

AiEnsured

Ensuring comprehensive validation of AI systems



An article on

Regression Analysis

Unveiling Relationships and Making Predictions

Submitted by: Rayapureddi Venkata Sri Sai Subhash

Intern in Machine Learning May 2023-July 2023

(rayapureddi.subhash@testaing.com)

INTRODUCTION:

Regression analysis is a statistical tool used to examine the connections between variables. Its main purpose is to investigate the cause-and-effect relationship between two variables, such as the impact of a price increase on demand or changes in the money supply on the inflation rate. To study these issues, researchers gather data on the variables of interest and employ regression analysis to estimate the quantitative effect of the causal variables on the influenced variable. Additionally, researchers typically evaluate the "statistical significance" of the estimated relationships, which reflects the level of confidence in the accuracy of the estimated relationship compared to the true relationship.



Source: aihr.com

Understanding Regression Analysis:

Regression analysis involves modelling the relationship between a dependent variable and one or more independent variables. The dependent variable is the one we seek to predict or explain, while the independent variables are the factors that may influence the dependent variable. The relationship is typically represented by a mathematical equation, and the goal is to estimate the equation's parameters to understand the relationship's nature and strength.



Source: medium.com

Terminologies Related to the Regression Analysis:

- **Dependent Variable:** The main factor in Regression analysis which we want to predict or understand is called the dependent variable. It is also called **target variable**.
- **Independent Variable:** The factors which affect the dependent variables or which are used to predict the values of the

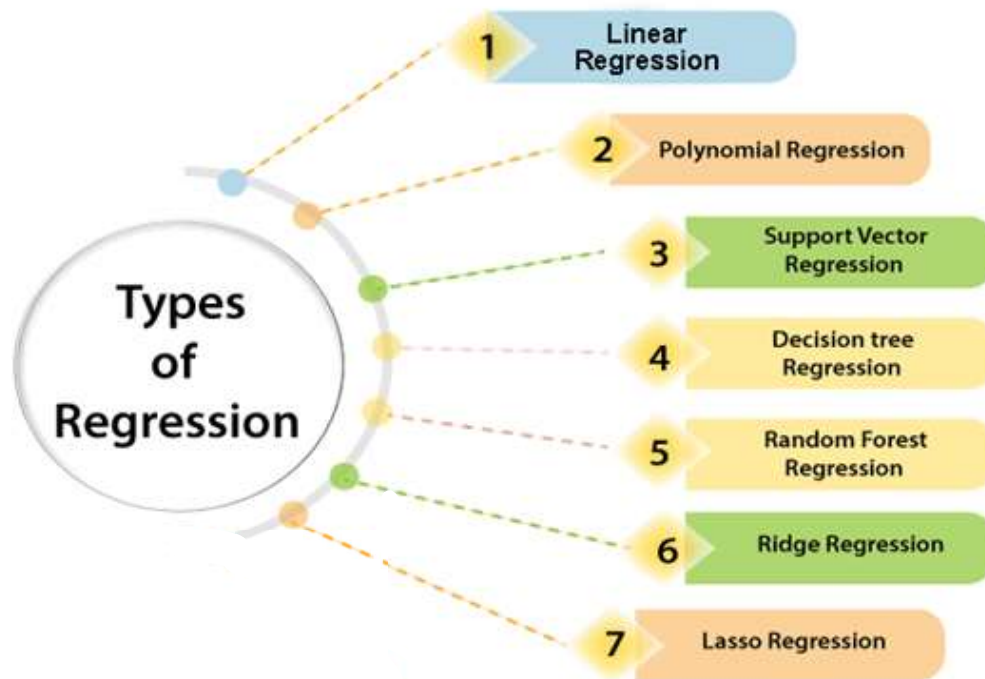
dependent variables are called independent variable, also called as a **predictor**.

- **Outliers:** Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.
- **Multicollinearity:** If the independent variables are highly correlated with each other than other variables, then such condition is called Multicollinearity. It should not be present in the dataset, because it creates problem while ranking the most affecting variable.
- **Underfitting and Overfitting:** If our algorithm works well with the training dataset but not well with test dataset, then such problem is called **Overfitting**. And if our algorithm does not perform well even with training dataset, then such problem is called **underfitting**.

Types of Regression:

There are various types of regressions which are used in data science and machine learning. Each type has its own importance on different scenarios, but at the core, all the regression methods analyse the effect of the independent variable on dependent variables.

- Linear Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression
- Ridge Regression
- Lasso Regression



Source: javatpoint

Interpreting Regression Results:

Regression analysis provides valuable insights through the estimated coefficients and statistical measures. The coefficients indicate the strength and direction of the relationship between the variables. Positive coefficients suggest a positive relationship, while negative coefficients imply a negative relationship. The statistical measures, such as R-squared and p-values, help assess the model's goodness-of-fit and the significance of the independent variables, respectively.

Applications of Regression Analysis:

Regression analysis has various applications across different fields. Here are some common applications:

- **Financial forecasting:** It helps analyse the impact of factors like interest rates, inflation, and GDP on stock prices, bond yields, and consumer spending.
- **Sales and promotions forecasting:** Regression analysis assists in understanding how advertising expenditures, pricing strategies, and customer demographics influence sales and market share.
- **Testing automobiles:** Regression analysis helps in testing automobiles by analysing the relationships between variables such as performance, safety, efficiency, and customer satisfaction to make informed decisions for optimization and improvement.
- **Weather analysis and prediction:** Regression analysis helps in weather analysis and prediction by identifying relationships between weather variables (such as temperature, humidity, pressure) and historical data, enabling the development of models that can forecast future weather conditions with reasonable accuracy.
- **Time series forecasting:** Regression analysis can aid in time series forecasting by identifying trends, patterns, and relationships between variables over time. It helps to model and predict future values based on historical data, allowing for accurate predictions, trend analysis, and understanding the underlying factors influencing the time series data.

Conclusion:

Regression analysis is a versatile statistical technique that helps uncover relationships between variables, make predictions, and understand the impact of independent variables on a dependent variable. By applying regression analysis appropriately, researchers, analysts, and decision-makers can gain valuable insights and make informed decisions in diverse fields.

References:

- <https://www.saedsayad.com/regression.htm>
- <https://www.javatpoint.com/regression-analysis-in-machine-learning>
- <https://www.sciencedirect.com/topics/mathematics/regression-analysis>
- <https://www.scribbr.com/statistics/simple-linear-regression/>