

# Introduction to Logistic Regression

## Understanding Binary Classification

**Kazeem Adesina DAUDA**

AI4WIA BOOTCAMP

September 9, 2023

# Outline

- 1 Introduction to Logistic Regression
- 2 The Logistic Regression Model
- 3 Estimating Parameters
- 4 Assumptions of Logistic Regression
- 5 Interpretation of Results
- 6 Conclusion

# Logistic Regression

- Logistic regression is a statistical method used for binary classification.
- It models the probability of an event occurring (e.g., 1) as a function of one or more independent variables.
- It's commonly used in areas like medical diagnosis, spam detection, and more.

# The Logistic Regression Model

## Model

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

$P(Y = 1)$  : Probability of event occurring

$X$  : Independent Variable(s)

$\beta_0$  : Intercept

$\beta_1$  : Coefficient(s)

$e$  : Base of natural logarithm

# Estimating Parameters

- The goal of logistic regression is to estimate the parameters  $\beta_0$  and  $\beta_1$  that best fit the data.
- The estimation process typically involves maximum likelihood estimation.
- The logistic function models the relationship between the independent variables and the log-odds of the event.

# Assumptions of Logistic Regression

Logistic regression makes several assumptions about the data:

- 1 Linearity in log-odds: The relationship between independent variables and the log-odds of the event is linear.
- 2 Independence of errors: Observations are independent.
- 3 No multicollinearity: Independent variables are not highly correlated.

# Interpretation of Results

- Coefficients ( $\beta_0$  and  $\beta_1$ ) indicate the strength and direction of the relationship between independent variables and the log-odds of the event.
- Odds ratios are often used to interpret the impact of independent variables on the odds of the event.

# Conclusion

- Logistic regression is a valuable method for binary classification tasks.
- The logistic regression model estimates probabilities using a logistic function.
- Assumptions about linearity, independence, and multicollinearity should be considered.
- Interpretation of coefficients and odds ratios helps understand the impact of independent variables.



# Questions?