Module3

Machine Learning

Syllabus

1. **Similarity-based Learning:** Nearest-Neighbor Learning, Weighted K-Nearest-Neighbor Algorithm, Nearest Centroid Classifier, Locally Weighted Regression (LWR).
2. **Regression Analysis:** Introduction to Regression, Introduction to Linear Regression, Multiple Linear Regression, Polynomial Regression, Logistic Regression.
3. **Decision Tree Learning:** Introduction to Decision Tree Learning Model, Decision Tree Induction Algorithms.

Regression Analysis

- Introduction to Regression,
- Introduction to Linear Regression,
- Multiple Linear Regression,
- Polynomial Regression,
- Logistic Regression.

Regression Analysis

- Regression analysis is one of the most widely used and oldest supervised learning techniques.
- Regression analysis is a supervised learning technique used to predict continuous variables.
- Unlike classification methods that work with **categorical data**, regression focuses on estimating **numerical values**.
- It helps **identify linear and non-linear relationships** between variables within a dataset.

Regression Analysis

- It is primarily employed to model relationships between **variables and make predictions**.
- Given a training dataset **D** consisting of **N** data points (xi,yi), where i=1,2,...,N, regression helps establish a mathematical relationship between one or more **independent variables** xi and a **dependent variable** yi.
- This relationship is generally expressed as: $y=f(x)$
- Here, the independent variable x is also referred to as an **explanatory variable, predictor variable, covariate, or domain point**. The dependent variable y is often called a **label, target variable, or response variable.**

Regression Analysis

- Regression analysis measures **how the dependent variable changes** when an **independent variable is altered while keeping other factors constant**.
- This method helps in understanding the effect of exploratory variables, making regression essential for **prediction and forecasting**.

Key Objectives of Regression Analysis

Regression is used to predict **continuous** or **quantitative** variables such as price and revenue. The main goals of regression analysis include answering the following questions:

1. What is the relationship between the variables?
2. How strong is this relationship?
3. Is the relationship **linear** or **non-linear**?
4. What is the significance of each variable?
5. How much does each variable contribute to the outcome?

Applications of Regression Analysis

Regression is widely applied across various industries. Some common use cases include:

1. Predicting **sales** of goods or services.
2. Estimating the **value of bonds** in portfolio management.
3. Determining **insurance premiums** for companies.
4. Forecasting **agricultural crop yields**.
5. Evaluating **real estate prices**.

Introduction to Linearity, Correlation, and Causation

- The accuracy of regression analysis depends on factors such as **correlation** and **causation**.

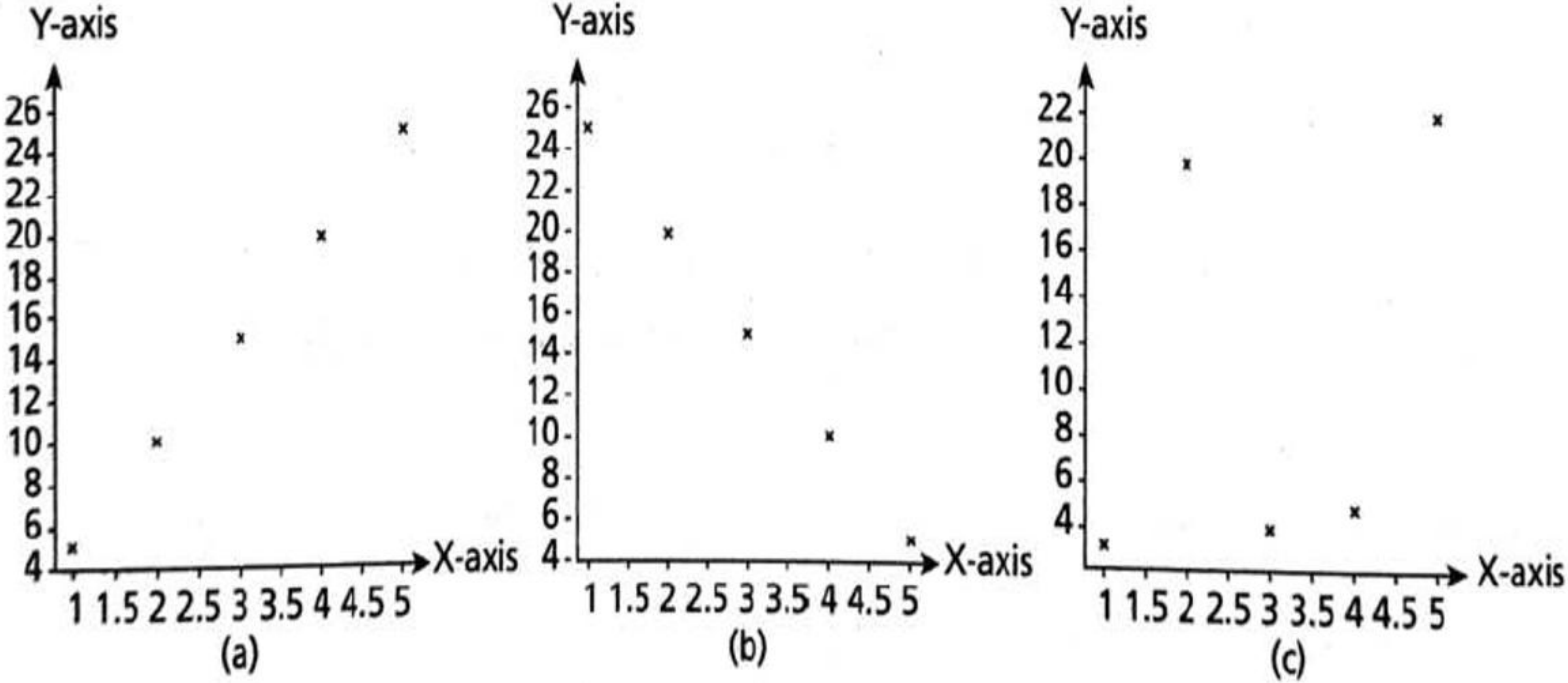
Regression and Correlation

- **Correlation** describes the relationship between two variables, such as **x** and **y**, **regression** focuses on predicting one variable based on another.
- To measure correlation strength, the **Pearson correlation coefficient (r)** is commonly used.
- This coefficient determines the degree of **association** between two variables:
  - **Positive correlation**: As one variable increases, the other also increases.
  - **Negative correlation**: As one variable increases, the other decreases.
  - **No correlation (random points)**: There is no observable relationship between variables.

