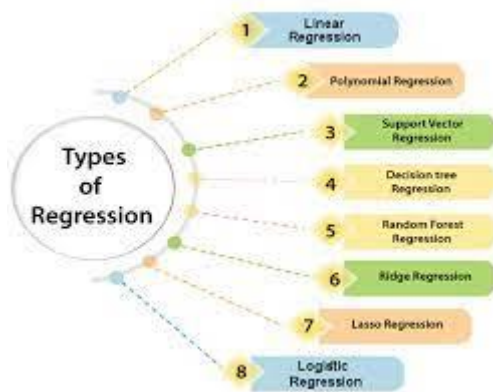


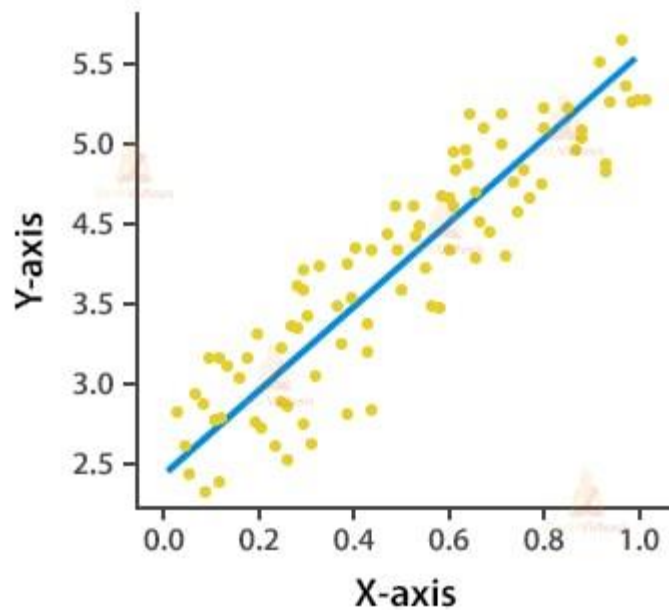
- **Regression:**
- The evaluation of relationship between two or more variables is called Regression Analysis. It is a statistical technique.
- Regression Analysis helps enterprises to understand what their data points represent, and use them wisely in coordination with different business analytical techniques in order to make better decisions.
- Regression Analysis helps an individual to understand how the typical value of the dependent variable changes when one of the independent variables is varied, while the other independent variables remain unchanged. Therefore, this powerful statistical tool is used by Business Analysts and other data professionals for removing unwanted variables and choosing only the important ones.
- The benefit of regression analysis is that it allows data crunching to help businesses make better decisions. A greater understanding of the variables can impact the success of a business in the coming weeks, months, and years in the future.
- Regression is a fundamental statistical technique used in data analytics to model and analyze the relationship between a dependent variable and one or more independent variables. It is commonly employed for tasks such as prediction, hypothesis testing, and understanding the impact of variables on an outcome. In regression analysis, you try to find the best-fitting linear or nonlinear equation that describes the relationship between the variables.

**Types of Regression:** There are different types of regression models, including:

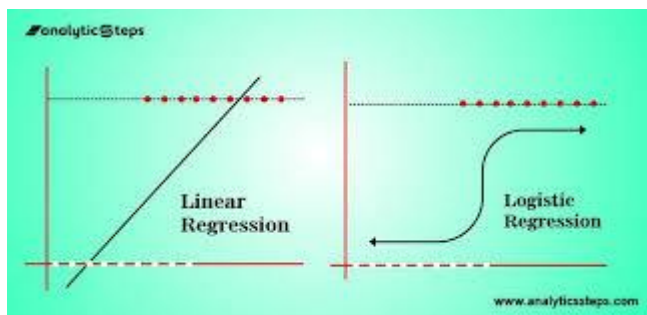


**Linear Regression:** Assumes a linear relationship between the dependent and independent variables. Simple linear regression involves one independent variable, while multiple linear regression involves two or more independent variables.

# Linear Regression



**Logistic Regression:** Used when the dependent variable is binary (two classes) and you want to predict probabilities or classify data into categories.



**Polynomial Regression:** Models relationships that are not strictly linear but can be approximated using polynomial equations.

**Ridge and Lasso Regression:** Used for regularization in linear regression to prevent overfitting by adding penalty terms to the coefficients.

**Time Series Regression:** Applied when the data has a time component, and you want to model how variables change over time.

**Assumptions:** Linear regression, in particular, relies on several assumptions, such as linearity, independence of errors, constant variance of errors (homoscedasticity), and normal distribution of errors. Violations of these assumptions can affect the validity of the model.

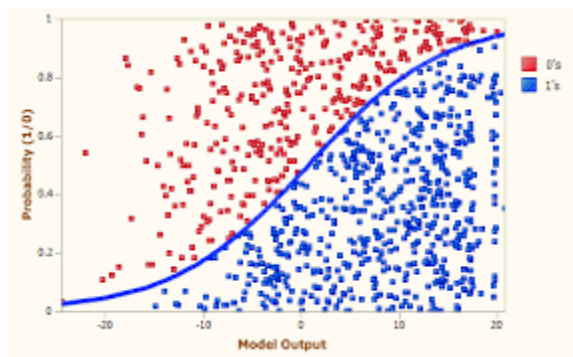
**Model Evaluation:** To assess the quality of a regression model, you typically use metrics such as R-squared (coefficient of determination), mean squared error (MSE), root mean squared error (RMSE), and others, depending on the specific problem and model type.

**Interpretation:** Regression models provide coefficients for each independent variable, indicating the strength and direction of their relationship with the dependent variable. Coefficients help you interpret how changes in the independent variables affect the outcome.

**Overfitting and Underfitting:** It's crucial to strike a balance between a model that fits the training data well but doesn't generalize to new data (overfitting) and a model that is too simple to capture the underlying patterns (underfitting). Techniques like cross-validation and regularization can help address these issues.

**Applications:** Regression analysis is used in various fields, including economics (e.g., predicting stock prices), social sciences (e.g., studying the impact of education on income), healthcare (e.g., predicting patient outcomes), and many others.

**Software:** There are numerous software tools and libraries available for performing regression analysis, such as Python's scikit-learn, R, and Microsoft Excel.



S