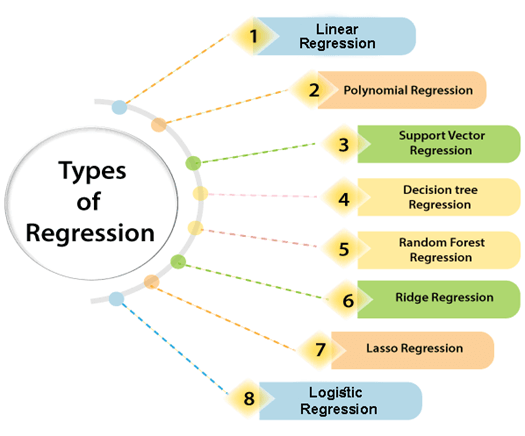
Types of Regression

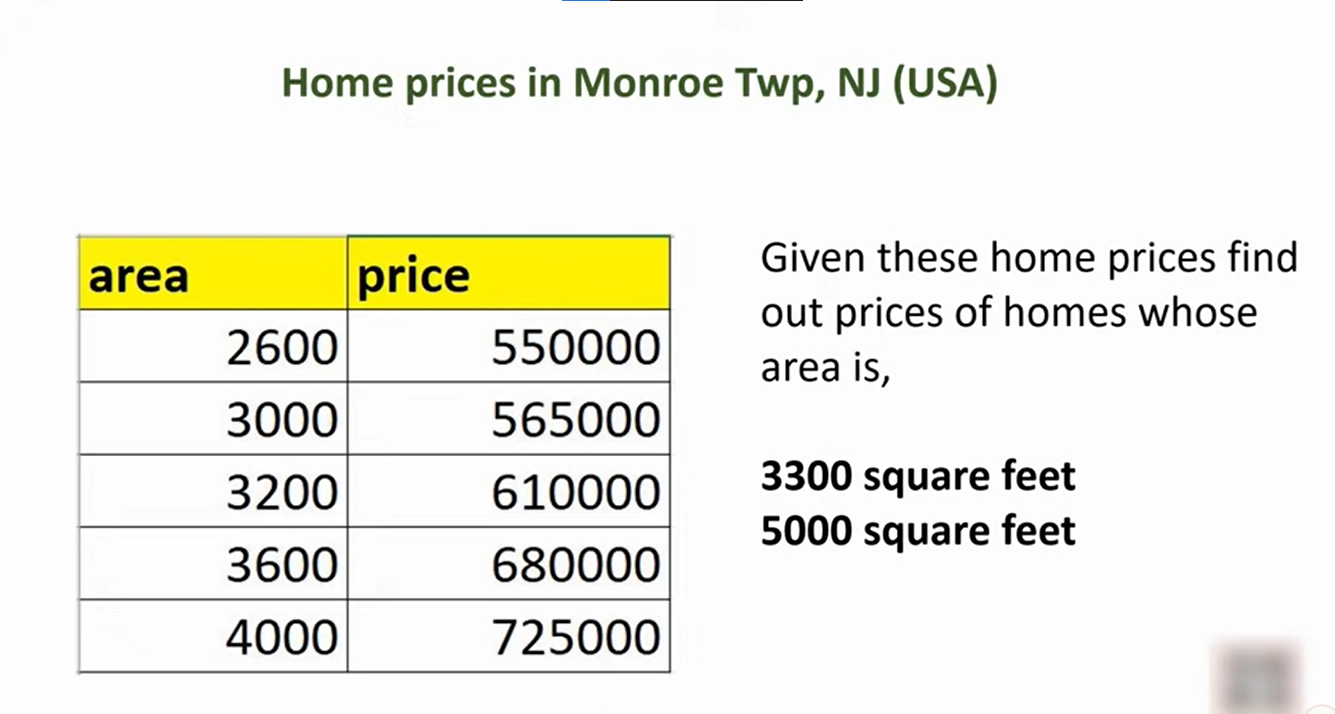
There are various types of regressions which are used in data science and machine learning. Each type has its own importance on different scenarios, but at the core, all the regression methods analyze the effect of the independent variable on dependent variables. Here we are discussing some important types of regression which are given below:

* **Linear Regression**
* **Logistic Regression**
* **Polynomial Regression**
* **Support Vector Regression**
* **Decision Tree Regression**
* **Random Forest Regression**
* **Ridge Regression**
* **Lasso Regression:**