Regression and Correlation

- Correlation between two variables can be effectively visualized using a **scatter plot**, which is a **2D graph** representing the relationship between **explanatory (independent) variables** and **response (dependent) variables**.
- In a **scatter plot**:
  - The **x-axis** represents the **independent** or **predictor variables**.
  - The **y-axis** represents the **dependent** or **predicted variables**.
- Scatter plots help in **data exploration** by identifying patterns between variables.

Types of Correlations (+ve, -ve and No)



Regression and Causation

- **Causation** refers to a direct cause-and-effect relationship between variables, where **x causes y** to occur or vice versa.
- It is often represented as **x → y**.
- However, **correlation and regression do not necessarily imply causation**.

Regression and Causation

- For instance, a correlation between **economic background** and **exam scores** does not mean that a better economic background directly causes higher marks.
- Similarly, an increase in **cool drink sales** due to rising temperatures does not establish a **direct causal link—other factors may also influence sales**.
- While high temperature **contributes** to increased demand, it is not the **sole cause**.

Linearity and Non-Linearity Relationships

- A linear relationship between variables means that the dependent and independent variables are related in a way that can be represented by a straight line.
- This relationship is expressed using the equation: $y=ax+b$ where a and b are constants.
- In a linear relationship, as one variable increases, the corresponding variable also changes in a predictable linear manner.

Linearity and Non-Linearity Relationships

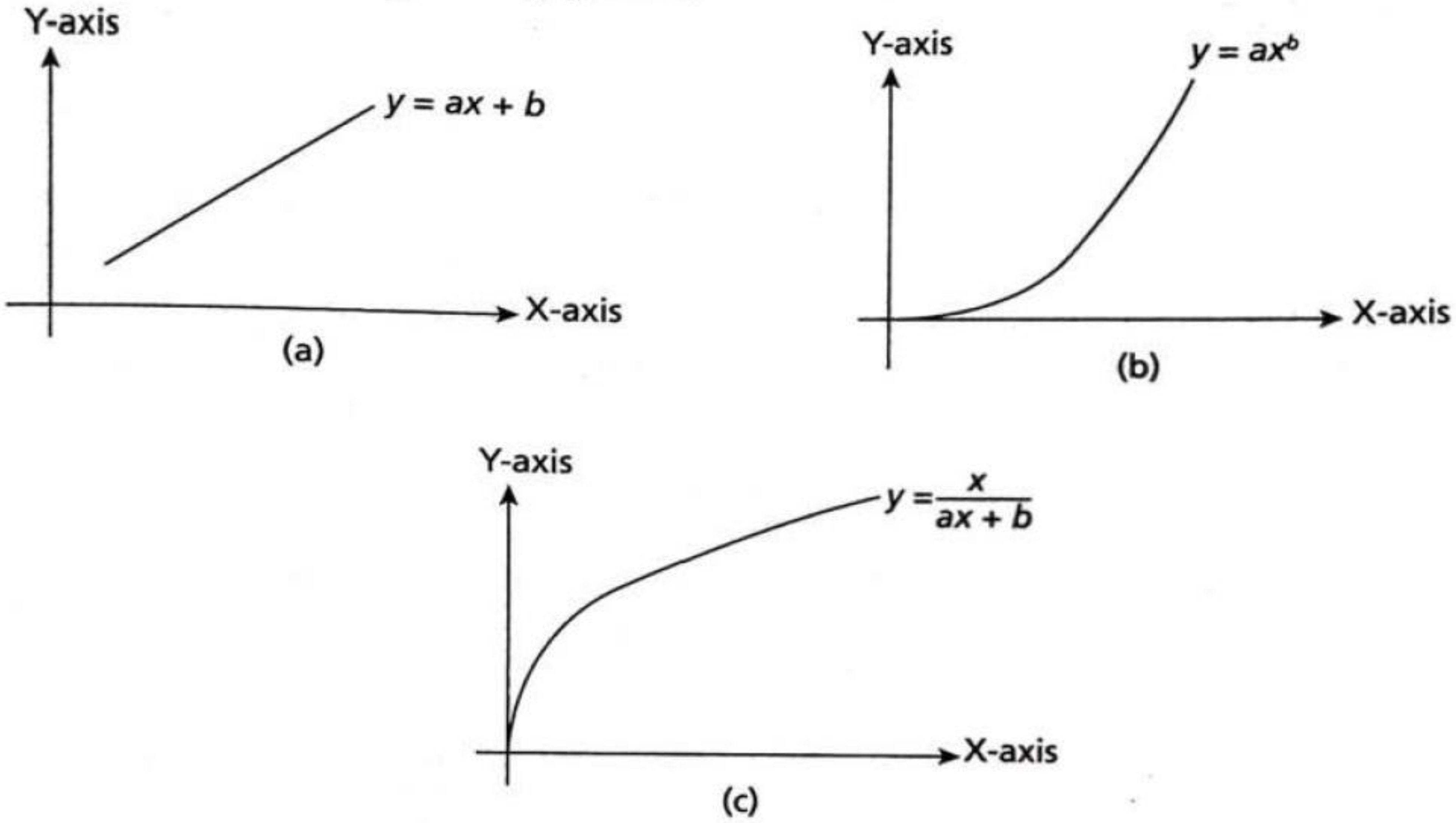
- On the other hand, non-linear relationships exist in cases where the data does not follow a straight-line pattern.
- These relationships are often found in exponential and power functions.
- In these cases, the x-axis represents the independent variable, while the y-axis represents the dependent variable.

Linearity and Non-Linearity Relationships

- Figure provides three graphical examples:

- **(a) Linear Relationship:** Represented by the equation $y = ax + b$, which forms a straight line.
- **(b) Non-Linear Relationship (Power Function):** Represented by $y = ax^b$, showing an exponential curve.
- **(c) Non-Linear Relationship (Fractional Function):** Represented by $y = \frac{x}{ax+b}$, which forms a curved graph.

