DSEB K61, MFE, NEU

Machine Learning Assignment #6

Tô Đức Anh | 11196328

October 2021

1 Problems

1. Tự biến đổi lại thuật toán logistic regression, từ xây dựng công thức, likelihood, maximize likelihood, đạo làm negative log likelihood theo ma trận hệ số.

2. Tìm hàm f(x), biết $f'(x) = f(x)(1 - f(x))$

3. Dùng thuật toán gradient descent tối ưu hàm $f(x) = x^2$, với giá trị khởi tạo $x = 2$ và thử các learning rate khác nhau $(0.2, 1, 2)$, vẽ đồ thị hàm loss để hiện sự thay đổi hàm $f(x)$ sau các bước update.

4. Dùng thuật toán logistic regression phân loại hồ sơ cho vay hay không với dữ liệu ở đây dataset, input là lương, thời gian làm việc, output là cho vay hay không. xây dựng model, tìm tham số.
   dự đoán với các hồ sơ sau: (lương 2, thời gian: 3), (lương: 1, thời gian: 8), (lương: 5, thời gian: 5).

2 Answers

.

2.1 Question 1

Assume that we are classifying data into 2 classes, C_1 and C_2, the one stays in C_1 will not be able to stays in C_2.
Assume that we are tossing a coin:

$$p(Head) = \frac{p(Head)}{p(Head) + p(Tail)} = \frac{1}{1 + \frac{p(Tail)}{p(Head)}}$$

With $p(Head) = 1 - p(Tail)$ and otherwise.
So that we have our $Posterior Probability$ of class C_1 is:

$$p(C_1|x) = \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1) + p(p(x|C_2)p(C_2))} = \frac{1}{1 + e^{-x}} = \sigma(\alpha) \quad (1)$$

With x is our data, and C_1 is the probability that data lie in the first group
Keep transforming we have:

$$\frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1) + p(p(x|C_2)p(C_2))}$$

$=$

$$\frac{1}{1 + \frac{p(x|C_2)p(X_2)}{p(x|C_1)p(C_1)}}$$

Since $\frac{p(x|C_2)p(X_2)}{p(x|C_1)p(C_1)}$ is not negative, we let that $= e^{-a}$.
Rewrite the formula, we have:

$$\frac{1}{1 + \frac{p(x|C_2)p(X_2)}{p(x|C_1)p(C_1)}} = \frac{1}{1 + e^{-a}}$$

Just like (1) above!

Next, we calculate the derivative of the function above:

$$\begin{aligned}
f'(x) &= \frac{1}{1 + e^{-x}}' \\
&= ((1 + e^{-x})^{-1})' \\
&= -(1 + e^{-x})^{-2}(-e^{-x}) \\
&= \frac{e^{-x}}{(1 + e^{-x})^2} \\
&= \frac{1}{1 + e^{-x}} \cdot \frac{e^{-x}}{1 + e^{-x}} \\
&= \frac{1}{1 + e^{-x}} \cdot \frac{(1 + e^{-x}) - 1}{1 + e^{-x}} \\
&= \frac{1}{1 + e^{-x}} \cdot \left(\frac{1 + e^{-x}}{1 + e^{-x}} - \frac{1}{1 + e^{-x}}\right) \\
&= \frac{1}{1 + e^{-x}} \cdot \left(1 - \frac{1}{1 + e^{-x}}\right) \\
&= \sigma(x) \cdot (1 - \sigma(x))
\end{aligned} \tag{1}$$