Linear Regression:

* Linear regression is a statistical regression method which is used for predictive analysis.
* It is one of the very simple and easy algorithms which works on regression and shows the relationship between the continuous variables.
* It is used for solving the regression problem in machine learning.
* Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), hence called linear regression.
* If there is only one input variable (x), then such linear regression is called **simple linear regression**. And if there is more than one input variable, then such linear regression is called **multiple linear regression**.
* The relationship between variables in the linear regression model can be explained using the below image. Here we are predicting the salary of an employee on the basis of **the year of experience**.



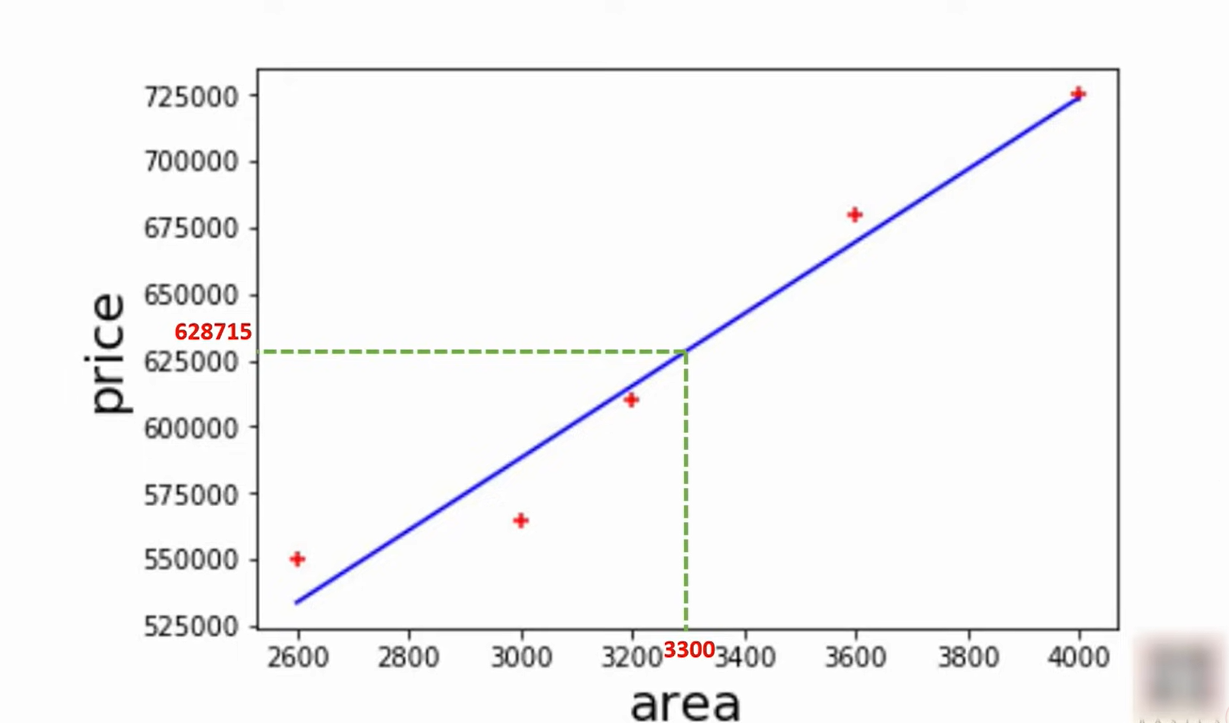
* Below is the mathematical equation for Linear regression:

