ML TUẦN 5: LOGISTIC REGRESSION

NGÔ ĐOÀN KIÊN

October 2022

1 Problem 1: Tính vector calculus dl/dw

We already have the Log likelihood function with parameter θ as below:

$$l(\theta) = \text{Log } L(\theta) = \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)} + (1 - y^{(i)}) \log(1 - h(x^{(i)}))$$

Firstly, we define:

$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$ (1)
$g(z) = \frac{1}{1+e^{-z}}$
g(z) is called the sigmoid function

For the later use, we take derivative of g(z):

$$\begin{aligned} g'(z) &= \frac{1}{(1+e^{-z})^2}(e^{-z}) \\ &= \frac{1}{1+e^{-z}}.(1 - \frac{1}{1+e^{-z}}) \\ &= g(z)(1 - g(z)) \qquad (2) \end{aligned}$$

We take partial derivative of $l(\theta)$ by θ

$$\begin{aligned} \frac{\partial}{\partial \theta_j} l(\theta) &= \left(y \frac{1}{g(\theta^T x)} - (1-y) \frac{1}{1-g(\theta^T x)} \right) \frac{\partial}{\partial \theta_j} g(\theta^T x) \\ &= \left(y \frac{1}{g(\theta^T x)} - (1-y) \frac{1}{1-g(\theta^T x)} \right) g(\theta^T x)(1 - g(\theta^T x)) \frac{\partial}{\partial \theta_j} \theta^T x \qquad \text{by (2)} \\ &= \left(y(1 - g(\theta^T x)) - (1-y) g(\theta^T x) \right) x_j \end{aligned}$$

$= (y - g(\theta^T x))x_j$

$= (y - h_\theta(x))x_j \qquad \text{by (1)}$

2 Problem 2: Loss function of Logistic regression

2.1 Chứng minh với model logistic thì loss binary crossentropy là convex function với W

We define f(x) is the binary crossentropy loss function of Logistic regression

And $\hat{y} = h_\theta(x)$ has been mentioned from Problem 1. Change for shorter form (easier to read the formula)

$$
\begin{aligned}
=> -f(x) &= y\log(\hat{y}) + (1-y)\log(1-\hat{y}) \\
&= y\log(\tfrac{1}{1+e^{-\theta x}}) + (1-y)\log(1 - \tfrac{1}{1+e^{-\theta x}}) \\
&= y\log(\tfrac{e^{\theta x}}{1+e^{\theta x}}) + (1-y)\log(\tfrac{1}{1+e^{\theta x}}) \\
&= y\log(e^{\theta x}) - y\log(1+e^{\theta x}) + (1-y)\log(1) - (1-y)\log(1+e^{\theta x}) \\
&= y(\theta x) - y\log(1+e^{\theta x}) - \log(1+e^{\theta x}) + y\log(1+e^{\theta x}) \\
&= xy\theta - \log(1+e^{\theta x})
\end{aligned}
$$

$=> f(x) = \log(1+e^{\theta x}) - xy\theta$

Now we have f(x) in the very simple form.

To prove f(x) is convex we have to point out that f'(x) >= 0 for any x

$$
\begin{aligned}
\frac{\partial f}{\partial \theta} &= \tfrac{1}{1+e^{\theta x}} x.e^{\theta x} - xy \\
&= \tfrac{x}{1+e^{-\theta x}} - xy
\end{aligned}
$$

$$\frac{\partial^2 f}{\partial^2 \theta} = \frac{-x.e^{-\theta x}.(-x)}{(1+e^{-\theta x})^2}$$

$= \frac{x^2.e^{-\theta x}}{(1+e^{-\theta x})^2} \quad \geq 0 \ \forall x$

=> proved

2.2 Chứng minh loss mean square error không là convex function với W

For simplicity, assume that we only have 1 feature "x" and binary labels for "y"

Define f(x) = MSE and $\hat{y}$ is the predicted value obtained after applying sigmoid function

We have:

$\hat{y} = \frac{1}{1+e^{-\theta x}}$

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

First order derivative:

$g(x) = \frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta}$

We already have $\hat{y}' = \hat{y}(1 - \hat{y})$ proved from problem 1

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

$= -2[y\hat{y} - y\hat{y}^2 - \hat{y}^2 + \hat{y}^3]x$

Second order derivative:

$\frac{\partial^2 f}{\partial^2 \theta} = \frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial \theta}\right) = \frac{\partial g}{\partial \theta} = \frac{\partial g}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta}$

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

Because the term $x^2 \geq 0 \ \forall x$

And $\hat{y}(1 - \hat{y}) \in [0, \frac{1}{4}]$ since $\hat{y} \in [0, 1]$

We consider $H(\hat{y}) = -2[y - 2y\hat{y} - 2\hat{y} + 3\hat{y}^2]$

and y is binary labels that is equal to 0 or 1

When y = 0

$H(\hat{y}) = -2[3\hat{y}(\hat{y} - \frac{2}{3}]$

It is clearly that $H(\hat{y}) < 0$ when x $\in (\frac{2}{3}, 1]$ => Proved MSE is not convex