LOGISTIC REGRESSION 2

$$\text{log loss} = -\frac{1}{m}\sum_{i=1}^{m} \underbrace{y_i}_{\text{actual}} \log(\overbrace{\hat{y}_i}^{\text{predicted.}}) + (1-y_i)\log(1-\hat{y}_i)$$

$$\underset{w_j}{\text{Arg min}}\,[\text{loss}]$$

Gradient Desunt

$$w_{new} = w_{old} - \eta \frac{\partial L}{\partial w}$$

$$\frac{d}{dz_i}\hat{y}_i = \frac{d}{dz_i}\sigma(z_i) = \frac{d}{dz_i}\sigma(w^Tx+w_0) = \frac{d}{dz_i}\sigma(w_1x_1 - \cdots w_0)$$

$$= \frac{1}{1+e^{-(w^Tx+w_0)}}$$

$$f(x) = \sigma(x) = \frac{1}{1+e\cdot x}$$

$$f'(x) = f(x)\left[1-f(x)\right]$$

$$\boxed{\frac{d}{dz_i}\hat{y}_i = \hat{y}_i(1-\hat{y}_i)} \quad \Bigg| \quad \frac{d}{dz_i}\sigma(z_i) = \sigma(z_i)\left[1-\sigma(z_i)\right]$$

$$\underset{w_j}{\mathrm{Argmin}} \; -\frac{1}{n}\sum_{i=1}^{n} \underbrace{y_i \log(\hat{y}_i)}_{A} + \underbrace{(1-y_i)\log(1-\hat{y}_i)}_{B}$$

for Only 1 point

$$L = -\left[\underbrace{y_i \log(\hat{y}_i)}_{A} + \underbrace{(1-y_i)\log(1-\hat{y}_i)}_{B}\right]$$

$$L = -\left[A + B \right]$$

$$\frac{\partial L}{\partial w_j} = -\left[\overset{A}{\frac{\partial A}{\partial w_j}} + \overset{B}{\frac{\partial B}{\partial w_j}} \right]$$

$$y_i \log(\hat{y}_i)$$

$$y_i \log(\sigma(z_i))$$

$$y_i \log\left(\frac{1}{1+e^{-(w^T x + w_0)}}\right)$$

A

$$\frac{dA}{dw_j} = \frac{\partial A}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_j}$$

$$= y\left(\frac{1}{\cancel{\hat{y}}}\right) \cancel{\hat{y}} \cdot (1-\hat{y}) \, (x_j)$$

$$z_i = w^T x + w_0$$

$$\frac{\partial z_i}{\partial w_j}$$

$$\hat{y}_i = \sigma(z_i)$$

$$\frac{\partial \hat{y}_i}{\partial z_i} = f(x)\,(1-f(x))$$

$$\boxed{\frac{\partial A}{\partial w_j} = y \cdot (1-\hat{y}) \quad (x_j)}$$

$$B = (1-y_i)\log(1-\hat{y}_i)$$

$$\frac{\partial B}{\partial w_j} = \frac{\partial B}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_j}$$

$$= \frac{1-y_i}{\cancel{1-\hat{y}_i}} \cdot (-1) \cdot \hat{y}_i \cancel{(1-\hat{y}_i)} \cdot (x_j)$$

$$\underbrace{\sigma(z_i)}_{\hat{y}}\underbrace{(1-\sigma(z))}_{(1-\hat{y})}$$

$$A = -y\log\hat{y}$$

$$\frac{\partial A}{\partial \hat{y}} = -\frac{y}{\hat{y}}$$

$$A = -\overset{y}{3}\log\overset{\hat{y}_i}{x}$$

$$\frac{\partial A}{\partial x} = -\frac{3}{x}$$

$$\frac{\partial A}{\partial \hat{y}_i} = -\frac{3}{\hat{y}_i}$$

$$B = \underset{\text{Const}}{\boxed{(1-y_i)}}\log(1-\boxed{\hat{y}_i})$$

$$\frac{\partial B}{\partial \hat{y}_i} = (1-y_i)$$

$$f(x) = \log(1-x)$$

$$f'(x) = \frac{1}{1-x}\cdot(-1)$$

$$B = (1-y_i)(\log(1-\hat{y}_i))$$

$$\frac{\partial B}{\partial \hat{y}_i} = \frac{1-y_i}{1-\hat{y}_i}\cdot(-1)$$

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

$$\frac{\partial \hat{y}_i}{\partial z_i} = f(x)(1-f(x))$$

$$= \hat{y}_i(1-\hat{y}_i)$$

$$\frac{d}{dz_i}\sigma(z_i) = \sigma(z_i)\left[1-\sigma(z_i)\right]$$

$$\frac{d\hat{y}_i}{\partial z_i} = \hat{y}_i(1-\hat{y}_i)$$

$$\boxed{\frac{\partial B}{\partial w_j} = (1-y)(x_j)(-1)(\hat{y}_i)}$$

$$\frac{\partial L}{\partial w_j} = -\left[\frac{\partial A}{\partial w_j} + \frac{\partial B}{\partial w_j}\right]$$

