$Cross \quad Entropy \quad Loss \quad in \quad Binary \quad Logistic \quad \mathrm{Re}gression$

رگرسیون لجستیک یک الگوریتم تحت نظارت ($\sup ervised$) یادگیری ماشین است که برای مسایل طبقه بندی ($classification$) بکارمیرود.برخلاف رگرسیون خطی که مقادیرپیوسته را پیش بینی می کند این الگوریتم احتمال اینکه یک ورودی به یک کلاس مشخص تعلق داشته باشد را پیش بینی می کندوبرای طبقه بندی باینری بکارمیرودجاییکه خروجی یکی از دوحالت ممکن(بله یاخیر،درست یا نادرست،صفر یایک و...)باشد.دراین مقاله یک معرفی اجمالی از پایه ها ومفاهیم این الگوریتم را خواهیم دیدونکته اصلی ، یافتن مشتق تابع هزینه است که من به روشی متفاوت وبا استفاده از خواص تابع سیگموید آنرا محاسبه کرده ام.

هدف رگرسیون لجستیک باینری ، آموزش دادن یک کلاسیفایر ($classifier$) , (y) است که بتواند درمورد کلاس یک ورودی مشاهده شده ، یک تصمیم باینری بگیرد.

یک ورودی مشاهده شده ($features$) , $x=(x_1,x_2,...,x_n)$ را در نظر بگیرید.خروجی کلاسیفایر y میتواند مقدار 1 راداشته باشد(یعنی ورودی مشاهده شده عضوی از کلاس است) ویا مقدار 0 (یعنی ورودی مشاهده شده عضوی از کلاس نیست).می خواهیم $P(y=1|x)$

یعنی احتمال اینکه ورودی مشاهده شده عضوی از کلاس باشد را بدانیم.

رگرسیون لجستیک می توانداین مساله را با یادگیری از یک مجموعه آموزش یافته،یک بردار

از وزن ها ویک جمله بایاس ($Intercept$) or ($Bias \quad Term$) حل نماید.

با فرض اینکه $x=(x_1,x_2,...,x_n)$ بردار متغیرهای ورودی و $w=(w_1,w_2,...,w_n)$ بردار وزن های مربوطه(ویاضرایب) و b عرض از مبدا(ویاجمله بایاس) باشد بابکارگیری تابع خطی روی متعیرورودی x ، مقدار پیوسته از رگرسیون خطی عبارت خواهدبوداز:

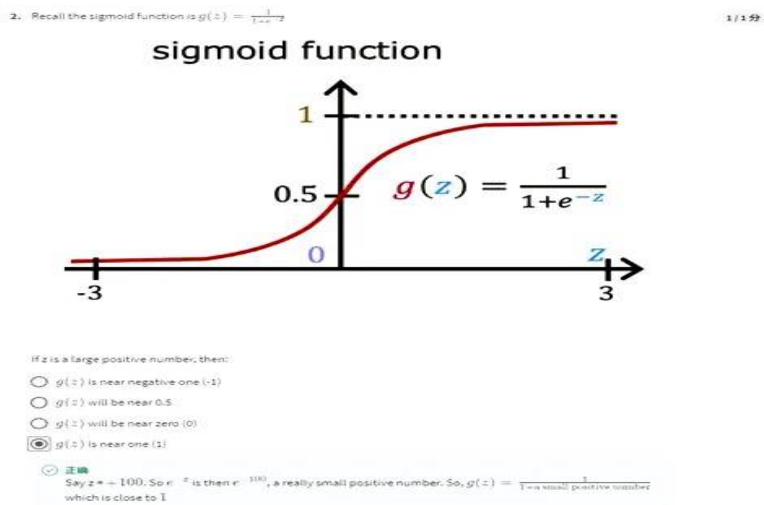
$$z=(\sum_{i=1}^{n}w_ix_i)+b=w.x+b$$

وزن w_i ضریب متغیر x_i است.

رگرسیون خطی، به منظورتبدیل z به یک احتمال بین صفرویک ،تابع سیگموید را بکارمی گیرد.