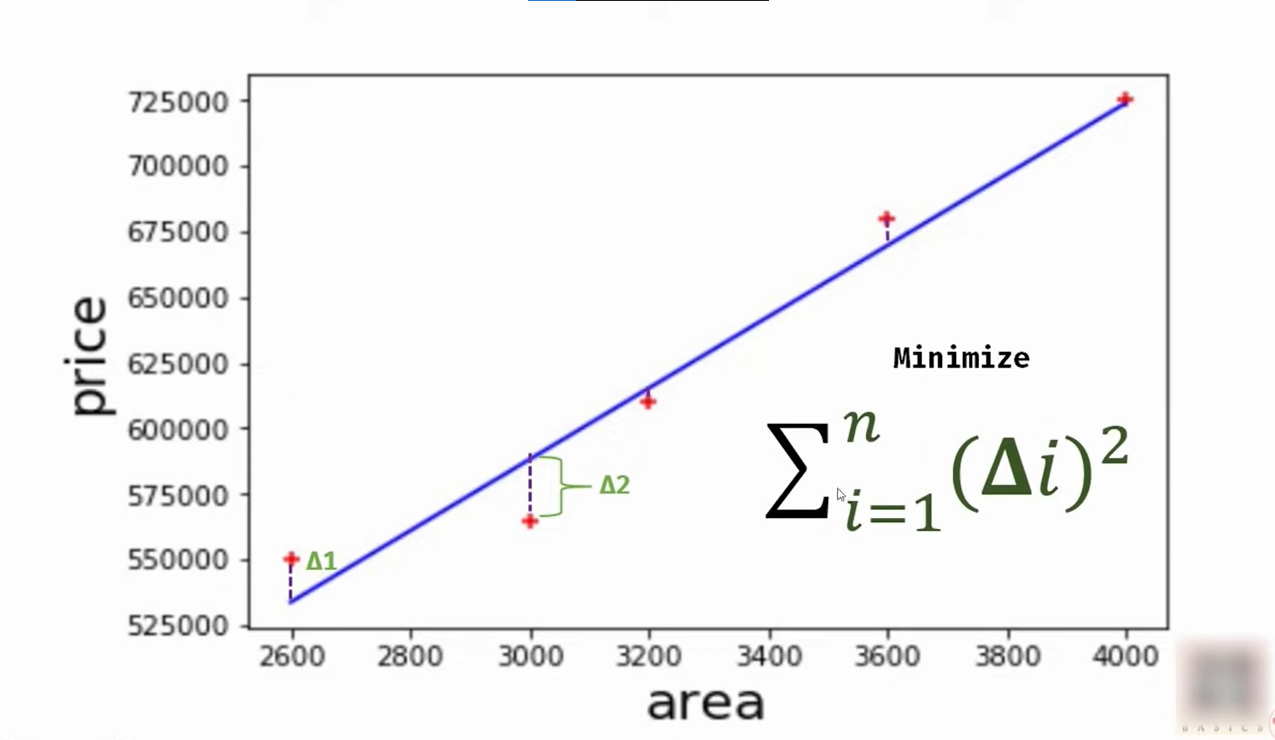
1. Y= aX+b

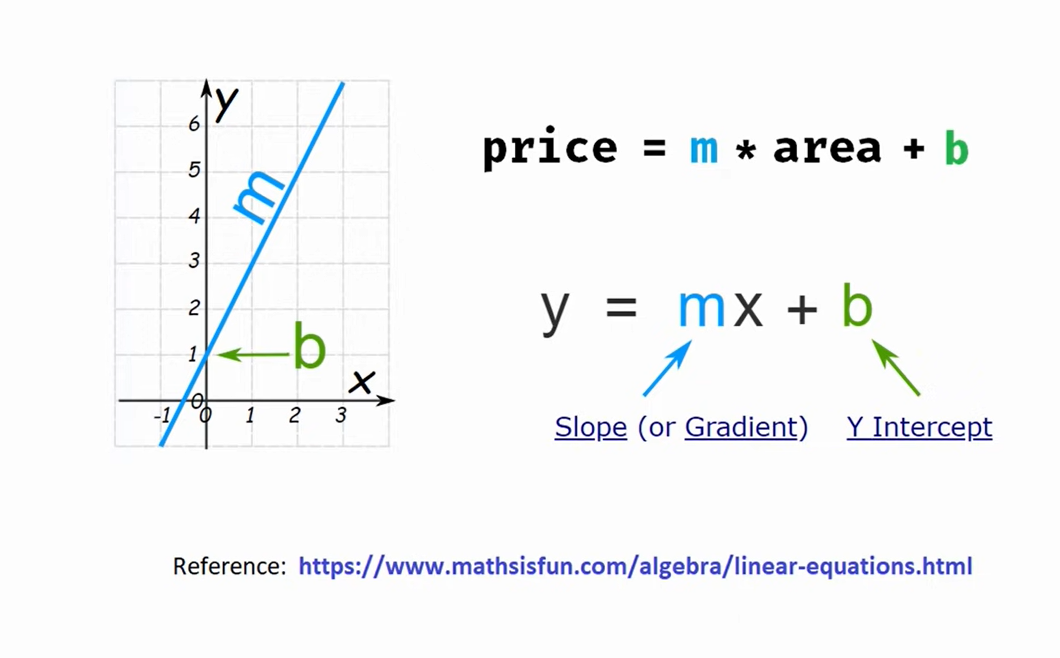
**Here, Y = dependent variables (target variables),**  
**X= Independent variables (predictor variables),**  
**a and b are the linear coefficients**

