Gradient Descent

11 April 2023 09:56

$$y = \underset{?}{b} + \underset{?}{w}\, x$$

y x

$$\begin{bmatrix} \vdots \end{bmatrix} \begin{bmatrix} \vdots \end{bmatrix}$$

say, $b = 0.4$

$w = -0.6$

$$\hat{y} = 0.4 + (-0.6)x$$

(prediction of y)

Steps:-

1) Values of b & w are assigned some random values $(-1, 1)$

2) Calculate $\hat{y}$ & also calculate the error f^n

$$L = \frac{1}{2n}\sum(y_i - \hat{y}_i)^2$$

when y is cont.

Log Loss

$$L = -\frac{1}{n}\left[\sum\left(y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i)\right)\right]$$

when y is binary

3) We update the values of b & w by the formulae;

$$\text{new } w = \text{old } w - \eta \frac{\partial L}{\partial w}$$

$$\text{new } b = \text{old } b - \eta \frac{\partial L}{\partial b}$$

where η is called learning rate $\eta \in (0,1)$

We update b & w only if $L >$ tolerance (say 0.001) otherwise assume w & b values as final values

4) Repeat step (2) & (3) until $L <$ tolerance.

$$X = [-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2]$$

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

$$y = 3x + 5 \qquad [-1, 0.5, 2, 3.5, 5, 6.5, 8, 9.5, 11]$$

$$\frac{d}{dx} mx = m$$

initial_b = 4

initial_w = 2

$$\hat{y}_i = b + w x_i$$

$$L = \frac{1}{2n}\sum(y_i - \hat{y}_i)^2 = \frac{1}{2n}\sum\left[y_i - b - w x_i\right]^2$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial b} = \frac{1}{\cancel{2}n}\cancel{2}(-1)\sum(y_i - \hat{y}_i).(1)$$

$$= -\frac{1}{n}\sum(y_i - \hat{y}_i)$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial w} = \frac{1}{2n}2(-1)\sum(y_i - \hat{y}_i).x_i$$

$$= -\frac{1}{n}\sum(y_i - \hat{y}_i)x_i$$

Assume $\eta = 0.01$

$$\underset{w}{\text{new}} = \underset{w}{\text{old}} - 0.01\left(-\frac{1}{n}\sum(y_i - \hat{y}_i)x_i\right)$$

$$\underset{b}{\text{new}} = \underset{b}{\text{old}} - 0.01\left(-\frac{1}{n}\sum(y_i - \hat{y}_i)\right)$$

i_b = 4

i_w = 2

$$\begin{array}{cc} X & \hat{y} \\ \begin{bmatrix} -2 \\ -1.5 \\ -1 \\ -0.5 \\ 0 \end{bmatrix} & \begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix} \end{array}$$

$$\begin{bmatrix} \vdots \\ -0.5 \\ 0 \\ 0.5 \\ 1 \\ 1.5 \\ 2 \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \\ \\ \end{bmatrix}$$

Min Max Scaling : $\dfrac{X - \min(X)}{\max(X) - \min(X)}$

$\min(x) = 29$
$\max(x) = 102$

X	
43	$(43-29)/(102-29)$
84	$(84-29)/(102-29)$
29	⋮
102	
78	

$$y = b + w_1 x_1 + w_2 x_2 \quad ; \quad \hat{y}_i = b + w_1 x_{1i} + w_2 x_{2i}$$

$$L = \frac{1}{2n}\left[\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right]$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_1} = \frac{1}{2n}(-2)\sum(y_i - \hat{y}_i)x_{1i}$$

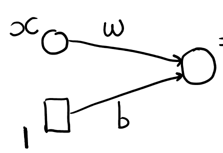
$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_2} = \frac{1}{2n}(-2)\sum(y_i - \hat{y}_i)x_{2i}$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b} = \frac{1}{2n}(-2)\sum(y_i - \hat{y}_i)$$

Logistic Regression

y	x
1	⋮
0	
0	
1	
0	

y : 1 occurence
: 0 non-occurence

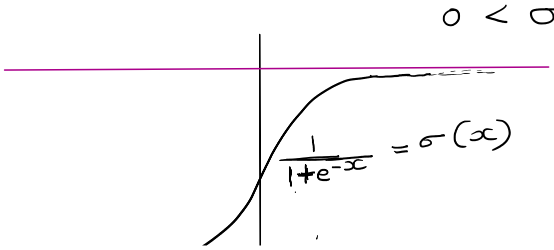


$z = wx + b$

Sigmoid

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

$$0 < \sigma(z) < 1$$

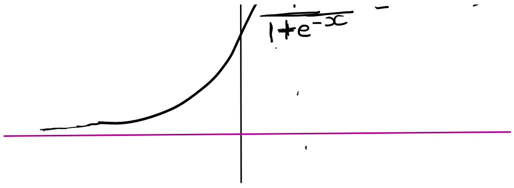


w, b

$$\hat{y} = \sigma(wx + b)$$

Probability

1



σ

Probability

$$= \frac{1}{1+e^{-(wx+b)}}$$

Age

1 : Purch
0 : non Purch

x	y
23	1
18	1
25	1
28	1
60	0
56	0
67	0
50	0

$w = ?$

$b = ?$

$$L = \frac{-1}{n}\sum\left(y_i \ln \hat{y}_i + (1-y_i)\ln(1-\hat{y}_i)\right)$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial w} \quad ; \quad \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial b}$$