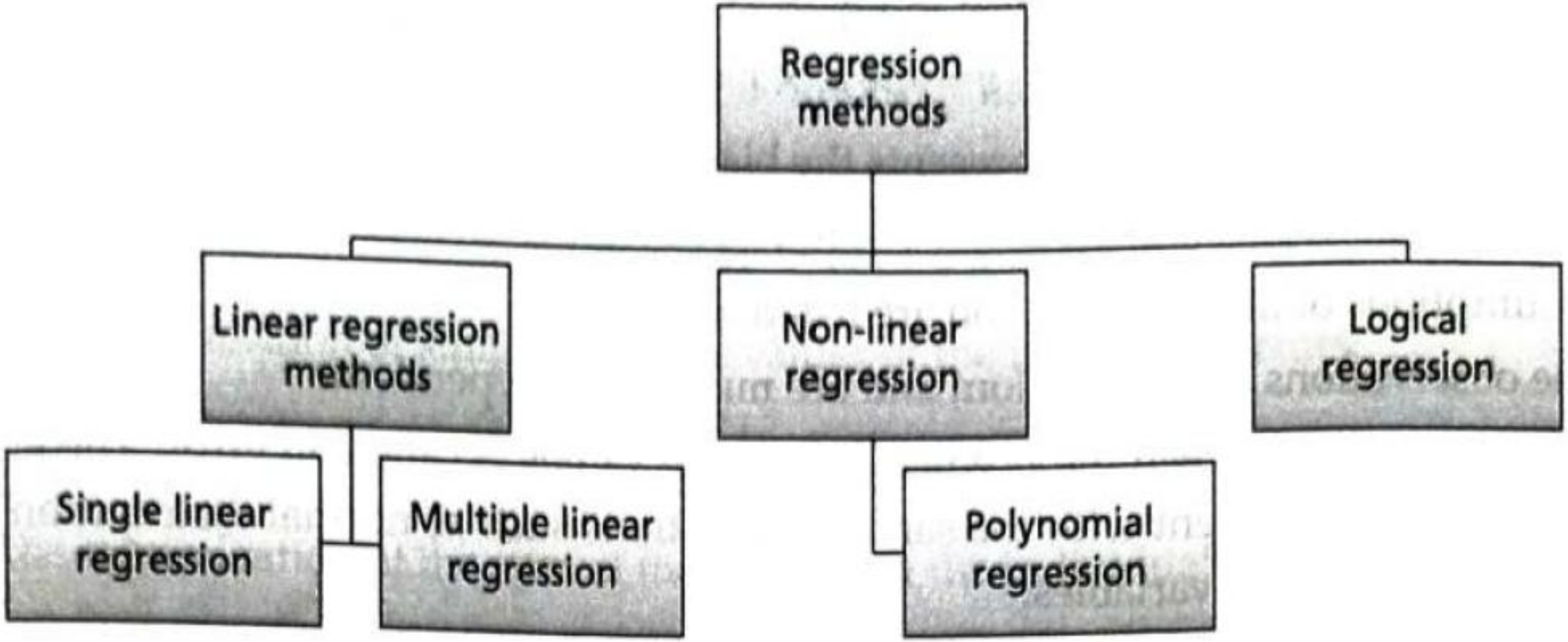
Linearity and Non-Linearity Relationships



What is Regression?

- In machine learning, **regression** is a type of supervised learning technique used to predict a **continuous target variable** (also known as the dependent variable) based on **one or more input features** (independent variables).
- The **main goal of regression** is to model the relationship between the input variables and the target variable so that you can make **predictions on new, unseen data**.
- **Continuous Output**: Unlike classification, where the output is a category (**e.g., yes/no or class labels**), regression deals with continuous outputs (**e.g., price, temperature, sales figures**).

Types of Regression Methods



Types of Regression Methods

1. **Linear Regression : Linear regression fits a straight line** to the given data to determine the relationship between one independent variable and one dependent variable. This method is useful for identifying and describing linear relationships.
2. **Multiple Regression :** Multiple regression **extends linear regression by considering two or more independent variables to predict a single dependent variable**. It helps in analyzing relationships among multiple variables.
3. **Polynomial Regression : Polynomial regression is a type of non-linear regression where an Nth-degree polynomial is used to model the relationship between the independent and dependent variables**. This method is particularly useful when data patterns exhibit curvature rather than a straight-line trend. Polynomial multiple regression extends this approach to cases with multiple independent variables.
4. **Logistic Regression: Logistic regression is used for classifying categorical variables**. It models the relationship between one or more independent variables and a dependent variable that falls into categories, such as "yes/no" or "pass/fail." It is also known as a **binary classifier** in cases where there are only two possible outcomes.
5. **Lasso and Ridge Regression:** Lasso and Ridge regression methods apply **regularization** techniques to limit the size and number of coefficients of the independent variables. These methods help in preventing overfitting and improving model performance.

1. Linear Regression

$$y = \beta_0 + \beta_1 x + \epsilon$$

Example: Predicting house price (y) based on square footage (x):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \epsilon$$

2. Multiple Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

Example: Predicting house price (y) based on square footage (x_1) and number of bedrooms (x_2):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \beta_2 \times \text{Bedrooms} + \epsilon$$

3. Polynomial Regression

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \epsilon$$

Example: Predicting house price (y) using a quadratic relationship with square footage (x):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \beta_2 \times \text{SquareFootage}^2 + \epsilon$$

4. Logistic Regression

$$P(y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}}$$

Example: Predicting whether a customer will buy a product (y) based on income (x_1) and age (x_2):

$$P(\text{Buy}=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \times \text{Income} + \beta_2 \times \text{Age})}}$$

a.k.a. **Log Odds**

or **Logit**

Intercept

$$\boxed{\log\left(\frac{P}{1-P}\right)} = \beta_0 + \beta_1 X$$

5. Lasso Regression

$$\min_{\beta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n} |\beta_j|$$

where

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij}$$

Example: Predicting house price while reducing unimportant features using L1 regularization.

6. Ridge Regression

$$\min_{\beta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n} \beta_j^2$$

where

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij}$$

Example: Predicting house price while penalizing large coefficients using L2 regularization.

Regularization in Regression

- **Regularization** is a technique used in regression models to prevent **overfitting** by adding a penalty to the model's complexity.
- It introduces a **regularization term** (also called a **penalty term**) to the loss function, which helps in controlling the magnitude of regression coefficients (β).

Why is Regularization Needed?

- In high-dimensional datasets, models may fit the training data too well, capturing noise **rather than true patterns (overfitting).**
- Regularization reduces the effect of **less important features** by **shrinking coefficients toward zero**.
- It helps in improving the **model's generalization ability on unseen data**.

Types of Regularization in Regression

1. Lasso Regression (L1 Regularization)

- Uses the **L1 norm penalty**, adding the absolute values of the coefficients to the loss function.
- Some coefficients can become exactly **zero**, leading to feature selection.

$$\min_{\beta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n} |\beta_j|$$

Types of Regularization in Regression

2. Ridge Regression (L2 Regularization)

- Uses the **L2 norm penalty**, adding the squared values of the coefficients to the loss function.
- Shrinks coefficients but does not make them exactly zero.

$$\min_{\beta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n} \beta_j^2$$

Types of Regularization in Regression

3. Elastic Net (Combination of L1 & L2)

- A hybrid approach that combines **Lasso** and **Ridge** regression.
- Helps when features are highly correlated.

$$\min_{\beta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^{n} |\beta_j| + \lambda_2 \sum_{j=1}^{n} \beta_j^2$$

Choosing the Right Regularization

- **Use Lasso** if feature selection is needed.
- **Use Ridge** when all features contribute but should be regularized.
- **Use Elastic Net** when features are correlated.

