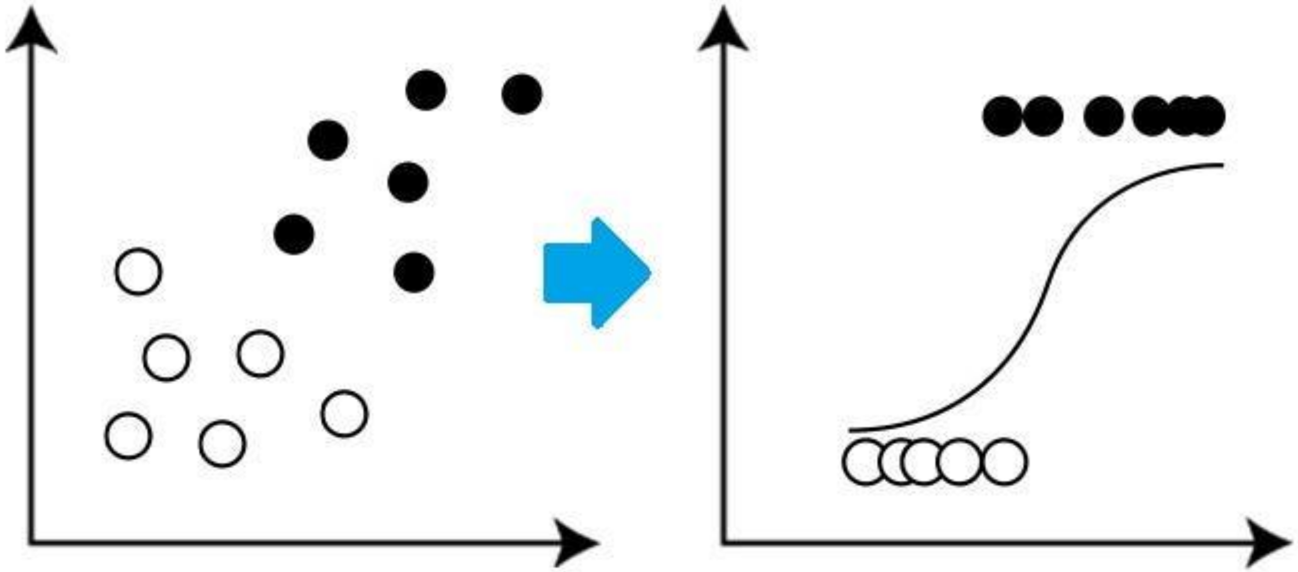
Logistic Regression

Logistic Regression

- Logistic regression is a statistical method used for modeling the probability that a given observation belongs to a certain category or class.
- It's particularly well-suited for binary classification tasks, where the outcome variable has two possible categories (e.g., "yes" or "no", "positive" or "negative").
- In Logistic regression, instead of fitting a regression line, we fit an “S” shaped logistic function, which predicts two maximum values (0 or 1).

LOGISTIC REGRESSION



Terminologies

- **Independent variables:** The input characteristics or predictor factors applied to the dependent variable's predictions.
- **Dependent variable:** The target variable in a logistic regression model, which we are trying to predict.
- **Logistic function:** The formula used to represent how the independent and dependent variables relate to one another. The logistic function transforms the input variables into a probability value between 0 and 1, which represents the likelihood of the dependent variable being 1 or 0.

Terminologies

- **Odds:** It is the ratio of something occurring to something not occurring. it is different from probability as the probability is the ratio of something occurring to everything that could possibly occur.
- **Coefficient:** The logistic regression model's estimated parameters, show how the independent and dependent variables relate to one another.
- **Intercept:** A constant term in the logistic regression model, which represents the log odds when all independent variables are equal to zero.
- **Maximum likelihood estimation:** The method used to estimate the coefficients of the logistic regression model, which maximizes the likelihood of observing the data given the model.

