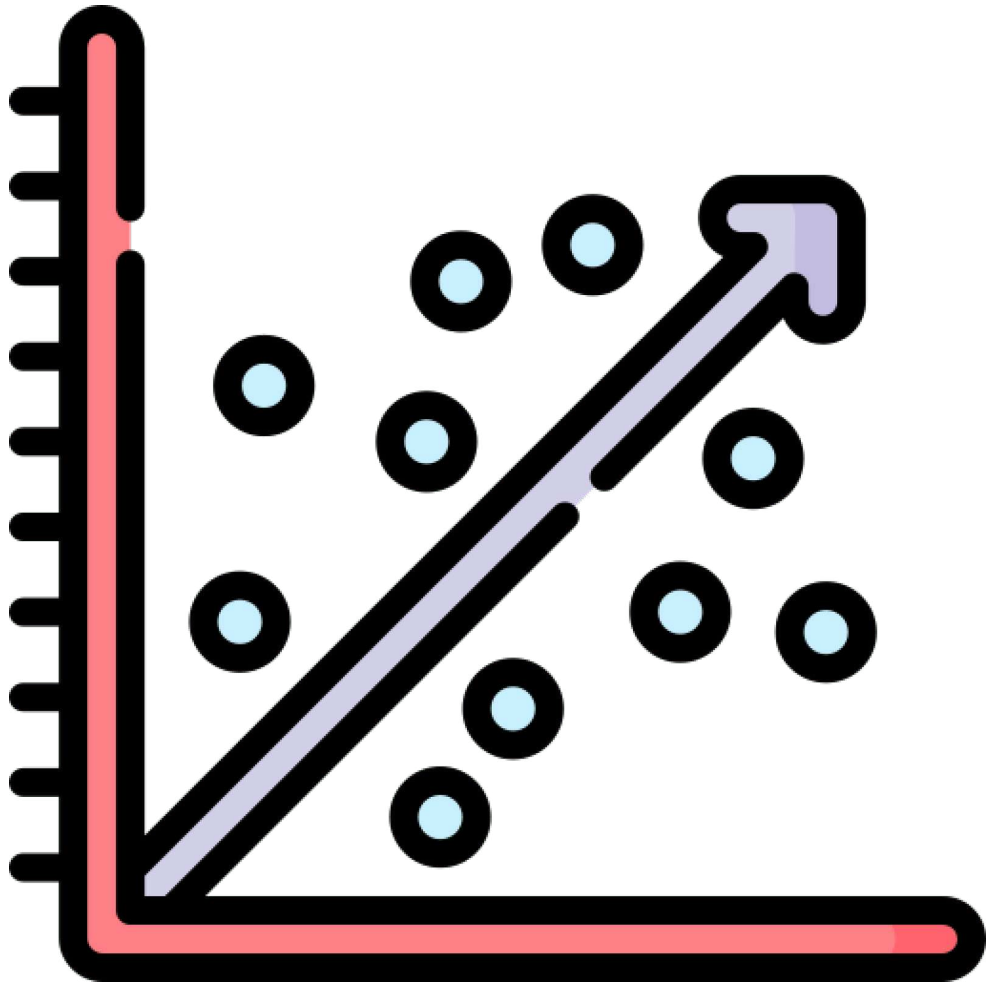
Regression Analysis



DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE

What is Regression?

Regression is a method for understanding the relationship between independent variables or features and a dependent variable or outcome.

Outcomes can then be predicted once the relationship between independent and dependent variables has been estimated

For example, You want to build a regression model to predict the hourly wages of a worker using variables like education level, gender, experience, skillset, etc.