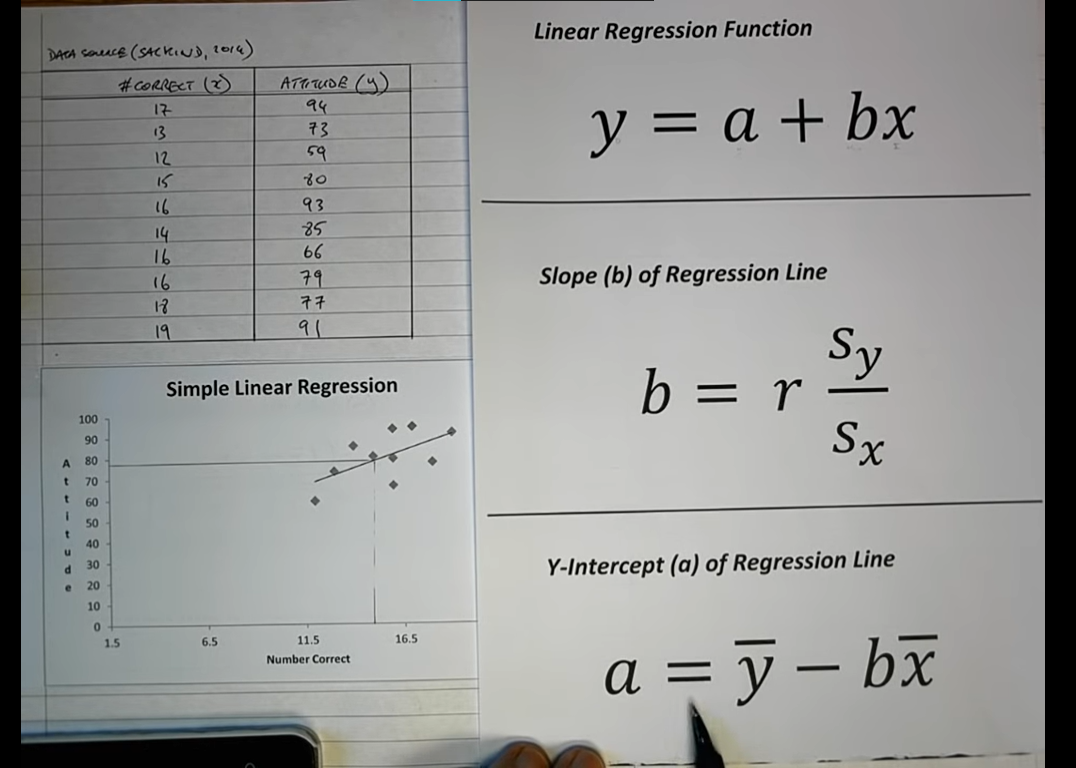
Some popular applications of linear regression are:

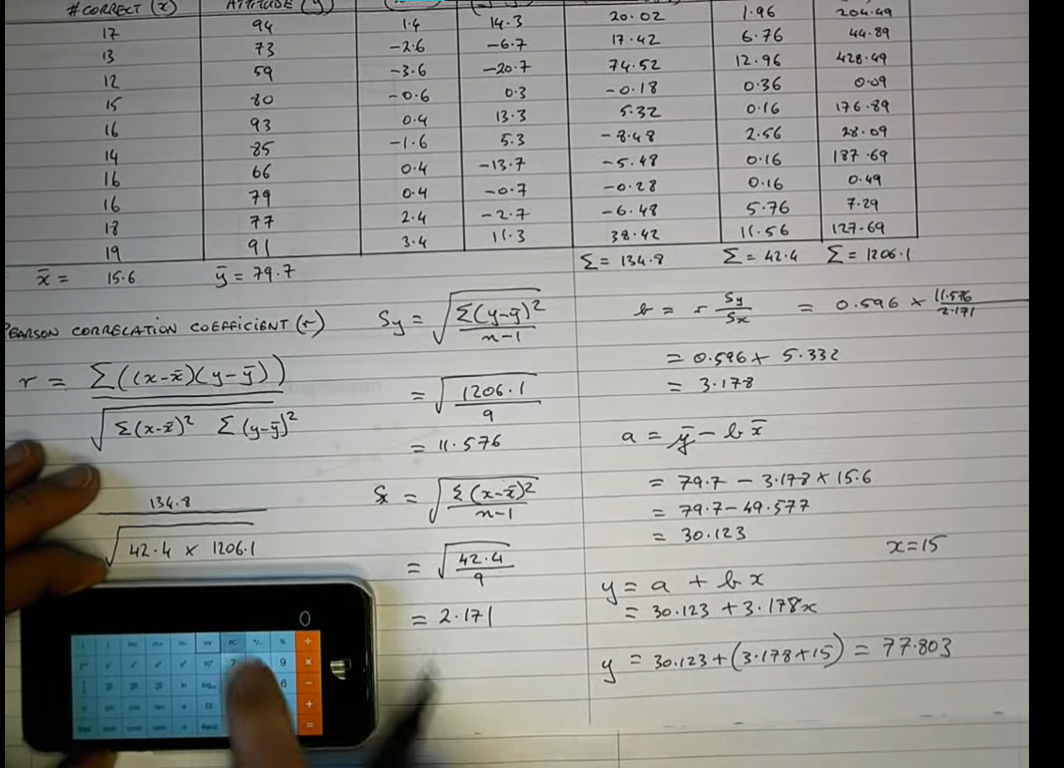
* **Analyzing trends and sales estimates**
* **Salary forecasting**
* **Real estate prediction**
* **Arriving at ETAs in traffic.**

****

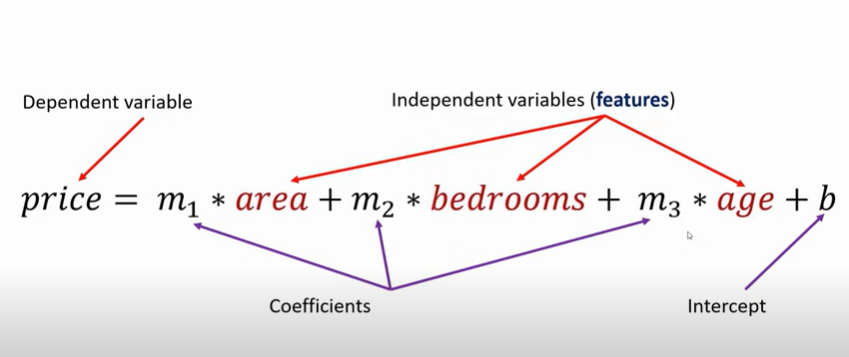
****

****





**multiple linear regression:**



# ML Polynomial Regression

* Polynomial Regression is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial. The Polynomial Regression equation is given below:

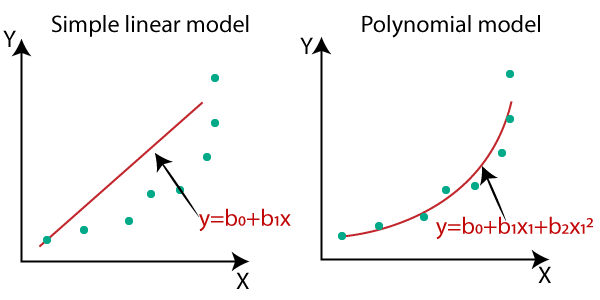
y= b0+b1x1+ b2x12+ b2x13+...... bnx1n

* It is also called the special case of Multiple Linear Regression in ML. Because we add some polynomial terms to the Multiple Linear regression equation to convert it into Polynomial Regression.
* It is a linear model with some modification in order to increase the accuracy.
* The dataset used in Polynomial regression for training is of non-linear nature.
* It makes use of a linear regression model to fit the complicated and non-linear functions and datasets.
* **Hence, "In Polynomial regression, the original features are converted into Polynomial features of required degree (2,3,..,n) and then modeled using a linear model."**