Odds Ratio and Logit

- Odds ratio is defined as the ratio of the odds in presence of **B** and odds of **A** in the absence of **B** and vice versa.
- In other words,***Odds are the ratio of the probability of success to the probability of failure and Logit is Just the Log of the Odds Ratio.***

  Assume the probability of success is 0.6.
  So,probability of failure will be (1–0.6) = 0.4
  Odds are determined from probabilities and range between 0 to **∞**.
  So,Now odds(Success) = p/(1-p) or p/q = 0.6/0.4 = 1.5
  Also,odds(Failure) = 0.4/0.6 = 0.66667

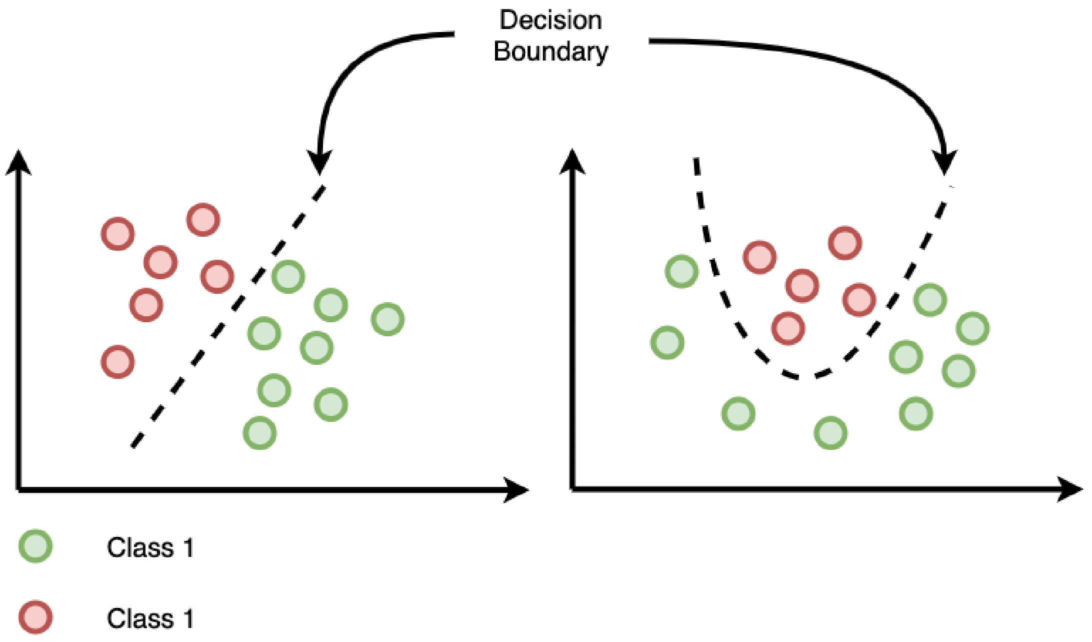
Odds

$$\text{Odds} = \frac{P}{1-P}$$

$$\text{Logit Function} = \log\left(\frac{p}{1-p}\right)$$

Decision Boundary

- A decision Boundary is a line or margin that separates the classes.
- Classification algorithm is all about finding the decision boundary that helps distinguish between the classes perfectly or close to perfect.
- Logistic Regression decides a proper fit to the decision boundary so that we will be able to predict which class a new data will correspond to.



Working
Step1: Model Representation

In logistic regression, you start with a linear equation, similar to linear regression:

$$z = w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n$$

$(x_1, x_2, \ldots, x_n)$ are the respective weights

$(w_1, w_2, \ldots, w_n)$term

w_0

The bias term (also known as the intercept term) represents the baseline probability of the positive class when all input features are zero. It is denoted by *w*0 in the model equation.

- Let the independent input features be:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ x_{21} & \cdots & x_{2m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

- The dependent variable is Y having only binary value i.e. 0 or 1.

$$Y = \begin{cases} 0 & \text{if } Class\ 1 \\ 1 & \text{if } Class\ 2 \end{cases}$$

- multi-linear function to the input variables X is

$$z = \left(\sum_{i=1}^{n} w_i x_i\right) + b \qquad z = w \cdot X + b$$

Step 2. **Sigmoid Function Transformation**

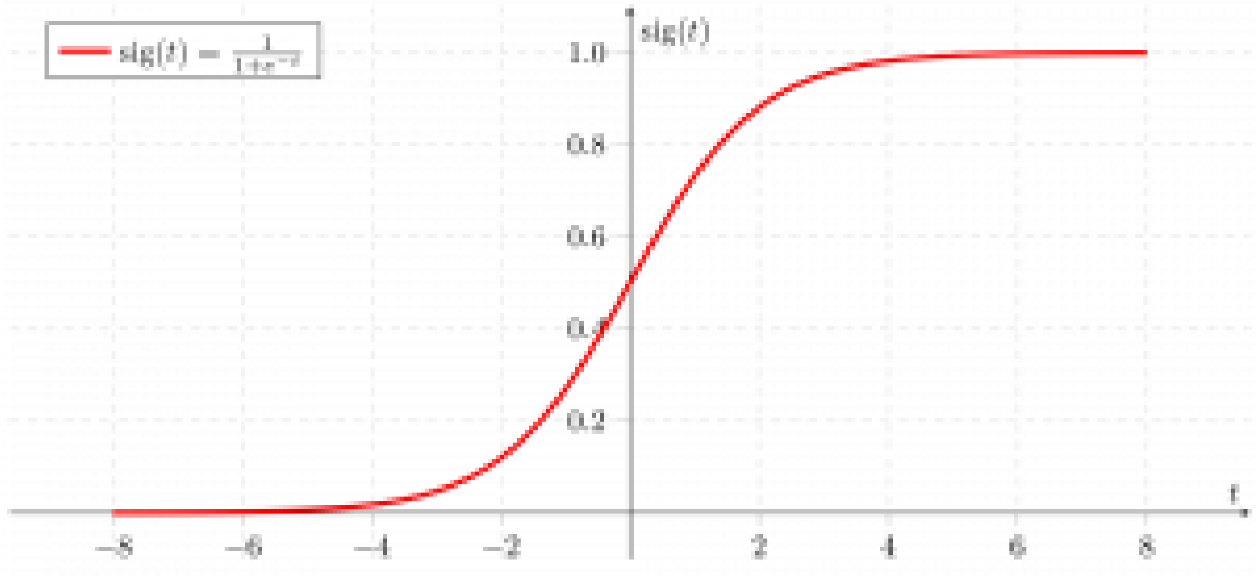
- Instead of directly using the linear equation to predict the output, logistic regression applies a sigmoid (logistic) function to transform the linear combination into a probability value between 0 and 1. The sigmoid function is defined as:

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

where *z* is the linear combination calculated in the previous step

Sigmoid Function

- we use the sigmoid function where the input will be z and we find the probability between 0 and 1. i.e. predicted y



- Odd- odd is the ratio of something occurring to something not occurring

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

- Applying natural log on odd. then log odd will be

$$\log\left[\frac{p(x)}{1-p(x)}\right] = z$$

$$\log\left[\frac{p(x)}{1-p(x)}\right] = w \cdot X + b$$

$$\frac{p(x)}{1-p(x)} = e^{w\cdot X+b} \quad \cdots \text{Exponentiate both sides}$$

