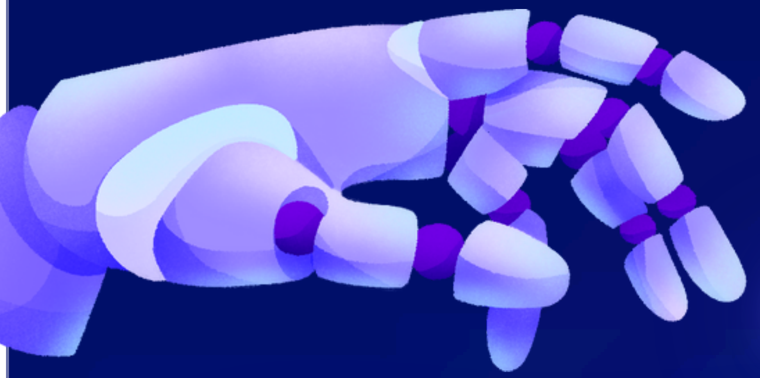
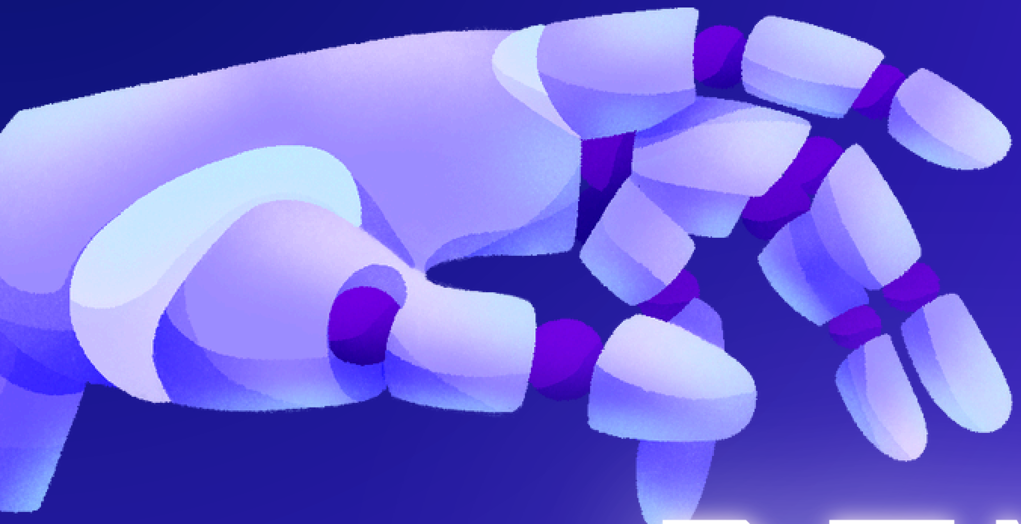