Types of Regression Analysis

DATARANCH.org

VISUALIZE | ANALYZE | CAPITALIZE

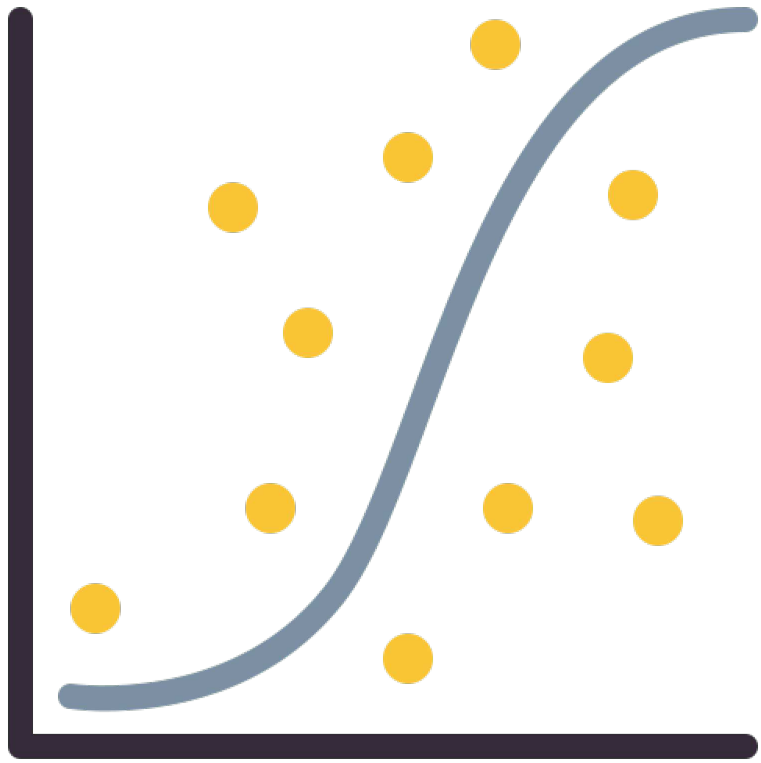
Simple Linear Regression

It is a type of regression that is used to model the relationship between a dependent variable and a **single** independent variable. E.g. The usage could be to predict an employee's salary based on his/her qualification.

Multiple Linear Regression

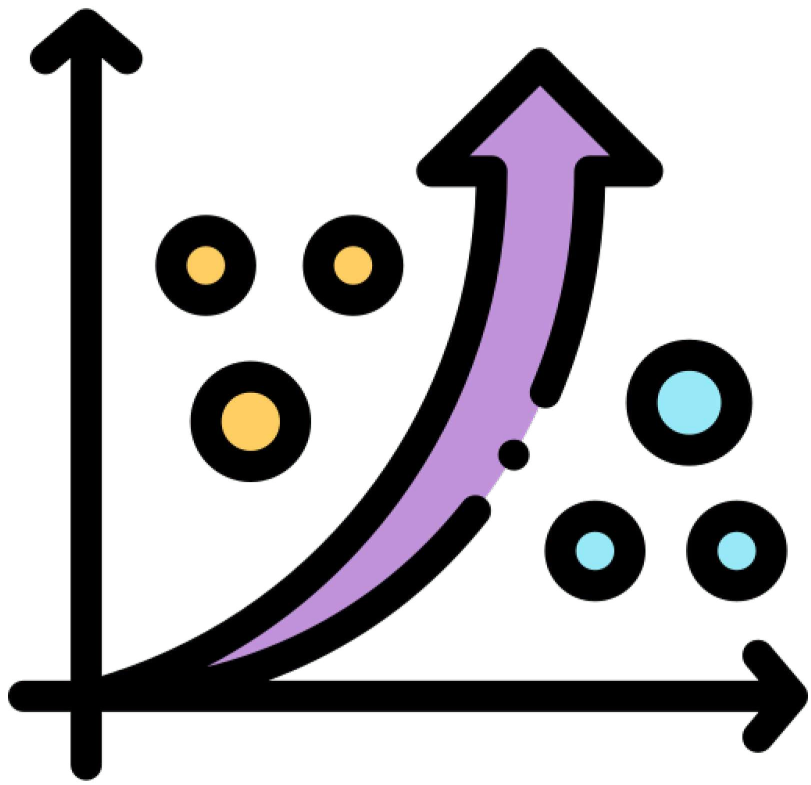
It is a type of regression that is used to model the relationship between a dependent variable and **two or more** independent variables. E.g. The usage could be to predict an employee's salary based on his/her qualification, experience, city.

Polynomial Regression



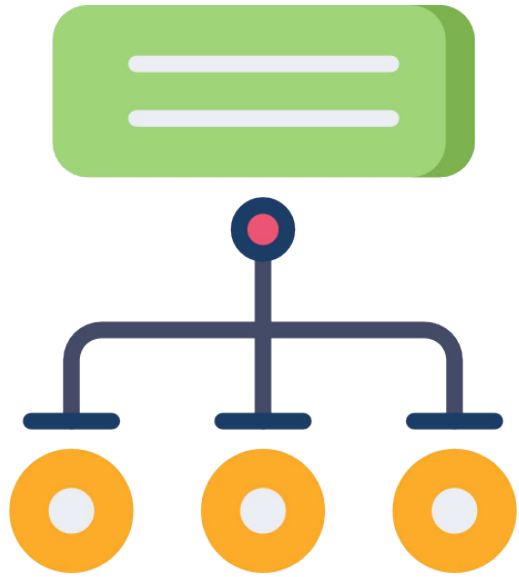
It is a type of regression that models the relationship between a dependent variable and an independent variable by fitting a polynomial equation to the data. E.g. It is widely applied to predict the spread rate of COVID-19 and other infectious diseases.

Logistic Regression



It is a type of regression used to model the relationship between a dependent variable and one or more independent variables when the dependent variable is binary. E.g. Deciding whether to admit a student in class or not based on his/her grade, skill set, and experience.

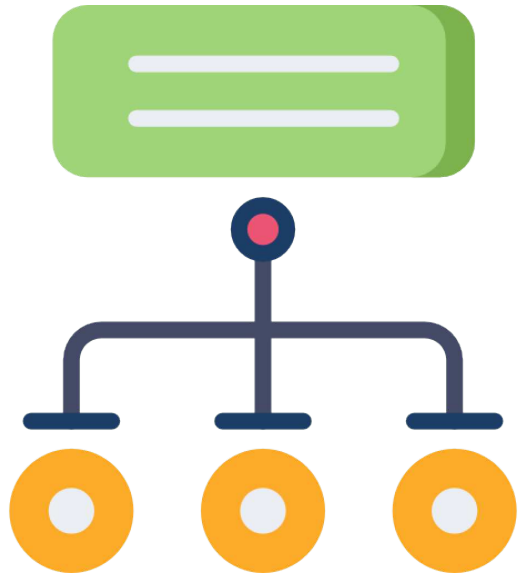
Other Types of Regression



Ridge Regression: It is a type of regression that is used to prevent overfitting in multiple linear regression by adding a penalty term to the cost function.

Lasso Regression: It is a type of regression that is used to perform feature selection in multiple linear regression by adding a penalty term to the cost function that encourages the coefficients of some of the independent variables to be exactly zero.

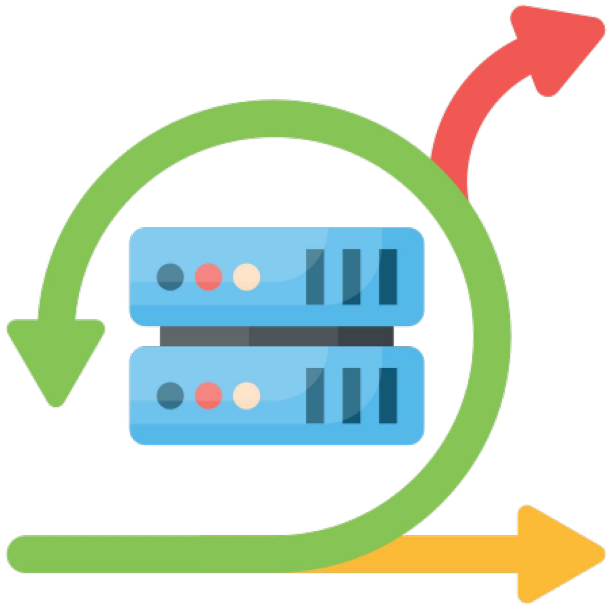
Other Types of Regression



ElasticNet Regression: It is a type of regression analysis that combines the L1 and L2 regularization methods of Lasso and Ridge regression, respectively. It is used for linear regression problems where the number of independent variables is greater than the number of observations or when the independent variables are highly correlated with each other.

Bayesian Regression: In Bayesian Regression, the prior distribution of the parameters is specified before the data is observed. This prior distribution represents the researcher's belief about the parameters before any data is collected. Once the data is observed, the prior distribution is updated to become the posterior distribution, which represents the researcher's belief about the parameters after observing the data.

