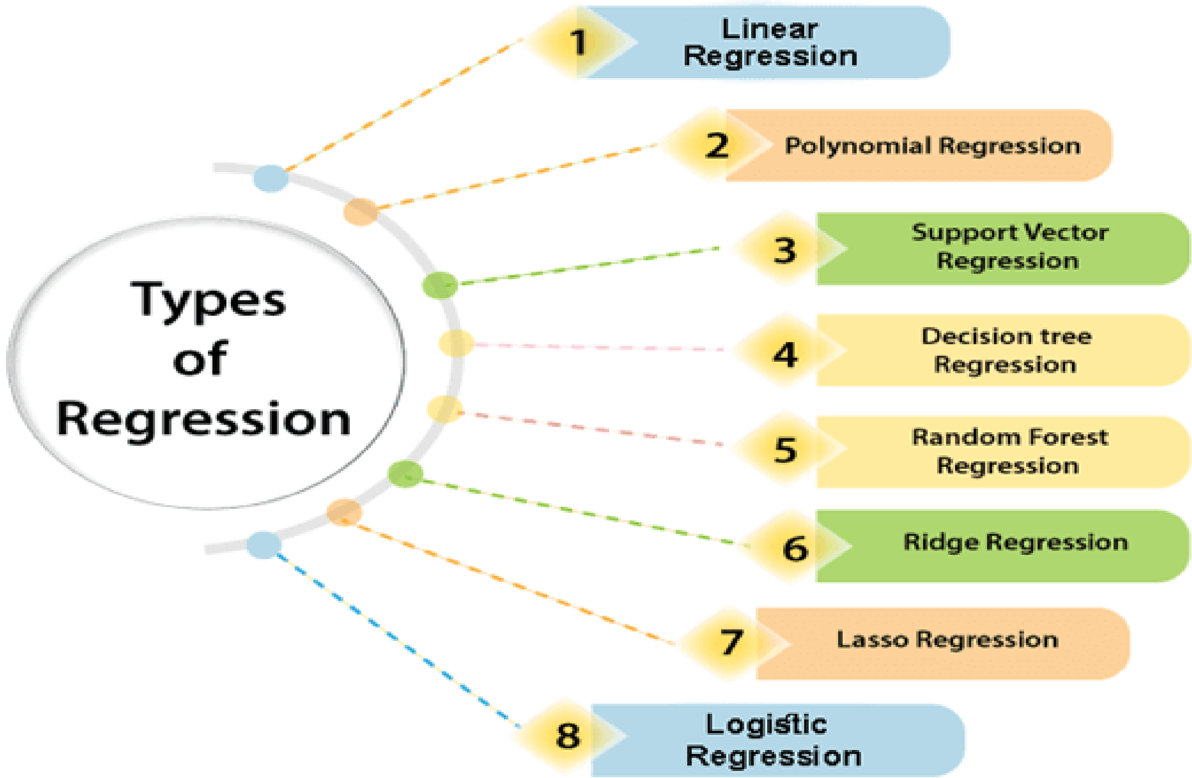
# Supervised Learning-Regression

## Types of regression




---

**1. Linear Regression**

The most extensively used modelling technique is linear regression, which assumes a linear connection between a dependent variable (Y) and an independent variable (X).

It employs a regression line, also known as a best-fit line.