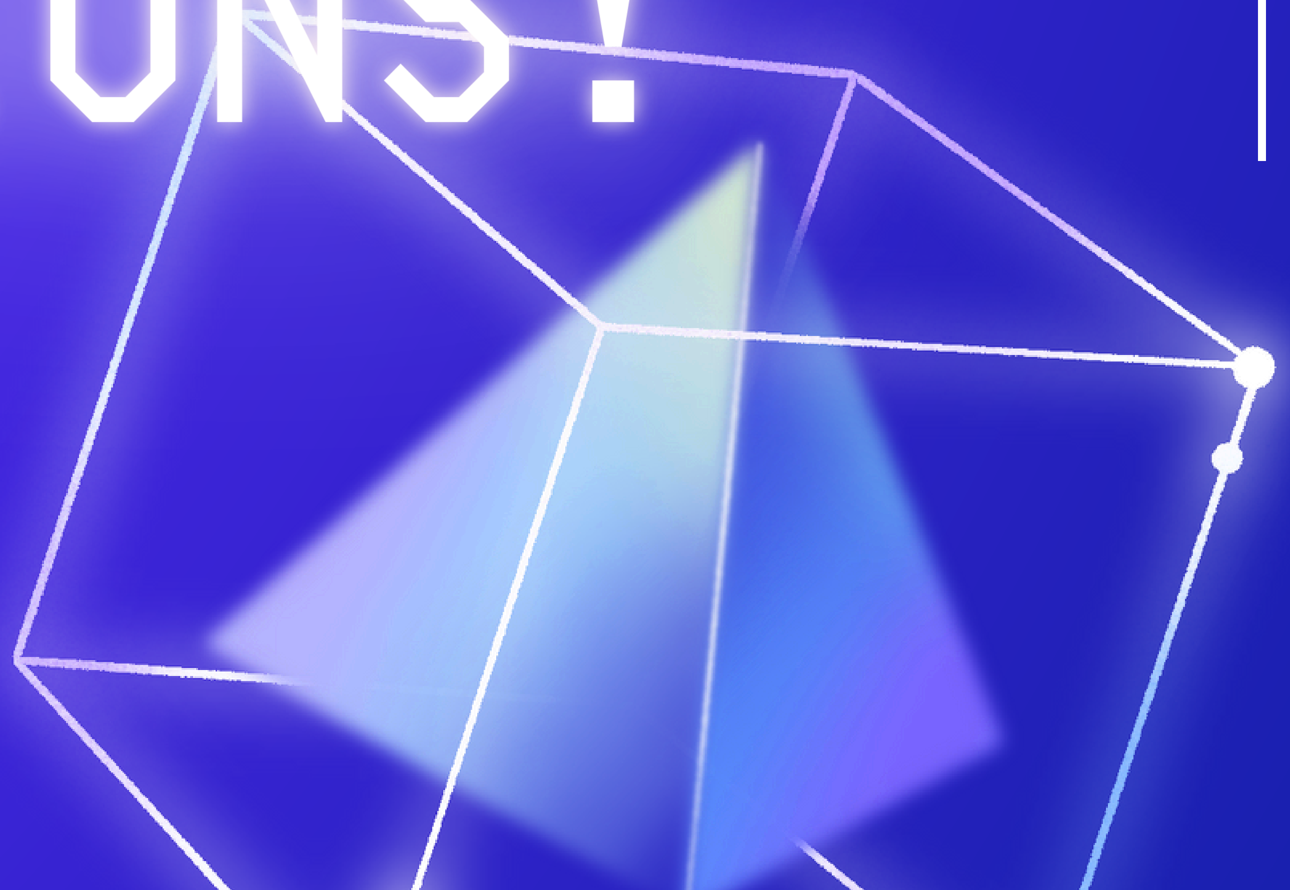
AI/MACHINE LEARNING WORKSHOP

DAY 2:INTRODUCTION TO AI & ML: LOGISTIC REGRESSION

Youth Oppurtunties in Tech Inovation



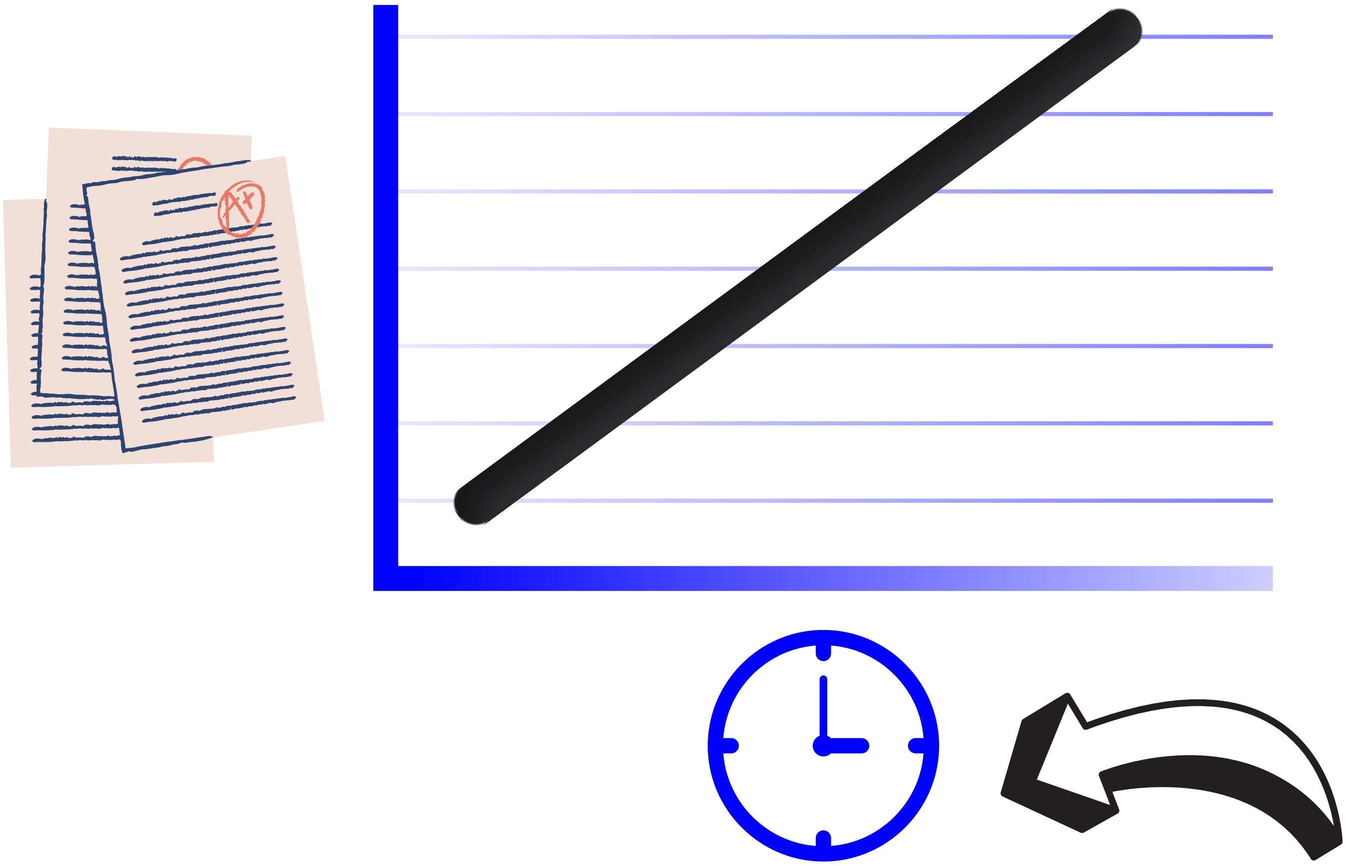
REMINDER PLEASE ASK QUESTIONS!



