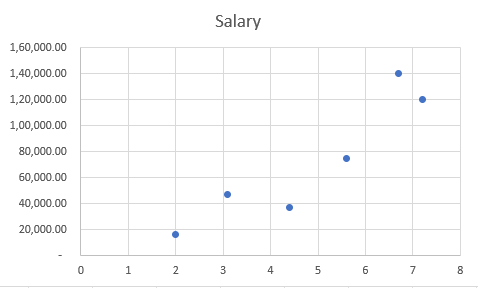
# Linear regression

Linear regression is a supervised machine learning algorithm used for predicting a continuous outcome variable (also called the dependent variable) based on one or more predictor variables (independent variables). In the context of linear regression, the relationship between the independent variables and the dependent variable is assumed to be linear.

**Simple linear regression**: If we have one independent variable and one dependent variable then it is called as simple linear regression

**Multiple linear regression**: If we have more than one independent variable then it is multiple linear regression.

|  |  |
| --- | --- |
| Experience | Salary |
| 5.6 | 75,000.00 |
| 7.2 | 1,20,000.00 |
| 3.1 | 47,000.00 |
| 4.4 | 37,000.00 |
| 2 | 16,000.00 |
| 6.7 | 1,40,000.00 |



## Key Concept

* Linear Relationship:
  + In a linear regression model, it is assumed that the relationship between the independent variable(s) and the dependent variable is linear. This implies that a change in the predictor variables will result in a proportional change in the outcome variable.
* Best Fit Line:
  + The goal of linear regression is to find the best-fit line that minimizes the sum of squared differences (residuals) between the predicted and actual values of the dependent variable. This line is often referred to as the "best fit line" or the "regression line."
* Residuals:
  + Residuals are the differences between the actual values of the dependent variable and the values predicted by the linear regression model. The sum of squared residuals is minimized to find the best fit line.
* OLS (Ordinary Least Squares) Method:
  + The Ordinary Least Squares method is a common approach used to estimate the parameters (slope and intercept) of the linear regression model. It minimizes the sum of squared residuals to find the best-fit line.
* Gradient Descent:
  + Gradient descent is an optimization algorithm used to find the minimum of a function iteratively. In the context of linear regression, it can be employed to find the optimal values of the model parameters (slope and intercept) by iteratively adjusting them in the direction that minimizes the cost function (sum of squared residuals).
* Linear Regression Equation:
  + The linear regression equation is often represented as: . where y is the dependent variable, x is the independent variable, m is the slope, and b is the y-intercept when independent variable is zero.
* Assumption of Linear regression:
  + Linearity: The relationship between variables is linear.
  + Independence: Residuals are independent of each other.
  + Homoscedasticity: Residuals have constant variance. (Outlier treatment can fix the issue)
  + Normality: Residuals are normally distributed.
  + Multi-collinearity: Two or more independent variables have no co-relation.
  + Auto-correlation: There is no auto-correlation

|  |  |  |  |
| --- | --- | --- | --- |
| Experience | Profile | Gender | Salary |
| 7.5 | Web | M | 2,10,000.00 |
| 2.5 | Android | M | 60,000.00 |
| 6 | ML | F | 1,80,000.00 |
| 8 | Web | F | 2,50,000.00 |
| 2.5 | ML | F | 80,000.00 |

Since, we have 3 independent variables in above example the equation would be

a, b, c can be considered as weights of independent variables. Linear regression is finding the weights to predict the new variable. D is the offset value in the linear regression model

## Ordinary Least square:

We can calculate the value of m and b with the below formulae in OLS method:

We have to find the line where we can reduce the Sum of square. Hence, we have to differentiate the above Sum of square with m and b.

Calculus

Calculus is a branch of mathematics that focuses on the study of rate of change and the accumulation of quantities. It provides framework for understanding how things change and how they accumulate over time and space. There are two main branches of it, differential calculus and integral calculus.

In calculus we mostly study things which change according to a pattern. These patterns of change are mathematical represent by concept of functions. We will understand the ideas of limits, differentiation and integration.

1. **Differential Calculus:**
   * **Concept of Derivatives:** Differential calculus deals with the concept of derivatives, which represent the rate at which a quantity changes. It is the slope of a curve at a specific point. Derivatives are used to analyse motion, growth, decay, and other dynamic phenomena.
   * **Notation:** The derivative of a function f(x) with respect to x is often denoted as f'(x) or dy/dx.
2. **Integral Calculus:**
   * **Concept of Integrals:** Integral calculus focuses on the concept of integrals, which represent the accumulation of quantities. Integrals are used to calculate areas under curves, total accumulated change, and other measures of accumulation.
   * **Notation:** The integral of a function f(x) with respect to x is denoted as ∫f(x)dx.