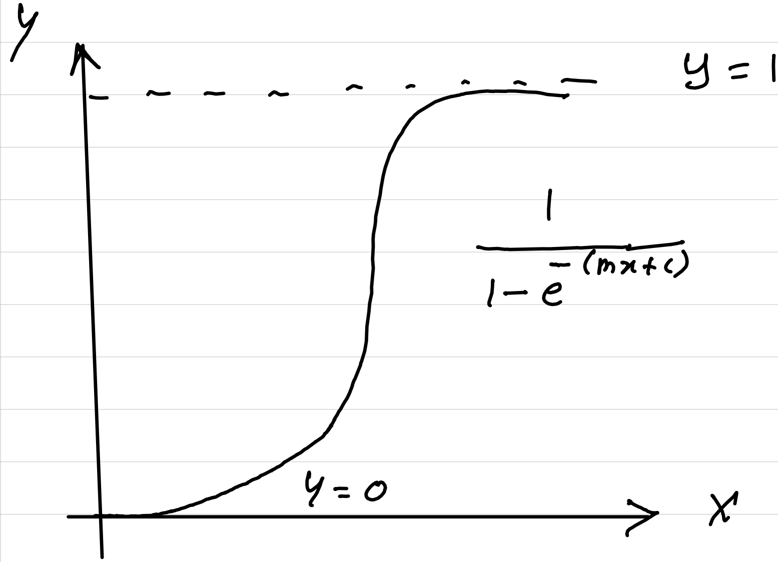
Logistic Regression



Used for classification of binary variables

Binary variables are dependant & many other variables are independant

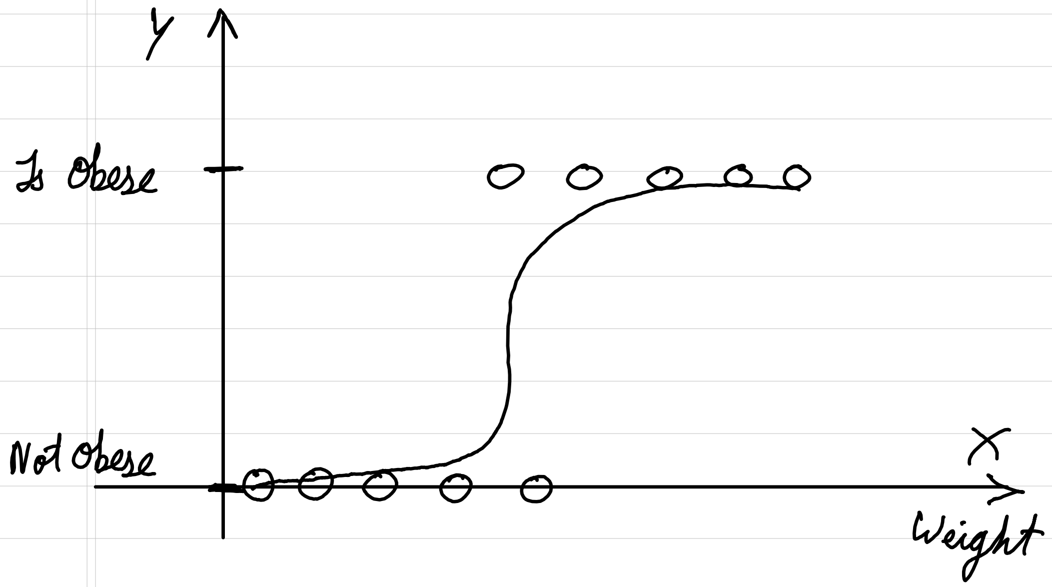
$$y = \frac{1}{1 - e^{-(mx+c)}}$$

Task is to find m & c such that the curve fits the data points well

Logistic Regression

Logistic Regression is just like linear regression where we fit a curve to a dataset

In logistic regression, we fit a logistic curve (sigmoid)

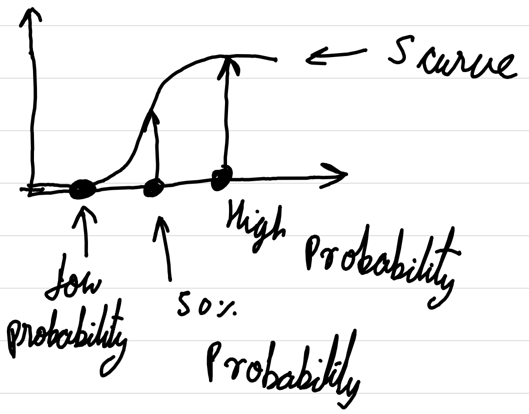


Logistic regression is used for classification

We want to predict if a data point is in class A or class B

Y is a discrete variable with two outcomes (Obese or not obese) unlike in linear regression where the Y variable was continuouse (like size)

Logistic regression gives the probability that a mouse is obese or not.



