* Example1 ⇒

Hours Studies (X)	Result (Y)
29	0 (No)
15	0 (No)
33	1 (Yes)
28	1 (Yes)
39	1 (Yes)

(1) Calculate the probability of pass for the student who studied 33 hours.

(2) At least how many hours the student should study that make sure will pass the course with probability of 95%..?

⇒ $Y = aX + b$

$$a = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{5(100) - (144)(3)}{5(4460) - (144)^2}$$

$$= \frac{500 - 432}{22300 - 20736}$$

$$= \frac{68}{1564} = \underline{\underline{0.04}}$$

$a = 0.04$

X	Y	XY	X^2
29	0	0	841
15	0	0	225
33	1	33	1089
28	1	28	784
39	1	39	1521
Σ 144	3	100	4460

$$b = \frac{1}{n}\left[\sum y - a\sum x\right]$$

$$= \frac{1}{5}\left[3 - 0.04(144)\right]$$

$$= \frac{1}{5}\left[3 - 5.76\right]$$

$$= -\frac{2.76}{5} = \underline{\underline{-0.552}}$$

∴ Our Model is

$$Y = 0.04X + -0.552$$

For, $X = 33$

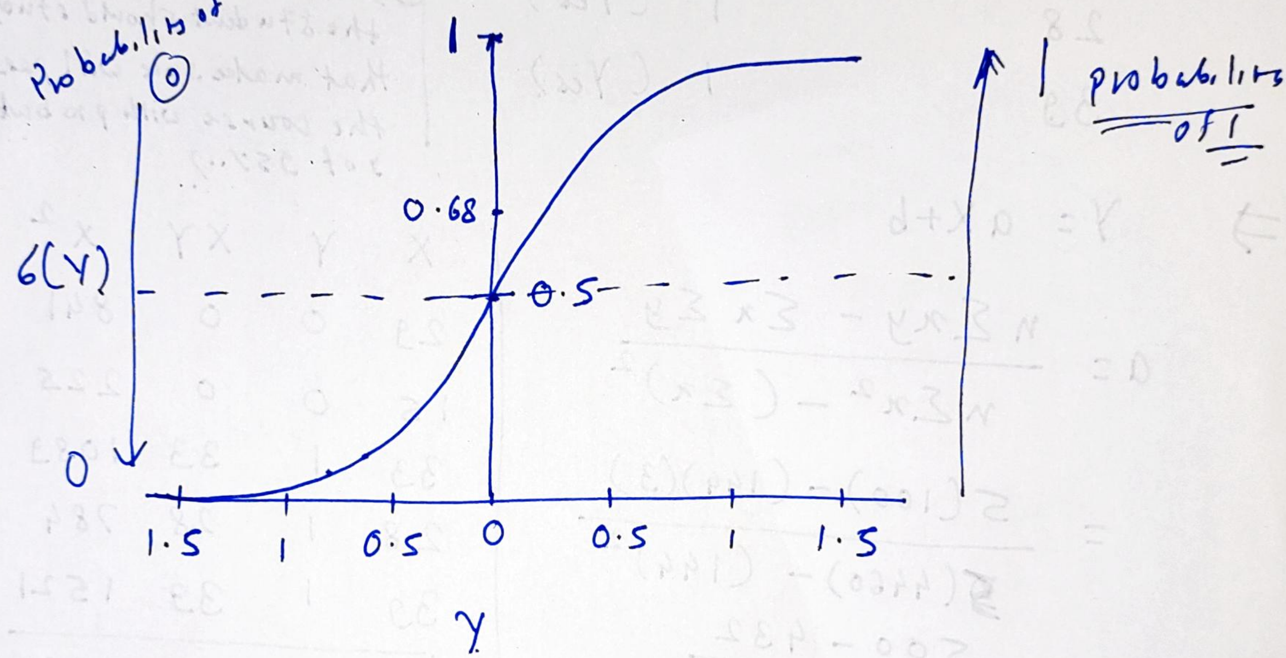
$$Y = 0.04(33) - 0.552.$$

$$= 1.32 - 0.552$$

$$= \underline{0.76}$$

$$\underset{\sigma(Y)}{\underline{\underline{P}}} = \frac{1}{1+e^{-Y}} = \frac{1}{1+e^{-0.76}}$$

$$= \frac{1}{1+0.46} = \frac{1}{1.46} = \underline{\underline{0.68}}$$



P(Result = Yes / Study Hours)
(1)

If P(Result = Yes / Study Hours) $\geq$ 0.5, then
(1)
we say that person may pass the exam.

If P(Resul = Yes / Study Hours) < 0.5, then.
(1)
we say that person may not pass the exam

$\therefore$ If student will study for 33 hours student ~~will~~ may pass the exam.

2) $P = \dfrac{1}{1+e^{-y}}$

$0.95 = \dfrac{1}{1+e^{-y}}$

$0.95 * (1+e^{-y}) = 1 \Rightarrow 0.95 + 0.95e^{-y} = 1$

~~0.95 * (1+e~~

$0.95e^{-y} = 1 - 0.95$

$0.95 e^{-y} = 0.05$

$e^{-y} = \dfrac{0.05}{0.95}$

$e^{-y} = 0.0526$

$\ln(e^{-y}) = \ln(0.0526)$

$-y = -2.95$

$\boxed{y = 2.95}$

$y = 0.04x - 0.0552$

$2.95 = 0.04x - 0.0552$

$2.95 + 0.0552 = 0.04x.$

$3.0052 = 0.04x$

$x = \dfrac{3.0052}{0.04}$

$\boxed{x = 75.13}$ hours

* Gradiant Descent :-

~~Example~~ ⇒

- During the training process (Gradiant Descent) we update. the weight as per following formula,

$$\theta_{new} = \theta_{prv} - \eta \nabla J$$

Here, ∇J ⇒ Derivative (Partial) of cost function.