$$= -\left[(1-\hat{y}_i)\, y \cdot x_j - (1-y)\,\hat{y}_i\, x_j\right]$$

$$= -\left[y x_j - \cancel{y \hat{y}_i x_j} - \hat{y}_i x_j + \cancel{y \hat{y}_i x_j}\right]$$

$$\frac{\partial L}{\partial w_j} = -\left[x_j \left(y_i - \hat{y}_i \right) \right]$$

$$\boxed{\frac{\partial L}{\partial w_j} = x_j \left(\hat{y} - y \right)}$$

$$\frac{\partial L}{\partial w} = -\frac{1}{n} \sum_{i=1}^{n} \left(\hat{y}_i - y_i \right) x_j$$

final expression.

Linear Reg

$$\hat{y} = w^T x + w_0$$

Log Reg

$$\hat{y} = \frac{1}{1+e^{-(w^T x + w_0)}}$$

Gradient Descent

$\Rightarrow \quad w^{t+1} = w^t - \eta \boxed{\frac{\partial L}{\partial w}}$

① Pick random values of $w_1 \quad w_2 \quad w_3 \; \ldots \; w_d$

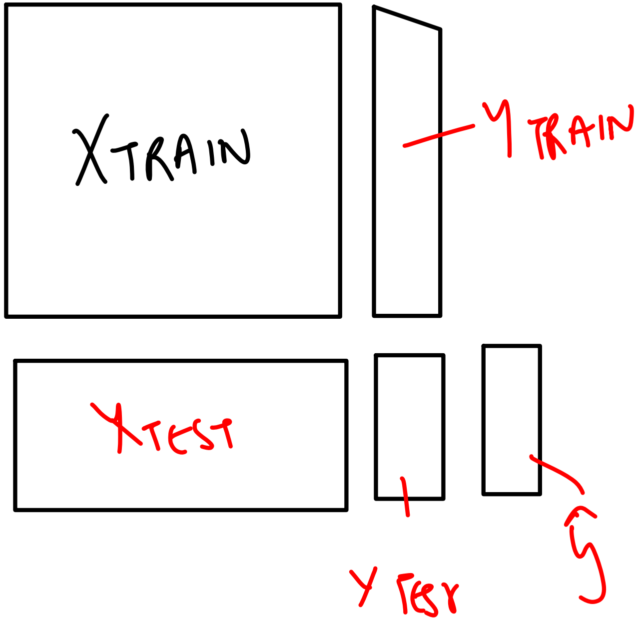
② $\frac{\partial L^0}{\partial w_0}, \quad \frac{\partial L^0}{\partial w_1}, \quad \frac{\partial L^0}{\partial w_2} \quad \ldots \quad \frac{\partial L^0}{\partial w_d}$

③ $w_1^{t+1} = w_1^t - \boxed{\eta} \frac{\partial L^0}{\partial w_1}$ → hyperparameter.

$\boxed{w_2^{t+1}}$ $\quad$ $\boxed{w_3^{t+1}}$

$\boxed{w_1^{t+1}}$

Evaluation Metric → R^2, R^2 adjusted,



Y_{TEST}	$\hat{y}$	
0	0	✓
1	0	✗
1	1	✓
0	0	✓
1	0	✗
0	1	✗

$$y_i = \hat{y}_i$$

$$\text{Accuracy} = \frac{\text{no. of correct predictions}}{\text{Total no. of prediction}} = \frac{3}{6} = 0.5$$

$$= \underline{\underline{50\%}}$$

---

y_{true}	$\hat{y}$
0.5	0.49
0 2	0.22

$$(\hat{y} - y)$$

LOG ODDS

odds of winning . 4:1

Odds of horse winning a race ⇒ 4:1

$$\text{Odds} = \frac{\text{Probability of Success}}{\text{Probability of failure}} \Rightarrow$$

