



Machine Learning and its Algorithms

Data Science, Machine Learning Course



What is ML

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning** focuses on the development of computer programs that can access data and use it learn for themselves.

Types of Machine Learning – At a Glance

Supervised Learning

- Makes machine Learn explicitly
- Data with clearly defined output is given
- Direct feedback is given
- Predicts outcome/future
- Resolves classification and regression problems



Unsupervised Learning

- Machine understands the data (Identifies patterns/structures)
- Evaluation is qualitative or indirect
- Does not predict/find anything specific

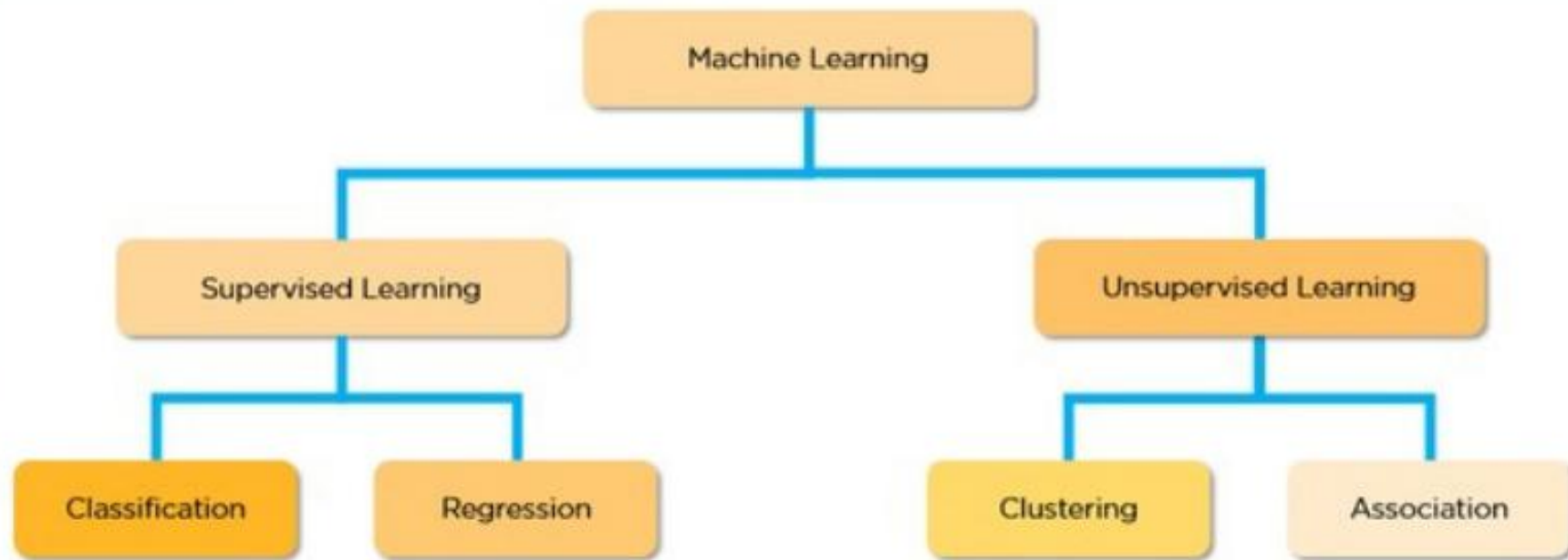


Reinforcement Learning

- An approach to AI
- Reward based learning
- Learning form +ve & +ve reinforcement
- Machine Learns how to act in a certain environment
- To maximize rewards



AI related Subjects





What is Regression

Regression models (both linear and non-linear) are used for predicting a real value, like salary for example.

If your independent variable is time, then you are forecasting future values, otherwise your model is predicting present but unknown values. Types of Machine Learning Regression models:

Simple Linear Regression

Multiple Linear Regression

Polynomial Regression

Support Vector for Regression (SVR)

Decision Tree Regression

Random Forest Regression



Data Set

	A	B
1	YearsExperience	Salary
2	1.1	39343
3	1.3	46205
4	1.5	37731
5	2	43525
6	2.2	39891
7	2.9	56642
8	3	60150
9	3.2	54445
10	3.2	64445



