ML dseb62 w5

Nguyen Tuan Duy - 11204971 - dseb62

October 2022

1 Problem 1

$$\begin{aligned} L &= -\sum_{i=1}^{N} y_i log(\hat{y}_i) + (1-y_i)log(1-\hat{y}_i) \\ &= Y^T log(Y) + (1-Y)^T log(1-\hat{Y}) \\ \frac{\delta L}{\delta W} &= (\frac{\delta \hat{Y}}{\delta W})^T \frac{\delta L}{\delta \hat{Y}} \end{aligned}$$

$$W = \begin{bmatrix} w_0 \\ w_1 \\ . \\ . \\ w_n \end{bmatrix} x = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & .. & x1d \\ 1 & x_{21} & x_{22} & x_{23} & .. & x2d \\ . & . & . & . & . & . \\ 1 & x_{n1} & x_{n2} & x_{n3} & .. & xnd \end{bmatrix} y = \begin{bmatrix} y_0 \\ y_1 \\ . \\ . \\ y_n \end{bmatrix}$$

we have $\hat{Y} = \frac{1}{1+e^{-z}}$ where $z = w_0 + w_1 x1_1 + w_2 x_2 + ... + w_n x_n = XW$ then we can calculate

$$\frac{\delta \hat{Y}}{\delta z} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1+e^{-z}-1}{(1+e^{-z})^2} = \frac{1}{(1+e^{-z})} - \frac{1}{(1+e^{-z})^2} = \hat{Y} - \hat{Y}^2 = \hat{Y}(1-\hat{Y}) \tag{1}$$

$$\frac{\delta z}{\delta W} = X \tag{2}$$

$$\frac{\delta L}{\delta \hat{y}} = \frac{Y}{\hat{Y}} - \frac{1-Y}{1-\hat{Y}} = \frac{Y - Y\hat{Y} - \hat{Y} + Y\hat{Y}}{\hat{Y}(1-\hat{Y})} = \frac{Y-\hat{Y}}{\hat{Y}(1-\hat{Y})} \tag{3}$$

from (1) (2) and (3),

$$\frac{\delta L}{\delta W} = (\frac{\delta \hat{Y}}{\delta W})^T \frac{\delta L}{\delta \hat{Y}} = (\frac{\delta \hat{Y}}{\delta z}\frac{\delta z}{\delta W})^T \frac{\delta L}{\delta \hat{Y}} = X^T(\hat{Y} - Y) \tag{4}$$

2 Problem 5

to show that a function is convex, we can show that the its second order derivative is non negative
loss binary cross entropy:

$$L_{bce} = -\sum_{i=1}^{N} y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)$$
$$\frac{\delta L}{\delta W} = \frac{\delta L}{\delta \hat{y}} * \frac{\delta \hat{y}}{\delta z} * \frac{\delta z}{\delta W}$$
$$\frac{\delta \hat{y}}{\delta z} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1+e^{-z}-1}{(1+e^{-z})^2} = \frac{1}{(1+e^{-z})} - \frac{1}{(1+e^{-z})^2} = \hat{y} - \hat{y}^2 = \hat{y}(1-\hat{y})$$
$$\frac{\delta z}{\delta W} = X$$

$$\begin{aligned}
\frac{\delta L}{\delta \hat{y}} &= -\sum_{i=1}^{N} \frac{y_i}{\hat{y}_i} - \frac{1-y_i}{1-\hat{y}_i} \\
&= -\sum_{i=1}^{N} \frac{y_i - y_i\hat{y}_i - \hat{y}_i + y_i\hat{y}_i}{\hat{y}_i(1-\hat{y}_i)} \\
&= -\sum_{i=1}^{N} \frac{y_i - \hat{y}_i}{\hat{y}_i(1-\hat{y}_i)}
\end{aligned}$$

$$\frac{\delta L}{\delta W} = -\sum_{i=1}^{N} \frac{y_i - \hat{y}_i}{\hat{y}_i(1-\hat{y}_i)} x\hat{y}_i(1-\hat{y}_i) = \sum_{i=1}^{N} (\hat{y}_i - y_i)x \tag{5}$$

$$\frac{\delta^2 L}{\delta^2 W} = \frac{\delta^2 L}{\delta^2 \hat{y}} * \frac{\delta \hat{y}}{\delta z} * \frac{\delta z}{\delta W} = \sum_{i=1}^{N} x^2(1-\hat{y}_i)\hat{y}_i$$

because $e^{-z} \geq 0 \ \forall z$ then $0 \leq \frac{1}{1+e^{-z}} \leq 1$ *or* $0 \leq \hat{y} \leq 1$ and $(1-\hat{y}) \geq 0$ so $\frac{\delta^2 L}{\delta^2 W} \geq 0$ and the loss binary cross entropy is convex