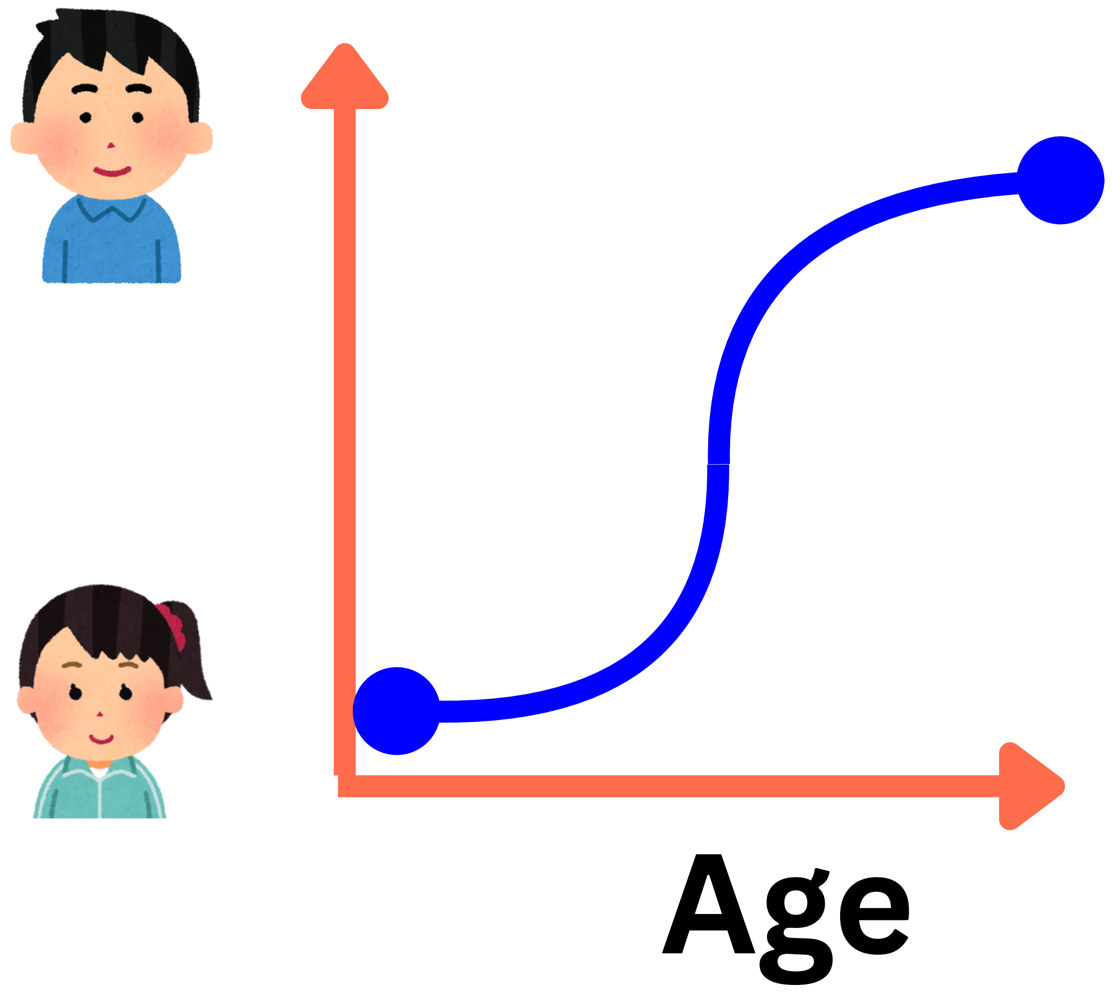
WHAT IS LOGISTIC REGRESSION:

PREDICTS THE PROBABILITY THAT A GIVEN INPUT BELONGS TO A PARTICULAR CATEGORY BY FITTING A CURVE TO DATA USING THE LOGISTIC (SIGMOID) FUNCTION.