Simple Linear Regression

Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables

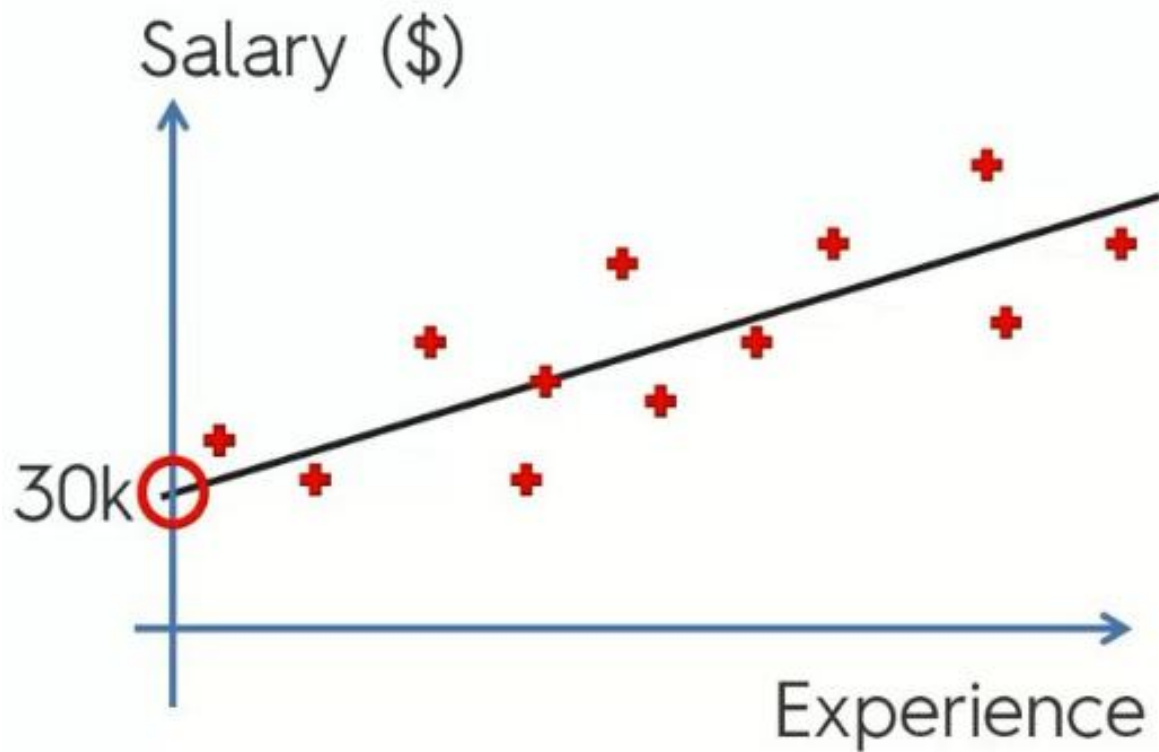
$$y = b_0 + b_1 * x_1$$

Diagram illustrating the components of the Simple Linear Regression equation:

- Constant**: Points to b_0
- Coefficient**: Points to b_1
- Dependent variable (DV)**: Points to y
- Independent variable (IV)**: Points to x_1

What is Data Science

Simple Linear Regression:



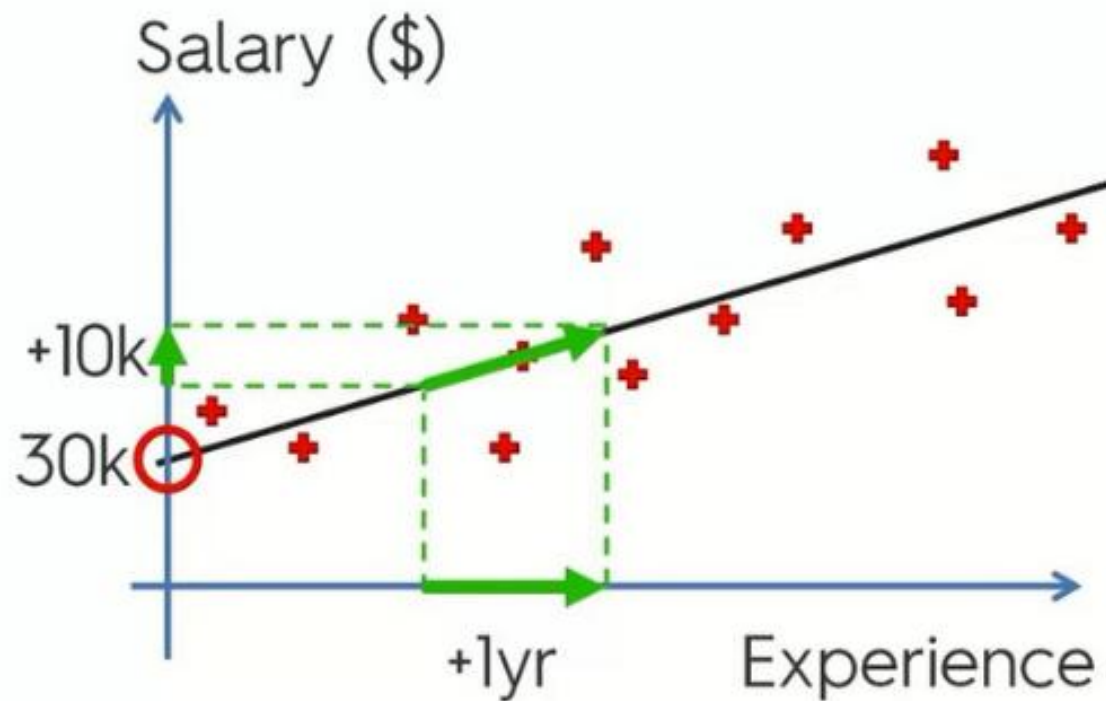
$$y = b_0 + b_1 * x$$



$$\text{Salary} = \textcircled{b_0} + b_1 * \text{Experience}$$

What is Data Science

Simple Linear Regression:



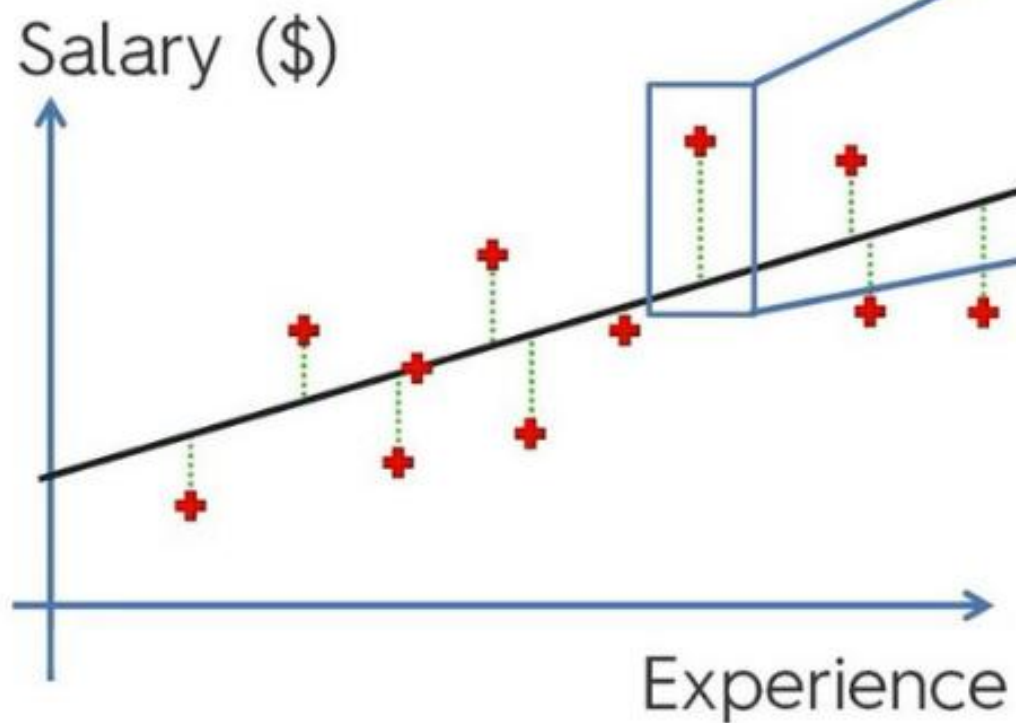
$$y = b_0 + b_1 * x$$



$$\text{Salary} = b_0 + b_1 * \text{Experience}$$

What is Data Science

Simple Linear Regression:



$$\text{SUM } (y - \hat{y})^2 \rightarrow \min$$

Polynomial linear regression

Simple
Linear
Regression

$$y = b_0 + b_1 x_1$$

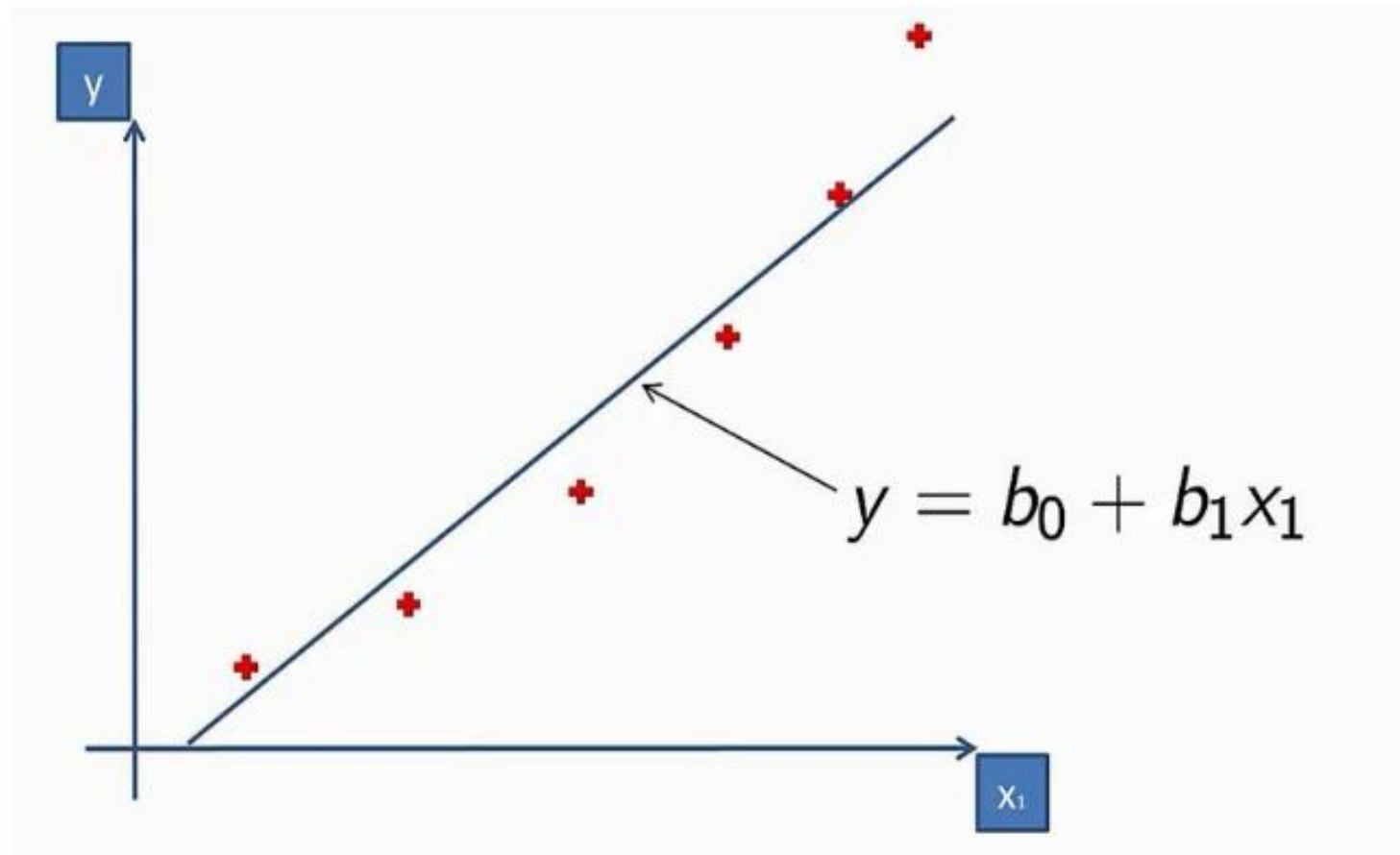
Multiple
Linear
Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