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad \hat{y} = \frac{1}{1+e^{-(xw+b)}}$$

$$\frac{d}{dz}\sigma(z) = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right)$$

$$\frac{d}{dx}x^n = nx^{n-1}$$

$$= \frac{d}{dz}(1+e^{-z})^{-1} = -1\,(1+e^{-z})^{-2}\,e^{-z}(-1)$$

$$\frac{d}{dx}x^{-1} = -1x^{-1-1}$$
$$= -x^{-2}$$

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

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

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

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

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

$$\frac{\partial}{\partial \hat{y}_i}\left(\frac{-1}{n}\sum\left[y_i \ln \hat{y}_i + (1-y_i)\ln(1-\hat{y}_i)\right]\right)$$

$$= \frac{-1}{n}\sum\left(\frac{y_i}{\hat{y}_i} + \frac{1-y_i}{1-\hat{y}_i}(-1)\right) = \frac{-1}{n}\sum\left[\frac{y_i(1-\hat{y}_i) - \hat{y}_i(1-y_i)}{\hat{y}_i\,(1-\hat{y}_i)}\right]$$

$$= -\frac{1}{n}\sum\left[\frac{y_i - \cancel{y_i\hat{y}_i} - \hat{y}_i + \cancel{y_i\hat{y}_i}}{\hat{y}_i(1-\hat{y}_i)}\right]$$

$$\frac{\partial L}{\partial \hat{y}_i} = -\frac{1}{n}\sum\left[\frac{y_i - \hat{y}_i}{\hat{y}_i(1-\hat{y}_i)}\right]$$

$$\frac{\partial \hat{y}_i}{\partial w} = \frac{\partial}{\partial w}\left[\sigma(wx_i+b)\right] = \sigma(wx_i+b)\left[1-\sigma(wx_i+b)\right](x_i)$$

$$\frac{\partial \hat{y}_i}{\partial w} = \frac{\partial}{\partial w}\left[\sigma(wx_i + b)\right] = \sigma(wx_i+b)\left[1-\sigma(wx_i+b)\right](x_i)$$
$$= \hat{y}_i(1-\hat{y}_i)\,x_i$$

$$\frac{\partial \hat{y}_i}{\partial b} = \frac{\partial}{\partial b}\left[\sigma(wx_i+b)\right] = \sigma(wx_i+b)\left(1-\sigma(wx_i+b)\right)(1)$$
$$= \hat{y}_i(1-\hat{y}_i)$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial w} = -\frac{1}{n}\sum \frac{(y_i-\hat{y}_i)}{\cancel{\hat{y}_i(1-\hat{y}_i)}}\,\cancel{\hat{y}_i(1-\hat{y}_i)}\,x_i = \frac{-1}{n}\sum(y_i-\hat{y}_i)x_i$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial b} = -\frac{1}{n}\sum \frac{(y_i-\hat{y}_i)}{\cancel{\hat{y}_i(1-\hat{y}_i)}}\,\cancel{\hat{y}_i(1-\hat{y}_i)} = \frac{-1}{n}\sum(y_i-\hat{y}_i)$$

$$\underset{w}{\text{new}} = \underset{w}{\text{old}} - \eta\frac{\partial L}{\partial w} \;;\quad \underset{b}{\text{new}} = \underset{b}{\text{old}} - \eta\frac{\partial L}{\partial b}$$

x_1 = Income

x_2 = Lot_Size

y = Response

$$L = -\frac{1}{n}\sum\left[y_i\ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i)\right]$$

$$\hat{y}_i = \sigma(w_1x_1 + w_2x_2 + b)$$

$$\frac{\partial \hat{y}_i}{\partial w_1} = \hat{y}_i(1-\hat{y}_i)x_{1_i} \;;\quad \frac{\partial \hat{y}_i}{\partial w_2} = \hat{y}_i(1-\hat{y}_i)x_{2_i}$$

$$\frac{\partial \hat{y}_i}{\partial b} = \hat{y}_i(1-\hat{y}_i)$$

$$\frac{\partial L}{\partial w_1} = -\frac{1}{n}\sum\left[\frac{y_i-\hat{y}_i}{\hat{y}_i(1-\hat{y}_i)}\right]\hat{y}_i(1-\hat{y}_i)x_{1_i} = \frac{-1}{n}\sum(y_i-\hat{y}_i)x_{1_i}$$

$$\frac{\partial L}{\partial w_2} = -\frac{1}{n}\sum\frac{(y_i-\hat{y}_i)}{\hat{y}_i(1-\hat{y}_i)}\hat{y}_i(1-\hat{y}_i)x_{2_i} = \frac{-1}{n}\sum(y_i-\hat{y}_i)x_{2_i}$$

$$\frac{\partial L}{\partial b} = -\frac{1}{n}\sum\frac{(y_i-\hat{y}_i)}{\hat{y}_i(1-\hat{y}_i)}\hat{y}_i(1-\hat{y}_i) = -\frac{1}{n}\sum(y_i-\hat{y}_i)$$