Limitations of Regression Methods

- **Outliers** – **Abnormal data points** can distort the regression model by shifting the regression line toward them, leading to **biased results.**
- **Sample Size** – The ratio of **independent to dependent variables** should be at least **20:1**. Each explanatory variable should have at least **20 samples**, with a minimum of five samples required in extreme cases.
- **Missing Data** – Incomplete training data can reduce the reliability and accuracy of the model, making it unfit for predictions.
- **Multicollinearity** – When explanatory variables are highly correlated (above 0.9), the model becomes biased. Perfect correlation (1.0) leads to singularity issues. To mitigate this, highly correlated variables should be removed. If multiple variables have the same correlation, the tolerance **measure (1 - R^2) is used to eliminate the variable with the highest value.**

What is Regression?

- In machine learning, **regression** is a type of supervised learning technique used to predict a **continuous target variable** (also known as the dependent variable) based on **one or more input features** (independent variables).
- The **main goal of regression** is to model the relationship between the input variables and the target variable so that you can make **predictions on new, unseen data**.
- **Continuous Output**: Unlike classification, where the output is a category (**e.g., yes/no or class labels**), regression deals with continuous outputs (**e.g., price, temperature, sales figures**).

Introduction to Linear Regression

Linear regression is a statistical method used to model the relationship between a dependent variable and an independent variable by fitting a straight line to the data points. The equation of the regression line is given by:

$$y = a_0 + a_1 \cdot x + e$$

where:

- a_0 is the **intercept**, representing the bias.
- a_1 is the **slope**, indicating the rate of change of y with respect to x.
- e represents the **error** in prediction.

Assumptions of Linear Regression

- The observations (y) are **random and independent** of each other.
- The error (difference between predicted and actual values) follows a normal distribution with a **zero mean and constant variance**.
- The distribution of the **error term is independent** of the explanatory variables.
- The parameters of the regression model ($a0$,$a1$) **remain constant**.

Ordinary Least Squares (OLS) Method

- Linear regression is typically implemented using the **Ordinary Least Squares (OLS)** method, also known as the **least squares method**. This approach finds the best-fitting line by minimizing the sum of squared errors.

Linear regression is solved by minimizing the sum of squared residuals:

$$J(a_0, a_1) = \sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i)]^2$$

Estimation of Parameters

By solving the minimization problem, the coefficients a_0 and a_1 are calculated as:

$$a_1 = \frac{\overline{xy} - (\bar{x} \cdot \bar{y})}{\overline{x^2} - (\bar{x}^2)}$$

$$a_0 = \bar{y} - a_1 \cdot \bar{x}$$

where $\bar{x}$ and $\bar{y}$ are the means of x and y, respectively.

Ordinary Least Squares (OLS) Method

- Each **data point** has a corresponding predicted value from the **regression line**.
- The **vertical distance** between a data point and the predicted value is called the **error (residual).**
- The sum of **squared residuals** is computed to measure how well the **line fits the data**.
- The **best-fit line is the one that minimizes this sum of squared errors**.

Ordinary Least Squares (OLS) Method

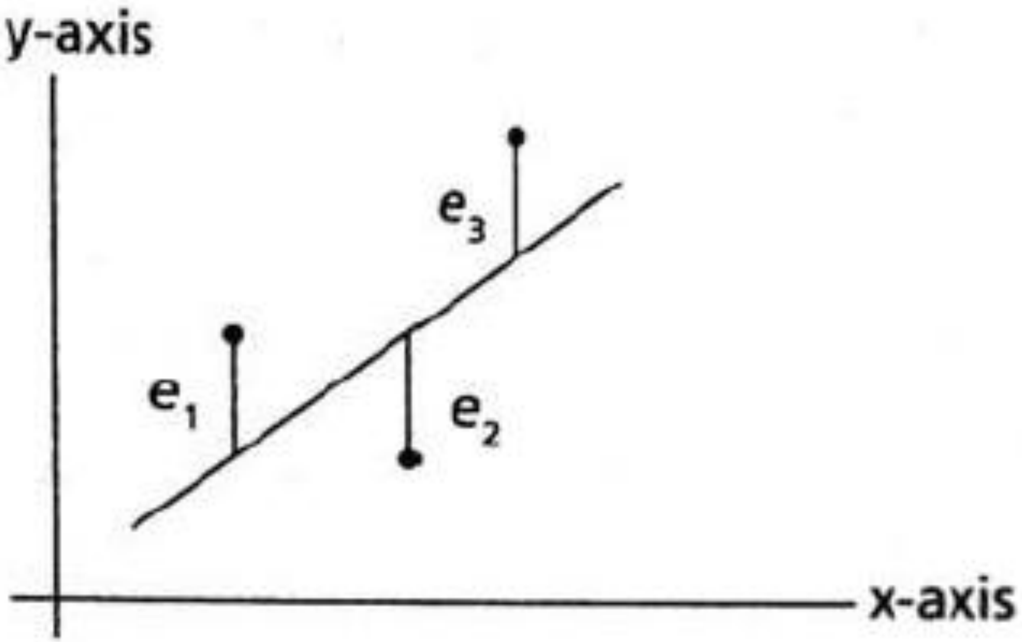
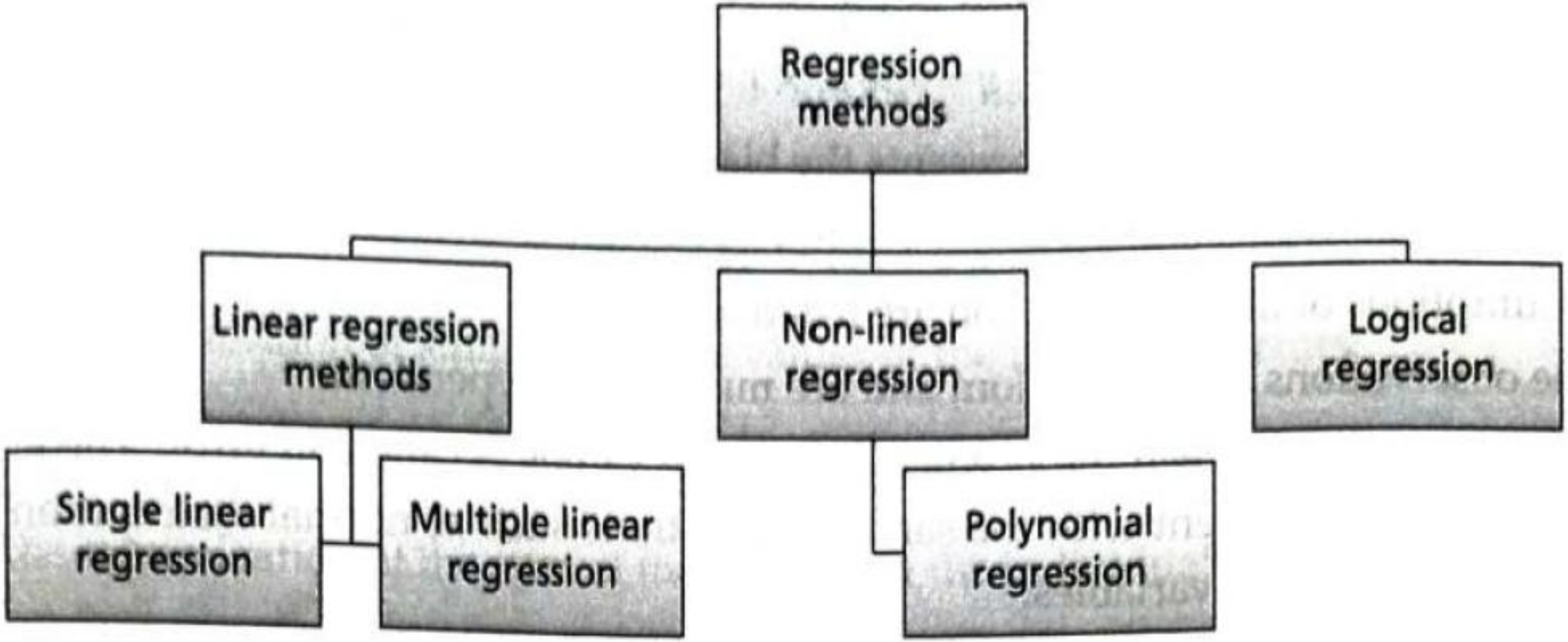