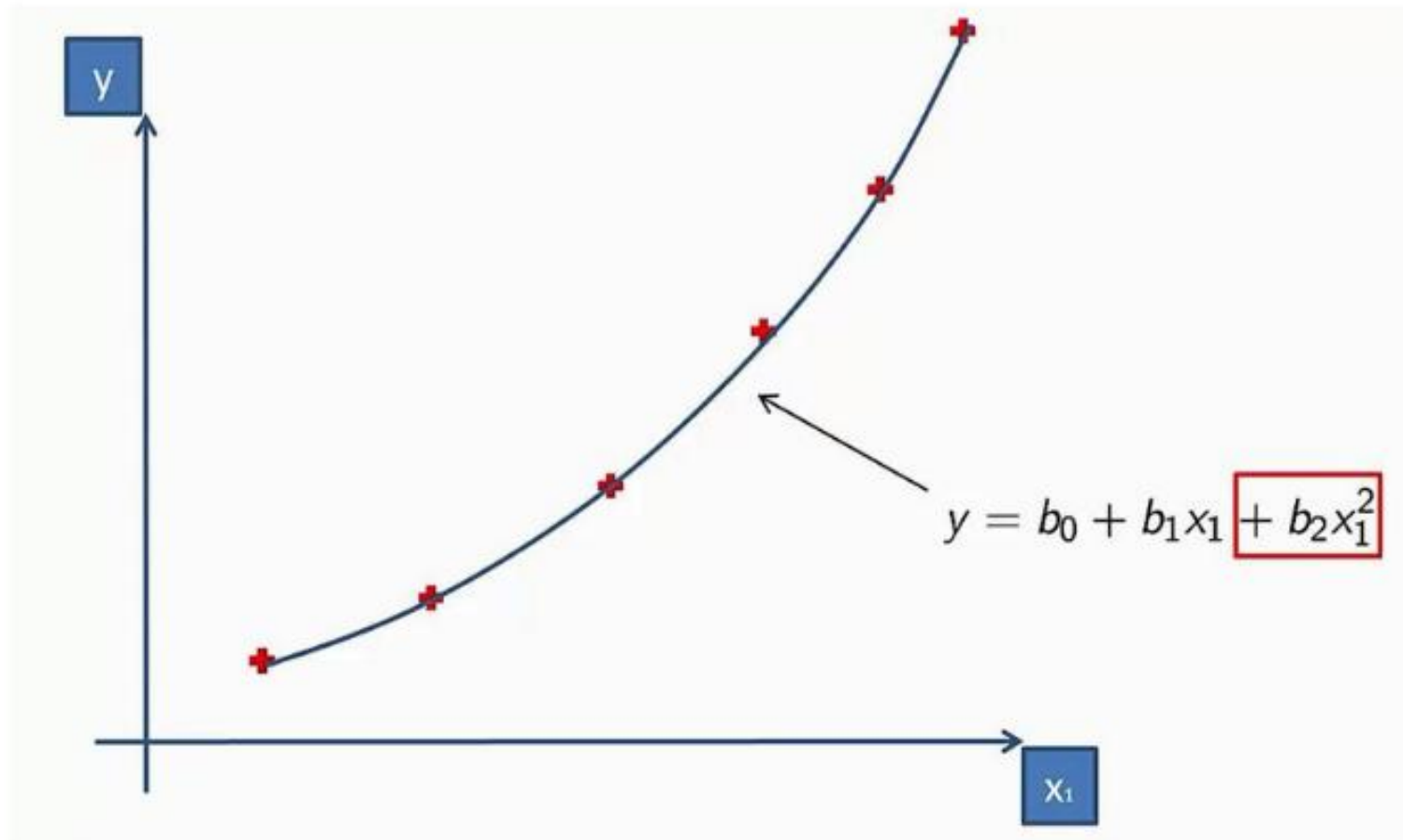
Polynomial
Linear
Regression

$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \dots + b_n x_1^n$$

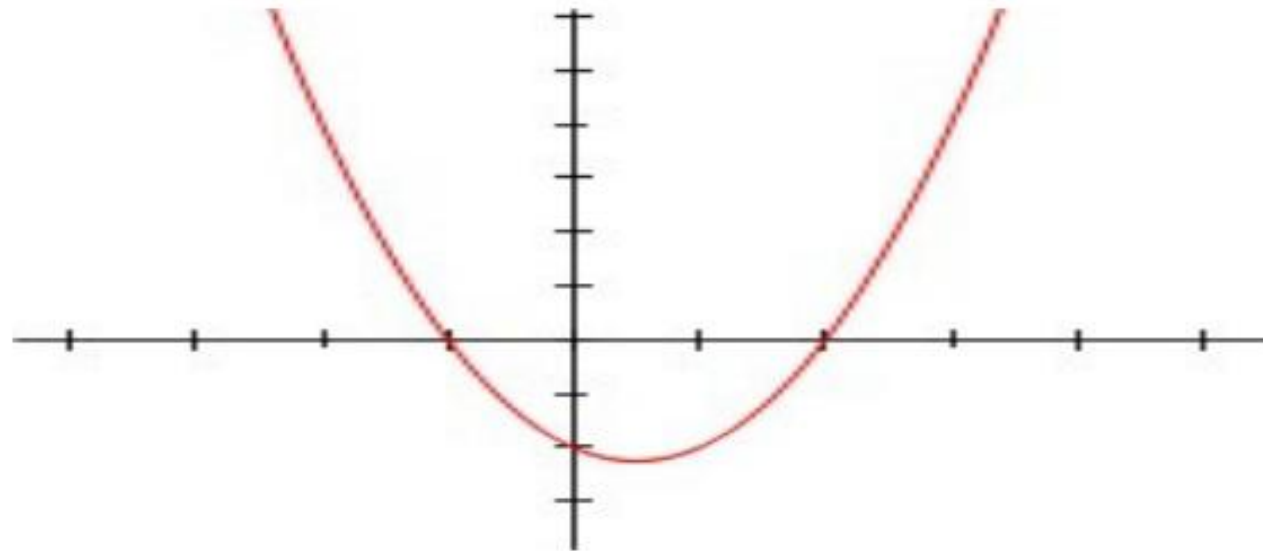
Polynomial linear regression



Polynomial linear regression



Non-linear regression



Nonlinear regression involves curves. This is partly true, and if you want a loose definition for the difference, you can probably stop right there. However, *linear equations can sometimes produce curves.*

In order to understand why, you need to take a look at the linear regression equation form.