Usage of Regression

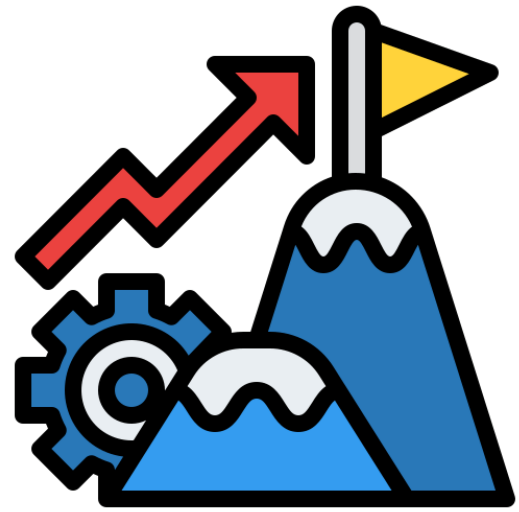


1. Forecasting continuous outcomes like house prices, stock prices, or sales.

2. Predict the success of future retail sales or marketing campaigns to ensure resources are used
effectively.

3. Predict customer or user trends, such as on e-commerce websites.

Challenges

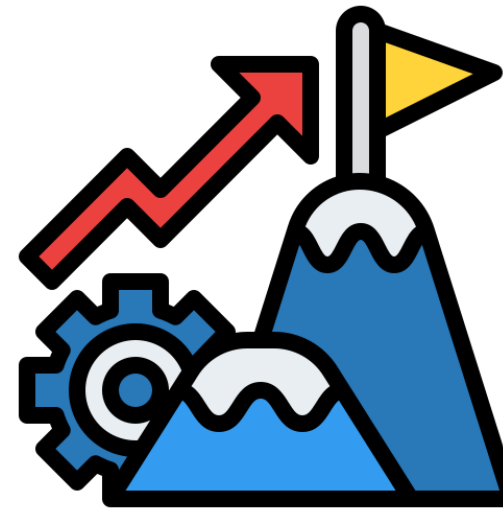


Overfitting: Overfitting occurs when the model is too complex and fits the training data too closely, resulting in poor performance on new or unseen data. It can be addressed using regularization techniques like L1 and L2 regularization or early stopping.

Underfitting: Underfitting occurs when the model is too simple and fails to capture the underlying patterns in the data. It can be addressed by increasing the complexity of the model or by adding more relevant features.

Multicollinearity: Multicollinearity occurs when two or more independent variables are highly correlated with each other. This can make it difficult to estimate the individual effects of each variable and can lead to unstable parameter estimates.

Challenges



Outliers: Outliers are data points that are significantly different from the rest of the data. They can have a large influence on the regression model and can lead to inaccurate estimates of the parameters

.

Non-linearity: Regression models assume a linear relationship between the independent and dependent variables. However, in some cases, the relationship may be non-linear, and this can lead to inaccurate predictions.

Missing data: Missing data can make it difficult to estimate the parameters of the regression model and can lead to biased estimates if not handled properly.

Follow **#DataRanch** on LinkedIn for more...

DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE

Follow **#DataRanch** on LinkedIn for more...

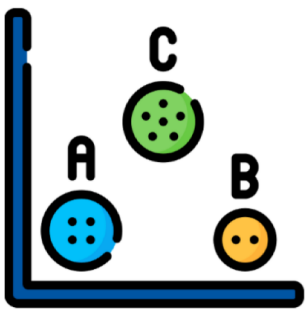
What is Supervised Learning?

DATARANCH.org

What is Unsupervised Learning?

DATARANCH.org

Clustering

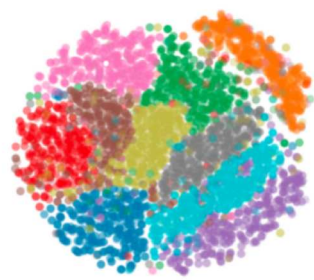


DATARANCH.org

Principal Component Analysis

DATARANCH.org

t-Distributed Stochastic Neighbour Embedding (t-SNE)



DATARANCH.org

DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE