LOGISTIC REGRESSION

AI CLUB

REGRESSION

Regression is used to predict continuous numerical values

CLASSIFICATION

Classification is used to predict categorical labels.

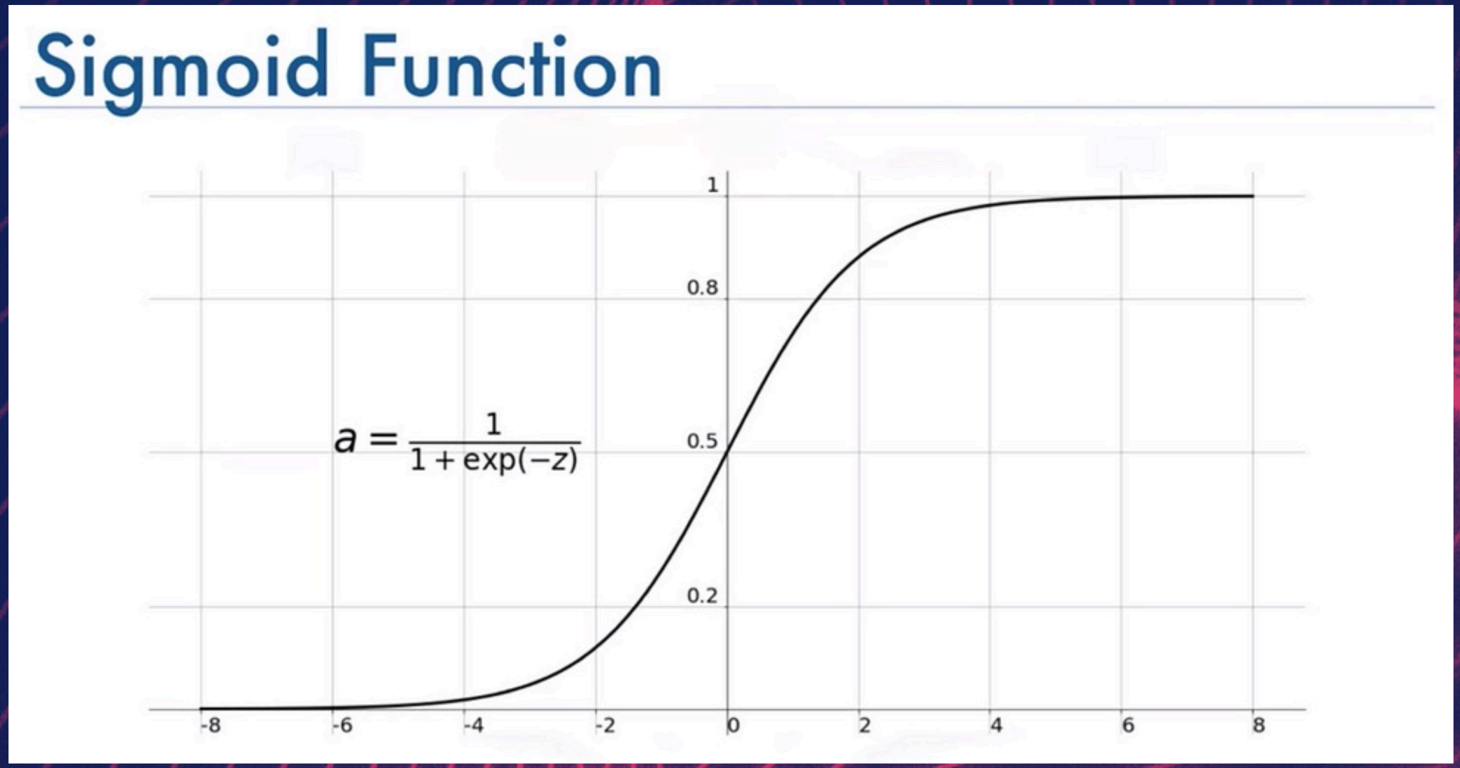
WHY IS IT NOT CALLED LOGISTIC CLASSIFICATION?

In linear regression, our main aim is to estimate the values of Y-intercept and weights, minimize the cost function, and predict the output variable Y.

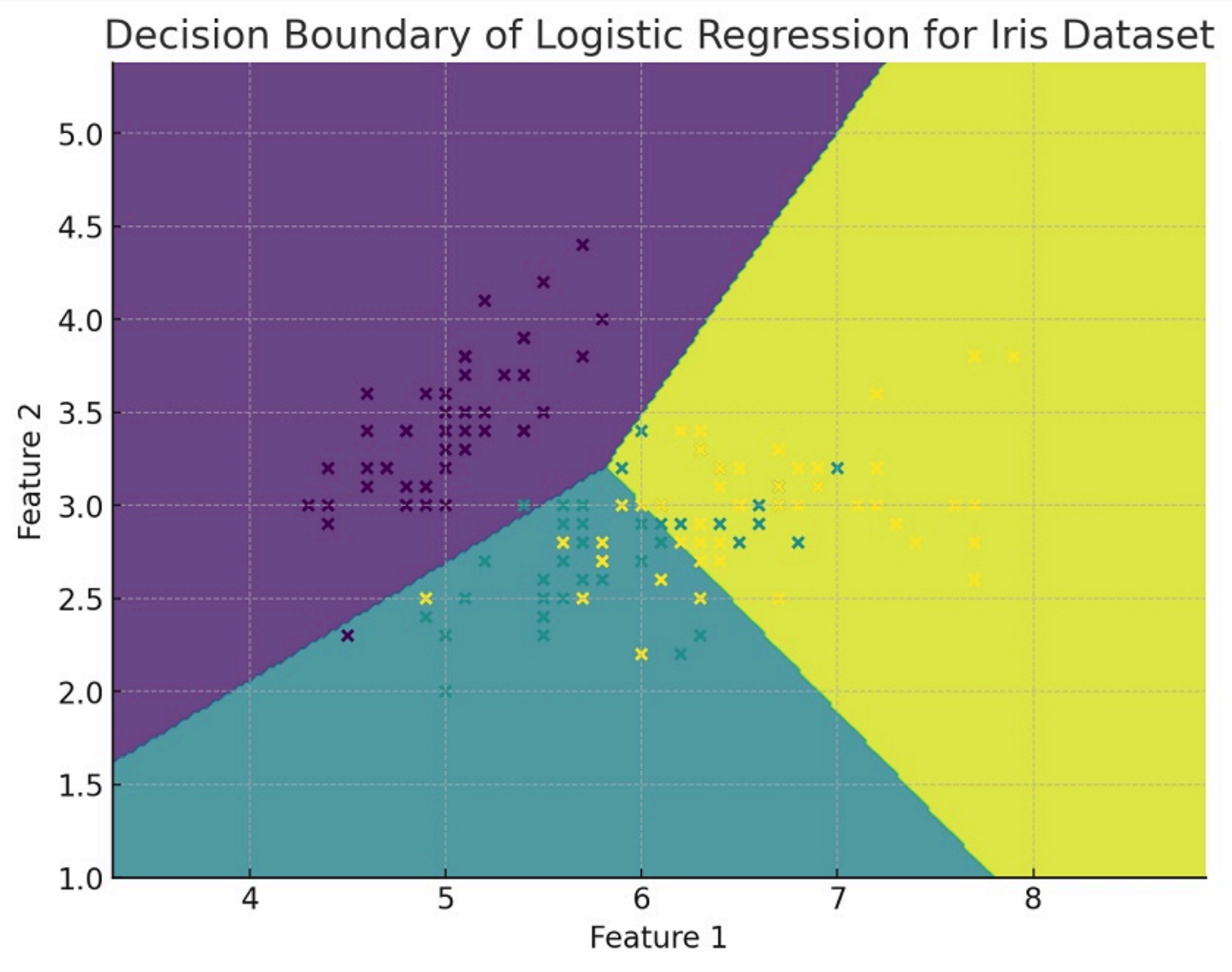
In logistic regression, we perform the exact same thing but with one small addition. We pass the result through a special function known as the Sigmoid Function to predict the output Y.