The linear connection is defined as **Y = c+m*X + e**

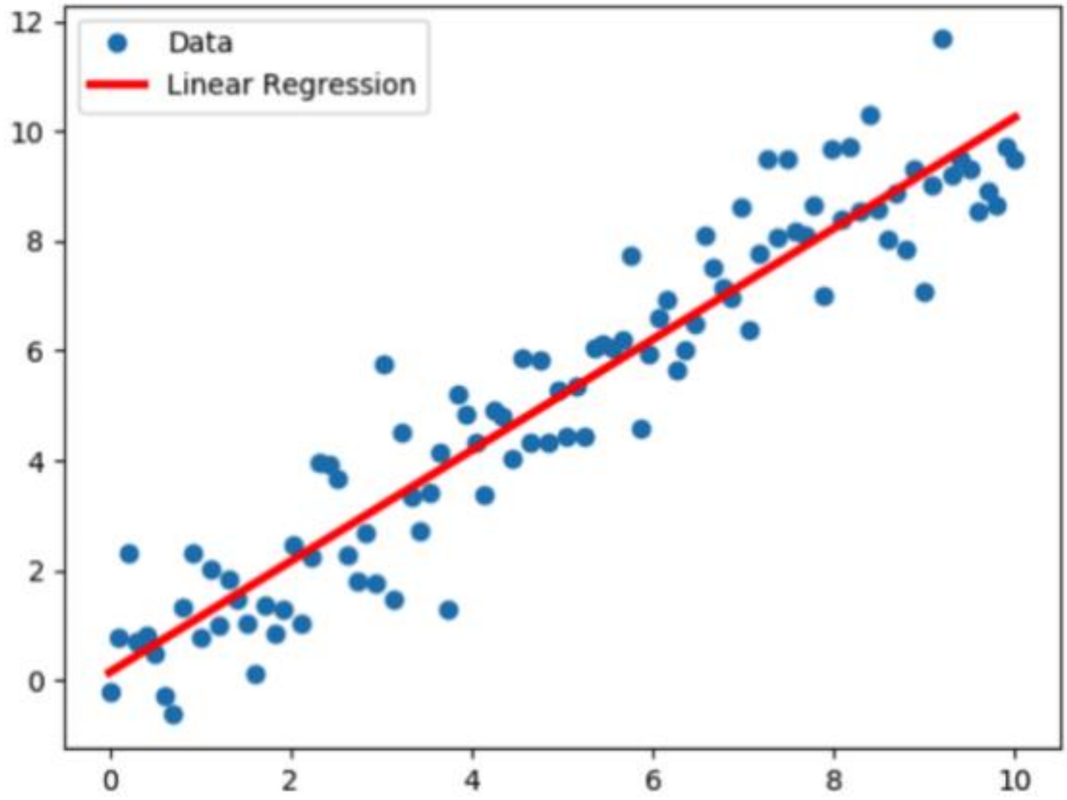
Where,

'c' denotes the intercept,

'm' denotes the slope of the line, and

'e' is the error term.

The linear regression model can be simple (with only one dependent and one independent variable) or complex (with numerous dependent and independent variables) (with one dependent variable and more than one independent variable).




**Y = c+m*X + e**

The best fit line is determined by varying the values of m and c.

The predictor error is the difference between the observed values and the predicted value.

The values of m and c get selected in such a way that it gives the minimum predictor error.

There are different types of linear regression. The two major types of linear regression are

- simple linear regression and
- multiple linear regression.

For a single variable linear regression: $Y = \beta_0 + \beta_1 X + \epsilon$

For multiple variables (multiple linear regression): $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$

Where:

- $Y$ is the dependent variable (what you are predicting),
- $X_1, X_2, \ldots, X_n$ are the independent variables (predictors),
- $\beta_0$ is the intercept (value of $Y$ when all predictors are 0),
- $\beta_1, \ldots, \beta_n$ are the coefficients for the predictors,
- $\epsilon$ is the error term.

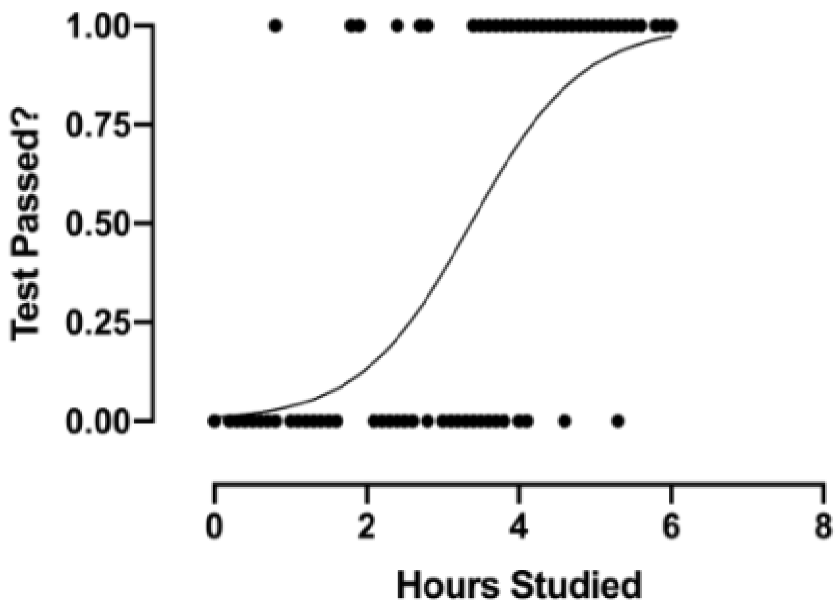
## 2. Logistic Regression

Logistic regression is used when the dependent variable is binary (e.g., 0/1, True/False). Instead of predicting the actual value of the dependent variable, logistic regression predicts the probability that the outcome belongs to a particular class (e.g., probability of "Yes" or "No"). It uses a sigmoid function to map the predicted values into probabilities between 0 and 1.