Figure 5.4: Data Points and their Errors

Types of Regression Methods



1. Linear Regression

$$y = \beta_0 + \beta_1 x + \epsilon$$

Example: Predicting house price (y) based on square footage (x):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \epsilon$$

2. Multiple Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

Example: Predicting house price (y) based on square footage (x_1) and number of bedrooms (x_2):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \beta_2 \times \text{Bedrooms} + \epsilon$$

3. Polynomial Regression

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \epsilon$$

Example: Predicting house price (y) using a quadratic relationship with square footage (x):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \beta_2 \times \text{SquareFootage}^2 + \epsilon$$

4. Logistic Regression

$$P(y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}}$$

Example: Predicting whether a customer will buy a product (y) based on income (x_1) and age (x_2):

$$P(\text{Buy} = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \times \text{Income} + \beta_2 \times \text{Age})}}$$

a.k.a. **Log Odds**

or **Logit**

Intercept

$$\boxed{\log\left(\frac{P}{1-P}\right)} = \beta_0 + \beta_1 X$$

5. Lasso Regression

$$\min_{\beta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n} |\beta_j|$$

where

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij}$$

Example: Predicting house price while reducing unimportant features using L1 regularization.

6. Ridge Regression

$$\min_{\beta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n} \beta_j^2$$

where

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij}$$

Example: Predicting house price while penalizing large coefficients using L2 regularization.

Linear Regression : Estimation of Parameters using OLS methods

By solving the minimization problem, the coefficients a_0 and a_1 are calculated as:

$$a_1 = \frac{\overline{xy} - (\bar{x} \cdot \bar{y})}{\overline{x^2} - (\bar{x}^2)}$$

$$a_0 = \bar{y} - a_1 \cdot \bar{x}$$

where $\bar{x}$ and $\bar{y}$ are the means of x and y, respectively.

Example 5.1:

- Consider a scenario where the sales data for five weeks (in thousands) is provided in Table 5.1. Using this dataset, apply the linear regression technique to predict the sales for the 7th and 12th weeks.

Table 5.1: Sample Data

Week (x_i)	Sales (y_i) (in Thousands)
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

Solution:

Since there are 5 data points (i=1,2,3,4,5), we compute the necessary statistical values in Table 5.2.

Table 5.2: Computation Table

x_i	y_i	x_i^2	$x_i \times y_i$
1	1.2	1	1.2
2	1.8	4	3.6
3	2.6	9	7.8
4	3.2	16	12.8
5	3.8	25	19

Table 5.2: Computation Table

x_i	y_i	x_i^2	$x_i \times y_i$
1	1.2	1	1.2
2	1.8	4	3.6
3	2.6	9	7.8
4	3.2	16	12.8
5	3.8	25	19

Step 1: Compute the Averages

- Mean of x: $\bar{x} = \frac{15}{5} = 3$
- Mean of y: $\bar{y} = \frac{12.6}{5} = 2.52$
- Mean of x^2: $\overline{x^2} = \frac{55}{5} = 11$
- Mean of $x_i \times y_i$: $\overline{xy} = \frac{44.4}{5} = 8.88$

Step 2: Compute the Slope (a_1) and Intercept (a_0)

Using the formula for slope:

$$a_1 = \frac{\overline{xy} - (\bar{x} \cdot \bar{y})}{\overline{x^2} - (\bar{x}^2)}$$

$$a_1 = \frac{8.88 - (3 \times 2.52)}{11 - 3^2} = \frac{8.88 - 7.56}{11 - 9} = \frac{1.32}{2} = 0.66$$

Using the formula for the intercept:

$$a_0 = \bar{y} - a_1 \cdot \bar{x}$$

$$a_0 = 2.52 - (0.66 \times 3) = 2.52 - 1.98 = 0.54$$

Step 3: Construct the Regression Model

The equation of the fitted line is:

$$y = 0.54 + 0.66x$$

Step 4: Predict Sales for the 7th and 12th Weeks

Using the equation:

- For $x = 7$:

$$y = 0.54 + 0.66 \times 7 = 0.54 + 4.62 = 5.16$$

- For $x = 12$:

$$y = 0.54 + 0.66 \times 12 = 0.54 + 7.92 = 8.46$$

Thus, the predicted sales for the 7th week are **5.16 thousand**, and for the 12th week, **8.46 thousand**.

Example: Finding the Inverse of a 2×2 Matrix

For a given 2×2 matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The **inverse** of A, denoted as A^{-1}, is calculated using the formula:

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where **det(A)** is the determinant of matrix A, given by:

$$\det(A) = (ad - bc)$$

Example Calculation

Consider the matrix:

$$A = \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}$$

1. **Calculate the Determinant**

$$\det(A) = (4 \times 6) - (7 \times 2) = 24 - 14 = 10$$

Since the determinant is non-zero ($\det(A) \neq 0$), the matrix is invertible.