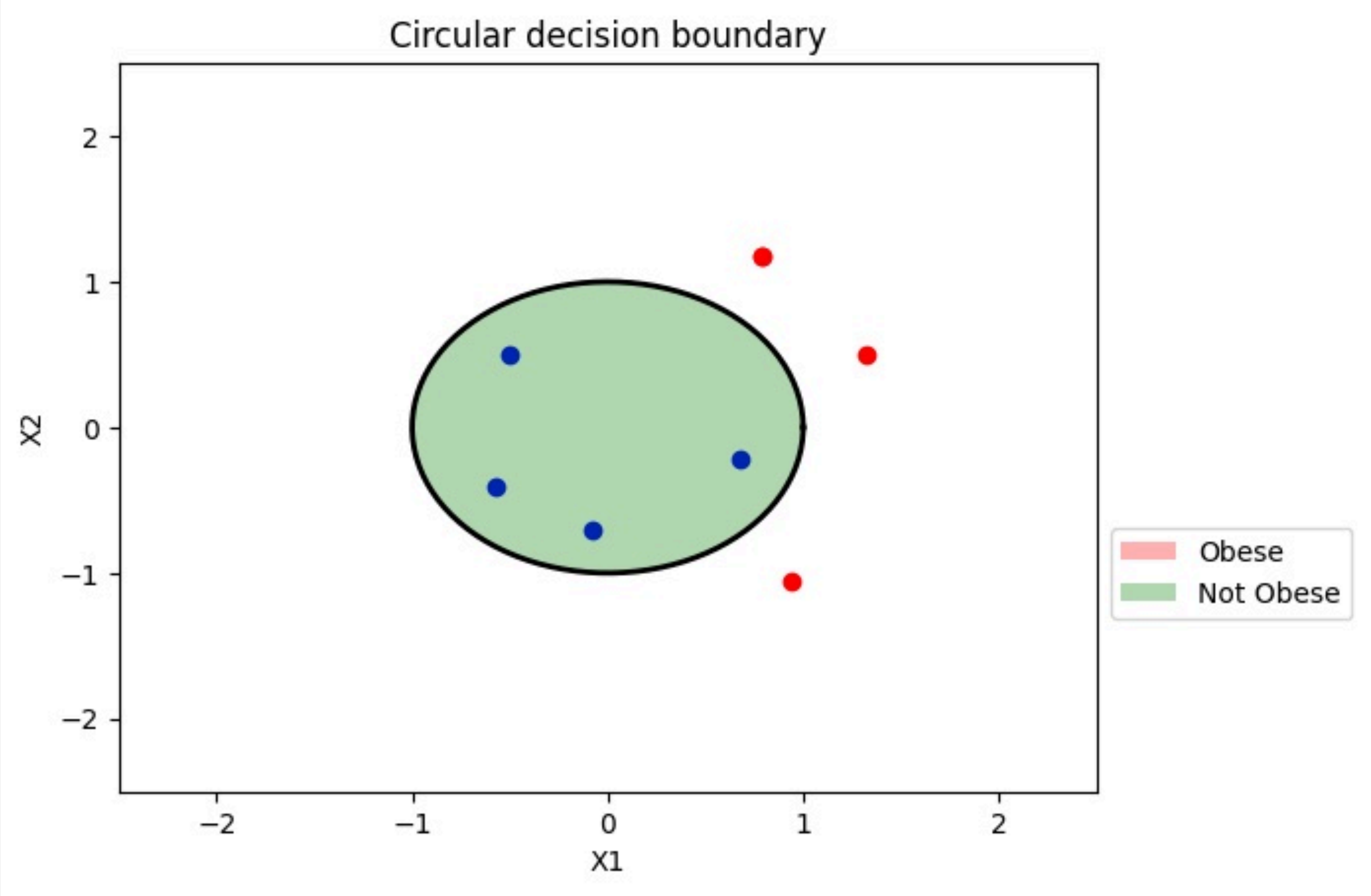
SIGMOID OR LOGISTIC FUNCTION

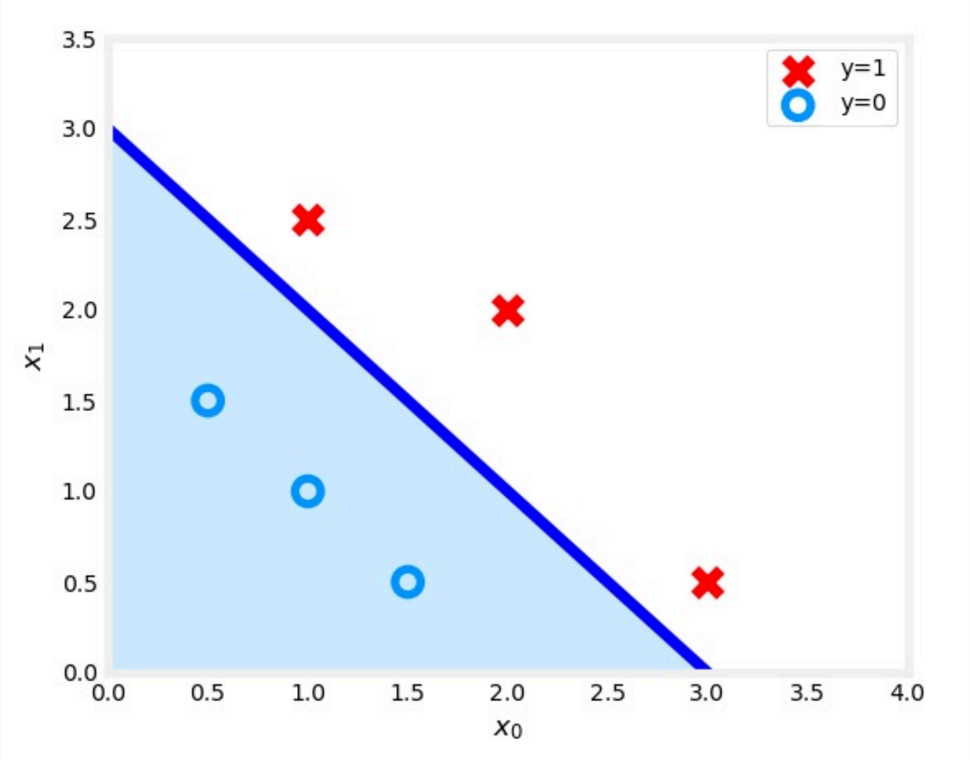
- The Linear regression model would predict a number between -infinity to +infinity
- However, we would like the predictions of our classification model to be between 0 and 1 since our output is either 0 or 1.
- This can be accomplished by using a "sigmoid function" which maps all input values to values between 0 and 1.



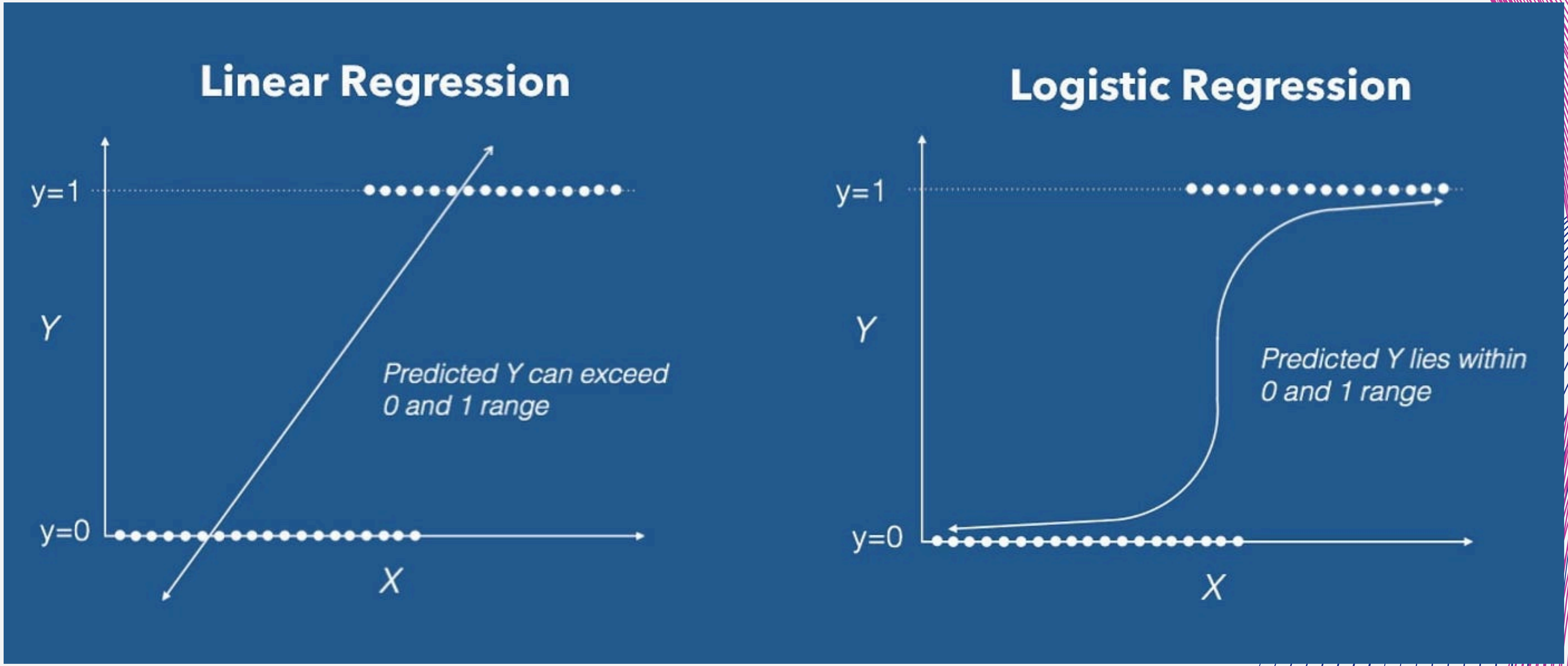
DECISION BOUNDARY



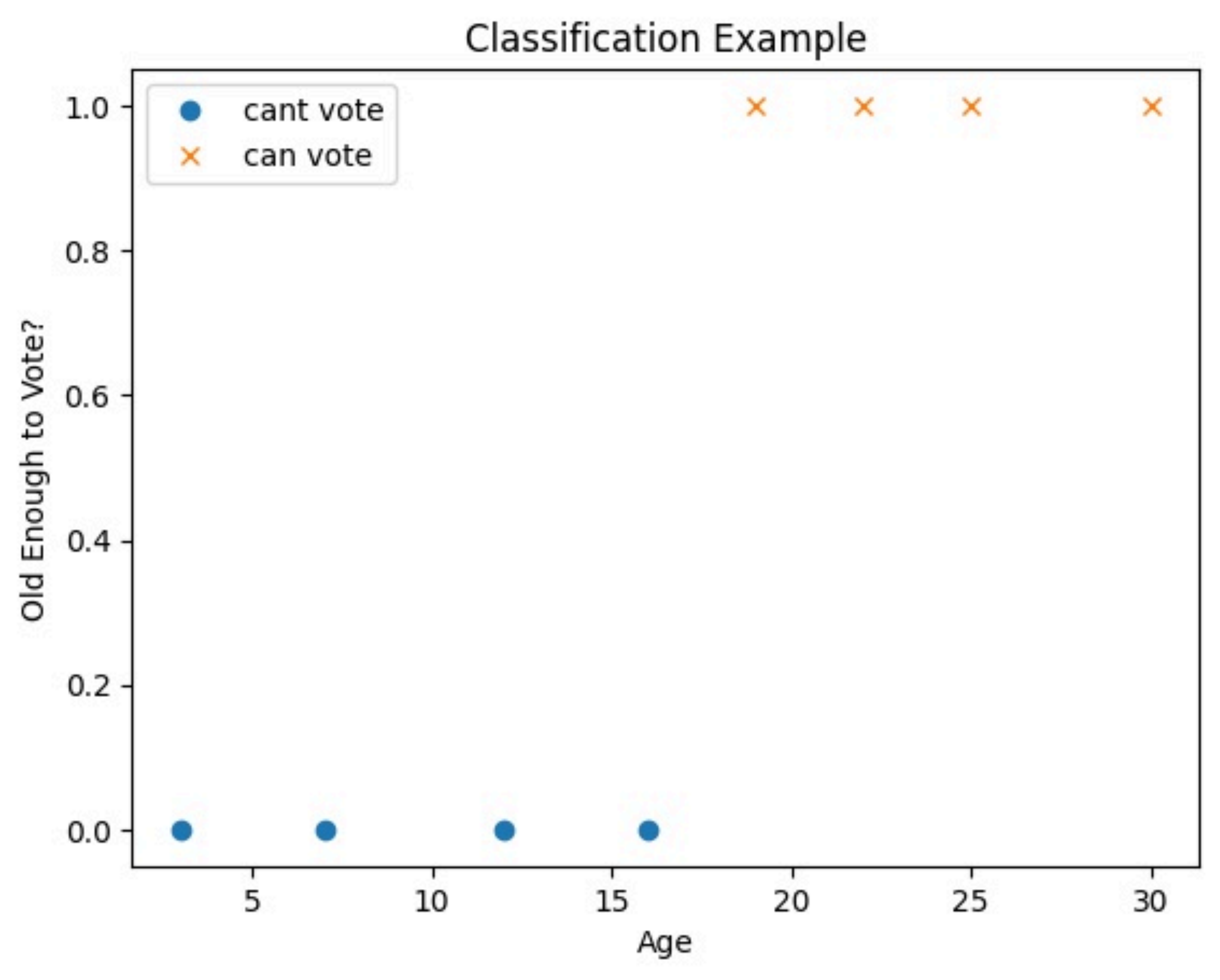


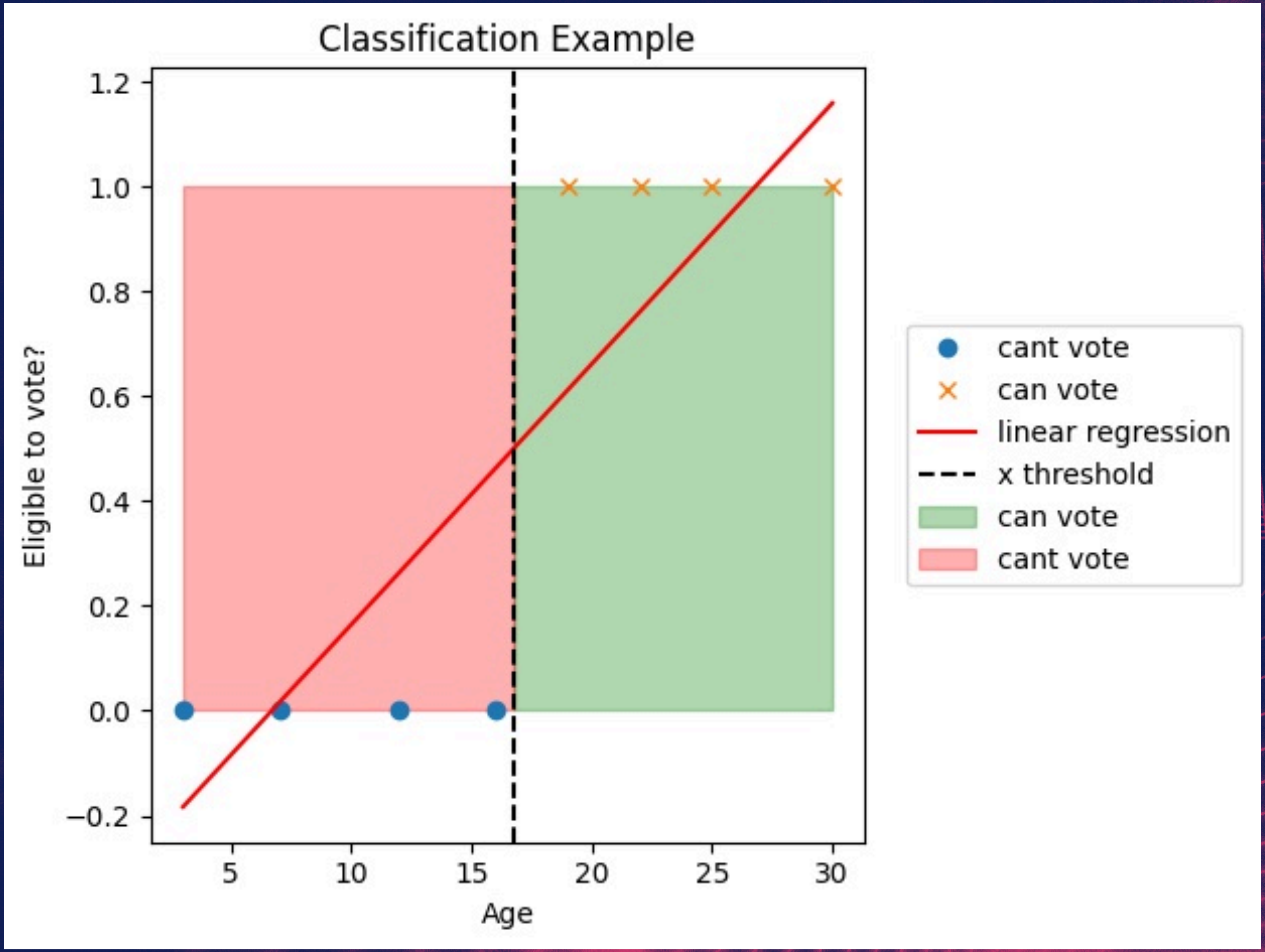


WHY DO WE NEED A LOGISTIC REGRESSION?



Linear Regression
y=1
Y
Predicted Y can exceed
0 and 1 range
y=0
X
Logistic Regression
y=1
Y
Predicted Y lies within
0 and 1 range
y=0
X


Classification Example
cant vote
can vote
1.0
0.8
