Nguyễn Quốc Hưng - 11205484

Ex 1:

with $t \in \{0,1\}$, $y(x) = \sigma(w^T x)$ and $\sigma(z) = \frac{1}{1+e^{-z}}$

$\Rightarrow y(x) = \sigma(w^T x) = \frac{1}{1+e^{-w^T x}}$

$$L(w) = -\sum_{i=1}^{N} (t_i \log(y_i) + (1-t_i)\log(1-y_i))$$

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

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

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

$$= \sigma(z)(1-(z))$$

For each $(x_i, y_i)$, we have the loss function

$$l = -(t_i, \log(y_i) + (1-t_i)\log(1-y_i))$$

$$\frac{\partial l}{\partial w_j} = \frac{\partial l}{\partial y_i}\frac{\partial y_i}{\partial w_j}$$

$$= -\left(t_i \frac{1}{y_i} - (1-t_i)\frac{1}{1-y_i}\right)\frac{\partial}{\partial w_j}\sigma(w^T x_i)$$

$$= -\left(t_i \frac{1}{\sigma(w^T x_i)} - (1-t_i)\frac{1}{1-\sigma(w^T x_i)}\right)\sigma(w^T x_i)(1-\sigma(w^T x_i))\frac{\partial}{\partial w_j} w^T x_i$$

$$= -(t_i(1-\sigma(w^T x_i)) - (1-t_i)\sigma(w^T x_i))x_{ij}$$

$$= -(t_i - y_i)x_{ij}$$

$$\Rightarrow \frac{\partial L}{\partial w} = -\sum_{i=1}^{N}(t_i - y_i)x_i = x^T(t-y)$$

Nguyễn Quốc Huy - 11205484

Ex5:

Binary-cross entropy: From exercise 1, we got

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

Since

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

So that

$$\frac{\partial^2 l}{\partial w^2} = x_i \frac{\partial \hat{y}_i}{\partial w}$$

$$= x_i^2 \hat{y}_i (1 - \hat{y}_i) \geq 0$$

Then, the loss binary-cross entropy with logistic mode is convex

Mean square error:

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

we remove $(i)$ to simplify $\rightarrow$ $L(MSE) = (y - \hat{y})^2$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} = -2(y - \hat{y}).x.\hat{y}.(1 - \hat{y})$$

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

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

The second deri

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

$$= -2x\left(yx\hat{y}(1-\hat{y}) - y2\hat{y}x\hat{y}(1-\hat{y}) - y2\hat{y}x\hat{y}(1-\hat{y}) - 2\hat{y}x\hat{y}(1-\hat{y}) + 3\hat{y}^2 x\hat{y}(1-\hat{y})\right)$$

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

Ex5: continue.

since $x^2\hat{y}(1-\hat{y}) \geq 0$, consider only

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

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

In case of (1), $f(\hat{y}) \leq 0$ when $\frac{2}{3} \leq \hat{y} < 1$

In case of (2), $f(\hat{y}) \leq 0$ when $0 \leq \hat{y} \leq \frac{1}{3}$

Then, the loss MSE with logistic is

Not convex

Scanned with CamScanner