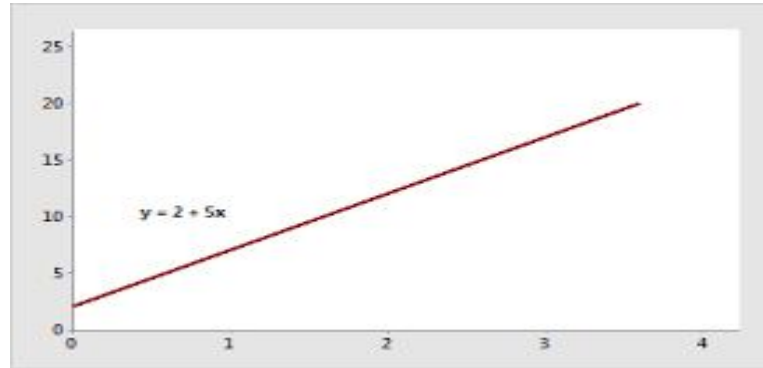# Linear Regression

Agenda

- Introduction to Regression and types
- Simple and Multivariable Linear Regression
- Implementing Linear Regression Model
- Analyzing Result Parameter & assumptions
- Autocorrelation
- MultiSomearity

What is Regression?

$y = 2 + 5x$

- Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable and a series of other variables (known as independent variables).

  Important terms related to regression analysis is

  - Dependent variable (Y) or target variable: Variable to predict.
  - Independent or predictor variable: Variables to estimate the dependent variable.
  - Multicollinearity: Situation in which two or more independent variables are highly linearly related.

Regression Vs Classification
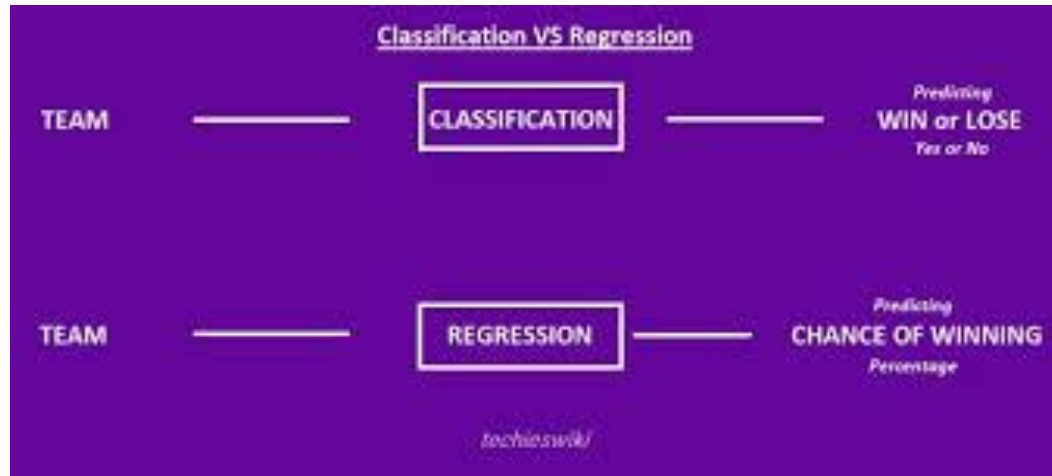
- Similarities- Both are methods of supervised learning and use labelled training data to train their models and make predictions. Therefore, those two tasks are often categorized under the same group in machine learning.
- The main difference between them is the output variable. While in regression, the output is numerical or continuous, in classification, the output is categorical or discrete. This defines the way classification and regression evaluate the predictions

Linear Regression

Regression Types

All the regression methods analyze the effect of the independent variables on dependent variables…

# Defining A Simple Linear Regression Model

Problem definition

Suppose you are heading data science division of a restaurant franchise and CEO is looking for the help in deciding cities for opening a new outlet. The chain already has presence in various cities and you have data for profits and populations from the cities. You would like to use this data to decide which city to expand to next..

- In this case our independent variable is □'x' which is city's population in 10,000s and
- The dependent variable is □ 'y', which is the profitability in Rs 10,000s
- Since the data is continuous and the prediction is based on the past labelled data, regression model is most suited here
- The pre processing of the data is already done and we have complete data on file 'Popln_profit.csv'

Implementing Linear Regression Model
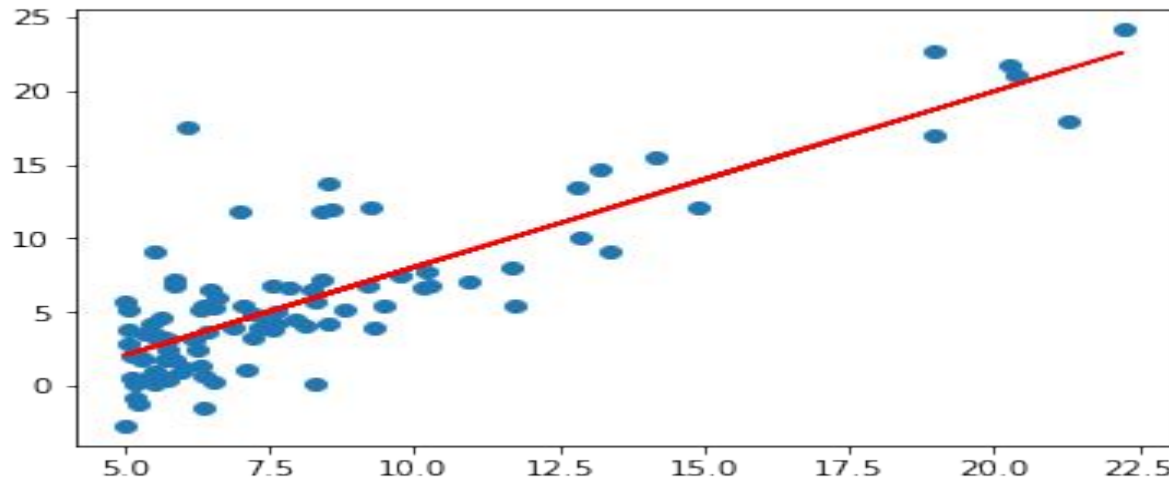
- We will be using basic regression model here, where given data (x, y) is represented by 'blue dots' here
- Our algorithm should be helping us to draw a regression line to predict the cities which can provide the best profitability
- We will use scikitlearn's Linear Regression to train our model on both the training and test sets.

Analysing Result Parameter

1. **Mean Absolute Error** (MAE) is the mean of the absolute value of the errors. It is calculated as:

$$MAE = \frac{1}{n} \sum_{j=1}^{n} |y_j - y_j|$$

Mean Absolute Error

2. **Mean Squared Error** (MSE) is the mean of the squared errors and is calculated as:

$$MSE = \frac{1}{N} \sum_{i}^{n} (Y_i - y_i)^2$$

Mean Squared Error

3. **Root Mean Squared Error** (RMSE) is the square root of the mean of the squared errors:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (y_j - \hat{y}_j)^2}$$

Root Mean Squared Error

Linear Regression

1. Define Regression and types of regression?

2. Explain the difference between regression and classification?

3. Explain simple linear regression along with Multiple Linear Regression?

4. What are the industry applications of regressions?

5. What are the ways to Analyzing Result Parameters?