## Need for Polynomial Regression:

The need of Polynomial Regression in ML can be understood in the below points:

* If we apply a linear model on a **linear dataset**, then it provides us a good result as we have seen in Simple Linear Regression, but if we apply the same model without any modification on a **non-linear dataset**, then it will produce a drastic output. Due to which loss function will increase, the error rate will be high, and accuracy will be decreased.
* So for such cases, **where data points are arranged in a non-linear fashion, we need the Polynomial Regression model**. We can understand it in a better way using the below comparison diagram of the linear dataset and non-linear dataset.



* In the above image, we have taken a dataset which is arranged non-linearly. So if we try to cover it with a linear model, then we can clearly see that it hardly covers any data point. On the other hand, a curve is suitable to cover most of the data points, which is of the Polynomial model.
* Hence, if the datasets are arranged in a non-linear fashion, then we should use the Polynomial Regression model instead of Simple Linear Regression.

#### Note: A Polynomial Regression algorithm is also called Polynomial Linear Regression because it does not depend on the variables, instead, it depends on the coefficients, which are arranged in a linear fashion.

## Equation of the Polynomial Regression Model:

**Simple Linear Regression equation:         y = b0+b1x         .........(a)**

**Multiple Linear Regression equation:         y= b0+b1x+ b2x2+ b3x3+....+ bnxn         .........(b)**

**Polynomial Regression equation:         y= b0+b1x + b2x2+ b3x3+....+ bnxn         ..........(c)**

When we compare the above three equations, we can clearly see that all three equations are Polynomial equations but differ by the degree of variables. The Simple and Multiple Linear equations are also Polynomial equations with a single degree, and the Polynomial regression equation is Linear equation with the nth degree. So if we add a degree to our linear equations, then it will be converted into Polynomial Linear equations.

Support Vector Regression:

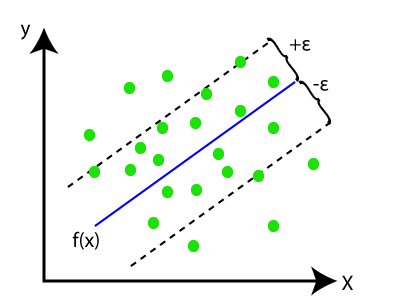
Support Vector Machine is a supervised learning algorithm which can be used for regression as well as classification problems. So if we use it for regression problems, then it is termed as Support Vector Regression.and outlier detection

The goal of sum algorithms is to create the best line or decision boundary on multiple dimension space the best decision boundary is called hyperplane

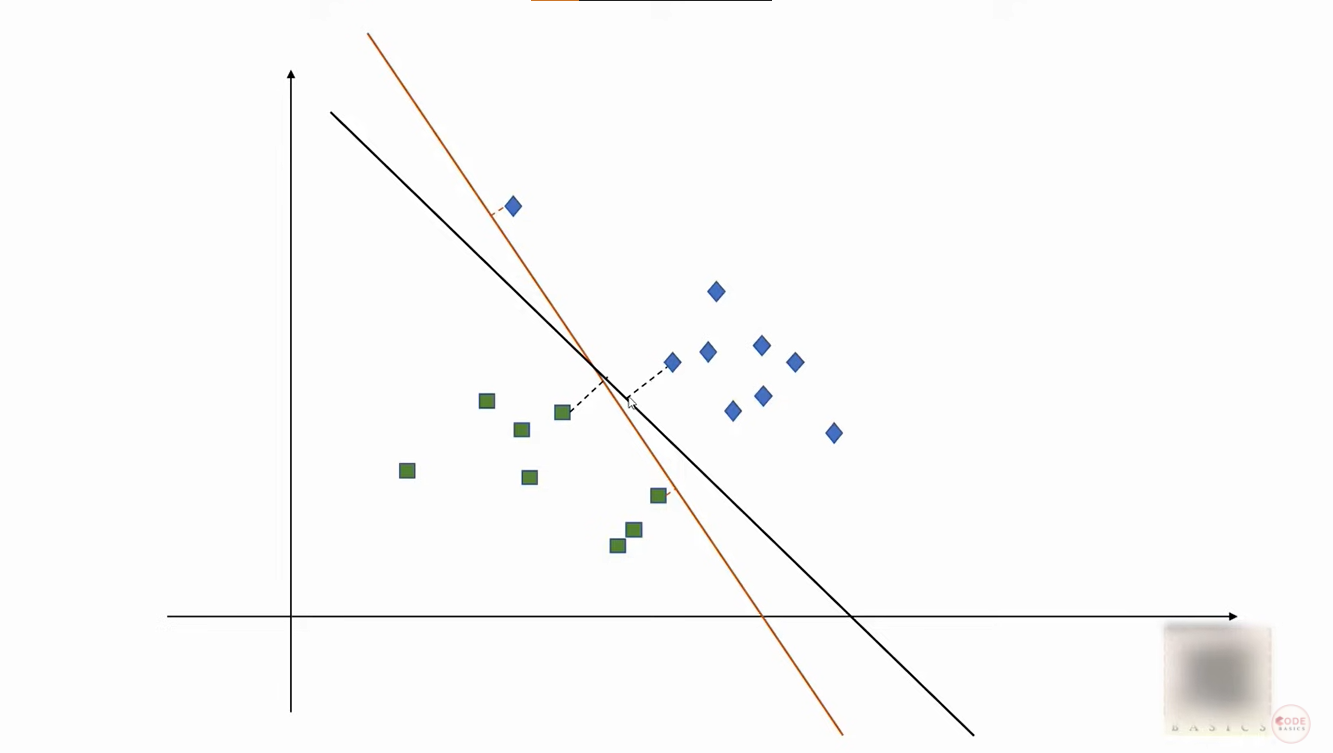
Support Vector Regression is a regression algorithm which works for continuous variables. Below are some keywords which are used in **Support Vector Regression**:

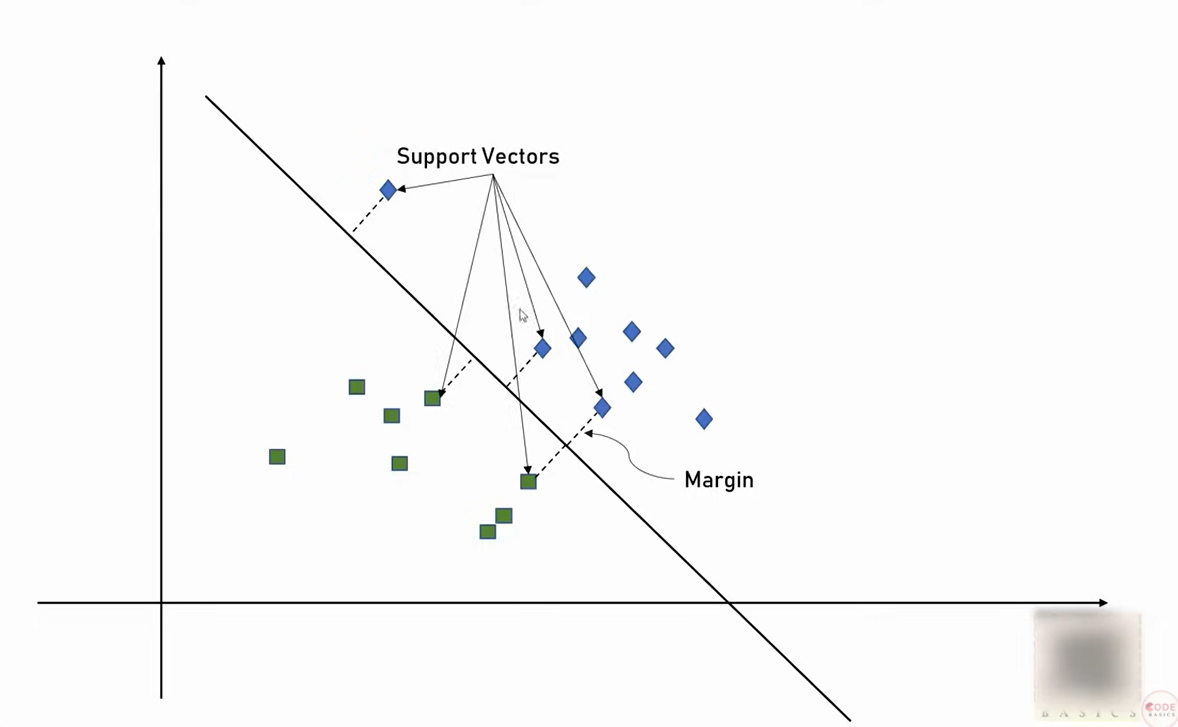
* **Kernel:** It is a function used to map a lower-dimensional data into higher dimensional data.
* **Hyperplane:** In general SVM, it is a separation line between two classes, but in SVR, it is a line which helps to predict the continuous variables and cover most of the datapoints.
* **Margin:** margine is the perpendicular distance b/w the closet data points line to the support vector . large margin is considered as good margin. Small margin consider a bad margin
* **Boundary line:** Boundary lines are the two lines apart from hyperplane, which creates a margin for datapoints.
* **Support vectors:** Support vectors are the datapoints which are nearest to the hyperplane and opposite class.
* **Gamma:** a parameter for non linear hyperplan gamma define how far and reach a single traning low value mean ‘fars’ and high values mean ‘close’