$$P(\text{winning}) = \frac{4}{4+1} = \frac{4}{5}$$

$$P(\text{failing}) = \frac{1}{1+4} = \frac{1}{5}$$

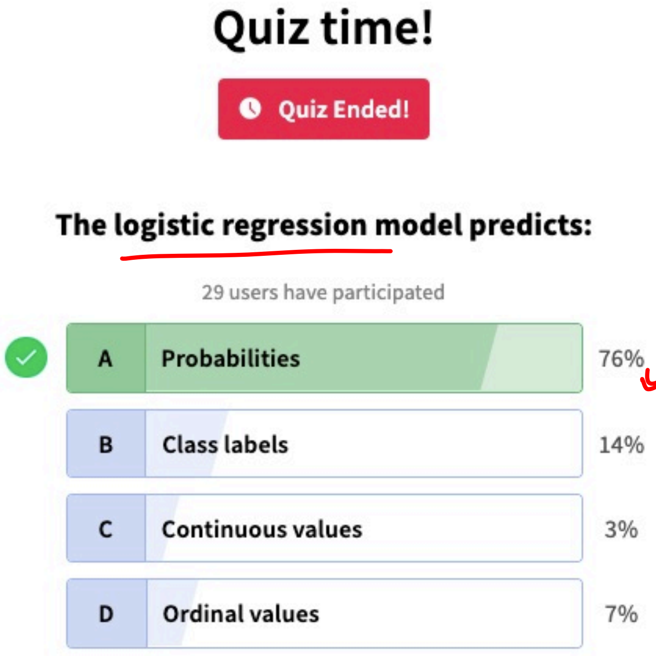
$P(\text{success}) = p$

$P(\text{failure}) = 1 - p$

$$\boxed{\text{odds} = \frac{p}{1-p}}$$

$$\sigma(z_i) = \underset{(0,1)}{\hat{y}_i} = P[\underset{(0,1)}{y_i = 1} \mid x_i] = p$$

$$P[y_i = 0 \mid x_i] = 1 - p$$



Quiz time!

Quiz Ended!

The logistic regression model predicts:

29 users have participated

✓	A	Probabilities	76%
	B	Class labels	14%
	C	Continuous values	3%
	D	Ordinal values	7%

✓ $\underline{\sigma(z_i)}$ $(0,1)$

$$P[y_i=1 \mid x_i] = P = \sigma(z_i)$$

$$P = \sigma(w^T x + w_0)$$

$$= \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(w^T x + w_0)}}$$

$$= \frac{1}{1+\frac{1}{e^{(w^T x + w_0)}}}$$

$$\frac{1}{\frac{e^{w^T x + w_0}+1}{e^{w^T x + w_0}}}$$

$$\Rightarrow \boxed{\frac{e^{w^T x + w_0}}{1+e^{w^T x + w_0}} = P}$$

$$1-p = 1 - \frac{e^{w^Tx+w_0}}{1+e^{w^Tx+w_0}} = \frac{1\cancel{+e^{w^Tx+w_0}}\cancel{-e^{w^Tx+w_0}}}{1+e^{w^Tx+w_0}}$$

$$1-p = \frac{1}{1+e^{w^Tx+w_0}}$$

$$\underline{\text{odds}} = \frac{p}{1-p} = \frac{\frac{e^z}{\cancel{1+e^z}}}{\frac{1}{\cancel{1+e^z}}} = e^z$$

$$= e^{w^Tx+w_0}$$

$$\frac{p}{1-p} = e^z$$

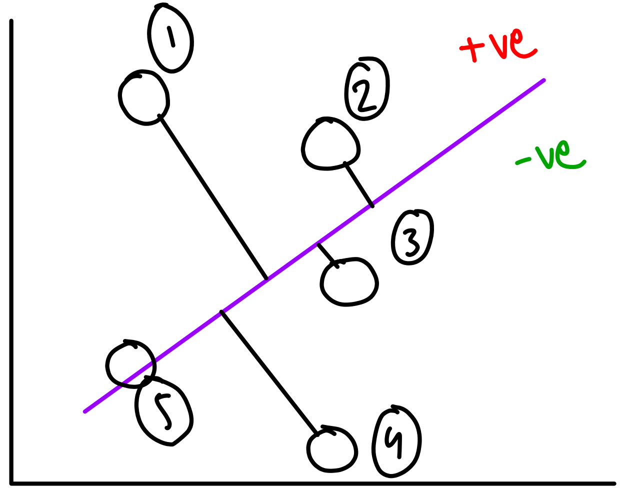
$$\log_e\left[\frac{p}{1-p}\right] = z = w^T x + w_0$$

↳ distance of point from line

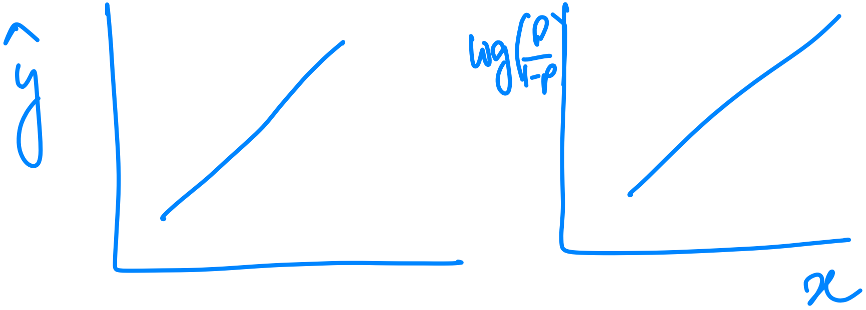
log(odds) = distance of point from line.

LOG ODDS INTERPRETATION OF LOG REG.

Linear Regression ⇒ $\hat{y} = w^T x + w_0$

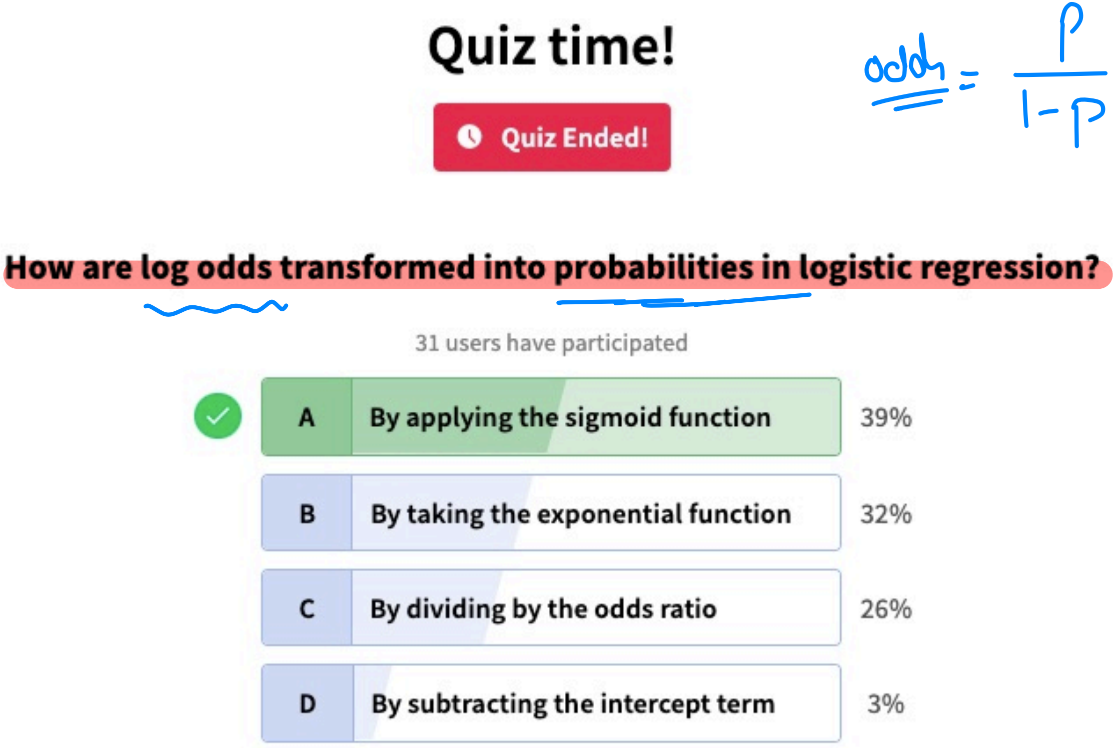


odds of x_1 $>$ odds of x_2



Log Rg $\Rightarrow$ $\log\left(\frac{p}{1-p}\right) = w^T x + w_0$

Non Linear Relationship



Quiz time!

How are log odds transformed into probabilities in logistic regression?

31 users have participated

✓	A	By applying the sigmoid function	39%
	B	By taking the exponential function	32%
	C	By dividing by the odds ratio	26%
	D	By subtracting the intercept term	3%

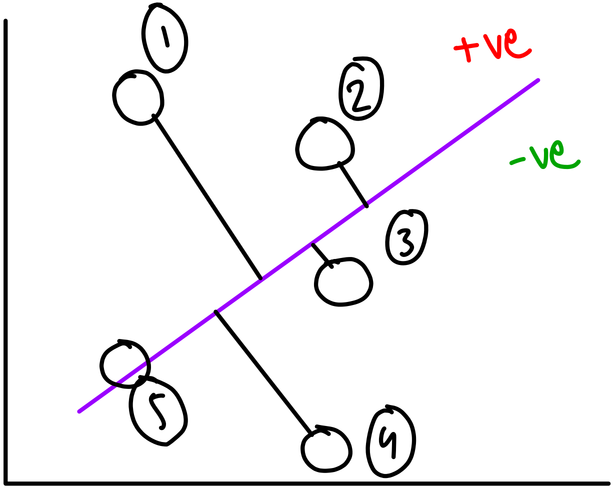
$$\text{odds} = \frac{p}{1-p}$$

$$\log\left(\frac{p}{1-p}\right)$$

Impact of Outliers

● 6 Outlier

① Outlier is on the correct side.



$$L = -\left[y\log\hat{y} + (1-y)\log(1-\hat{y})\right]$$

$y = 1$ $\qquad$ $\hat{y} = 1$

$L = -\log\hat{y} \Rightarrow \log 1 \rightarrow 0$

$\boxed{L = \text{Zero}}$.

NO IMPACT

② Outlier are present on the opposite side.

$$L = -\left[y\log\hat{y} + (1-y)\log(1-y_i)\right]$$

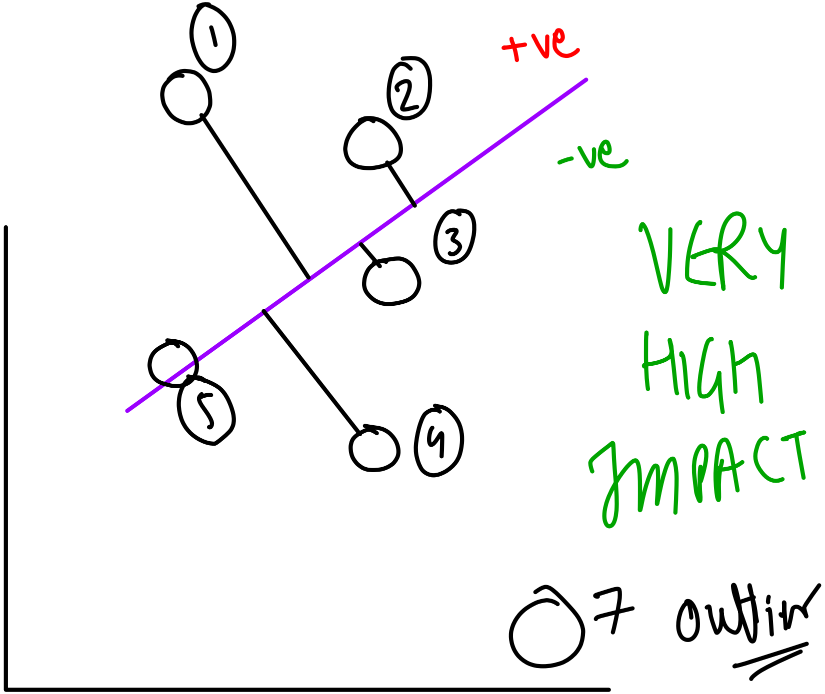
$y_i = 1$ $\quad \hat{y}_i = 0$

$\sigma(z_i) =$

$\hat{y} = \sigma(\underbrace{\boxed{z_i}}_{w^Tx+w_0})$

large Number

-ve number.

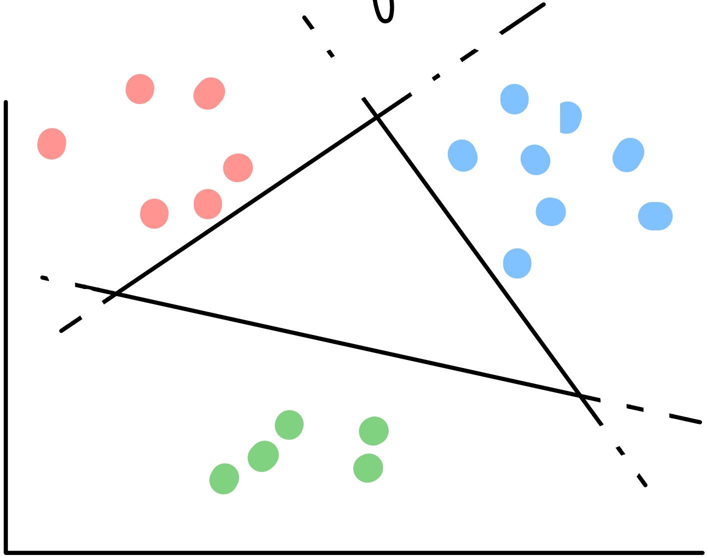


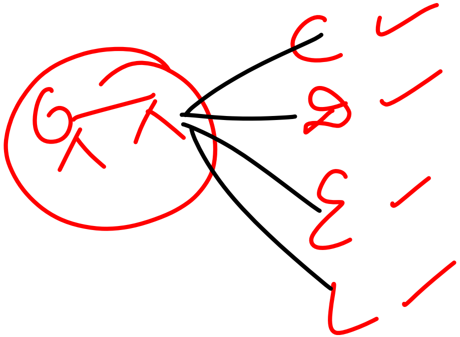
VERY HIGH IMPACT

$\sigma(z_i) \to 0 \Rightarrow \log 0 \to$ large

{0,1} Yes/No, CAT/DOG

Multi Class Classification





Class 1 → W

class 2 → S

class 3 → T

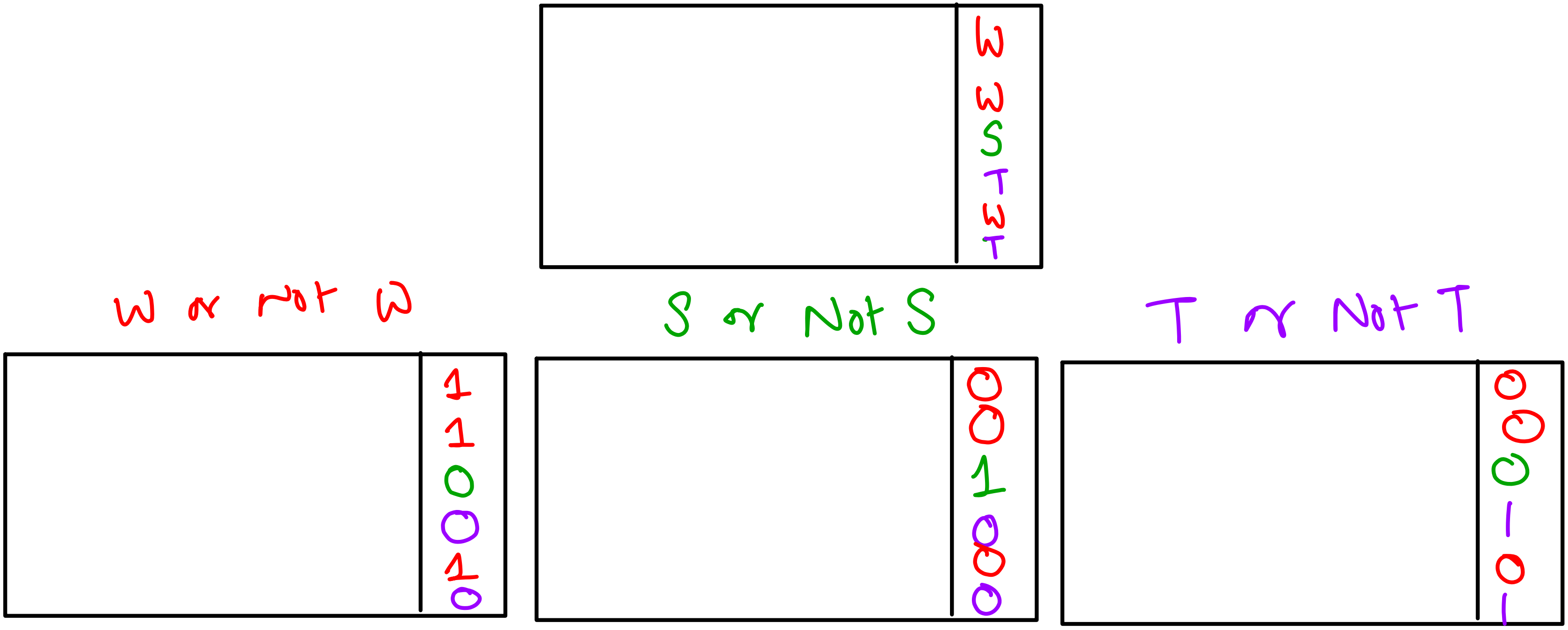
Sol^n

Create 3 diff Log Reg Models.