2. **Apply the Inverse Formula**

$$A^{-1} = \frac{1}{10}\begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix}$$

3. **Final Result**

$$A^{-1} = \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix}$$

Verification (Optional)

To verify, multiply A with A^{-1}:

$$A \cdot A^{-1} = \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix} \times \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix}$$

Perform matrix multiplication:

$$\begin{bmatrix} (4 \times 0.6) + (7 \times -0.2) & (4 \times -0.7) + (7 \times 0.4) \\ (2 \times 0.6) + (6 \times -0.2) & (2 \times -0.7) + (6 \times 0.4) \end{bmatrix}$$

$$= \begin{bmatrix} 2.4 - 1.4 & -2.8 + 2.8 \\ 1.2 - 1.2 & -1.4 + 2.4 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Linear Regression in Matrix Form

Matrix notation can be used to represent independent and dependent variables in regression analysis.

The equation for simple linear regression can be written in matrix form as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

This can be written compactly as:

$$Y = Xa + e$$

The equation for simple linear regression can be written in matrix form as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

This can be written compactly as:

$$Y = Xa + e$$

where:

- X is an $n \times 2$ matrix,
- Y is an $n \times 1$ vector,
- a is a 2×1 column vector of regression coefficients,
- e is an $n \times 1$ column vector of residuals.

Example 5.2: Linear Regression in Matrix Form

Find the linear regression equation for the **weekly product sales data** (in thousands) given in **Table 5.3**.

Table 5.3: Sample Data for Regression

x_i (Week)	y_i (Product Sales in Thousands)
1	1
2	3
3	4
4	8

Solution:

The independent variable X (week) and dependent variable Y (sales) are defined as:

$$X^T = [1 \quad 2 \quad 3 \quad 4]$$

$$Y^T = [1 \quad 3 \quad 4 \quad 8]$$

In matrix form, these are written as:

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

$$Y = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 8 \end{bmatrix}$$

where the first column of **X** is used to incorporate the bias term a_0.

Regression Calculation Formula:

The equation to compute the regression coefficients a is:

$$a = (X^T X)^{-1} X^T Y$$

The step-by-step computation of this equation follows.

Step-by-Step Computation of Linear Regression in Matrix Form

The process of computing the regression coefficients using the matrix formula $a = (X^T X)^{-1} X^T Y$ is outlined below:

1. Compute $X^T X$:

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$$

2. Compute the inverse of $(X^T X)$:

$$(X^T X)^{-1} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1} = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix}$$

2. Compute the inverse of $(X^T X)$:

$$(X^T X)^{-1} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1} = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix}$$

---

3. Compute $(X^T X)^{-1} X^T$:

$$(X^T X)^{-1} X^T = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 0 & -0.5 \\ -0.3 & -0.01 & 0.1 & 0.3 \end{bmatrix}$$

3. Compute $(X^T X)^{-1} X^T$:

$$(X^T X)^{-1} X^T = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 0 & -0.5 \\ -0.3 & -0.01 & 0.1 & 0.3 \end{bmatrix}$$

---

4. Compute $(X^T X)^{-1} X^T Y$:

$$(X^T X)^{-1} X^T Y = \begin{bmatrix} 1 & 0.5 & 0 & -0.5 \\ -0.3 & -0.01 & 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} -1.5 \\ 2.2 \end{bmatrix}$$

4. Compute $(X^T X)^{-1} X^T Y$:

$$(X^T X)^{-1} X^T Y = \begin{bmatrix} 1 & 0.5 & 0 & -0.5 \\ -0.3 & -0.01 & 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} -1.5 \\ 2.2 \end{bmatrix}$$

Thus, the computed regression coefficients are:

- **Intercept** (a_0) = **-1.5**
- **Slope** (a_1) = **2.2**

Thus, the computed regression coefficients are:

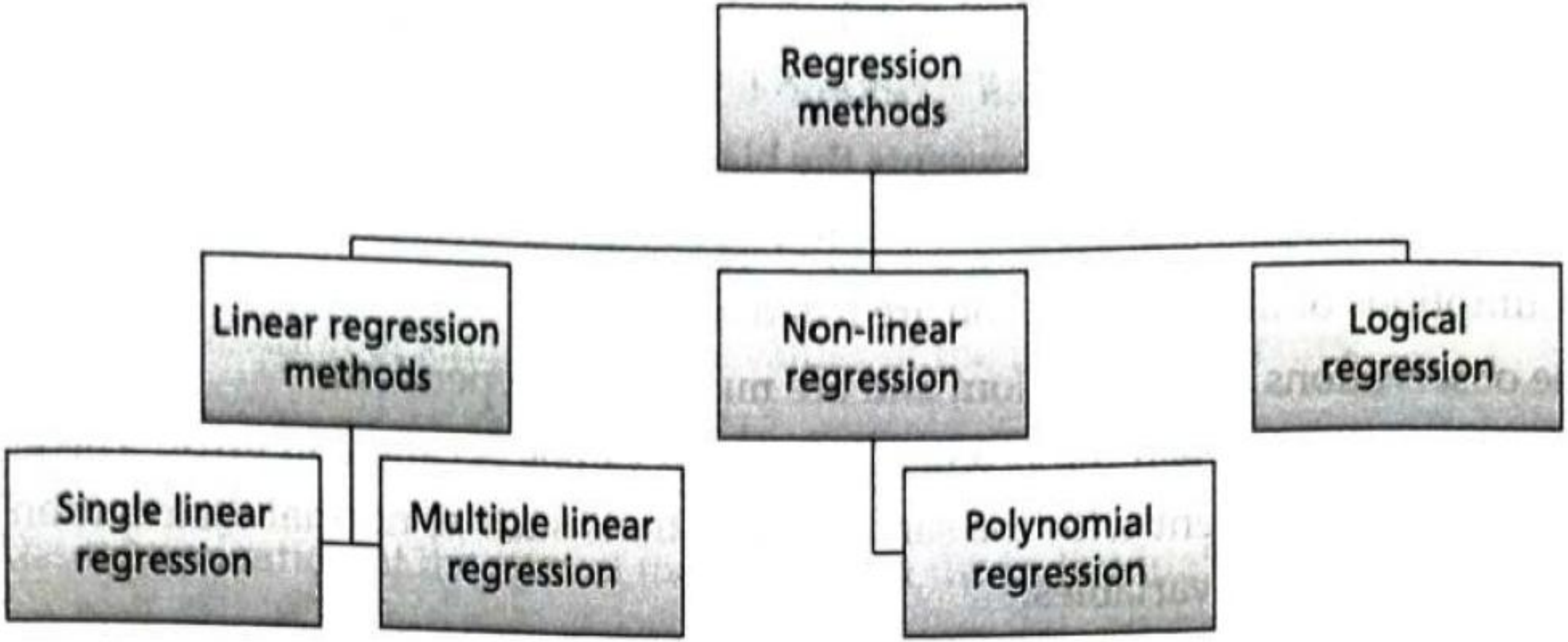
- **Intercept** (a_0) = **-1.5**
- **Slope** (a_1) = **2.2**

Final Regression Equation:

$$y = 2.2x - 1.5$$

This equation represents the fitted linear regression line.