$$p(x) = e^{w\cdot X+b} \cdot (1-p(x))$$

$$p(x) = e^{w\cdot X+b} - e^{w\cdot X+b} \cdot p(x))$$

$$p(x) + e^{w\cdot X+b} \cdot p(x)) = e^{w\cdot X+b}$$

$$p(x)(1+e^{w\cdot X+b}) = e^{w\cdot X+b}$$

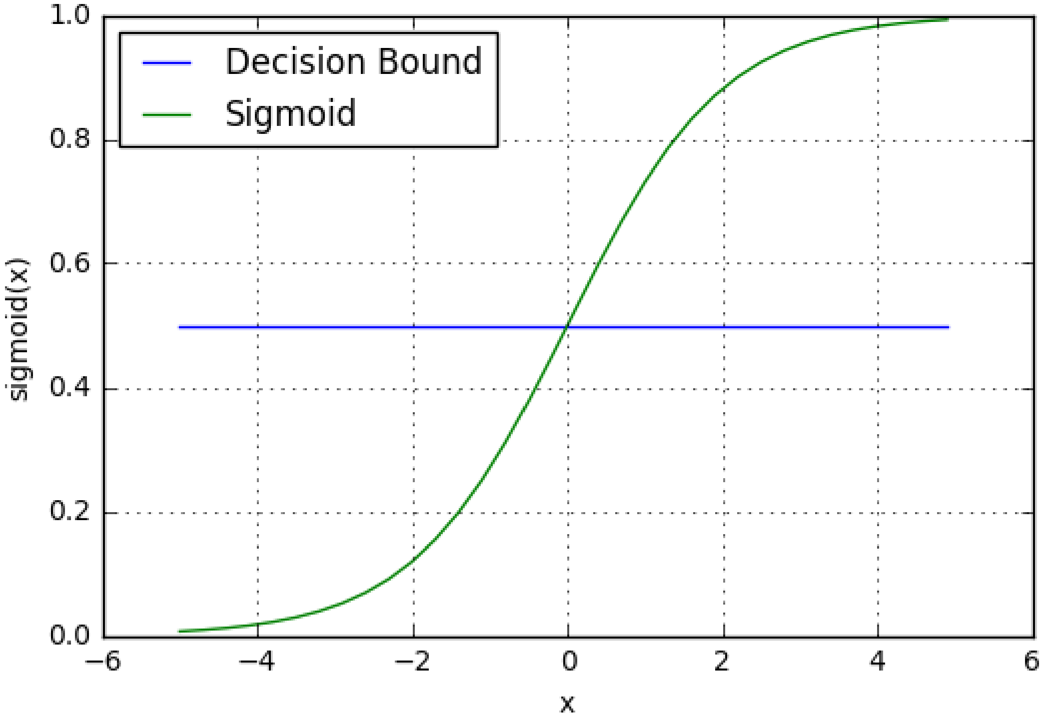
$$p(x) = \frac{e^{w\cdot X+b}}{1+e^{w\cdot X+b}}$$

then the final logistic regression equation will be:

$$p(X;b,w) = \frac{e^{w\cdot X+b}}{1+e^{w\cdot X+b}} = \frac{1}{1+e^{-w\cdot X+b}}$$

Step 4. Decision boundary

- The sigmoid function returns a probability value between 0 and 1. This probability value is then mapped to a discrete class which is either “0” or “1”. In order to map this probability value to a discrete class (pass/fail, yes/no, true/false), we select a threshold value. This threshold value is called Decision boundary. Above this threshold value, we will map the probability values into class 1 and below which we will map values into class 0.
- Mathematically, it can be expressed as follows:-
  - $p \geq 0.5 \Rightarrow class = 1$
  - $p < 0.5 \Rightarrow class = 0$


Decision Bound
Sigmoid
sigmoid(x)
x
1.0
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6

Step 5. Making predictions

- A prediction function in logistic regression returns the probability of the observation being positive, Yes or True.
- We call this as class 1 and it is denoted by P(class = 1).
- If the probability inches closer to one, then we will be more confident about our model that the observation is in class 1, otherwise it is in class 0.

Example

Feature 1 (X1)	Feature 2 (X2)	Feature 3 (X3)	Target (Y)
1.2	0.7	2.4	0
2.4	1.5	3.1	0
3.5	2.2	4.5	0
4.2	3.1	5.2	1
5.1	4.0	6.3	1

Take one data sample

Feature 1 (X1)	Feature 2 (X2)	Feature 3 (X3)	Target (Y)
1.2	0.7	2.4	0

- We'll assume random initial weights and bias:

$$w_0 = 0.5$$
$$w_1 = -0.3$$
$$w_2 = 0.8$$
$$w_3 = -0.2$$

1. **Feature values for the first data point**:

$$X_1 = 1.2$$
$$X_2 = 0.7$$
$$X_3 = 2.4$$

2. **Compute the linear combination**:

$$z = w_0 + w_1 \times X_1 + w_2 \times X_2 + w_3 \times X_3$$
$$z = 0.5 + (-0.3 \times 1.2) + (0.8 \times 0.7) + (-0.2 \times 2.4)$$
$$z = 0.5 - 0.36 + 0.56 - 0.48$$
$$z = 0.22$$

3. **Apply the sigmoid function**:

$$p = \frac{1}{1+e^{-z}}$$
$$p = \frac{1}{1+e^{-0.22}}$$
$$p \approx \frac{1}{1+0.8019}$$
$$p \approx \frac{1}{1.8019}$$
$$p \approx 0.5556$$

So, the input to the sigmoid function for the first data point in the dataset is approximately 0.22, and after applying the sigmoid function, the predicted probability *p* is approximately 0.5556.

