04- Logistic Regression

(Logistic Reg Graph)

To solve Classification problem → Binary Classification / Multiclass Classification

Can we solve this classification problem using Regression?

No, Beacuse when new outliers is added to the data the Best fit line changes and the value might give >1 or <0 which may be inacurate

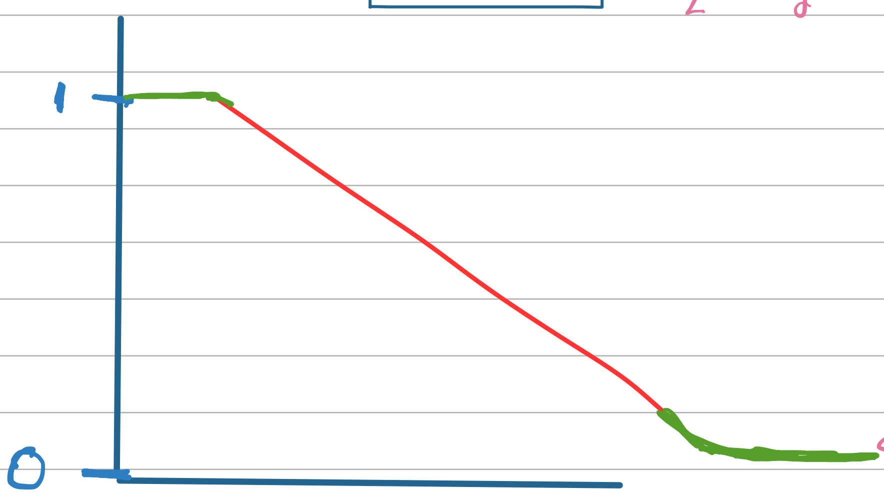
Why?

① Best fit line changes because of outliers ← Prediction goes wrong

② The outcomes >1 and <0

To solve this problem we we logistic Regression

$\boxed{0 \text{ to } 1}$ ← Squashing technique



← Squashing technique

$z = \boxed{h_\theta(x) = \theta_0 + \theta_1 x_1}$ → Best fit Line

⇓

[sigmoid Activation function]

⇓

0 to 1

$$\boxed{\sigma = \frac{1}{1+e^{-z}}}$$ => 0 to 1

$h_\theta(x) = \sigma(\theta_0 + \theta_1 x_1)$

We use sigmoid function so that we can squash the line.

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-z}}}$$

$z = \theta_0 + \theta_1 x_1$

→ logistic Regression

Linear Regression Cost function

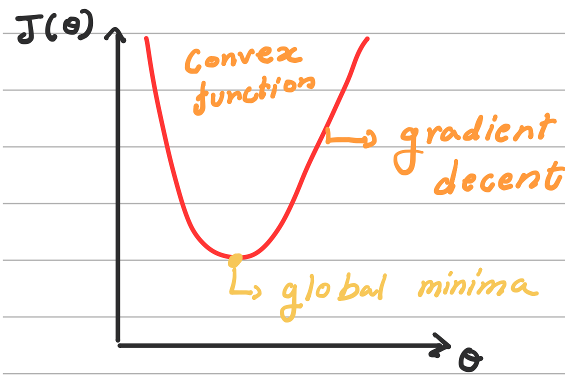
$$J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - h_\theta(x)_i)^2$$

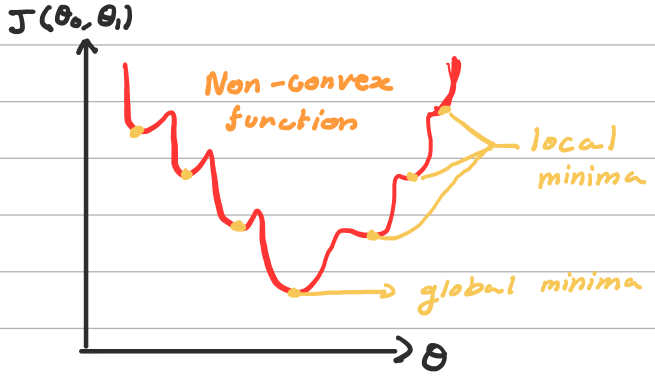
$$h_\theta(x) = \theta_0 + \theta_1 x_1$$

Logistic Regression Cost function

$$J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - h_\theta(x)_i)^2$$

$$h_\theta(x) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(\theta_0+\theta_1 x_1)}}$$





Non-Convex function: Function which may have many local minima is known as Non-convex function.

NoTe: Due to many Local minima present we dont use the the Logistic Regression cost function so we use Log Loss cost fun

Log Loss

$$h\theta(x) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{(\theta_0+\theta_1 x_1)}}$$

$$J(\theta_0,\theta_1) = \begin{cases} -\log(h\theta(x)) & \text{if } y=1 \\ -\log(1-h\theta(x)) & \text{if } y=0 \end{cases}$$

$$\boxed{J(\theta_0,\theta_1) = -y\log(h\theta(x)) - (1-y)\log(1-h\theta(x))} \Rightarrow \text{convex function}$$

condition 1:

if $\underline{y=1}$

$= -\log(h\theta(x)) \Rightarrow y=1$

if $\underline{y=0}$

$\Rightarrow -\log(1-h\theta(x))$

Logistic Regression with Regularization Parameters

cost function

$$J(\theta_0, \theta_1) = -y \log(h_\theta(x) - (1-y) - \log(1 - h_\theta(x))$$

$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1)}}$$

$$J(\theta_0, \theta_1) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

Reduce Overfitting

$$J(\theta_0, \theta_1) = -y \log(h_\theta(x) - (1-y) - \log(1 - h_\theta(x)) + L2 \text{ Reglarization}$$

$$J(\theta_0, \theta_1) = -y \log(h_\theta(x) - (1-y) - \log(1 - h_\theta(x)) + L1 \text{ Regularization} \Leftarrow \text{feature selection}$$

$$J(\theta_0, \theta_1) = -y \log(h_\theta(x) - (1-y) - \log(1 - h_\theta(x)) + L2 \text{ Reg} + L1 \text{ Reg}$$ (Elastic net)

L2 Regularization => Reduce Overfitting

$$J(\theta_0, \theta_1) = -y \log(h_\theta(x) - (1-y) - \log(1 - h_\theta(x)) + \lambda \sum_{i=1}^{n} (slope)^2$$

L1 Regularization => feature Selection

λ = hyperplane

$$J(\theta_0, \theta_1) = -y \log(h_\theta(x) - (1-y) - \log(1 - h_\theta(x)) + \lambda \sum_{i=1}^{n} |slope|$$

Elastic Net

$$J(\theta_0, \theta_1) = -y \log(h\theta(x) - (1-y) - \log(1 - h\theta(x)) + \lambda_1 \sum_{i=1}^{n} (slope)^2 + \lambda_2 \sum_{i=1}^{n} |slope|$$