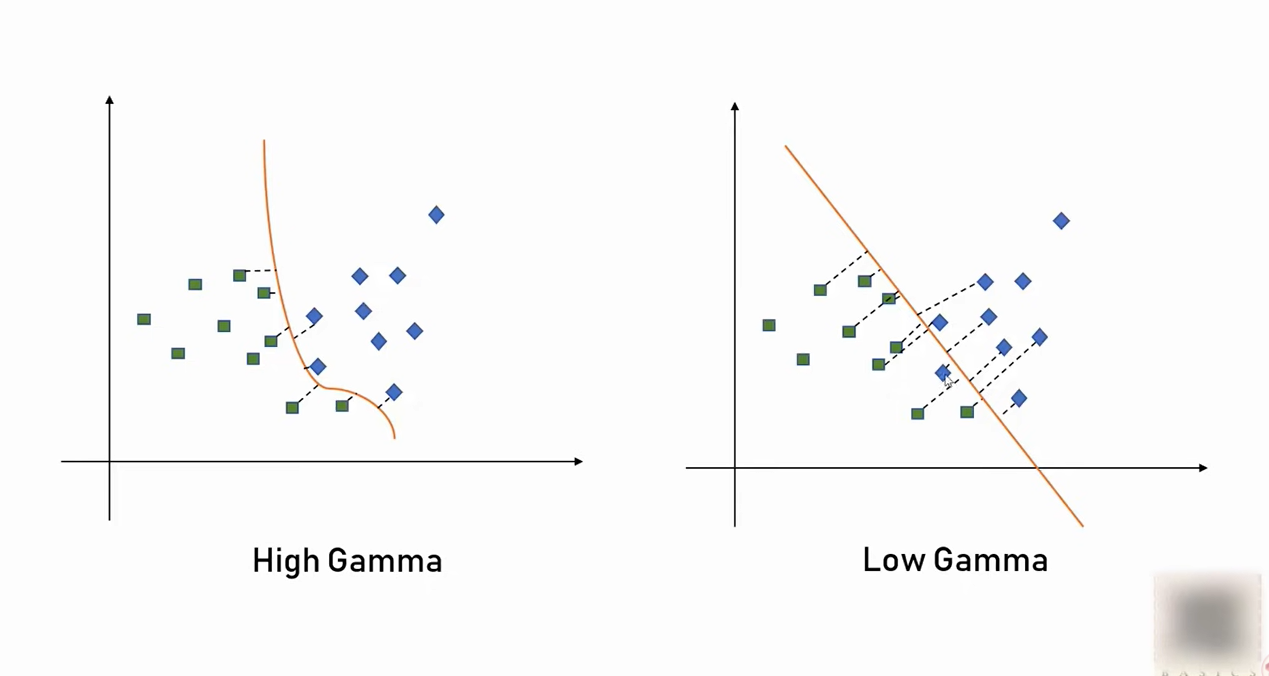
In SVR, we always try to determine a hyperplane with a maximum margin, so that maximum number of datapoints are covered in that margin. ***The main goal of SVR is to consider the maximum datapoints within the boundary lines and the hyperplane (best-fit line) must contain a maximum number of datapoints***. Consider the below image:



Here, the blue line is called hyperplane, and the other two lines are known as boundary lines.







Non linear:

