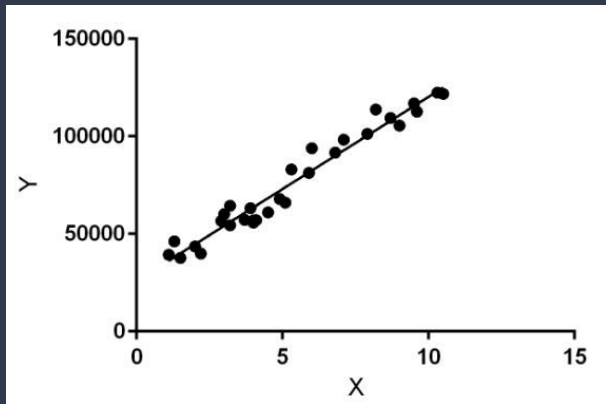


Machine Learning Investigations

HLTWK6 -Tasmin Ganger

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

Linear Regression



Supervised.

A linear regression is a technique used to determine a relationship between x and y. It uses the independent variable (x) as the input and predicts the dependent variable (y) by giving an output.

It is useful for statistically predicting the value of a variable (y) to the value of another (x). It is used to find relationship between two continuous variables.

Examples

- In Business, to make estimates or forecasts of business trends.
- Agriculture, by forecasting crop yields by determining the effect of water and fertiliser on crops.
- In medicine, to look at relationships in patients health such as drug dosage and blood pressure.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

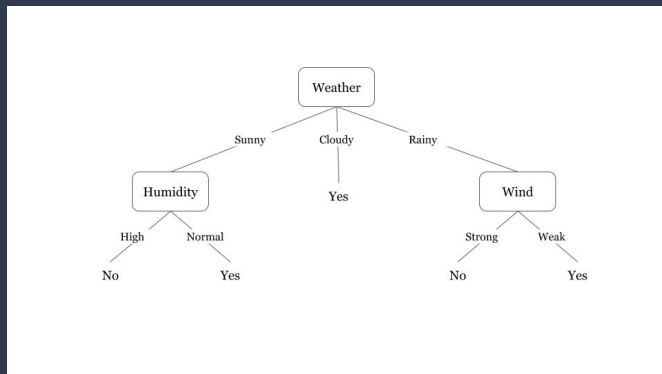
Diagram illustrating the components of the linear regression equation:

- Y_i : Dependent Variable
- β_0 : Population Y intercept
- β_1 : Population Slope Coefficient
- X_i : Independent Variable
- ϵ_i : Random Error term

The equation is structured as follows:

- The term $\beta_0 + \beta_1 X_i$ is labeled as the **Linear component**.
- The term ϵ_i is labeled as the **Random Error component**.

Decision Tree



Non-parametric supervised.

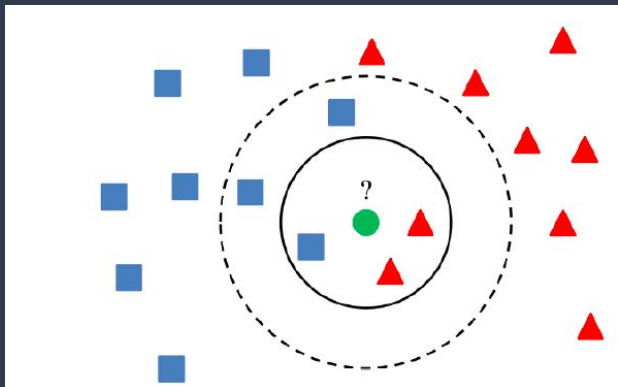
A Decision Tree creates predictions of target variables by making simple decisions from the data to determine the variables (features).

It is useful for evaluating options as it allows you to choose between the possible outcomes (outputs) of each option.

Examples

- Animals, to determine what category they would fall into by their characteristics
- For business projects, to predict whether factors such as time, risk and importance impact if the project taken up or disregarded.

KNN (K- Nearest Neighbor)



Supervised classification.

A KKN algorithm looks at the existence of similar things in close proximity of the value by calculating the distance between points on a graph

It is useful in solving both classification and regression problems as it is classifying a point based on other similar points.

Examples

- Stock market, looks a companies performance measures and economic data to predict the price of stocks.
- Medicine, by researching a diabetic persons glucose level from their infrared absorption spectrum.

$$sim(X, Y) = cos(X, Y) = \frac{\langle X, Y \rangle}{||X|| ||Y||} = \frac{\sum_i^d X_i Y_i}{\sqrt{\sum_i^d X_i^2} \sqrt{\sum_i^d Y_i^2}} \quad (1)$$