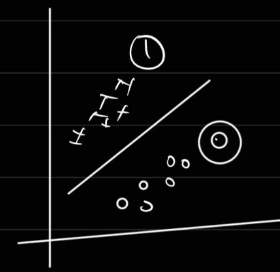
Logistic Regression for multiclass classification

* binary (dichotomous) classification problem < 1 / 0

  eg. diabetic | not diabetic

  eg. pass / fail



* multiclass classification (more than two) < 1 / 2 / 3

  ex. iris <

  eg. Student < A / B / C

  eg. Student < Science / Arts / Commerce

  eg. Placement < Yes / No / optout of the placement.

Solution :-

① OVR → (One Versus Rest)

② multinomial

① One Vs Rest

Till now $x_1 — x_n$ ⟶ Logistic Regression model → Probability value → Threshold

$\geq 0.5 \rightarrow 1$

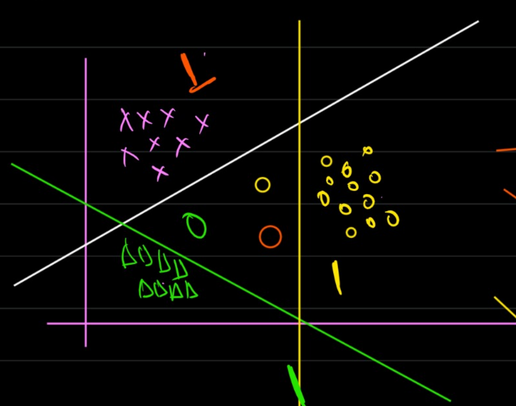
$< 0.5 \rightarrow 0$

$y <$ 1 / 2

Now

$y <$ 1 / 2 / 3

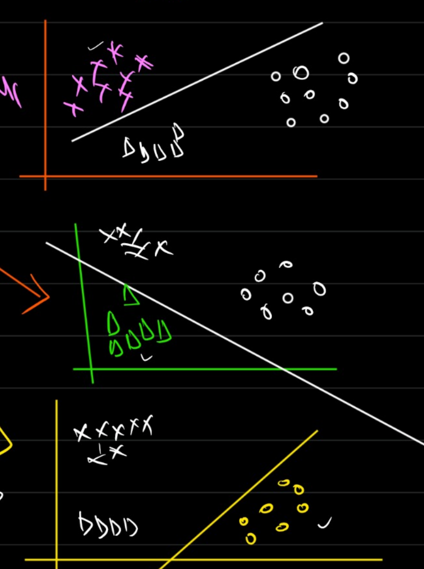
One Vs Rest → Logistic Regression model.



M1

M2

M3



X_1	X_2	O/P
—	—	1
—	—	2
—	—	3
—	—	1
—	—	1
—	—	2
—	—	2

* No of logistic regression model built = No of classes.

x_1	x_2	o/p	M_1 (1)	M_2 (2)	M_3 (3)
—	—	1	1	0	0
—	—	1	1	0	0
—	—	2	0	1	0
—	—	3	0	0	1
—	—	1	1	0	0
—	—	3	0	0	1

OHE

	1	2	3
1	1	0	0
2	0	1	0
3	0	0	1

One Vs Rest

M_1 - class 1 Vs Rest (1 vs (2,3))

M_2 — class 2 Vs Res (2 vs (1,3))

M_3 — class 3 Vs Rest → 3 vs (1,2)

$M_1 \rightarrow$ [x_1, x_2, M_1]

$M_2 \leftarrow$ [x_1 x_2 M_2]

$M_3 \leftarrow$ [x_1 x_2 M_3]

Logistic → Binary Class!

VK. — Batting, Bowling, Fielding

Bumrah — Bat, Bow, Fielding

Fielder — Bat, Bow, Fielding

One Vs Rest

x_1	x_2	---	x_n	M_1 (1)	M_2 (2)	M_3 (3)	
—	—		—	0.8	0.6	0.9	→ 3
—	—		—	0.6	0.8	0.5	→ 2
—		—	—				

M_1 (1): Expert for Class 1 — There is 60% changee that this row belongs to Class 1

M_2 (2): Expert for class 2 — 80% change class 2

M_3 (3): Expert for class 3 — 50% Chance belong 3

→ You will assign the predicted class whichever gives the highest probability.