Consider the mean square error loss function:

$$
\begin{aligned}
L &= (y-\hat{y})^2 \\
\frac{\delta L}{\delta W} &= \frac{\delta L}{\delta \hat{y}} * \frac{\delta \hat{y}}{\delta z} * \frac{\delta z}{\delta W} \\
&= -2(y-\hat{y})x\hat{y}(1-\hat{y}) \\
&= -2x(y\hat{y}-\hat{y}^2)(1-\hat{y}) \\
&= -2x(y\hat{y}-y\hat{y}^2-\hat{y}^2+\hat{y}^3) \\
\frac{\delta^2 L}{\delta^2 W} &= \frac{\delta^2 L}{\delta^2 \hat{y}} * \frac{\delta \hat{y}}{\delta z} * \frac{\delta z}{\delta W} \\
&= -2x(y-2y\hat{y}-2\hat{y}+3\hat{y}^2)x(1-\hat{y})\hat{y} \\
&= -2x^2(y-2y\hat{y}-2\hat{y}+3\hat{y}^2)(1-\hat{y})\hat{y}
\end{aligned}
$$

we have already proven $(1-\hat{y})\hat{y} \geq 0$ and $x^2 \geq 0$ then we only have to consider:

$$
\begin{aligned}
f &= -2(y-2y\hat{y}-2\hat{y}+3\hat{y}^2) \\
&= -2(y-2\hat{y}(y+1)+3\hat{y}^2)
\end{aligned}
$$

because y can only take the value 0 or 1, let's consider the case y = 0

$$
\begin{aligned}
f &= -2(-2\hat{y}+3\hat{y}^2) \\
&= -2(\hat{y}(-2+3\hat{y}))
\end{aligned}
$$

we can see that when $\hat{y}$ in the range [0, 2/3] f ≥ 0 when $\hat{y}$ in the range [2/3, 1] f ≤ 0, for exanple when $\hat{y} = 1$, f = -2 then $\frac{\delta^2 L}{\delta^2 W} < 0$, which proves L is not convex

3 Problem 2

Loss function:

$$
\begin{aligned}
L &= -\sum_{i=1}^{N} y_i log(\hat{y_i}) + (1-y_i)log(1-\hat{y_i}) \\
\hat{y_i} &= \frac{1}{1+e^{-z}} \\
z &= w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n
\end{aligned}
$$

The gradient descent algorithm to minimize the loss function with respect to W:
Step 1: choosing starting point for W
Step 2: update w based on $w_t = w_{t-1} - learning\_rate * \frac{\delta L}{\delta W}$
Step 3: if loss function L is small enough, stop the algorithm, else, repeat step 2

For step 2, we have already calculate $\frac{\delta L}{\delta W} = \sum_{i=1}^{N}(\hat{y}_i - y_i)x$ (5) in previous problem, from here, we can calculate

$$\frac{\delta L}{\delta W_0} = \sum_{i=1}^{N}(\hat{y}_i - y_i)$$
$$\frac{\delta L}{\delta W_1} = \sum_{i=1}^{N}(\hat{y}_i - y_i)x_1$$
$$\frac{\delta L}{\delta W_2} = \sum_{i=1}^{N}(\hat{y}_i - y_i)x_2$$

4 Problem 4

Suppose we want to accept the loan when $\hat{y} \geq t$ $(0 \leq t \leq 1)$, then:

$$\begin{aligned} \hat{y}_i &= \frac{1}{1+e^{-z}} \geq t \\ &\iff 1 + e^{-z} \leq \frac{1}{t} \\ &\iff e^{-z} \leq \frac{1}{t} - 1 \\ &\iff z \geq -\log\left(\frac{1}{t} - 1\right) \\ &\iff w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n \geq -\log\left(\frac{1}{t} - 1\right) \end{aligned}$$

And we can reject the loan when $\hat{y}_i < t$. Here we can see that the line separating two classes is $w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n = -\log\left(\frac{1}{t} - 1\right)$