Homework for all other data

Feature 1 (X1)	Feature 2 (X2)	Feature 3 (X3)	Target (Y)
1.2	0.7	2.4	0
2.4	1.5	3.1	0
3.5	2.2	4.5	0
4.2	3.1	5.2	1
5.1	4.0	6.3	1

Given the weights and bias:

$$w_0 = 0.5$$
$$w_1 = -0.3$$
$$w_2 = 0.8$$
$$w_3 = -0.2$$

- **Data point 1**:

$$z = 0.5 + (-0.3 \times 1.2) + (0.8 \times 0.7) + (-0.2 \times 2.4)$$
$$z \approx 0.22$$
$$p \approx \frac{1}{1+e^{-0.22}} \approx 0.5556$$

- **Data point 2**:

$$z = 0.5 + (-0.3 \times 2.4) + (0.8 \times 1.5) + (-0.2 \times 3.1)$$
$$z \approx -0.39$$
$$p \approx \frac{1}{1+e^{0.39}} \approx 0.4036$$

- **Data point 3**:

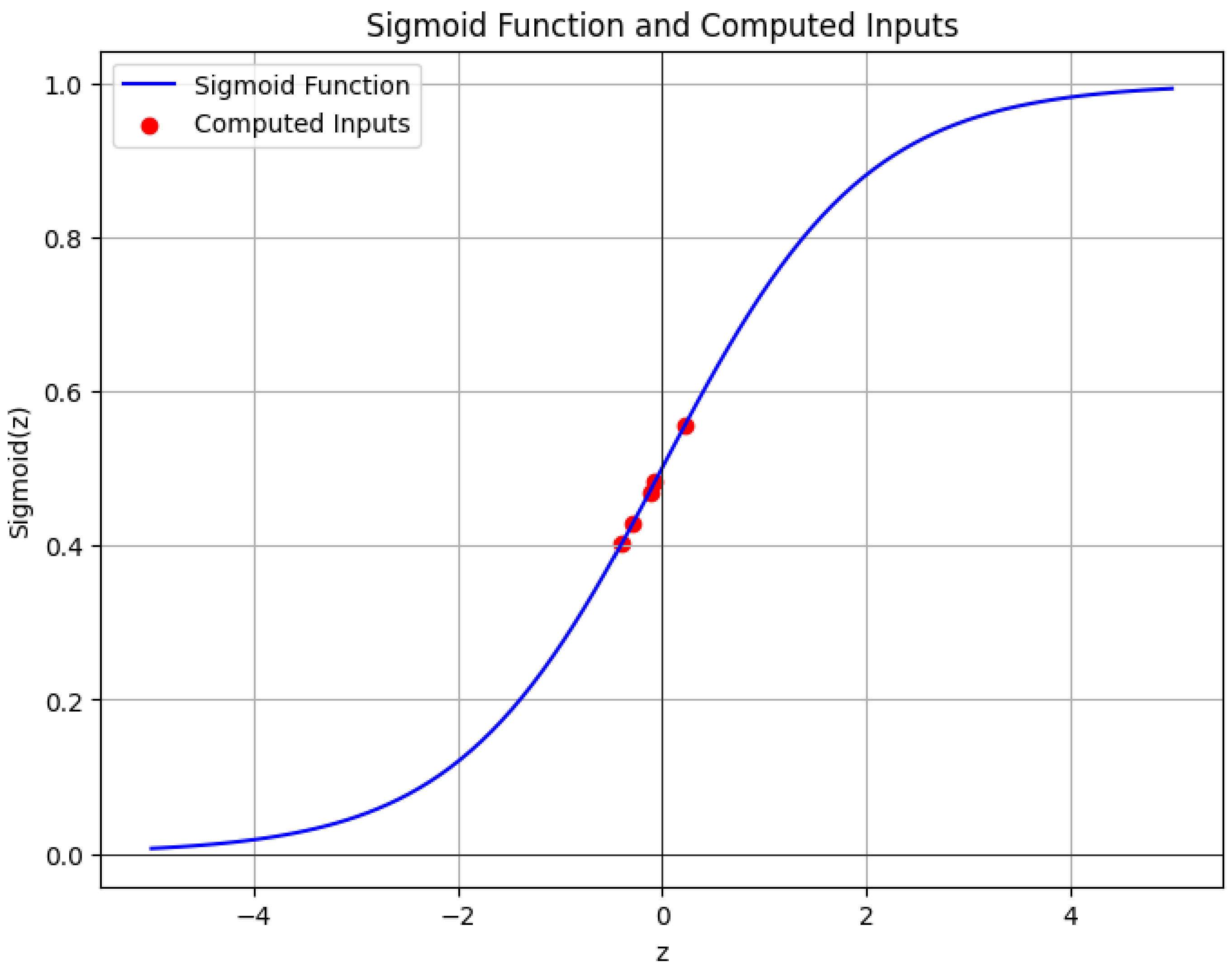
$$z = 0.5 + (-0.3 \times 3.5) + (0.8 \times 2.2) + (-0.2 \times 4.5)$$
$$z \approx -0.07$$
$$p \approx \frac{1}{1+e^{0.07}} \approx 0.4827$$