0.6
0.4
0.2
0.0
Old Enough to Vote?
5
10
15
20
25
30
Age



Classification Example
Eligible to vote?
Age
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
5
10
15
20
25
30
cant vote
can vote
linear regression
x threshold
can vote
cant vote

COST FUNCTION

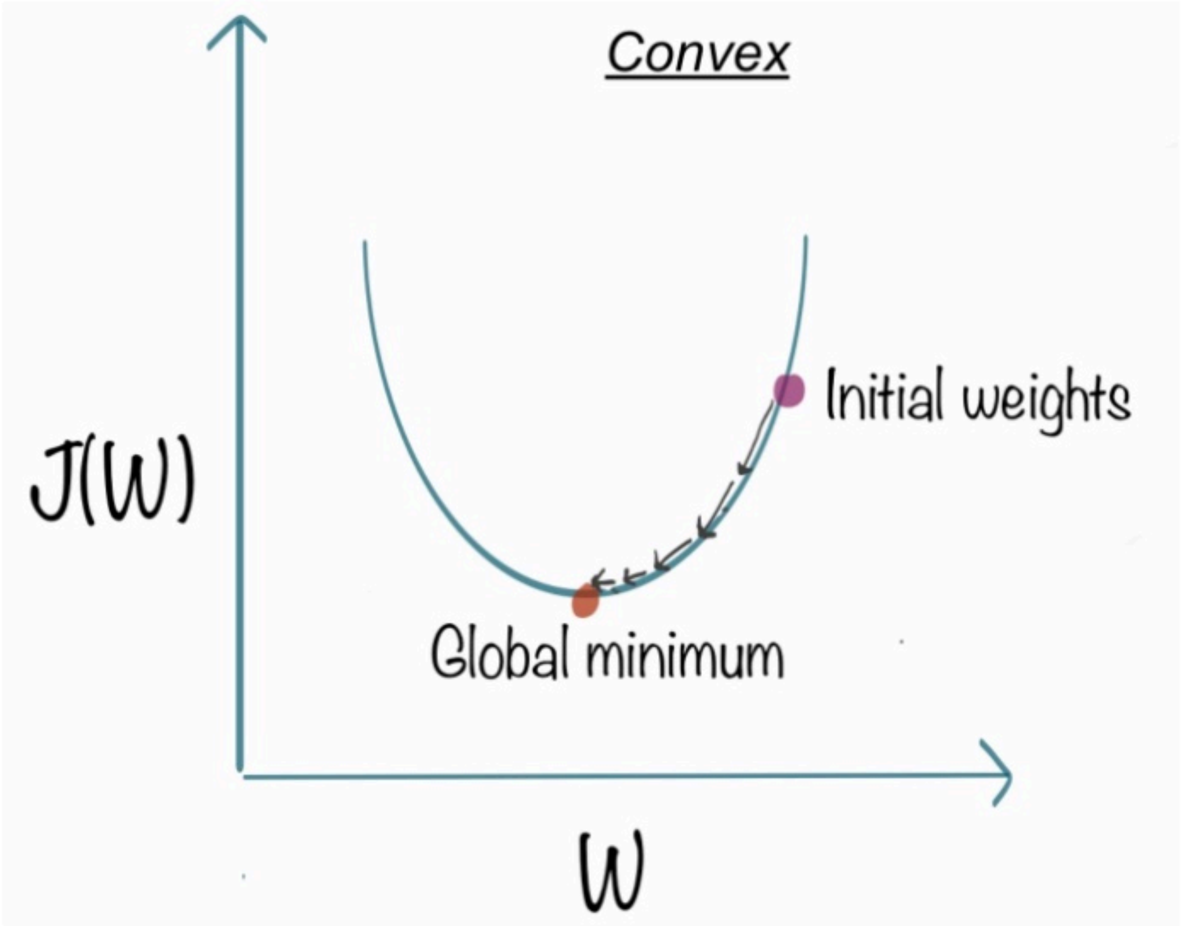
$$J(w,b) = \frac{1}{2m} \sum_{i=0}^{m-1} (f_{w,b}(x^{(i)}) - y^{(i)})^2$$