Types of Regression Methods



1. Linear Regression

$$y = \beta_0 + \beta_1 x + \epsilon$$

Example: Predicting house price (y) based on square footage (x):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \epsilon$$

2. Multiple Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

Example: Predicting house price (y) based on square footage (x_1) and number of bedrooms (x_2):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \beta_2 \times \text{Bedrooms} + \epsilon$$

3. Polynomial Regression

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \epsilon$$

Example: Predicting house price (y) using a quadratic relationship with square footage (x):

$$\text{Price} = \beta_0 + \beta_1 \times \text{SquareFootage} + \beta_2 \times \text{SquareFootage}^2 + \epsilon$$

4. Logistic Regression

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}}$$

Example: Predicting whether a customer will buy a product (y) based on income (x_1) and age (x_2):

$$P(\text{Buy} = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \times \text{Income} + \beta_2 \times \text{Age})}}$$

Multiple Linear Regression

- Multiple regression involves using **multiple predictor (independent) variables** to estimate a **dependent variable**.
- This **extends simple linear regression** by incorporating more explanatory factors.

The fundamental assumptions of multiple linear regression include:

- **Independent variables are not highly correlated** (to avoid multicollinearity).
- **Residuals follow a normal distribution**.

For two predictor variables x_1 and x_2, the multiple regression equation is:

$$y = f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2$$

More generally, for n independent variables:

$$y = f(x_1, x_2, x_3, ..., x_n) = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n + \varepsilon$$

where:

- $x_1, x_2, ..., x_n$ are predictor variables,
- y is the dependent variable,
- $a_0, a_1, ..., a_n$ are regression coefficients,
- ε represents the error term.

Example 5.5:

Apply multiple regression to the data in **Table 5.7**, which presents weekly sales y alongside sales of products x_1 and x_2. Use the matrix method for calculation.

Table 5.7: Sample Data

x_1 (Product One Sales)	x_2 (Product Two Sales)	y (Output Weekly Sales in Thousands)
1	4	1
2	5	6
3	8	8
4	2	12

Solution

We represent the data in matrix form:

$$X = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 \\ 6 \\ 8 \\ 12 \end{bmatrix}$$

The regression coefficients are given by:

$$\hat{a} = (X^T X)^{-1} X^T Y$$

Substituting the values, we obtain:

$$\hat{a} = \begin{bmatrix} -1.69 \\ 3.48 \\ -0.05 \end{bmatrix}$$

Thus, the final multiple regression equation is:

$$y = -1.69 + 3.48x_1 - 0.05x_2$$

This equation can be used to predict weekly sales based on product sales data.

Polynomial Regression

- When the relationship between the **independent and dependent variables is non-linear, traditional linear regression is ineffective and leads to significant errors.** To address non-linear regression problems, two methods can be used:
- **Data Transformation**: Converting **non-linear data into a linear form**, making it suitable for linear regression techniques.
- **Polynomial Regression**: Fitting a polynomial equation to model the non-linear relationship.

Transformations

The first approach involves transforming non-linear data into a linear form, allowing the use of standard linear regression techniques. For instance, consider the exponential function:

$$y = ae^{bx}$$

Applying the natural logarithm to both sides results in:

$$\ln y = bx + \ln a$$

This transformation converts the exponential equation into a linear form.

Similarly, a power function:

$$y = ax^b$$

can be transformed using the logarithmic function:

$$\log_{10} y = b \log_{10} x + \log_{10} a$$

Once transformed, linear regression techniques can be applied to estimate the parameters. The inverse transformation can then be used to obtain the final results.

Polynomial Regression

- Polynomial regression directly models non-linear relationships by using an **nth-degree** polynomial.
- Unlike data transformation, this approach fits **polynomial curves to capture different levels of curvature**.

Polynomial regression generates non-linear curves such as quadratic and cubic functions. A second-degree polynomial (quadratic transformation) takes the form:

$$y = a_0 + a_1 x + a_2 x^2$$

while a third-degree polynomial (cubic transformation) is given as:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

Polynomials of degree 4 and higher are generally avoided due to their tendency to overfit, resulting in unnecessarily complex curves.

Fitting a Second-Degree Polynomial

Given data points $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, the objective is to fit a quadratic function:

$$y = a_0 + a_1 x + a_2 x^2$$

The error function to minimize is:

$$E = \sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i + a_2 x_i^2)]^2$$

Error Function (Least Squares Method)

To find the best-fit polynomial, the error function (also called the residual sum of squares, RSS) is minimized:

$$E = \sum_{i=1}^{n} (y_i - (a_0 + a_1 x_i + a_2 x_i^2))^2$$

This function measures the difference between the predicted values from the polynomial model and the actual observed values y_i. The objective is to determine a_0, a_1, a_2 such that this error is minimized.

$$E = \sum_{i=1}^{n} (y_i - (a_0 + a_1 x_i + a_2 x_i^2))^2$$

Deriving the Normal Equations

To minimize the error function, we take the partial derivatives of E with respect to each coefficient a_0, a_1, a_2 and set them to zero:

$$\frac{\partial E}{\partial a_0} = 0, \quad \frac{\partial E}{\partial a_1} = 0, \quad \frac{\partial E}{\partial a_2} = 0$$

Partial Derivative with Respect to a_0

$$\frac{\partial E}{\partial a_0} = \sum_{i=1}^{n} 2(y_i - (a_0 + a_1 x_i + a_2 x_i^2)) \cdot (-1) = 0$$

Simplifying:

$$\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2) = 0$$

Rearrange:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 x_i + \sum_{i=1}^{n} a_2 x_i^2$$

Since a_0, a_1, and a_2 are constants, they can be factored out:

$$n a_0 + \left(\sum x_i\right) a_1 + \left(\sum x_i^2\right) a_2 = \sum y_i$$

This is the **first normal equation.**

To determine the coefficients a_0, a_1, and a_2, we take the partial derivatives with respect to each coefficient, set them to zero, and solve the resulting system of equations:

$$na_0 + \left(\sum x_i\right) a_1 + \left(\sum x_i^2\right) a_2 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 + \left(\sum x_i^3\right) a_2 = \sum x_i y_i$$

$$\left(\sum x_i^2\right) a_0 + \left(\sum x_i^3\right) a_1 + \left(\sum x_i^4\right) a_2 = \sum x_i^2 y_i$$

Matrix Form Representation

Rewriting the above equations in matrix form:

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum (x_i y_i) \\ \sum (x_i^2 y_i) \end{bmatrix}$$

Matrix Form Representation

Rewriting the above equations in matrix form:

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum(x_i y_i) \\ \sum(x_i^2 y_i) \end{bmatrix}$$

This equation is in the form:

$$Xa = B$$

Solving for a:

$$a = X^{-1}B$$

Thus, polynomial regression enables accurate modeling of non-linear relationships by fitting a polynomial curve that minimizes errors while avoiding excessive complexity.

