MACHINE LEARNING

Nghiem Thi Ngoc Minh
MSV 11202545

HW week5

1 Exercise 1:

Calculate vector calculus $\dfrac{\partial L}{\partial w} = X^T(\hat{y} - y)$

$$\begin{aligned} L = -log\ p(t|w) &= -\sum_{i=1}^{N} y_i log(\hat{y_i}) + (1 - y_i) log(1 - \hat{y_i}) \\ &= -(y\ log\hat{y} + (1-y) log(1-\hat{y})) \end{aligned}$$

$$\hat{y} = \sigma(X^T w) = \frac{1}{1 + e^{-X^T w}}\ ,\ z = e^{-X^T w}$$

We have:

$\dfrac{\partial L}{\partial w} = \dfrac{\partial L}{\partial \hat{y}}.\dfrac{\partial \hat{y}}{\partial z}.\dfrac{\partial z}{\partial w}$ (chain rule)

- $\dfrac{\partial L}{\partial \hat{y}} = -\Big(y.\dfrac{1}{\hat{y}} - (1-y).\dfrac{1}{1-\hat{y}}\Big) = -\Big(\dfrac{y}{\hat{y}} - \dfrac{1-y}{1-\hat{y}}\Big)$
- $\dfrac{\partial \hat{y}}{\partial w} = \dfrac{\partial \hat{y}}{\partial z}.\dfrac{\partial z}{\partial w}$
  $= -\dfrac{1}{(1+z)^2}.(-Xe^{-X^T w}) = X.\dfrac{1}{1+z}.\dfrac{z}{1+z} = X\hat{y}(1-\hat{y})$

$$\begin{aligned} \Rightarrow \frac{\partial L}{\partial w} &= -\Big(\frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}}\Big).X\hat{y}(1-\hat{y}) \\ &= \frac{-y + y\hat{y} + \hat{y} - y\hat{y}}{\hat{y}(1-\hat{y})}.X\hat{y}(1-\hat{y}) \\ &= X(\hat{y} - y) \end{aligned}$$

Under the matrix form: $\dfrac{\partial L}{\partial w} = X^T(\hat{y} - y)$

2 Problem 5

a. Binary cross entropy:

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

Since:

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

So that:

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

$\longrightarrow$ the loss binary-cross entropy with logistic model is convex

b. MSE:

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

Remove index (i) we have:
$L(MSE) = (y - \hat{y})^2$
So:

$$
\begin{aligned}
\frac{\partial L}{\partial w} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} \\
&= -2(y - \hat{y}) \cdot x \cdot \hat{y} \cdot (1 - \hat{y}) \\
&= -2 \cdot x \cdot (y \cdot \hat{y} - \hat{y}^2) \cdot (1 - \hat{y}) \\
&= -2 \cdot x \cdot (y \cdot \hat{y} - y \cdot \hat{y}^2 - \hat{y}^2 + \hat{y}^3)
\end{aligned}
$$

The second derivate:

$$
\begin{aligned}
\frac{\partial^2 L}{\partial w^2} &= -2 \cdot x \cdot (y \cdot \frac{\partial \hat{y}}{\partial w} - y \cdot \frac{\partial \hat{y}^2}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} - \frac{\partial \hat{y}^2}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} + \frac{\partial \hat{y}^3}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} \\
&= -2 \cdot x(y \cdot x \cdot \hat{y} \cdot (1 - \hat{y}) - y \cdot 2 \cdot \hat{y} \cdot x \cdot \hat{y} \cdot (1 - \hat{y}) - 2 \cdot \hat{y} \cdot x \cdot \hat{y} \cdot (1 - \widehat{y)}) + 3 \cdot \hat{y}^2 \cdot x \cdot \hat{y} \cdot (1 - \hat{y})) \\
&= -2 \cdot x^2 \cdot \hat{y} \cdot (1 - \hat{y}) \cdot (y - 2y\hat{y} - 2\hat{y} + 3\hat{y}^2)
\end{aligned}
$$

Since $x^2 \cdot \hat{y}(1 - \hat{y}) \geq 0$, consider only: $f(\hat{y}) = -2(y - 2y\hat{y} - 2\hat{y} + 3\hat{y}^2$

$$
\begin{cases}
4\hat{y} - 6\hat{y}^2 = 2\hat{y}(2\hat{y} - 3\hat{y})(*) \qquad when\, y = 0 \\
-2 + 8\hat{y} - 6\hat{y}^2 = -2(3\hat{y} - 1)(\hat{y} - 1)(**) \qquad when\, y = 1
\end{cases}
$$

In case (*), $f(\hat{y}) \leq 0$ when $\frac{2}{3} \leq \hat{y} \leq 1$
In case (**), $f(\hat{y}) \leq 0$ when $0 \leq \hat{y} \leq \frac{1}{3}$
$\longrightarrow$ the loss mean square error with logistic model is NOT convex.