Difference between Linear Regression and Logistic Regression



The dependent variable is a metric value in a Linear Regression

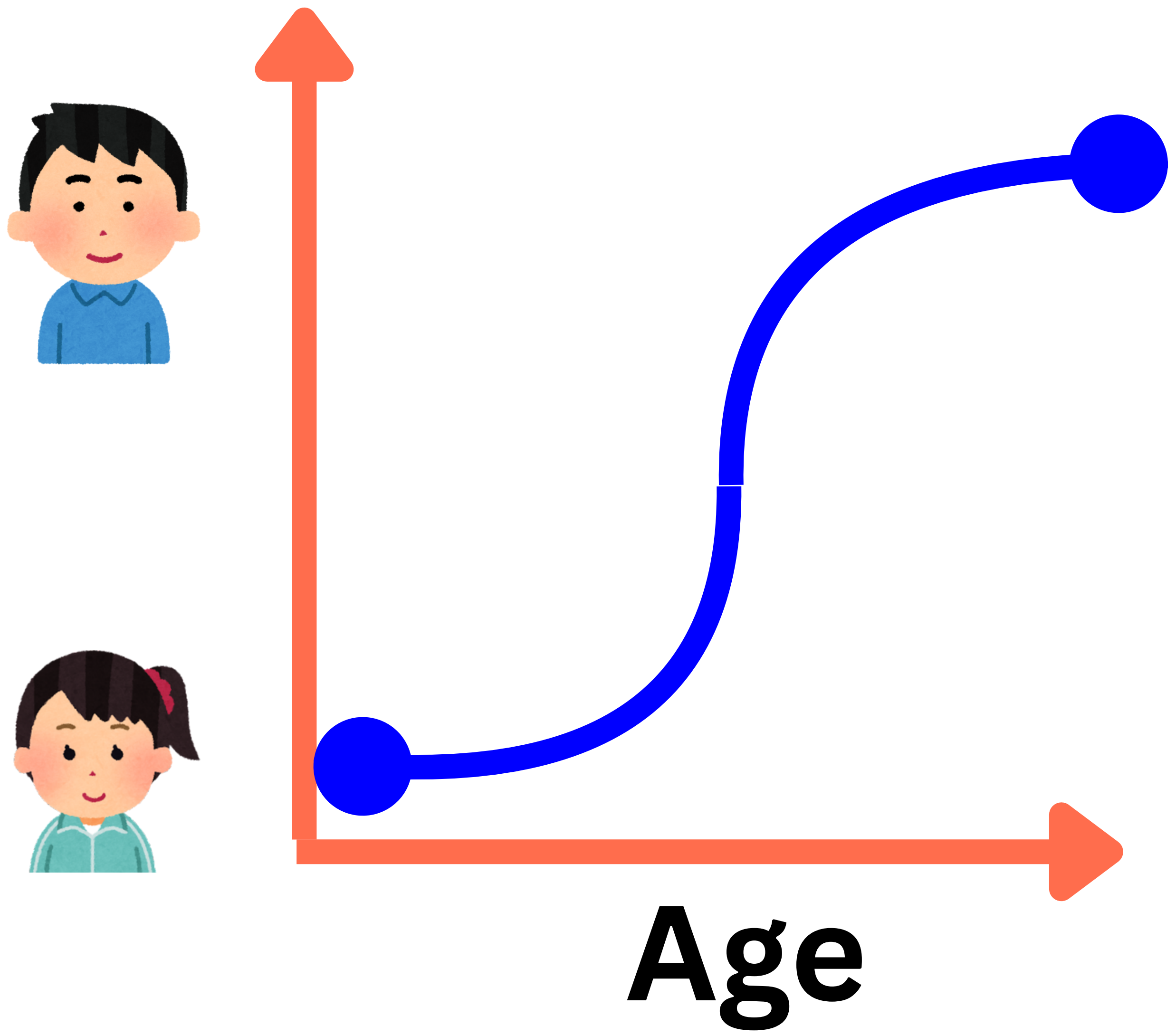


The dependent variable is a <u>dichotomous</u> value in a Linear Regression

Logistic Regression

Dichtomous Variable is a variable with 2 outcomes or values

For example: MALE or FEMALE

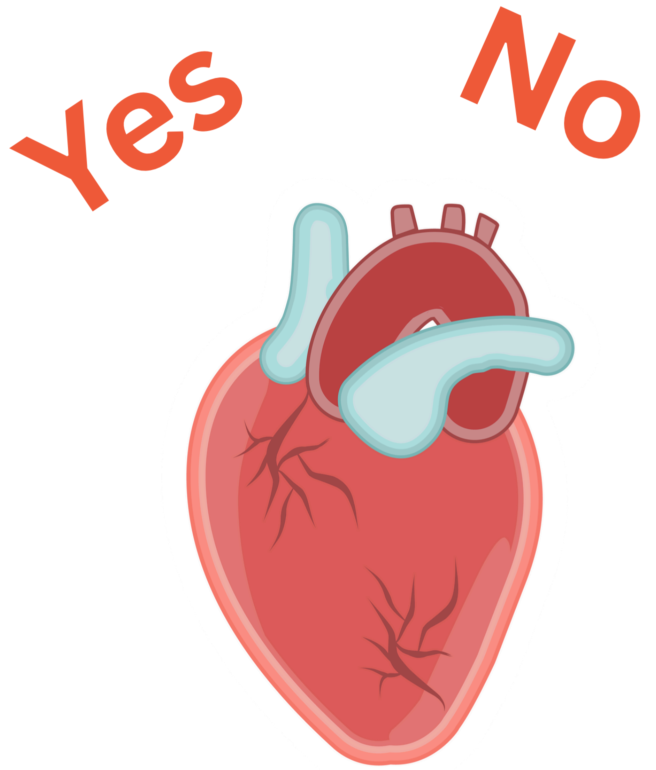


The dependent variable is a dichotomous value in a Linear Regression

Logistic Regression

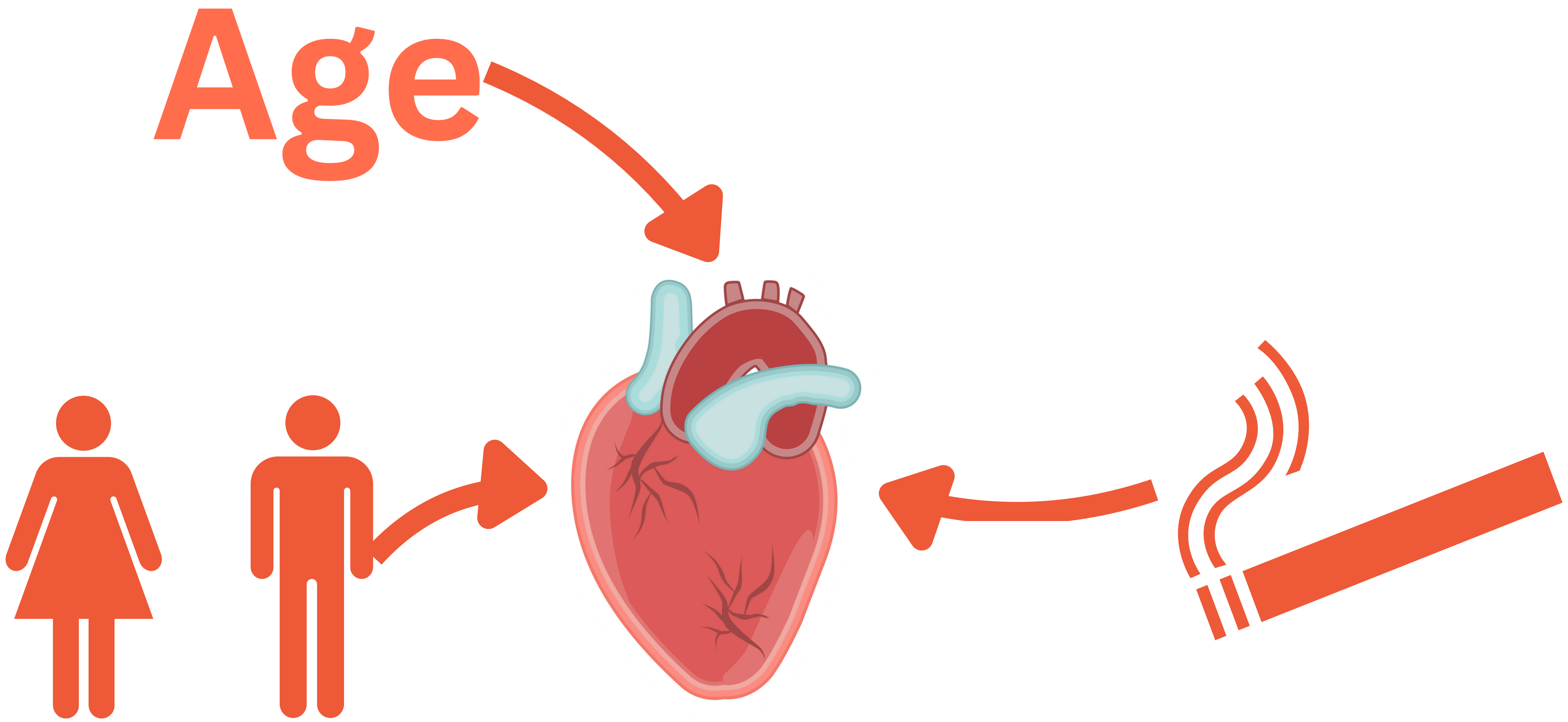
Dichtomous Variable is a variable with 2 outcomes or values

