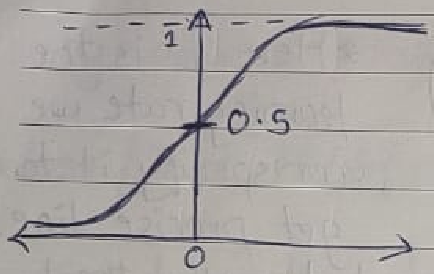
NOTES Date

Logistics Regression maths.



*This curved line is obtained by sigmoid function.

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

*We apply sigmoid on a straight line to fit the datapoints inorder to get the classification correct

$y = mx + c$

$\sigma(y) = \sigma(mx+c)$

* $y = \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n + b$

- We apply sigmoid on this whole eqn above.

* Here ω is weights and x are features.

$$X = \begin{bmatrix} x_1 \dots \\ x_2 \dots \\ \vdots \\ x_n \dots \end{bmatrix}_{n \times m} \quad W = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}_{n \times 1} \quad Y = [\dots\dots]_{1 \times m}$$

$Y = W^T \cdot X + b$ $\quad \hat{y} = \sigma(\omega^T X + b)$.

NOTES Date

* loss f^n i.e Log-Loss

$$\text{cost } f^n = \frac{-1}{m} \sum_{i=1}^{m} [y \log(\hat{y}) + (1-y)\log(1-\hat{y})]$$

where $\hat{y}$ = $y_{predicted}$

$\Downarrow$

$\hat{y} = \sigma(W^T x + b) = \sigma(z)$

$\therefore z = \omega^T x + b$.

* Now we calculate derivative of cost f^n with regard to ω & b.

* Loss/Error for single observation can be given by: Here $\hat{y} = a$.

$$L = -[y \log a + (1-y)\log(1-a)]$$

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

$$z = \omega^T x + b.$$

$\therefore$ By using chain rule:

$$\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial a} \times \frac{\partial a}{\partial z} \times \frac{\partial z}{\partial \omega}$$

NOTES Date

$$\frac{\partial L}{\partial a} = \frac{-y}{a} + \frac{(1-y)}{(1-a)}$$

α

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

Now we know,

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

$$\text{So, } \frac{e^{-z}}{(1+e^{-z})^2} = \left(\frac{1-a}{a}\right) \times a^2 = a(1-a)$$

α

$$\frac{\partial z}{\partial w} = x + 0 = x$$

α

$$\frac{\partial L}{\partial w} = \left[\frac{-y}{a} + \frac{(1-y)}{(1-a)}\right] \times (a)(1-a) \times x$$

$$= (a-y) \cdot x$$

$\therefore$ Now remeber this is w.r.t only one observation so to calculate for all observations we take into

NOTES Date

consideration matrix multiplication

$$\frac{\partial cost}{\partial w} = \underline{\underline{(A-Y) \cdot X}}$$

$\therefore$ Now to calculate for b.

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial a} \times \frac{\partial a}{\partial z} \times \frac{\partial z}{\partial b}$$

$$\frac{\partial z}{\partial b} = 1.$$

$$\frac{\partial L}{\partial b} = \frac{(a-y)}{\cancel{a(1-a)}} \times \cancel{a(1-a)} \times 1 = (a-y).$$

$$\frac{\partial cost}{\partial b} = \underline{\underline{(A-Y)}} \qquad \therefore dW = \frac{1}{m}(A-Y) \cdot X^T$$

α

$$dB = \frac{1}{m}(A-Y)$$

New $W_{new} = W_{old} - \alpha \cdot dW^T$

$B_{new} = B_{old} - \alpha \cdot dB$

α is the learning rate.