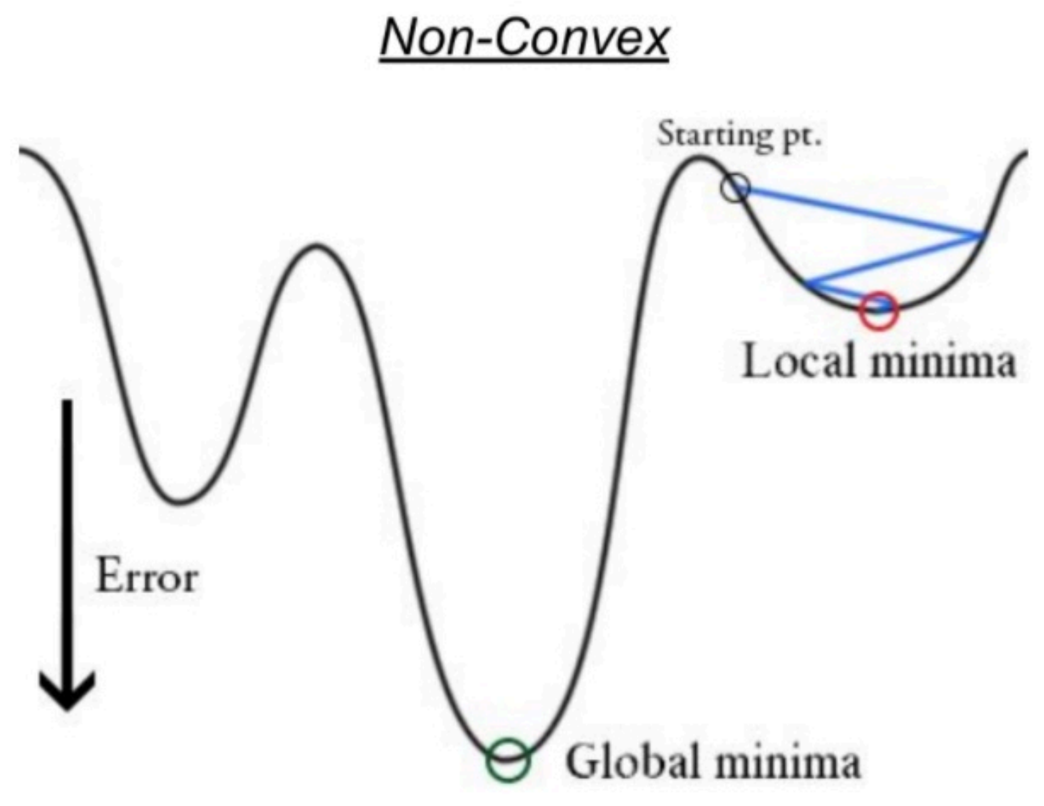
$$f_{w,b}(x^{(i)}) = wx^{(i)} + b$$

$$f_{w,b}(x^{(i)}) = sigmoid(wx^{(i)} + b)$$

slido.com
1118524

COST FUNCTION



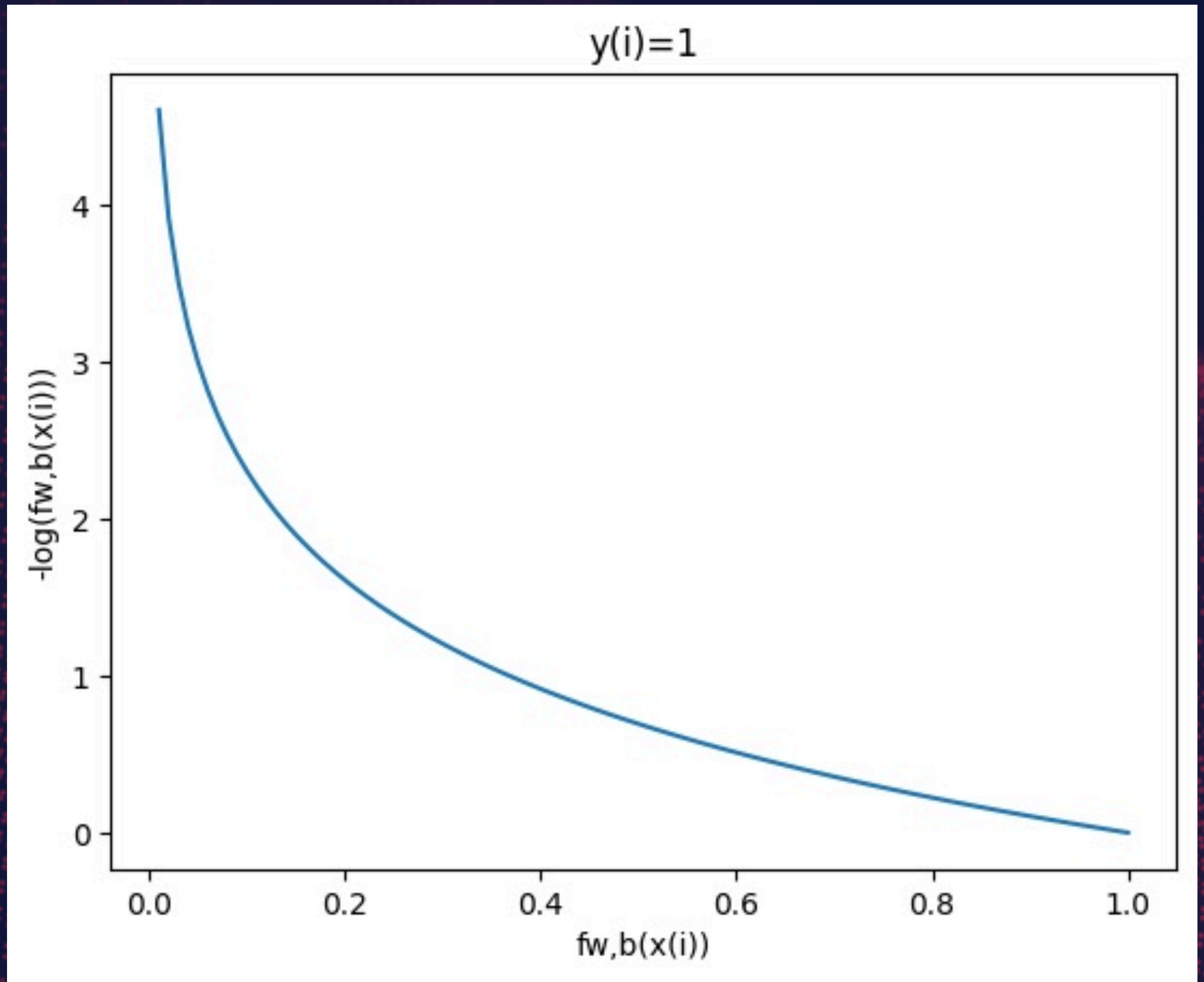


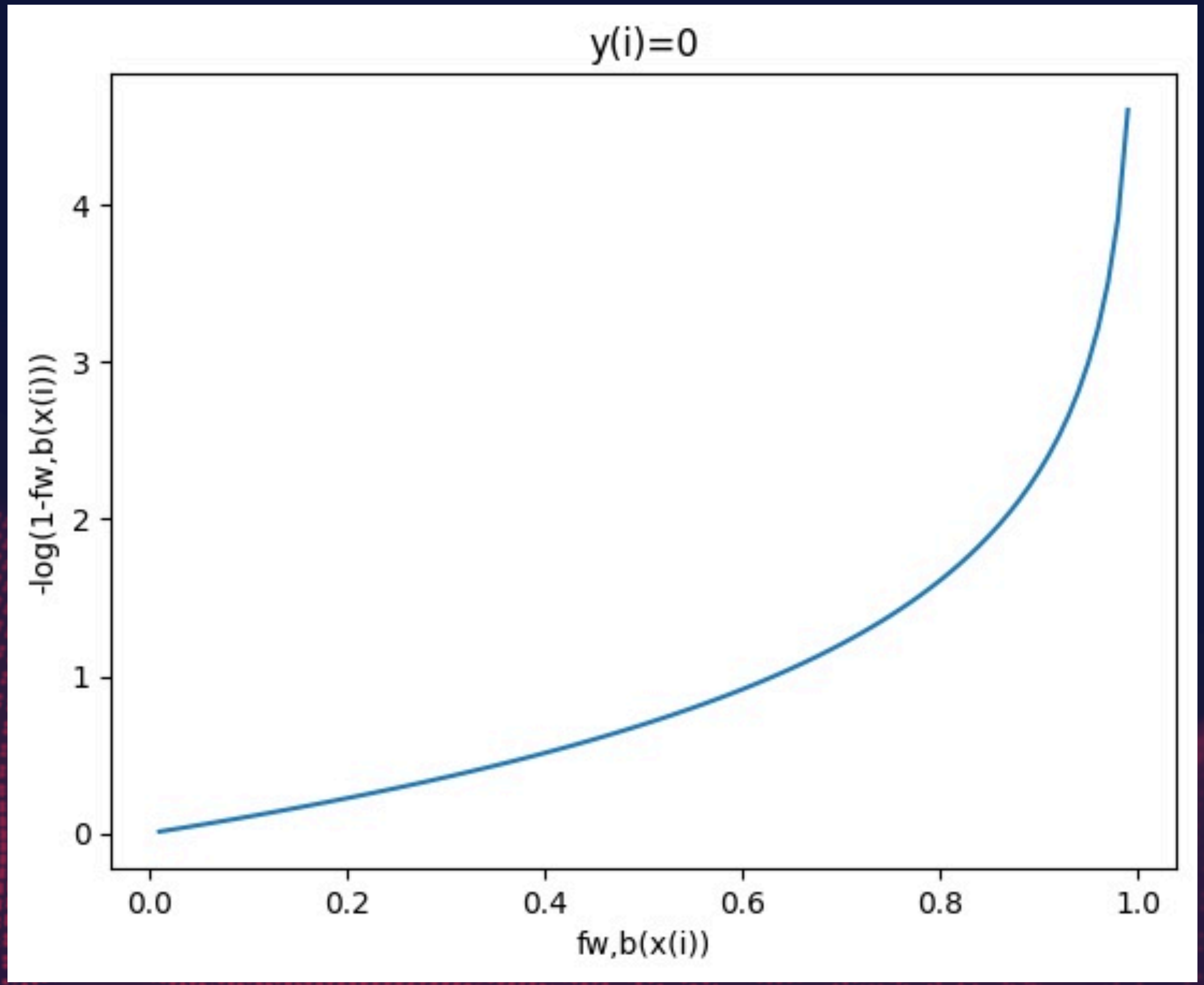
$$f_{w,b}(x^{(i)}) = wx^{(i)} + b$$

$$f_{w,b}(x^{(i)}) = sigmoid(wx^{(i)} + b)$$

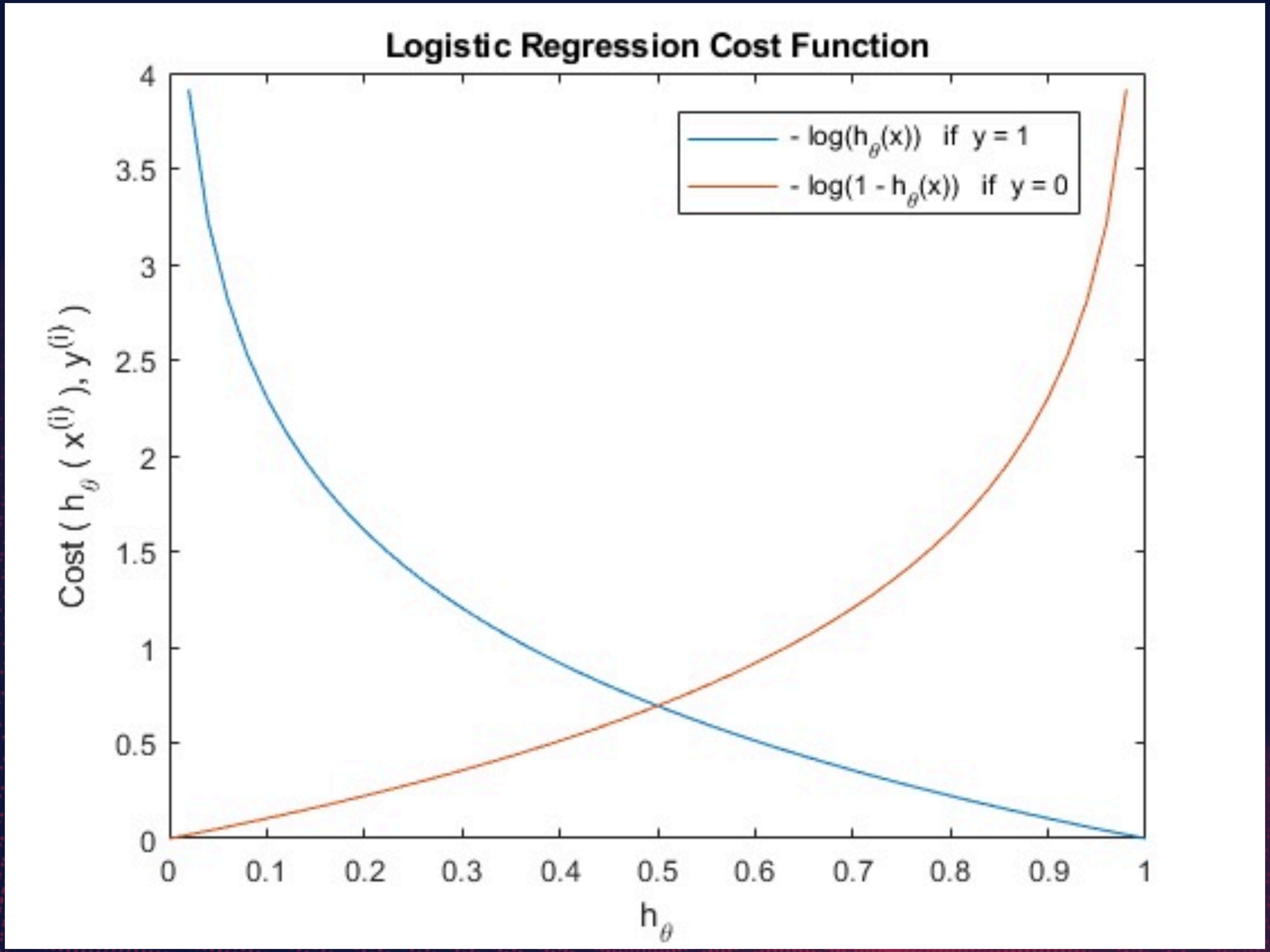
COST FUNCTION

$$loss(f_{\mathbf{w},b}(\mathbf{x}^{(i)}), y^{(i)}) = \begin{cases} -\log\left(f_{\mathbf{w},b}\left(\mathbf{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\mathbf{w},b}\left(\mathbf{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$





COST FUNCTION



OPTIMISATION

NOW THAT WE HAVE FOUND THE COST FUNCTION WE CAN NOW USE THE GRADIENT DESCENT TO UPDATE THE WEIGHTS(WJ) AND BIASES(B) TO MINMISE THE ERROR IN THE PREDICTION

OPTIMISATION

$$J(\vec{\mathrm{w}}, b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right) + \left(1 - y^{(i)}\right)\log\left(1 - f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right)\right]$$

repeat {

$$w_j = w_j - \alpha\frac{\partial}{\partial w_j}J(\vec{\mathrm{w}}, b)$$

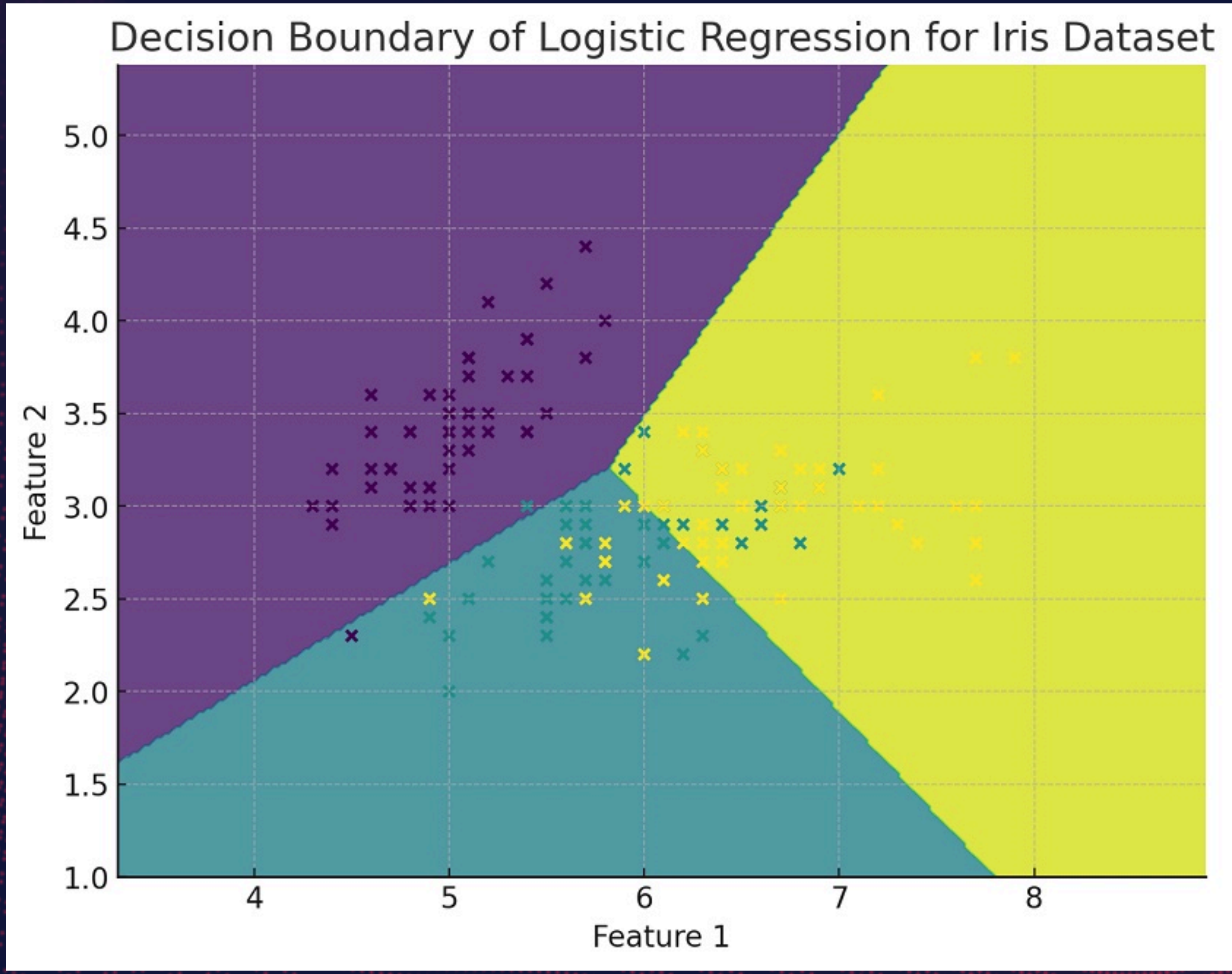
$$b = b - \alpha\frac{\partial}{\partial b}J(\vec{\mathrm{w}}, b)$$

}

$$\frac{\partial}{\partial w_j}J(\vec{\mathrm{w}}, b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)}\right)x_j^{(i)}$$