- **Data point 4**:

$$z = 0.5 + (-0.3 \times 4.2) + (0.8 \times 3.1) + (-0.2 \times 5.2)$$
$$z \approx -0.12$$
$$p \approx \frac{1}{1+e^{0.12}} \approx 0.4698$$

- **Data point 5**:

$$z = 0.5 + (-0.3 \times 5.1) + (0.8 \times 4.0) + (-0.2 \times 6.3)$$
$$z \approx -0.29$$
$$p \approx \frac{1}{1+e^{0.29}} \approx 0.4285$$

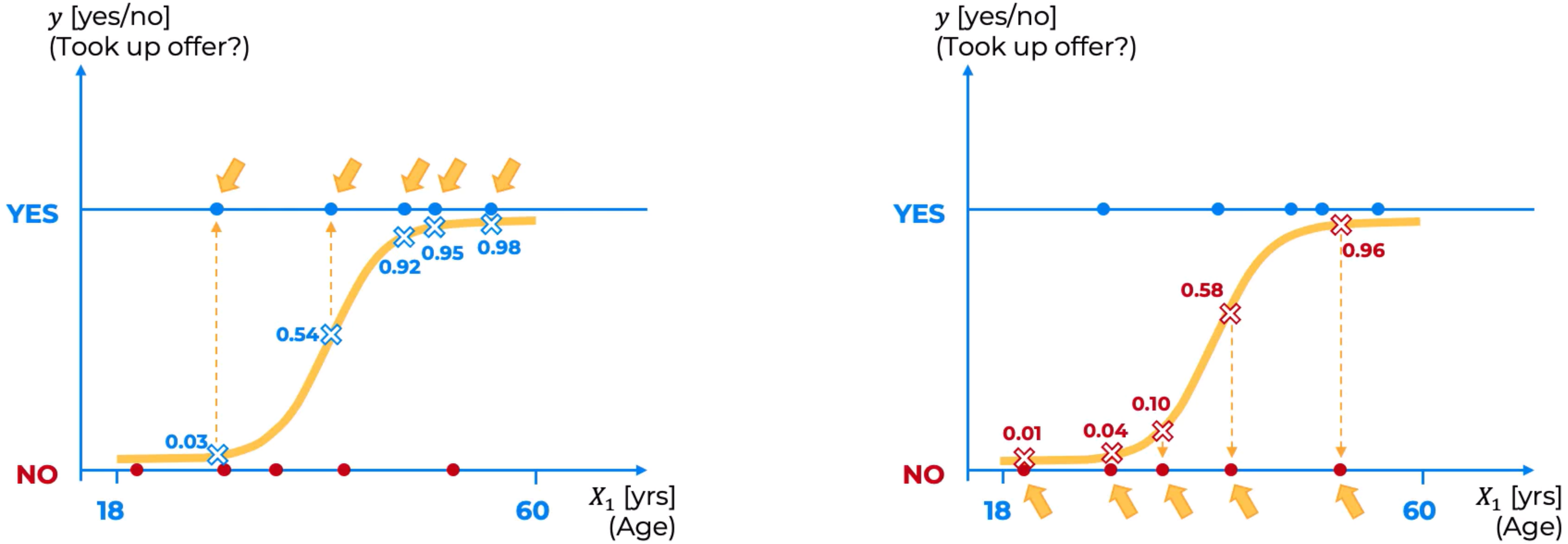


Sigmoid Function and Computed Inputs
Sigmoid Function
Computed Inputs
Sigmoid(z)
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
z