Based on the probability, the classification can be performed

This idea can be extended to multiple dimentions just like linear regression

Assumptions → ① Independant Observations (Non Time Series)
② Binary catagories
③ Logistic relationship
④ Large sample size
⑤ No outliers

Cost function

$$\begin{cases} = -\log(y') & y=1 \\ = -\log(1-y') & y=0 \end{cases}$$

$$j(\theta) = \frac{-1}{n} \Sigma\left(-y \log(y') - (1-y)\log(y')\right)$$

$$\frac{\partial j}{\partial \theta_i} = \frac{1}{n} \Sigma (y'_k - y_k) x^i_k$$

$$= \underbrace{\begin{bmatrix} y^i_k & \cdots\cdots \end{bmatrix}}_{\leftarrow k \text{ points} \rightarrow} \begin{bmatrix} x^2_k \\ x^2_k \\ x^2_k \end{bmatrix} \updownarrow k \text{ points for feature } 2$$

In linear regression, MSE is used

$$J = \frac{\sum (y - y)^2}{2}$$

Since $y = mx + c$

$J \propto m^2$

Global Minima = local Minima

But if we use the same steps for logistic regression, we get a non convex graph

Since $y = \frac{1}{1+e^{-mx}}$

J & m are related like this →

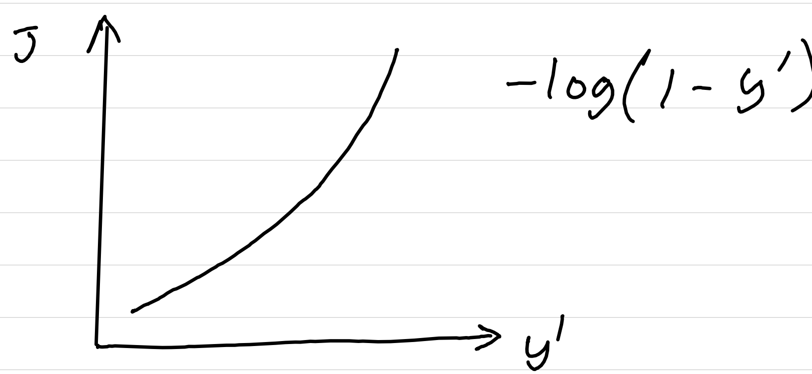
Global Minima ≠ local Minima

Hence a new error function is calculated called log loss

This is called as maximum likelyhood

If $y = 1$ ie class A

If $y = 0$ ie class B



Relation between m & J will be

$$J = -\log(y')$$

$$e^{-J} = y' = \frac{1}{1+e^{-mx+c}}$$

J & mx will be in the above curves.

Combining these curves into one equation we get

$$J = -y \log(y') - (1-y')\log(1-y')$$

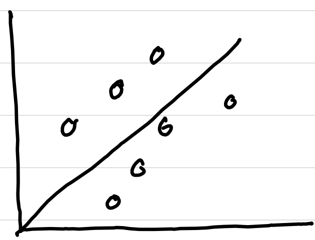
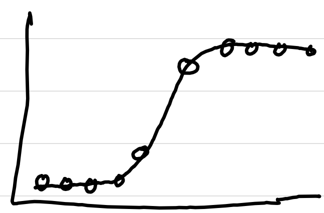
$$\frac{\partial J}{\partial w} = -y \frac{1}{y'} \frac{\partial}{\partial w}\left(\frac{1}{1+e^{-wx}}\right) + \cdots$$

$$= -\frac{y}{y'} \frac{x\, e^{-wx}}{(b+e^{-wx})^2} + \cdots$$

$$\vdots$$

$$\frac{\partial J}{\partial w} = \frac{1}{m}(y'-y)\,x \qquad [x \rightarrow \text{Matrix}]$$

In sum form, $\frac{\partial J}{\partial w} = \frac{1}{m} \sum_{i=0}^{m} (y'-y)x$

Linear Regression	Logistic Regression
Continuouse variables	Catagorical variables
Regression	Classification
Least square estimation	Maximum likelyhood estimation
Linear relationship required	Relationship based on threshold
	

Multiclass classification can be done using OvO and OvA streategies

Advantages → ① Simple
② Interpretable
③ Low variance model – lesser sensative to small variations in the dataset. When comparing with high variance models like decision trees that overfit.

Disadvantages → ① Sensative to outliers
② Cannot handle missing data
③ Assumes the logistic relationships in data

Applications of Logistic regression

→ Spam Detection, fraud detection, risk management

Q1 Hours of study vs exam percentage

Hours	0.5	1	1.5	2	2.5	2.5	2
exam %	70	65	75	60	65	70	60
Pass/fail	0	0	1	0	1	1	1

Calcutate b_0, b_1, b_2 for $\alpha = 0.3$

→ For logistic regression

$$y = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2)}}$$

$$\frac{\partial J}{\partial b_0} = \frac{1}{m} \Sigma (y' - y)$$

$$\frac{\partial J}{\partial b_1} = \frac{1}{m} \Sigma (y' - y) x_1$$

$$\frac{\partial J}{\partial b_2} = \frac{1}{m} \Sigma (y' - y) x_2$$

Step 1 : Initialize to 0

$$\left.\begin{array}{l} b_0 = 0 \\ b_1 = 0 \\ b_2 = 0 \end{array}\right\} \therefore \text{Prediction } y' = \frac{1}{1 + e^{-0}} = \frac{1}{2}$$

$$= 0.5$$

$$\frac{\partial J}{\partial b_0} = \frac{1}{m} \Sigma (y' - y)$$

$$= \frac{1}{7} \Big[(0.5 - 0) + (0.5 - 0) + (0.5 - 1) + (0.5 - 0) + (0.5 - 1) + (0.5 - 1) + (0.5 - 1) \Big]$$

$$= \frac{1}{7} \times -0.5$$