Step-1 # model = # class
Step-2 — prob of each model
Step-3 — Attach class with highest prob value.

disadvantage

→ A binary classification model is trained for each class. ⇒ Computationally expensive.

② multinomial method / softmax Regression.

→ we don't decompose the problem into binary classification.
→ Modify the loss / cost function.
→ Single model.

$$\text{Sigmoid} = \frac{1}{1+e^{-z}} \quad \Bigg| \quad \text{Softmax } \sigma(Z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

where j is no of class.

$$\sigma(z)_1 = \frac{e^{z_1}}{e^{z_1}+e^{z_2}+e^{z_3}}, \quad \sigma(z)_2 = \frac{e^{z_2}}{e^{z_1}+e^{z_2}+e^{z_3}}, \quad \sigma(z)_3 = \frac{e^{z_3}}{e^{z_1}+e^{z_2}+e^{z_3}}$$

Cost Fun / Loss fn (Log reg)

$$= -\frac{1}{n}\sum_{i=1}^{n} y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i$$

Modification in CF

$$\Rightarrow -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} y_k^{(i)} \log \hat{y}_k^{(i)}$$

(n – no of dp
K – no of class)

x_1	x_2	y	$y_{k=1}$	$y_{k=2}$	$y_{k=3}$
x_{11}	x_{12}	1	1	0	0
x_{21}	x_{22}	2	0	1	0
x_{31}	x_{32}	3	0	0	1

$$y_1^{(1)} \log \hat{y}_1^{(1)} + \underbrace{y_2^{(1)} \log \hat{y}_2^{(1)} + y_3^{(1)} \log \hat{y}_3^{(1)}}_{0} +$$

$$\underbrace{y_1^{(2)} \log \hat{y}_1^{(2)}}_{0} + y_2^{(2)} \log \hat{y}_2^{(2)} + \underbrace{y_3^{(2)} \log \hat{y}_3^{(2)}}_{0} +$$

$$\underbrace{y_1^{(3)} \log \hat{y}_1^{(3)} + y_2^{(3)} \log y_2^{(3)}}_{0} + y_3^{(3)} \log y_3^{(3)}$$

$$CF = y_1^{(1)} \log \hat{y}_1^{(1)} + y_2^{(2)} \log \hat{y}_2^{(2)} + y_3 \log \hat{y}_3^{(3)}$$

$$\hat{y}_1^{(1)}, \hat{y}_2^{(2)}, \hat{y}_3^{(3)}$$

$$\begin{bmatrix} \theta_0^{(1)} & \theta_1^{(1)} & \theta_2^{(1)} \\ \theta_0^{(2)} & \theta_1^{(2)} & \theta_2^{(2)} \\ \theta_0^{(3)} & \theta_1^{(3)} & \theta_3^{(3)} \end{bmatrix} \Leftarrow \left\{ \begin{aligned} \hat{y}_1^{(1)} &= \overset{\text{softmax}}{\sigma}\left(\theta_0^{(1)} + \theta_1^{(1)} x_{11} + \theta_2^{(1)} x_{12}\right) \\ \hat{y}_2^{(2)} &= \sigma\left(\theta_0^{(2)} + \theta_1^{(2)} x_{21} + \theta_2^{(2)} x_{22}\right) \\ \hat{y}_3^{(3)} &= \sigma\left(\theta_0^{(3)} + \theta_1^{(3)} x_{31} + \theta_3^{(3)} x_{33}\right) \end{aligned} \right.$$

→ gradient descent.

$$\frac{\partial L}{\partial \theta_0^{(1)}}, \frac{\partial L}{\partial \theta_0^{(2)}} - - - \quad 9 \text{ differentials}$$

Convergence

Simultaneously update thetas.

$$\left\{ \begin{aligned} \theta_1^{(1)} &= \theta_1^{(1)} - \eta \frac{\partial L}{\partial \theta_1} \\ &\;\vdots \end{aligned} \right.$$