تابع سیگموید $(Sigmoid)$ یا $(Logistic\quad Function)$ عبارتست از:

$$\sigma(z)=\frac{1}{1+e^{-z}}$$



این تابع دارای این خاصیت است که مقدار آن همواره بین صفر ویک قراردارد .همچنین،

$$\sigma(-z)=1-\sigma(z)$$

زیرا :

$$\sigma(z)+\sigma(-z)=\frac{1}{1+e^{-z}}+\frac{1}{1+e^{z}}=\frac{1+e^{z}+1+e^{-z}}{(1+e^{-z})(1+e^{z})}=\frac{2+e^{z}+e^{-z}}{1+e^{z}+e^{-z}+e^{0}}=1$$

احتمال قرارگرفتن(یا قرارنگرفتن) در یک کلاس را میتوان بصورت زیر درنظرگرفت:

$$P(y=1)=\sigma(z) \qquad , \quad P(y=0)=1-\sigma(z)$$

وارون تابع سیگموید، تابع لاجیت $(Logit\quad Function)$ عبارتست از :

$$z=\log it(p)=\sigma^{-1}(p)=Ln\frac{p}{1-p}$$

تابع هزینه $(or \quad Cross \quad Entropy \quad Loss) \quad L_{CE}:$

مقدارواقعی متغیر خروجی $y=0 \quad or \quad y=1 \qquad (Correct \quad Output):$

مقدارپیش بینی برای متغیر خروجی $\hat{y}=\sigma(w.x+b) \qquad (classifier)$

(توزیع برنولی)

$$Max: \quad p(y|x)=(\hat{y})^{y}(1-\hat{y})^{1-y}$$

$$y=1 \rightarrow \hat{y}=p(1|x) \quad , \quad y=0 \rightarrow p(0|x)=1-\hat{y}$$

$$Ln(p(y|x))=y(Ln(\hat{y}))+(1-y)(Ln(1-\hat{y}))$$

قرینه عبارت فوق تابع هزینه نام دارد وآن را با L_{CE} نمایش می دهیم.

$$Min: \quad L_{CE}(\hat{y},y)=-Ln(p(y|x))=-\left[yLn(\hat{y})+(1-y)(Ln(1-\hat{y})\right]$$
$$=-\left[yLn(\sigma(w.x+b)+(1-y)(Ln(1-\sigma(w.x+b))\right]$$
$$=-\left[yLn(\sigma(w.x+b)+(1-y)(Ln(\sigma\left[-(w.x+b)\right])\right]$$

مشتق جزیی تابع هزینه نسبت به w_j :

در اینجا می خواهیم ثابت کنیم که مشتق جزیی تابع هزینه(یعنی L_{CE}) نسبت به w_j عبارتست از:

$$x_j\left[\sigma(w.x+b)-y\right]$$

Proof :

$$\frac{\partial L_{CE}}{\partial w_j}=-y\left[\frac{\sigma'(w.x+b)}{\sigma(w.x+b)}\right]-(1-y)(\frac{\sigma'\left[-(w.x+b)\right]}{\sigma\left[-(w.x+b)\right]})=$$
$$-y\left[\frac{\sigma'(w.x+b)}{\sigma(w.x+b)}-(\frac{\sigma'\left[-(w.x+b)\right]}{\sigma\left[-(w.x+b)\right]})\right]-(\frac{\sigma'\left[-(w.x+b)\right]}{\sigma\left[-(w.x+b)\right]})=$$
$$y\left[\frac{-x_j e^{w.x+b}}{1+e^{w.x+b}}-\frac{x_j e^{-(w.x+b)}}{1+e^{-(w.x+b)}}\right]-(\frac{-x_j e^{w.x+b}}{1+e^{w.x+b}})=y(-x_j)+x_j(\frac{1}{e^{-(w.x+b)}+1})=$$
$$-yx_j+x_j\sigma=x_j(\sigma-y)$$

کد پایتون مربوط به یک مثال ساده ازاجرای رگرسیون لجستیک و محاسبه هزینه لجستیک را میتوانید در لینک زیر مشاهده نمایید.

https://github.com/sharifzadeh173/my-page