Machine Learning 1 - Homework week 5: Logistic Regression

Mai Xuan Bach
ID 11200489

Ngày 26 tháng 10 năm 2022

1 Problem 1. Calculate:

To calculate the gradient of binary-cross entropy, we will compute gradient of loss at each point.

We have:

$$L = -(y_i.\log(\hat{y}_i) + (1-y_i).\log(1-\hat{y}_i)) \tag{1}$$

$$\begin{aligned}
\frac{\partial L}{\partial w} &= -\left(y_i.\frac{\partial \log(\hat{y}_i)}{\partial w} + (1-y_i).\frac{\partial \log(1-\hat{y}_i)}{\partial w}\right) &(2)\\
&= -\left(y_i.\frac{\partial \log(\hat{y}_i)}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial w} + (1-y_i).\frac{\partial \log(1-\hat{y}_i)}{\partial \hat{y}_i}.\frac{\partial \hat{y}_i}{\partial w}\right) &(3)\\
&= -\left(y_i.\frac{1}{\hat{y}_i} - (1-y_i).\frac{1}{(1-\hat{y}_i)}\right).\frac{\partial \hat{y}_i}{\partial w} &(4)\\
&= -\left(\frac{y_i.(1-\hat{y}_i) - \hat{y}_i.(1-y_i)}{\hat{y}_i.(1-\hat{y}_i)}\right).\frac{\partial \hat{y}_i}{\partial w} &(5)\\
&= -\left(\frac{y_i - \hat{y}_i}{\hat{y}_i.(1-\hat{y}_i)}\right).\frac{\partial \hat{y}_i}{\partial w} &(6)
\end{aligned}$$

We calculate $\frac{\partial \hat{y}_i}{\partial w}$ separately:

Let $z = \exp(-w^T x)$, we have:

$$
\begin{aligned}
\frac{\partial \hat{y}_i}{\partial w} &= \frac{\partial \frac{1}{1+z_i}}{\partial w} && (7)\\
&= \frac{\partial \frac{1}{1+z_i}}{\partial z_i}.\frac{\partial z_i}{\partial w} && (8)\\
&= \frac{-1}{(1+z_i)^2}.(-z_i x_i) && (9)\\
&= \frac{z_i x_i}{(1+z_i)^2} && (10)\\
&= \frac{z_i}{1+z_i}.x_i.\frac{1}{1+z_i} && (11)\\
&= x_i.\hat{y}_i.(1-\hat{y}_i) && (12)
\end{aligned}
$$

Put (12) into (6), we got:

$$
\begin{aligned}
\frac{\partial L}{\partial w} &= -x_i(y_i - \hat{y}_i) && (13)\\
&= x_i.(\hat{y}_i - y_i) && (14)
\end{aligned}
$$

In the form of matrix, we got:

$$
\frac{\partial L}{\partial w} = X^T(\hat{Y} - Y) \qquad (15)
$$

2 Problem 5. Chứng minh:

Chứng minh với model logistic thì loss binary-crossentropy là convex function với W, loss mean square error không là convex function với W.

Answer:

a. Binary-cross entropy: From exercise 1, we got:

$$\frac{\partial L}{\partial w} = x_i.(\hat{y}_i - y_i) \tag{16}$$

Since

$$\frac{\partial \hat{y}_i}{\partial w} = x_i.\hat{y}_i.(1 - \hat{y}_i) \tag{17}$$

So that:

$$\begin{aligned} \frac{\partial^2 L}{\partial w^2} &= x_i.\frac{\partial \hat{y}_i}{\partial w} &(18)\\ &= x_i^2.\hat{y}_i.(1 - \hat{y}_i) \geq 0 &(19) \end{aligned}$$

Then, the loss binary-crossentropy with logistic model is **convex**.

b. MSE:

$$L = \frac{1}{N}\sum_{n=1}^{N}(\hat{y}_i - y_i)^2 \tag{20}$$

For short, we remove the index (i) to simplify the notation. We got: L (MSE) = $(y - \hat{y})^2$ and from (17), we have:

$$\begin{aligned} \frac{\partial L}{\partial w} &= \frac{\partial L}{\partial \hat{y}}.\frac{\partial \hat{y}}{\partial w} &(21)\\ &= -2(y - \hat{y}).x.\hat{y}.(1 - \hat{y}) &(22)\\ &= -2.x.(y.\hat{y} - \hat{y}^2)(1 - \hat{y}) &(23)\\ &= -2.x.(y.\hat{y} - y\hat{y}^2 - \hat{y}^2 + \hat{y}^3) &(24) \end{aligned}$$

The second derivate:

$$\begin{aligned} \frac{\partial^2 L}{\partial w^2} &= -2.x.(y.\frac{\partial \hat{y}}{\partial w} - y\frac{\partial \hat{y}^2}{\partial \hat{y}}.\frac{\partial \hat{y}}{\partial w} - \frac{\partial \hat{y}^2}{\partial \hat{y}}.\frac{\partial \hat{y}}{\partial w} + \frac{\partial \hat{y}^3}{\partial \hat{y}}.\frac{\partial \hat{y}}{\partial w}) &(25)\\ &= -2.x.(y.x.\hat{y}.(1 - \hat{y}) - y.2.\hat{y}.x.\hat{y}.(1 - \hat{y}) - 2\hat{y}.x.\hat{y}.(1 - \hat{y}) + 3.\hat{y}^2.x.\hat{y}.(1 - \hat{y})) &(26)\\ &= -2.x^2.\hat{y}.(1 - \hat{y}).(y - 2y\hat{y} - 2\hat{y} + 3\hat{y}^2) &(27) \end{aligned}$$

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

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

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

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

Then, the loss mean square error with logistic model is **NOT** convex.