The probability $p$ that the dependent variable $Y$ equals 1 is modeled as: $p(Y = 1|X) = \frac{1}{1+e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}}$ Where:

- $p$ is the predicted probability that $Y$ equals 1.
- The term inside the exponential is the linear combination of the predictors.

Alternatively, the equation can be written as: $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$ This is called the **log-odds** or **logit function**.




Predicting whether a student will **pass (1)** or **fail (0)** an exam based on their study hours.

- Feature $X$: Number of study hours.
- Target $Y$: Pass (1) or fail (0).

**Use Case:**

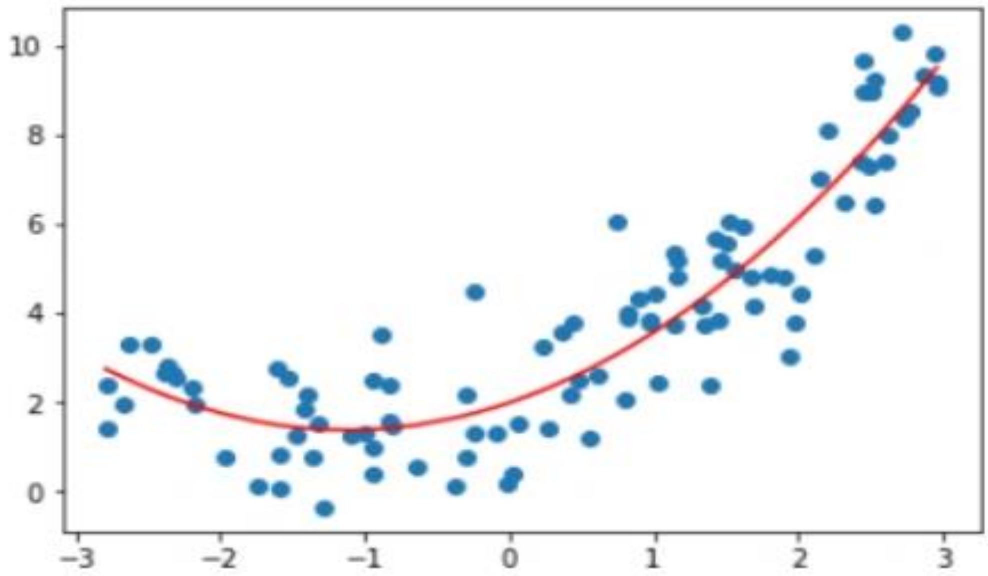
- Classifying whether an email is spam (1) or not spam (0).
- Predicting the likelihood of a customer buying a product (purchase or no purchase).

### 3. Polynomial Regression

Polynomial regression is an extension of linear regression where the relationship between the dependent variable and the independent variables is modeled as a polynomial. This allows for **non-linear relationships** between the variables but still uses the framework of linear regression to estimate the model coefficients.

For a degree $d$ polynomial: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_d X^d + \epsilon$ Where:

- $X^2, X^3, \ldots, X^d$ are higher-degree terms.
- $d$ is the degree of the polynomial.




Suppose you are predicting the **profit** of a company based on its **advertising budget**. The relationship might not be linear, but quadratic or cubic, meaning the return on investment decreases after a certain point.

- Feature $X$: Advertising budget.
- Target $Y$: Company profit.

The model could be quadratic, like: $\text{Profit} = \beta_0 + \beta_1 \times \text{Budget} + \beta_2 \times \text{Budget}^2 + \epsilon$

**Use Case:**

- Predicting the trajectory of a projectile where the height is a function of time squared.
- Modeling economic variables where relationships between variables are nonlinear.