For example: Disease Present or Not



The dependent variable is a <u>dichotomous</u> value in a Linear Regression

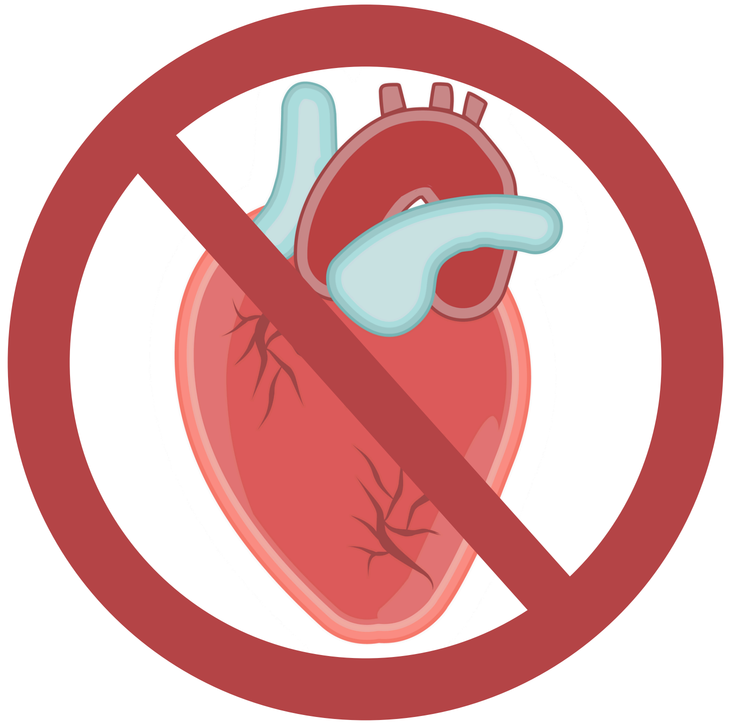
Logistic Regression



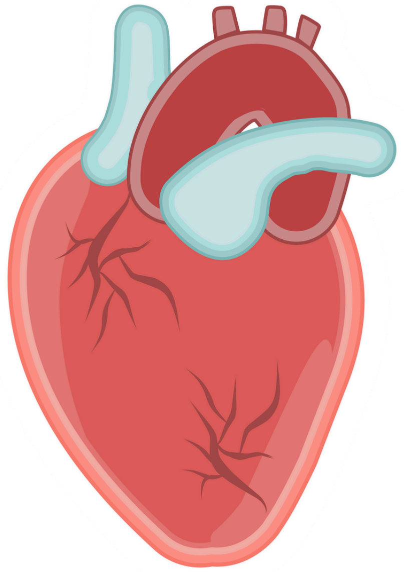
We can now understand the effects of other factors (Age, Smoking, Gender) on that particular disease

Logistic Regression

0 Stands for Non-Cancerous



1 Stands for Cancerous



Logistic Regression

Independent Variable

Dependent Variable

Age	Gender	Smoking	Cancer
18	Female	Smoker	0
45	Female	Non-Smoker	0
43	Female	Smoker	1
23	Female	Smoker	0
33	Female	Smoker	1
22	Male	Non-Smoker	1
25	Male	Smoker	1
12	Male	Smoker	0

Logistic Regression

Now we see what influence the Independent Variables will have on the cancer

Age	Gender	Smoking	Cancer
18	Female	Smoker	0
45	Female	Non-Smoker	0
43	Female	Smoker	1
23	Female	Smoker	0
33	Female	Smoker	1
22	Male	Non-Smoker	1
25	Male	Smoker	1
12	Male	Smoker	0

If there is an influence we can predict the likelihood of getting the cancer

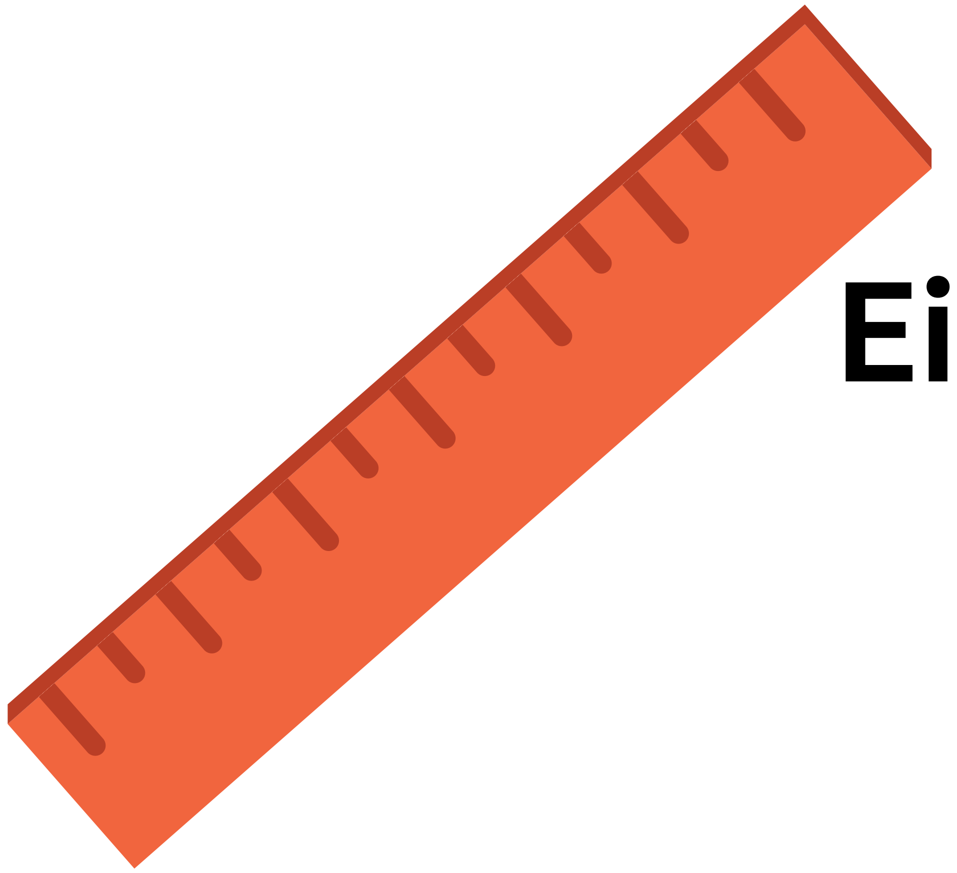
HOW DO WE USE LOGISTIC REGRESSION IN THIS CASE

LOGISTIC REGRESSION'S DEPENDENT VARIABLE MUST BE CATEGORICAL

Categorical vs Quantative Variables

Quantative Variables

Can Be Measured



Either Continous:

0.1, 0.992, 1/3

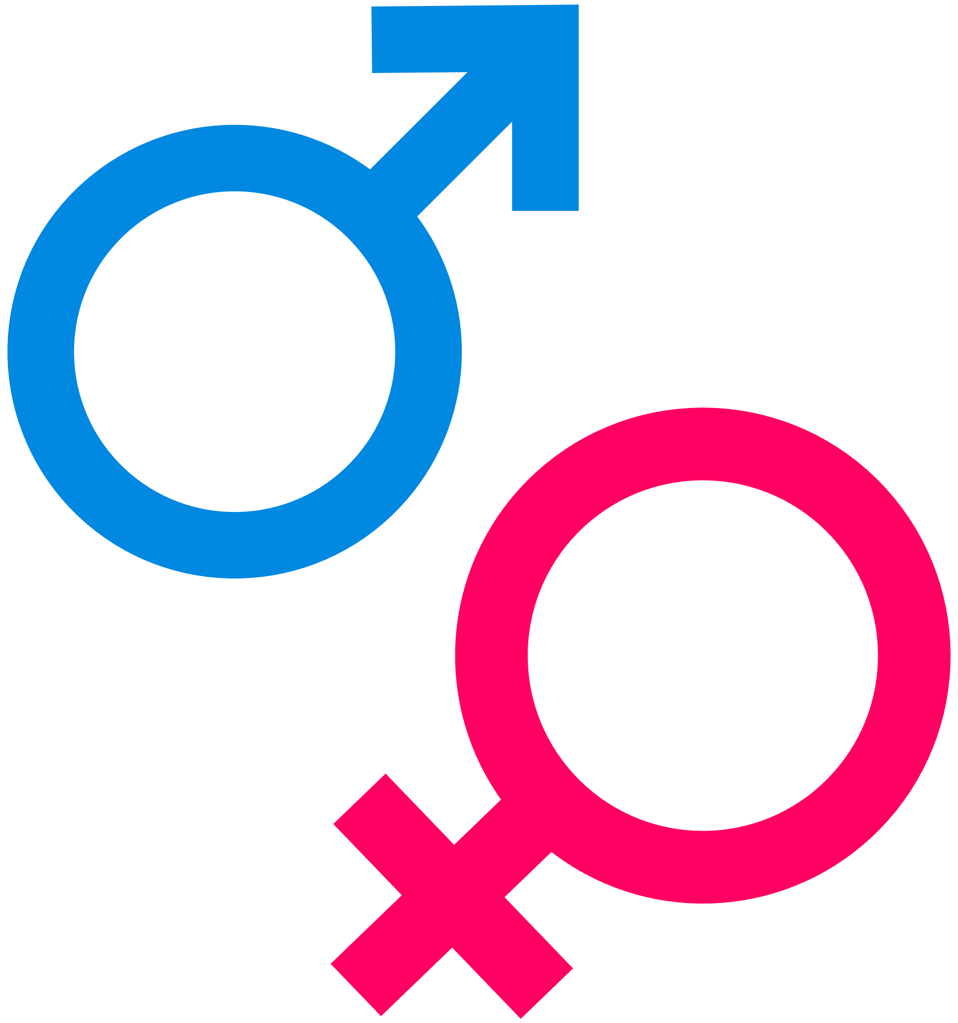
Or Discrete:

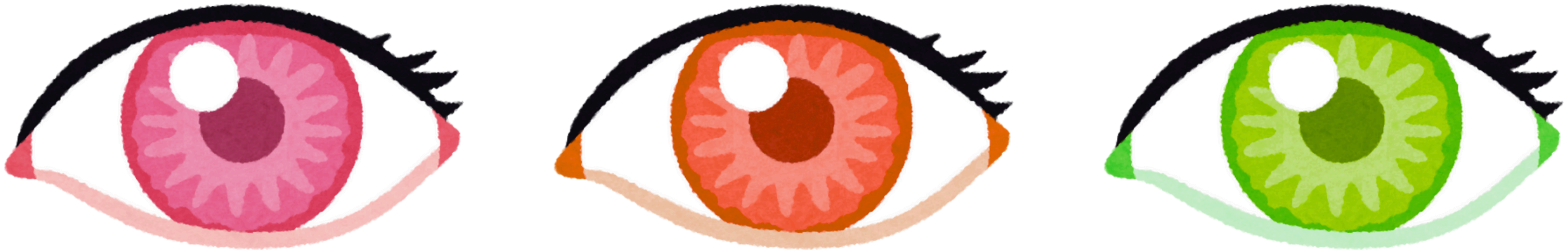
1, 2, 3, 4

Categorical vs Quantative Variables

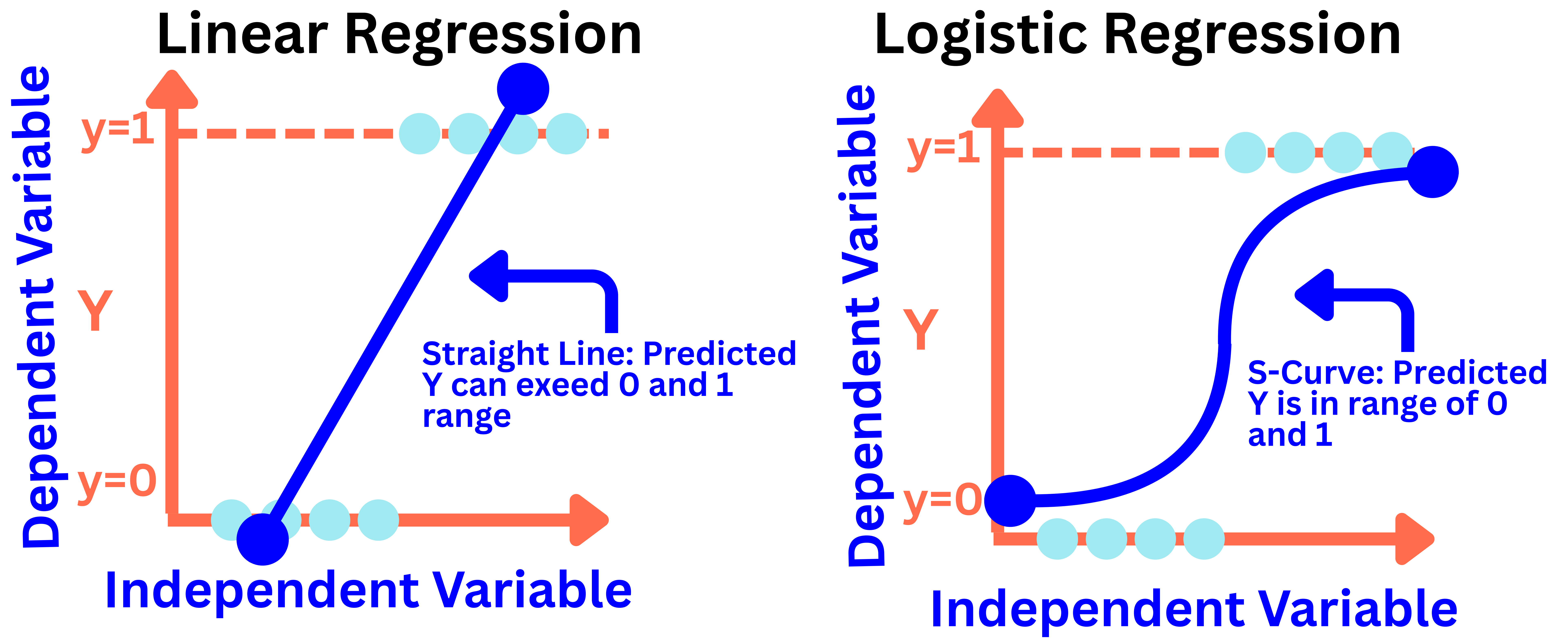
Categorical Values

These values describe qualities and cannot be meaningfully added or averaged

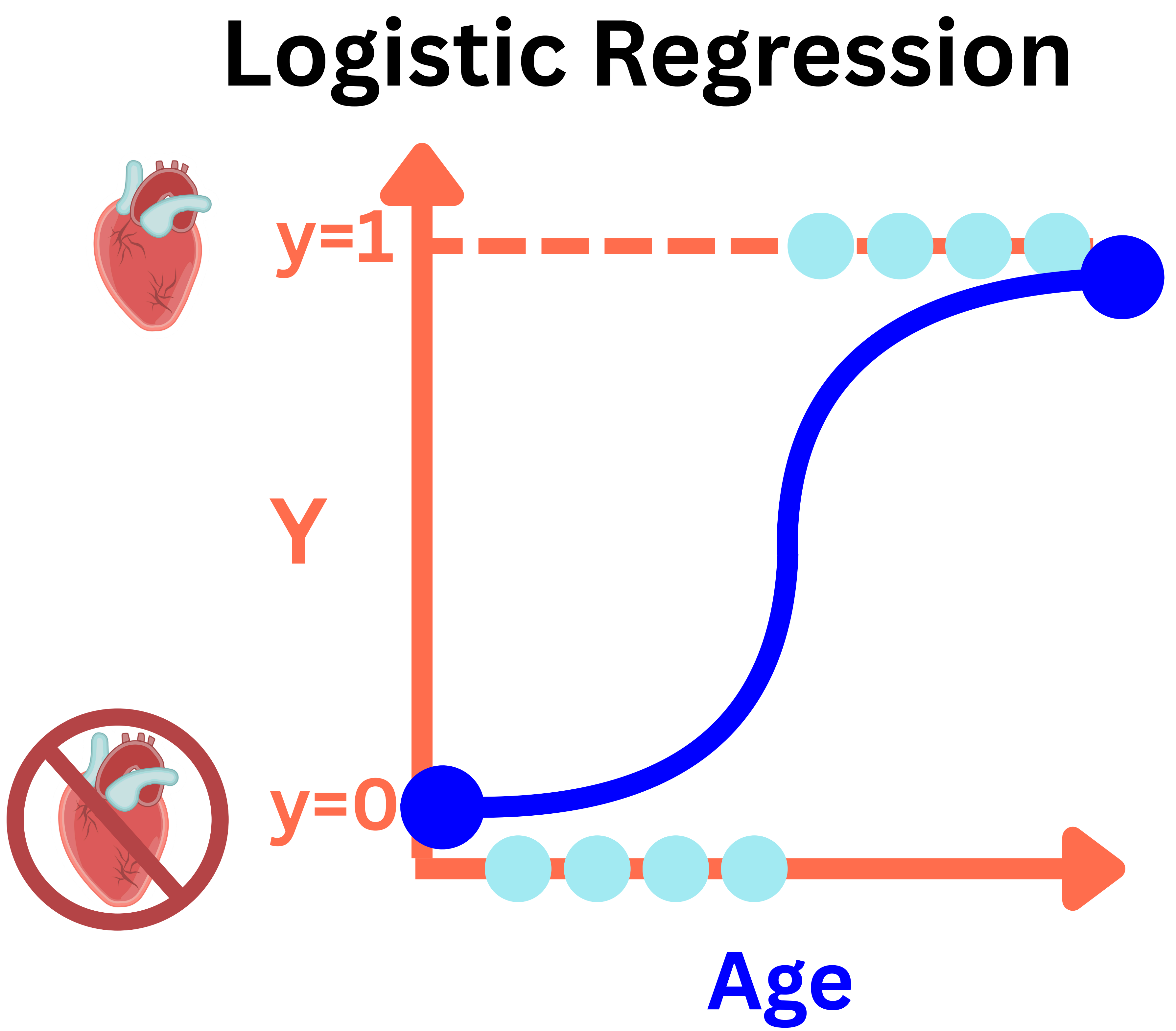




Linear vs Logistic

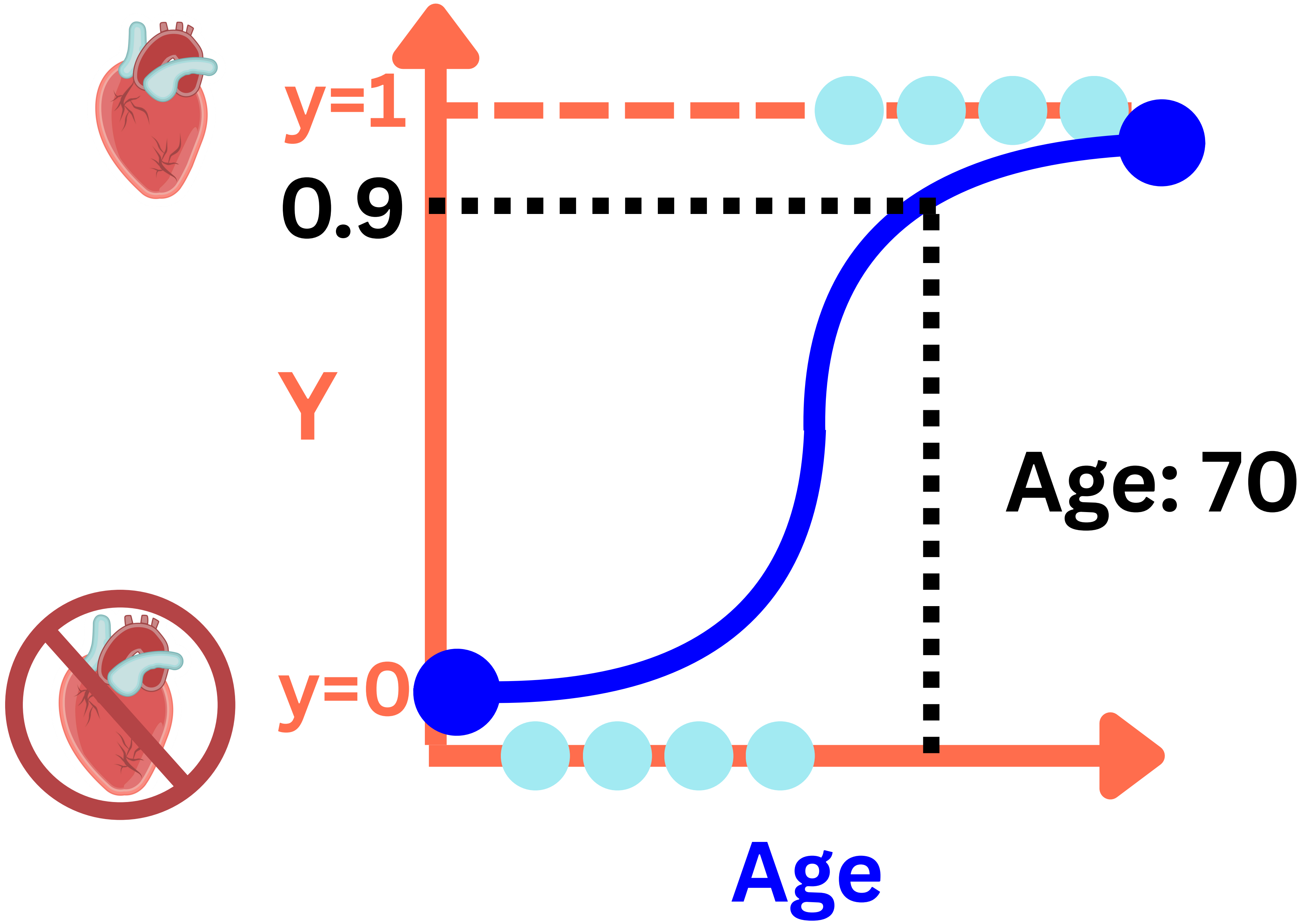


Logistic Regression



Logistic Regression

Logistic Regression



Linear Function from last time

$$\hat{y} = wx + b$$

- $\hat{y}$ = Carbon Emissions next week
- x = Todays Carbon Emissions
- (w,b) = Model Parameters

Stage 2

Linear Function from last time

$$z = \sum wx + b$$

Sigmoid Function

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

Inference: Stage 2

Linear Function from last time

$$z = \sum wx + b$$

Sigmoid Function

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

Stage 2

Sigmoid Function

$$\sigma(z) = \frac{1}{1 + e^{-(\Sigma wx + b)}}$$

This function “squashes” any real number z into the range (0,1), interpreting it as the probability of the positive class.

Stage 2

Sigmoid Function

$$P(y=1|x) = \frac{1}{1 + e^{-(\Sigma wx + b)}}$$

This also gives us the probability that the class y = 1

Stage 2

Sigmoid Function

$$P(y=0|x) = 1 - \left(\frac{1}{1 + e^{-(\Sigma wx + b)}}\right)$$

This now gives the probability of the negative class or $y = 0$

Stage 2

Sigmoid Function

$$P(y=0|x) = 1 - P(y=1|x)$$

This now gives the probability of the negative class or
y = 0