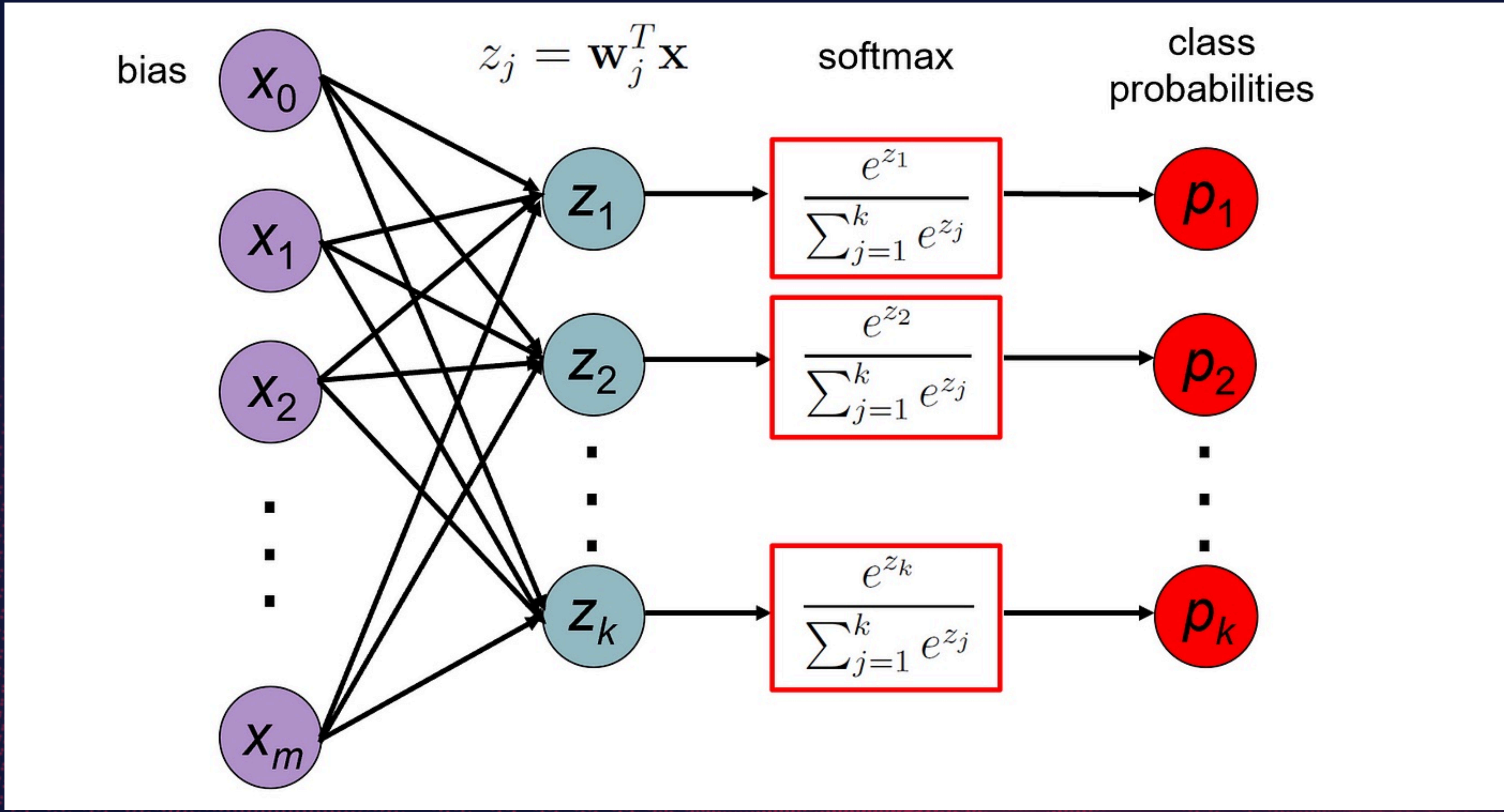
$$\frac{\partial}{\partial b}J(\vec{\mathrm{w}}, b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)}\right)$$

MULTICLASS REGRESSION



When the output has to be classified into more than 2 categories

SOFTMAX



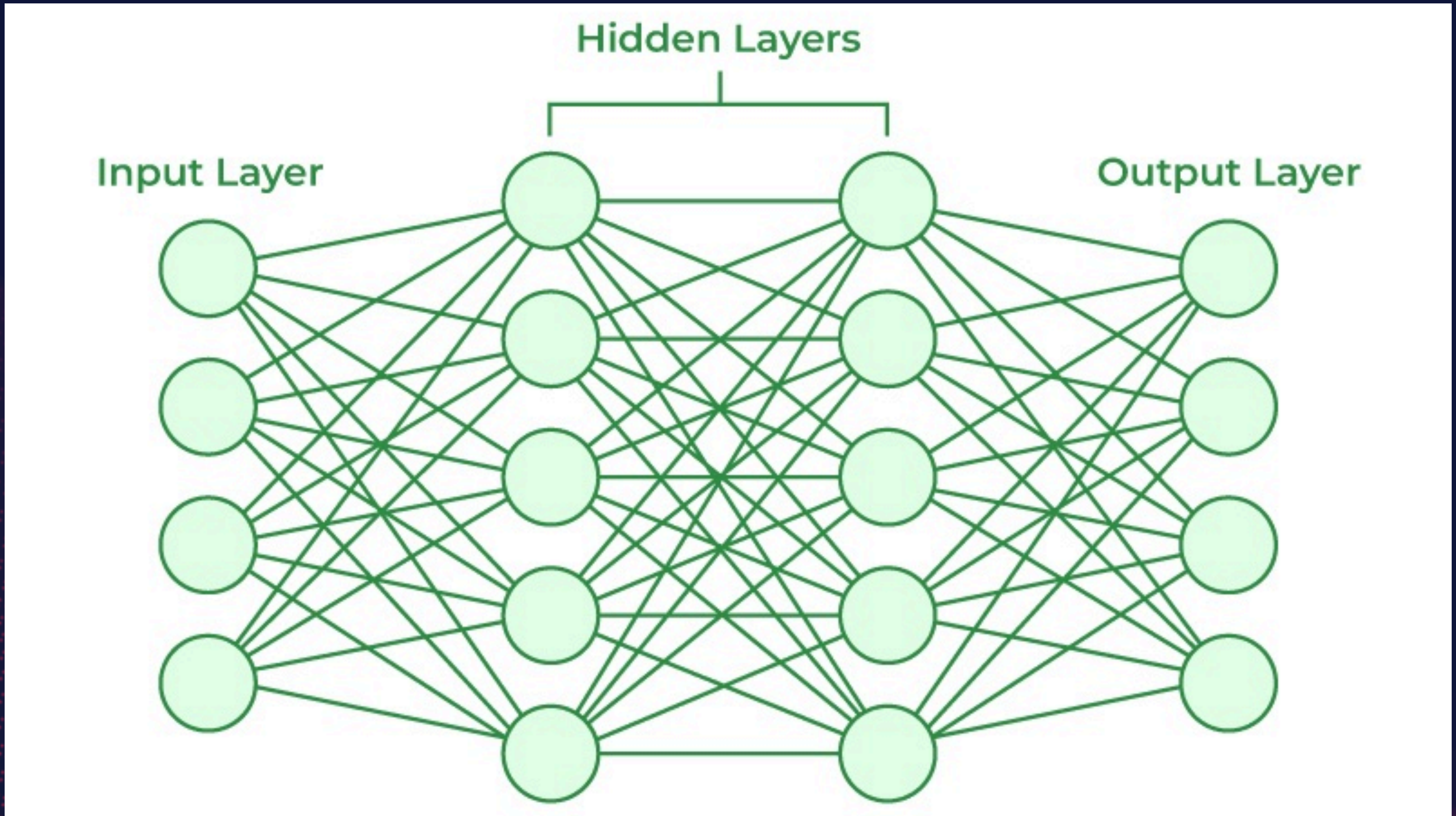
bias
x_0
x_1
x_2
x_m
$z_j = \mathbf{w}_j^T \mathbf{x}$
z_1
z_2
z_k
softmax
$\frac{e^{z_1}}{\sum_{j=1}^{k} e^{z_j}}$
$\frac{e^{z_2}}{\sum_{j=1}^{k} e^{z_j}}$
$\frac{e^{z_k}}{\sum_{j=1}^{k} e^{z_j}}$
class probabilities
p_1
p_2
p_k

LIMITATIONS

Logistic Regression assumes linearity between the input features and the binary outcome .

For example, predicting the likelihood of a customer making a purchase based on their age and income may not have a linear relationship.

In such cases, logistic regression may not capture the complex non-linear patterns in the data, and its performance may be limited.


Hidden Layers
Input Layer
Output Layer

CODE IMPLEMENTATION

AI
CLUB