| Aspect | Linear Regression | Logistic Regression | Polynomial Regression |
|---|---|---|---|
| **Type of Problem** | Regression (predicts a continuous value) | Classification (predicts a binary or categorical outcome) | Regression (predicts a continuous value) |
| **Dependent Variable (Target)** | Continuous (e.g., house price, salary) | Categorical/Binary (e.g., 0/1, Yes/No) | Continuous (e.g., sales, profit) |
| **Type of Relationship** | Linear (relationship between dependent and independent variables) | Logistic (log-odds, nonlinear relationship via the sigmoid function) | Nonlinear (uses polynomial terms of the independent variables) |
| **Output** | A continuous value (e.g., prediction of a price) | Probability (between 0 and 1) leading to classification (0 or 1) | A continuous value (but models nonlinear relationships) |
| **Linearity** | Assumes linear relationship between features and target | Assumes logit of the probability is linearly related to features | Models nonlinear relationships by using polynomial terms |
| **Example Use Case** | Predicting house prices based on size | Predicting if a customer will buy a product (Yes/No) | Predicting company profit based on advertising budget (nonlinear) |

## Other types of Regressions

- **Linear Regression**: Models the linear relationship between a dependent variable and one or more independent variables.
- **Logistic Regression**: Used for binary classification tasks, predicting the probability of a categorical outcome (e.g., yes/no).
- **Polynomial Regression**: Extends linear regression by fitting a polynomial curve to model nonlinear relationships.
- **Ridge Regression**: A type of linear regression with L2 regularization to prevent overfitting by shrinking coefficients.
- **Lasso Regression**: A linear regression model with L1 regularization that performs feature selection by shrinking some coefficients to zero.
- **Elastic Net Regression**: Combines both L1 and L2 regularization to balance between feature selection and coefficient shrinkage.

- **Ordinary Least Squares (OLS) Regression**: A type of linear regression that minimizes the sum of squared residuals to find the best-fitting line.

**What is Regularization?**

- **Regularization** is a technique used in machine learning and statistics to prevent models from becoming overly complex and overfitting the training data. Overfitting happens when a model learns not just the underlying patterns in the data but also the noise, making it perform poorly on unseen data.
- Regularization addresses this by adding a **penalty term** to the model's objective (loss) function. This penalty discourages the model from fitting the data too closely by shrinking the model's coefficients (weights). The goal is to create a model that generalizes better on new, unseen data.

There are two common types of regularization:

- **L1 regularization (Lasso)** and
- **L2 regularization (Ridge)**.

**4. L1 Regularization (Lasso Regression)**

L1 regularization adds a penalty equal to the **absolute value** of the coefficients (weights) to the loss function. This results in some coefficients being reduced to **exactly zero**, meaning that certain features are entirely removed from the model. This property of L1 regularization makes it useful for **feature selection**, as it automatically chooses the most important features and eliminates irrelevant ones.

**<u>Use Case:</u>**

- **Feature Selection**: In a dataset with many features, some of which may not be very important, L1 regularization can help simplify the model by reducing the less important feature coefficients to zero. This makes the model more interpretable and less prone to overfitting.

**Example:**

- Imagine you're predicting house prices using features like **size**, **number of rooms**, **location**, **year built**, etc. Some features, such as "year built," might have little effect compared to others like size. Using L1 regularization, the model might assign zero weight to "year built," effectively removing it, and keeping only the most relevant features.

**5. L2 Regularization (Ridge Regression)**

L2 regularization adds a penalty equal to the **square of the magnitude** of the coefficients (weights). Unlike L1, L2 regularization **shrinks the coefficients but does not eliminate them** (i.e., no coefficient becomes exactly zero). It makes all coefficients smaller, which reduces the model's complexity but keeps all features in the model.

- It is particularly effective when you have **multicollinearity** (when features are highly correlated) because it stabilizes the model by shrinking the weights uniformly.

**Use Case:**

- **Dealing with Multicollinearity**: When multiple features are correlated, the coefficients of a linear model can become unstable, leading to overfitting. L2 regularization helps by reducing the size of the coefficients, making the model more robust.

**Example:**

Suppose you are predicting **employee salaries** based on **experience**, **education**, **job role**, and **company size**. Some features, like "experience" and "job role," may be highly correlated. L2 regularization will shrink all the feature coefficients slightly, keeping all of them in the model, but preventing any one feature from having too much influence.

| Aspect | L1 Regularization (Lasso) | L2 Regularization (Ridge) |
|---|---|---|
| Penalty | Absolute value of coefficients | Square of the coefficients |
| Effect on Coefficients | Some coefficients become exactly zero (feature elimination) | Shrinks coefficients but does not set them to zero |
| Feature Selection | Yes, can eliminate irrelevant features | No, retains all features |
| Use Case | When feature selection is needed | When features are correlated (multicollinearity) |
| Simplicity | Leads to sparse models (fewer features) | Retains all features but reduces overfitting |
| Model Complexity | Simplifies the model by removing features | Reduces overfitting without feature elimination |