Example: Consider the data provided in the following Table

Table 5.8: Sample Data

x	y
1	1
2	4
3	9
4	15

Table 5.8: Sample Data

x	y
1	1
2	4
3	9
4	15

Solution: For applying polynomial regression, computation is done as shown in Table 5.9. Here, the order is 2 and the sample i ranges from 1 to 4.

Table 5.9: Computation Table

x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y$	x_i^3	x_i^4
1	1	1	1	1	1	1
2	4	8	4	16	8	16
3	9	27	9	81	27	81
4	15	60	16	240	64	256
$\Sigma x_i = 10$	$\Sigma y_i = 29$	$\Sigma x_i y_i = 96$	$\Sigma x_i^2 = 30$	$\Sigma x_i^2 y_i = 338$	$\Sigma x_i^3 = 100$	$\Sigma x_i^4 = 354$

It can be noted that, $N = 4$, $\Sigma y_i = 29$, $\Sigma x_i y_i = 96$, $\Sigma x_i^2 y_i = 338$. When the order is 2, the matrix using Eq. (5.28) is given as follows:

$$\begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 29 \\ 96 \\ 338 \end{bmatrix}$$

Therefore, using Eq. (5.29), one can get coefficients as:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix}^{-1} \times \begin{bmatrix} 29 \\ 96 \\ 338 \end{bmatrix} = \begin{pmatrix} -0.75 \\ 0.95 \\ 0.75 \end{pmatrix}.$$

This leads to the regression equation using Eq. (5.26) as:

$$y = -0.75 + 0.95\,x + 0.75\,x^2$$

Logistic Regression and Its Functionality

Logistic regression is a classification technique that predicts the probability of a categorical variable. It takes one or more input features (x) and estimates the response variable (y). If a probability were predicted using linear regression, it would be represented as:

$$p(x) = a_0 + a_1 x$$

However, logistic regression aims to model probability values that must range between 0 and 1. For instance, in an email classification problem (spam vs. normal mail), if the probability of an email being normal is 0.7, then there is a 70% chance that it is not spam.

The Logit Function in Logistic Regression

Linear regression can output values from $-\infty$ to $+\infty$, whereas probability values must be constrained between 0 and 1. To achieve this, a mapping function is required. The core mapping function in logistic regression is the **sigmoid function**, also known as the **logit function**, which is expressed as:

$$\text{logit}(x) = \frac{1}{1 + e^{-x}}$$

where x is the independent variable, and e is Euler's number. The purpose of the logit function is to ensure that any real number is mapped within the probability range of 0 to 1.

Probability and Odds

Logistic regression extends linear regression by transforming the output so that it lies in the 0-1 range. This transformation involves odds and log-odds.

Odds represent the likelihood of an event occurring compared to it not occurring:

$$\text{odds} = \frac{\text{probability of an event}}{\text{probability of a non-event}} = \frac{p}{1-p}$$

Taking the logarithm of the odds results in the **logit function**:

$$\log\left(\frac{p(x)}{1-p(x)}\right) = a_0 + a_1 x$$

Solving for $p(x)$, we obtain:

$$p(x) = \frac{\exp(a_0 + a_1 x)}{1 + \exp(a_0 + a_1 x)}$$

This function always produces values within the 0-1 range. By manipulating the numerator and denominator, it can be rewritten as:

$$p(x) = \frac{1}{1 + \exp(-(a_0 + a_1 x))}$$

Threshold Function

To make a classification decision, logistic regression applies a threshold function:

$$y = \begin{cases} 1, & \text{if } p(x) \geq 0.5 \\ 0, & \text{otherwise} \end{cases}$$

Example 5.7: Application of Logistic Regression

- Consider a binary logistic regression problem where the goal is to classify students as **"pass" or "fail"** based on their entrance exam scores. Historical data is used to determine the regression coefficients.

Given:

- a0=1
- a1=8
- x=60 (student's marks)

We calculate z as:

$$z = a_0 + a_1 x = 1 + 8(60) = 481$$

Applying the logistic function:

$$p(x) = \frac{1}{1 + \exp(-481)}$$

Since $\exp(-481)$ is an extremely small number, the probability approximates to:

$$p(x) \approx 0.44$$

Given the threshold of 0.5, and since $0.44 < 0.5$, the student with marks 60 is not selected.

Understanding Logistic Regression and Parameter Estimation

- In logistic regression, the relationship between dependent and independent variables is determined by estimating model parameters. These parameters are obtained using the **maximum likelihood estimation (MLE)** method, which uses training data to find the best values that minimize errors in predicted probabilities.
- Since multiple sets of coefficients can exist, the optimal set is chosen using the **MLE function**, which identifies the coefficients that maximize the probability of obtaining the observed data.

Likelihood Function in Logistic Regression

If π represents the probability of a successful outcome and $1 - \pi$ represents the probability of failure, then the likelihood function is given by:

$$L(a : y) = \prod_{i=1}^{n} \left(\frac{\pi_i}{1 - \pi_i} \right)^{y_i} (1 - \pi_i)$$

To estimate the parameters, the **log-likelihood function** is taken, and optimization techniques like the **Newton-Raphson method** can be used to maximize it.

Multinomial Logistic Regression

Logistic regression is mainly used for **binary classification**. However, it can be extended for multiple classes through **multinomial logistic regression**.

For example, if there are three classes: **Class 1, Class 2, and Class 3**, multinomial logistic regression creates three binary classification problems:

1. Class 1 vs. Not Class 1
2. Class 2 vs. Not Class 2
3. Class 3 vs. Not Class 3

These three classification models run simultaneously to determine the most probable class for a given input.

Advantages and Disadvantages of Logistic Regression

✅ Advantages:

- Logistic regression is a simple and efficient method for binary classification.
- The model is easy to interpret.

❌ Disadvantages:

- Multinomial logistic regression struggles with a **large number of attributes**.
- It can only handle **linear relationships** between variables.
- If attributes exhibit **multicollinearity** (high correlation among independent variables), the logistic regression model may perform poorly.