



**COMSATS University Islamabad,  
Sahiwal Campus  
Department of Computer Science**

---

**Artificial Intelligence – Assignment 04**

**Submitted by:**

**Muhammad Ibtisam Afzal**

**FA22-BCS-073**

**Submitted to:**

**Mr. Muhammad Jamil**

**Submission Date:**

**23-12-2024**

## Linear Regression

Linear regression is a statistical and machine learning technique used to model the relationship between a dependent variable (target) and one or more independent variables (predictors). It is a supervised learning algorithm widely used for predictive analysis and understanding relationships within data.

### Types of Linear Regression:

1. Simple Linear Regression: Involves one independent variable and one dependent variable. The relationship is represented as a straight line.
2. Multiple Linear Regression: Involves multiple independent variables influencing a single dependent variable.

### The Equation of Linear Regression

The relationship between variables is expressed as:

Where:

- : Dependent variable (target).
- : Independent variables (predictors).
- : Intercept of the line.
- : Coefficients of the predictors.
- : Error term accounting for deviations.

### Assumptions of Linear Regression

1. Linearity: The relationship between the dependent and independent variables is linear.
2. Independence: Observations are independent of each other.
3. Homoscedasticity: Constant variance of residuals (errors) across all levels of independent variables.
4. Normality: Residuals are normally distributed.

### Steps to Perform Linear Regression

1. Data Collection: Gather data containing dependent and independent variables.
2. Exploratory Data Analysis (EDA): Visualize and analyze the data to identify patterns and relationships.
3. Model Fitting: Use tools like Python's sklearn library to fit the linear regression model.
4. Evaluation: Assess model performance using metrics like Mean Squared Error (MSE), R-squared, etc.

## **Applications of Linear Regression**

1. Predictive Modeling: Predict future values (e.g., sales forecasting).
2. Trend Analysis: Identify trends in data.
3. Risk Assessment: Estimate risks in finance and insurance.
4. Medical Research: Analyze relationships between variables (e.g., drug dose and patient outcome).

## **Advantages and Limitations**

### **Advantages:**

1. Simple and easy to implement.
2. Interpretable results with clear relationships.
3. Efficient for small to medium-sized datasets.

### **Limitations:**

1. Assume linear relationships, which may not always hold.
2. Sensitive to outliers.
3. May underperform in complex scenarios with nonlinear relationships.

**By understanding linear regression, one can model relationships between variables, make predictions, and gain insights into data-driven phenomena.**