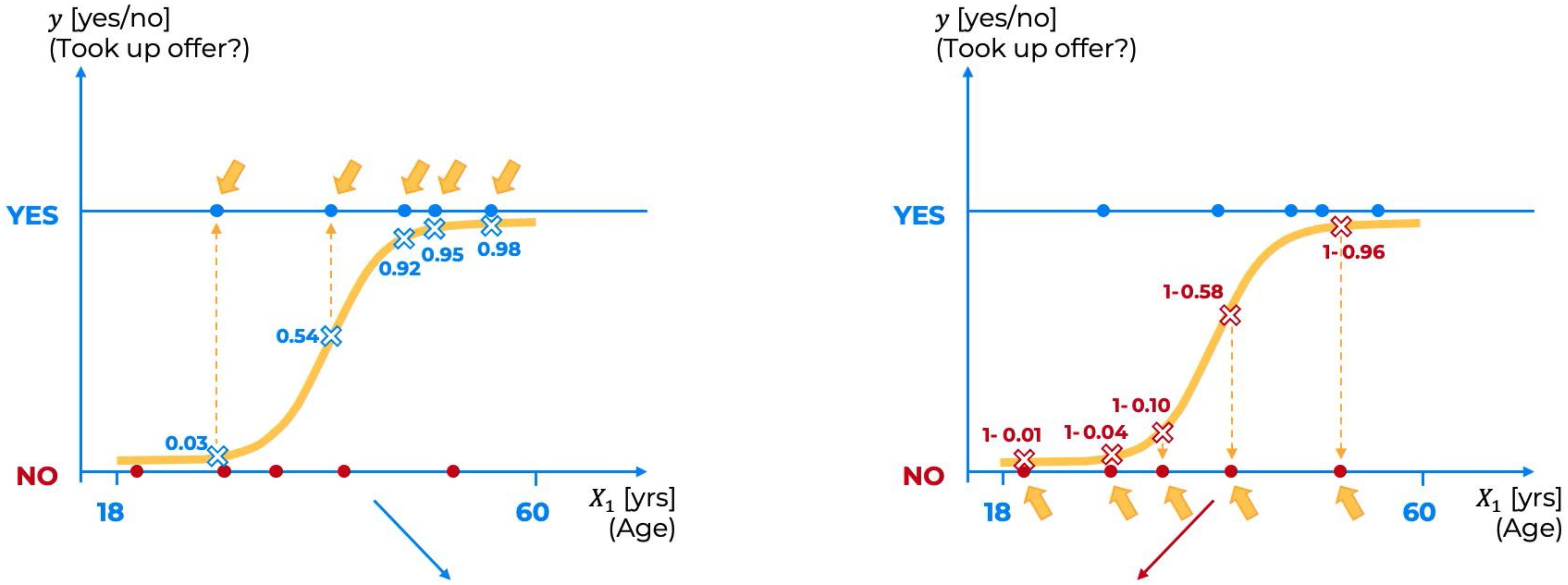
Likelihood Function for Logistic Regression

- the likelihood function is used to determine the optimal parameters (weights) of the model that maximize the likelihood of observing the given dataset. The likelihood function represents the probability of observing the data given the model parameters. Maximizing this likelihood is equivalent to minimizing the error between the predicted probabilities and the actual class labels.
- The predicted probabilities will be:
  - for y=1 The predicted probabilities will be: p(X;b,w) = p(x)
  - for y = 0 The predicted probabilities will be: 1-p(X;b,w) = 1-p(x)

$$L(b, w) = \prod_{i=1}^{n} p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}$$