Stage 2

Training (Stage 1)

Fit the Models prediction to True Value

Loss Function: Goal to get smallest loss

Binary Cross-Entropy

$$\text{Loss} = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(\hat{y}^{(i)}) + (1-y^{(i)})\log(1-\hat{y}^{(i)})\right]$$

m = number of training examples

$y^{(i)}$ = true label for example i

$\hat{y}^{(i)}$ = predicted probability for example i

Training (Stage 1)

Fit the Models prediction to True Value

Loss Function: Goal to get smallest loss

Binary Cross-Entropy

$$\text{Loss} = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(\underbrace{\hat{y}^{(i)}}_{\sigma(z)}) + (1-y^{(i)})\log(1-\underbrace{\hat{y}^{(i)}}_{\sigma(z)})\right]$$

m = number of training examples

$y^{(i)}$ = true label for example i

$\hat{y}^{(i)}$ = predicted probability for example i

Training (Stage 1)

Fit the Models prediction to True Value

Optimization Algorithim: Gradient Descent to adjust w and b

$$w = w - \alpha \frac{\partial \mathrm{MSE}}{\partial w}$$

$$b = b - \alpha \frac{\partial \mathrm{MSE}}{\partial b}$$

- α = Learning rate (controls step size)
- $\frac{\partial \mathrm{MSE}}{\partial w}$ and $\frac{\partial \mathrm{MSE}}{\partial b}$ are the gradients (slopes) of the loss function with respect to each parameter

We keep updating w and b until the error is minimized (or we reach a set number of iterations).

ACTIVITY: BUILD A LOGISTIC REGRESSION

QUESTIONS AND FINAL THOUGHTS!

THANK YOU!