Next, remind ourself to previous Linear Regression, we can see that we will transform x to $x^2, x^3, ..., x^n$,, but to shorten it, we will represent it as $\phi = [x^2, x^3, ..., x^n]$
So that, $y = w_0 + w_1\phi_1(x) + ... + w_d\phi_d(x)$ Remind our self back to Bernoulli distribution, we have that:

$$p(x) = t^x * (1 - t)^{1-x}$$

The logistics regression model can be defined as:

$$p(C_1|\phi) = y(\phi) = \sigma(w^T\phi)$$

$$p(C_2|\phi) = 1 - p(C_1|\phi)$$

Where $\sigma(w^T\phi)$ is the probability that ϕ or x is C_1. And since we are classifying data into 2 classes, so data either lie in C_1 or C_2, so that t only has 2 values, 1 or 0. ϕ_n is $\phi(x_n)$.
So that the likelihood of one data point is:

$$y_i = \sigma(\phi_i^T W)$$

$$p(t_i|w) = y_i^{t_1} * (1 - y_i)^{1-t_i)}$$

So that the likelihood of all data point is:

$$p(t|w) = \prod_{i=i}^{n} y_i^{t_1} * (1 - y_i)^{1-t_i)}$$

So we have to maximize $p(t|w)$, we take the negative *log* of $p(t|w)$:

$$-\log p(t|w) = -\sum_{i=1}^{n}(t_i \log y_i + (1 - t_i)\log(1 - y_i))$$

Read more why take the negative log here
We have to maximize the negative *log* of $p(t|w)$ with parameter w and $y_i = \sigma(\phi^T w)$
To maximize it we have to calculate the derivative and let it $= 0$.

$$MaxL = -(t \log y + (1 - t)\log(1 - y))$$

with

$$y = \sigma(w_0 + w_1\phi_1 + w_2\phi_2 + ... + w_D\phi_D)$$

and

$$z = w_0 + w_1\phi_1 + w_2\phi_2 + ... + w_D\phi_D$$

Using chain rule we have:

$$\frac{dL}{dw_i} = \frac{dL}{dy} * \frac{dy}{dz} * \frac{dz}{dw_i}$$

$$\begin{aligned} \frac{dL}{dy} &= -(\frac{t}{y} + -\frac{1-t}{1-y}) \\ &= \frac{t(1-y) - y(1-t)}{y(z-y)} \\ &= \frac{y}{1-y} \end{aligned} \tag{2}$$

$$\frac{dy}{dz} = \sigma(z) * (1 - \sigma(z)) \tag{3}$$

$$\frac{dz}{dw_i} = \phi_i \tag{4}$$

According to chainrule:

$$\frac{dL}{dw_i} = \frac{dL}{dy} * \frac{dy}{dz} * \frac{dz}{dw_i}$$

$=$

$$(y - t)\phi_i$$

So that:

$$\frac{dL}{dw} = \begin{bmatrix} \frac{dL}{dw_0} \\ \frac{dL}{dw_1} \\ .... \\ \frac{dL}{dw_D} \end{bmatrix} = \begin{bmatrix} y - t, \\ (y-t)\phi_1 \\ .... \\ (y-t)\phi_D \end{bmatrix}$$

2.2 Answer 2

We have that:

$$
\begin{aligned}
f'(x) &= f(x)(1-f(x)) \\
f'(x) &= f(x) - f(x)^2 \\
\frac{df}{dx} &= f(x) - f(x)^2 \\
\frac{df}{f(x)-f(x)^2} &= dx
\end{aligned}
\tag{5}
$$

Now, integrating both the sides of the differential equation $\int \frac{df}{f(x)-f(x)^2} = dx$, we get:

$$
\begin{aligned}
\int \frac{df}{f(x)-f(x)^2} &= \int dx \\
\int \frac{df}{f(1-f(x))} &= \int dx \\
\int (\frac{1}{f(x)} - \frac{1}{f(x)-1})df &= \int dx \\
\int \frac{1}{f}df - \int \frac{1}{f(x)-1}df &= \int dx \\
ln|f(x)| - ln|f-1| + C &= x + C \\
ln|\frac{f(x)}{f(x)-1}| &= x + C \\
\frac{f(x)}{f(x)-1} &= e^{x+C} \\
\frac{f(x)}{f(x)-1} &= C * e^{x} \\
\frac{f(x)-1}{f(x)} = e^{-x} * \frac{1}{C} 1 - \frac{1}{f(x)} &= e^{-x} \\
\frac{1}{f(x)} &= 1 - e^{-x} \\
f(x) &= \frac{1}{1-e^{-x}}
\end{aligned}
\tag{6}
$$

Theres problem here but i dont know why