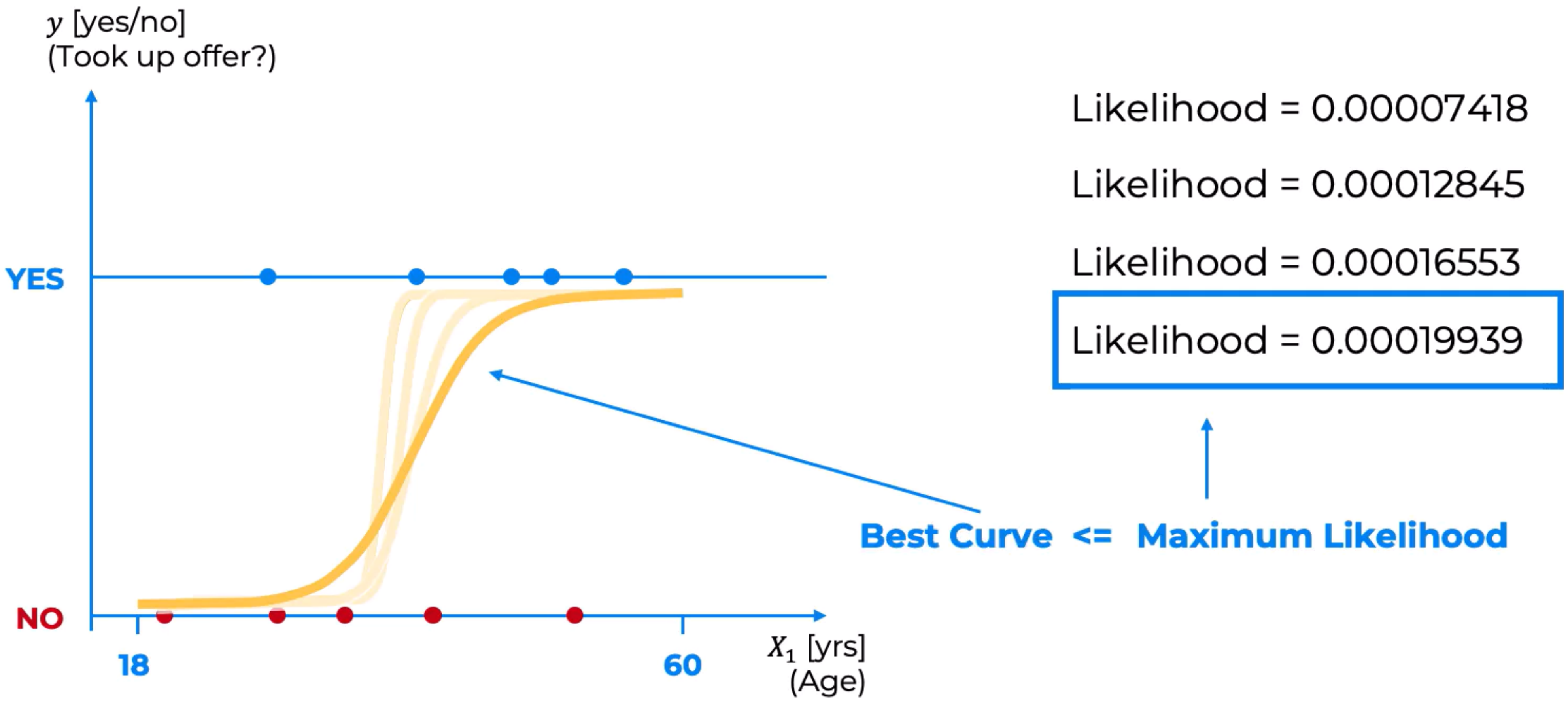


y [yes/no]
(Took up offer?)
YES
NO
0.03
0.54
0.92
0.95
0.98
18
60
X_1 [yrs]
(Age)
y [yes/no]
(Took up offer?)
YES
NO
0.01
0.04
0.10
0.58
0.96
18
60
X_1 [yrs]
(Age)



Likelihood = 0.03 x 0.54 x 0.92 x 0.95 x 0.98 x (1 – 0.01) x (1 – 0.04) x (1 – 0.10) x (1 – 0.58) x (1 – 0.96)

Likelihood = **0.00019939**



y [yes/no]
(Took up offer?)
YES
NO
18
60
X_1 [yrs]
(Age)
Likelihood = 0.00007418
Likelihood = 0.00012845
Likelihood = 0.00016553
Likelihood = 0.00019939
Best Curve <= Maximum Likelihood