x^T(\hat{y}-y)\right)$$

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

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

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

$$\frac{d}{d_{\theta_j}} L(\theta) = \left(y\frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)}\right)\frac{d}{d_{\theta_1}} g(\theta^T x)$$

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

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i),\ \hat{y}_i = \sigma(x_i^T w)$$

$$w = \begin{bmatrix} w_0 \\ w_1 \\ ... \\ w_d \end{bmatrix},\ x = \begin{bmatrix} 1\ x_{11}\ x_{12}\ ...\ x_{1d} \\ 1\ x_{21}\ x_{22}\ ...\ x_{2d} \\ ... \\ 1\ x_{n1}\ x_{n2}\ ...\ x_{nd} \end{bmatrix},\ y = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix},\ \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ ... \\ \hat{y}_n \end{bmatrix}$$

With every point (x^T, y_i), we have:

$$L = -(y_i \log(\hat{y}) + (1-y)\log(1-\hat{y}_i)),\ \hat{y}_i = \sigma(x_i^T w)$$

Using Chain Rule:

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

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

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

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

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

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

$$\frac{d\hat{y}_i}{dw} = \frac{\frac{1}{1+z_i}}{dw} = \frac{\frac{1}{1+z_i}}{dz_i}\frac{dz_i}{dw} = \frac{-1}{(1+z_i)^2}(z_i x_i) = -x\frac{z_i}{(1+z_i)^2} = x_i\hat{y}_i(1-\hat{y}_1) \ (2)$$

From (1) and (2):

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

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

5. Chứng minh với model logistic thì loss binary cross entropy là convex function với W, loss mean square error không là convex function với W

From Ex 1, we have:

$$\frac{dL}{dw} = x_i(\hat{y}_i - y_i)$$

Since:

$$\frac{d\hat{y}_i}{dw} = x_i\hat{y}_i(1-\hat{y}_i)$$

$$\Rightarrow \frac{d^2L}{dw^2} = x_i\frac{d\hat{y}_i}{dw} = x_i^2 \ \hat{y}_i(1-\hat{y}_1) \geq 0$$

So that the loss binary cross entropy with logistic model is convex.