η is constant factor, decide during start of training process.

θ_{new} ⇒ New parameter values

θ_{prv} ⇒ Old parameter value.

Example ⇒

Consider, a two independent features $x_1 = 3$, $x_2 = 2$ & third dependent feature $y = 1$. Assume, ~~$J(\theta)$~~ = cost function, $J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} y^i \log(\hat{y}^i) + (1-y^i) \log (1-\hat{y}^i)$.

Find the values of θ upto two descent.

⇒ ~~It~~ is ~~2 variable~~

Here, two independent variables & one dependent variable 'y' is considered.

∴ No. of parameters are 3 (~~one~~ two coefficient/weight & one bias)

∴ Assume, $\omega_1 = \omega_2 = b = 0$ (Initially)

& $\eta = 0.1$

* Descent! ⇒

Step 1: Calculate partial derivative of cost function (or consider) w.r.t. each parameter.

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y) x_j^i$$

Here, $\hat{y} = \sigma(\omega \cdot x + b)$

⇒ At actual it will be with multiple parameters

or $$\frac{\partial J(\theta)}{\partial \omega_j} = \frac{1}{m} \sum_{i=1}^{m} [\sigma(\omega \cdot x + b) - y] x_j$$

$$\therefore \nabla J(\theta) \text{ or } \nabla_{\omega, b} = \begin{bmatrix} \frac{\partial J(\theta)}{\partial \omega_1} \\ \frac{\partial J(\theta)}{\partial \omega_2} \\ \frac{\partial J(\theta)}{\partial b} \end{bmatrix} = \begin{bmatrix} (\sigma(\omega \cdot x + b) - y) x_1 \\ (\sigma(\omega \cdot x + b) - y) x_2 \\ (\sigma(\omega \cdot x + b) - y) x_3 \end{bmatrix}$$

$$= \begin{bmatrix} (\sigma(0) - 1) x_1 \\ (\sigma(0) - 1) x_2 \\ (\sigma(0) - 1) x_3 \end{bmatrix} = \begin{bmatrix} -0.5 x_1 \\ -0.5 x_2 \\ -0.5 \end{bmatrix}$$

$$\nabla_{\omega \cdot b} \text{ or } \nabla J(\theta) = \begin{bmatrix} -1.5 \\ -1.0 \\ -0.5 \end{bmatrix}$$

cep 2:

Now, we have a gradient, we compute the new parameter. vector θ^1 by moving θ^0 in the opposite direction from the gradient.

$$\theta_{new} = \theta_{prv} - \underbrace{\eta}_{\text{learning rate (how fast we can move on the surface?)}} \nabla J$$

$$\theta_{new} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} -1.5 \\ -1.0 \\ -0.5 \end{bmatrix}$$

$$\boxed{\theta_{new_1} = \begin{bmatrix} 0.15 \\ 0.1 \\ 0.05 \end{bmatrix}} \begin{matrix} w_1 \\ w_2 \\ b \end{matrix}$$

$\therefore$

Descent 2 $\Rightarrow$

Our multiple linear regression model is,

$$\hat{y} = x_1 w_1 + x_2 w_2 + b$$

$\therefore$ ~~New~~, Now we will calculate new values of $\nabla J(\theta)$

$$\nabla J(\theta) = \begin{bmatrix} \frac{\partial J(\theta)}{\partial w_1} \\ \frac{\partial J(\theta)}{\partial w_2} \\ \partial J(\theta)/\partial b \end{bmatrix} = \begin{bmatrix} (\sigma(x_1 w_1 + x_2 w_2 + b) - y)\, x_1 \\ (\sigma(x_1 w_1 + x_2 w_2 + b) - y)\, x_2 \\ (\sigma(x_1 w_1 + x_2 w_2 + b) - y) \end{bmatrix}$$ +1

$$\therefore \sigma(x_1 w_1 + x_2 w_2 + b) = \frac{1}{1 + e^{-(x_1 w_1 + x_2 w_2 + b)}}$$

$$= \frac{1}{1 + e^{-(3(0.15) + 2(0.1) + 0.05)}} = \frac{1}{1 + e^{-0.7}} = \frac{1}{1 + 0.49}$$

$$= \frac{1}{1.49} = \underline{0.67}$$

$$\nabla J(\theta) = \begin{bmatrix} -0.33x_1 \\ -0.33x_2 \\ -0.33 \end{bmatrix} = \begin{bmatrix} -0.99 \\ -0.66 \\ -0.33 \end{bmatrix}$$

Step 2 :

$$\theta_{new} = \theta_{prv} - \eta \nabla J(\theta)$$

$$= \begin{bmatrix} 0.15 \\ 0.1 \\ 0.05 \end{bmatrix} - 0.1 \begin{bmatrix} -0.99 \\ -0.66 \\ -0.33 \end{bmatrix}$$

$$= \begin{bmatrix} 0.15 \\ 0.1 \\ 0.05 \end{bmatrix} + \begin{bmatrix} 0.099 \\ 0.066 \\ 0.033 \end{bmatrix}$$

$$\theta_{new_2} = \begin{bmatrix} 0.249 \\ 0.166 \\ 0.083 \end{bmatrix}$$