① W or Not W

② S or Not S

③ T or Not T



if $y_i == W$
⇒ 1

⇒ 0

if $y_i == S$
⇒ 1

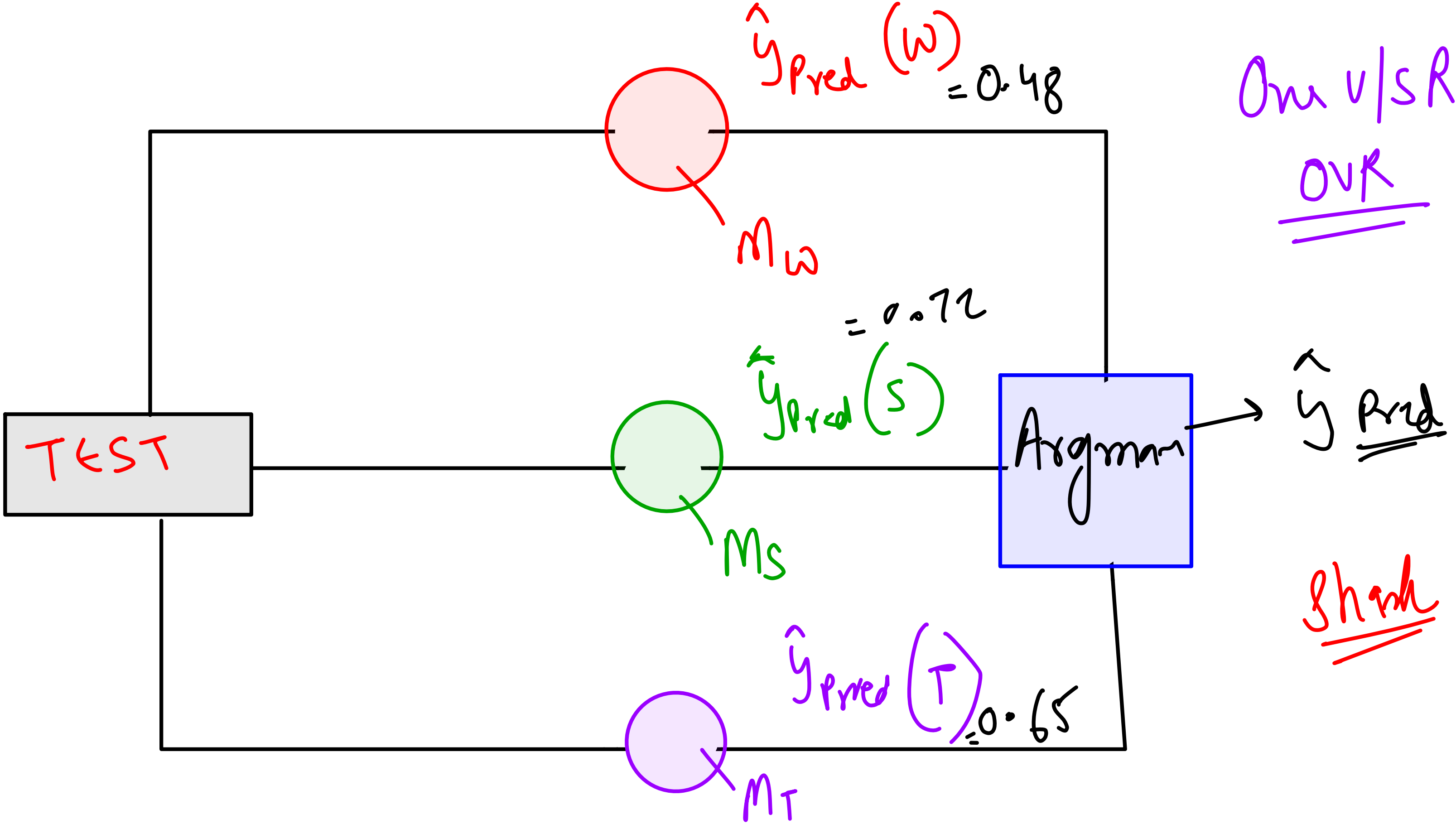
⇒ 0

if $y_i = T$
⇒ 1

⇒ 0

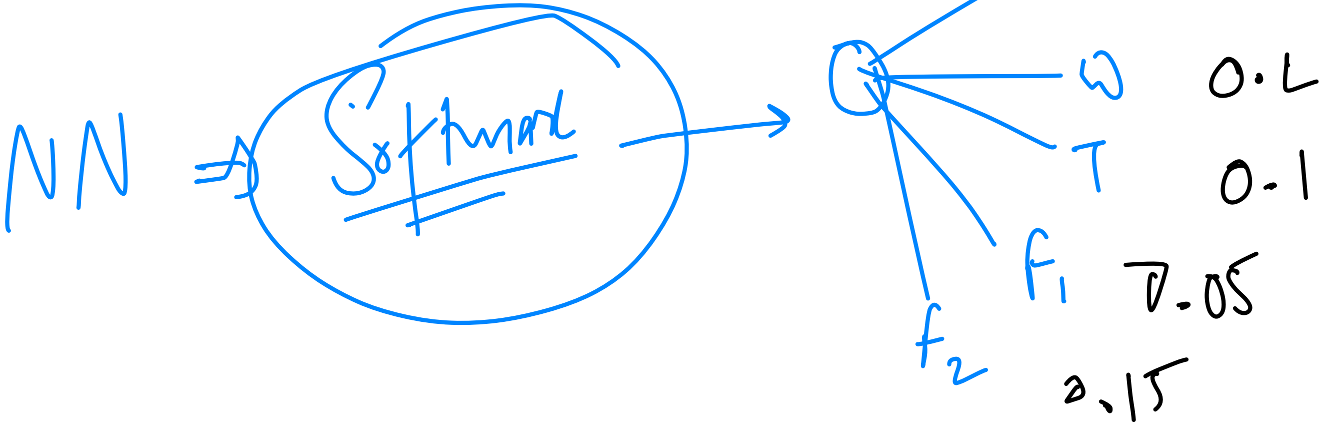
One v/s Rest

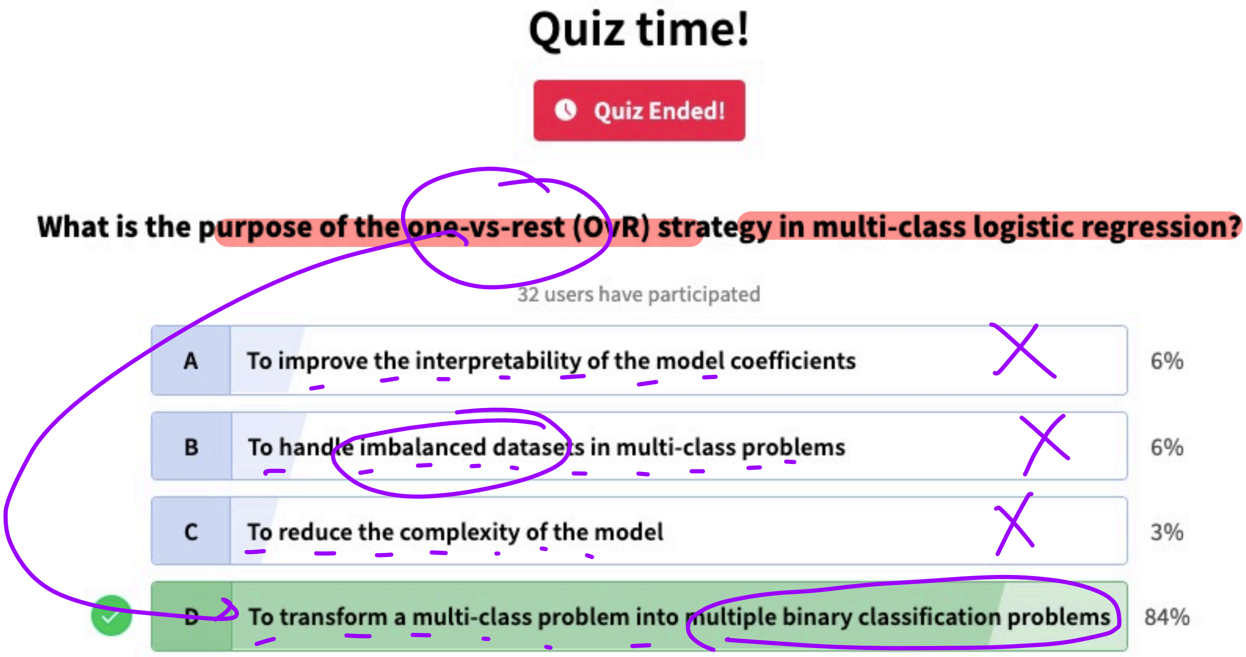
OvR



Shah

logistic $= \dfrac{1}{1+e^{-x}}$ $(0 \quad 1)$





Quiz time!

What is the purpose of the one-vs-rest (OvR) strategy in multi-class logistic regression?

32 users have participated

	A	To improve the interpretability of the model coefficients	6%
	B	To handle imbalanced datasets in multi-class problems	6%
	C	To reduce the complexity of the model	3%
✓	D	To transform a multi-class problem into multiple binary classification problems	84%

① Metrics

Precision
Recall
Confusion Matrix
F1 Score
ROC AUC
Type of error

② Code → ~~log reg~~ → walk through

① 👍
② 👎
walk through