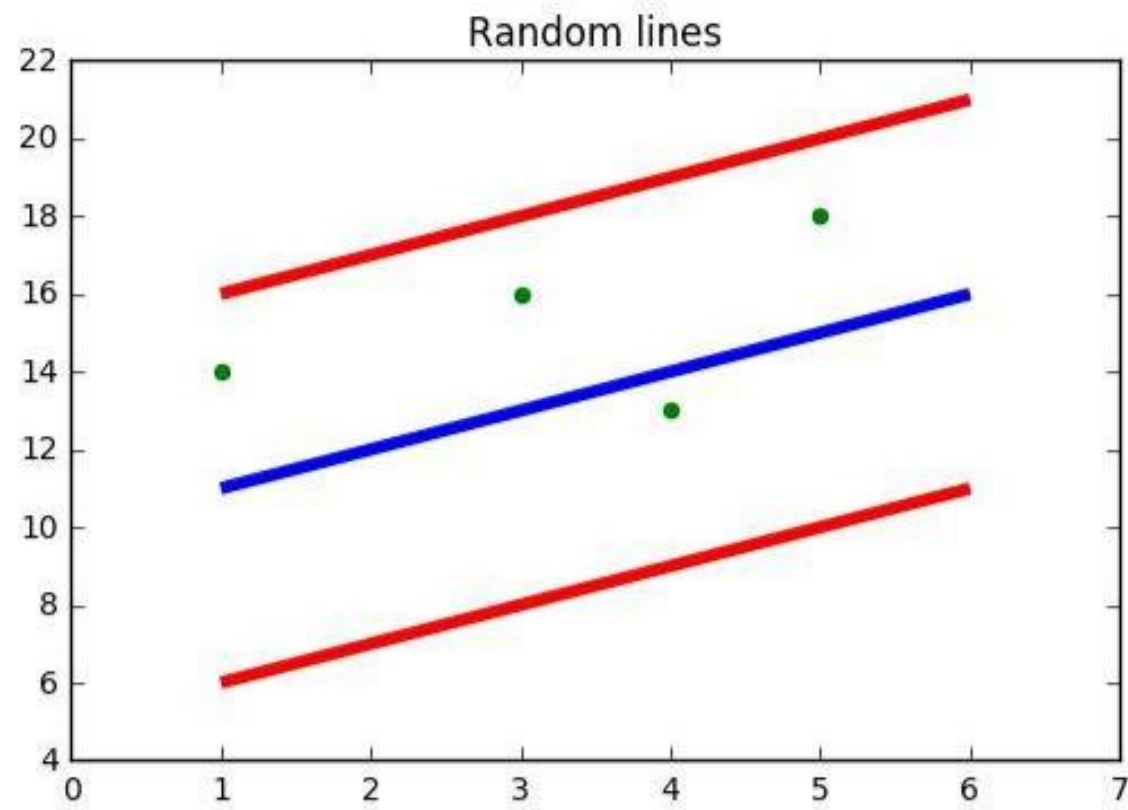
SVM, which stands for Support Vector Machine, is a classifier. Classifiers perform classification, predicting discrete categorical labels. SVR, which stands for Support Vector Regressor, is a regressor. Regressors perform regression, predicting continuous ordered variables. Both use very similar algorithms, but predict different types of variables

In simple regression we try to minimise the error rate. While in SVR we try to fit the error within a certain threshold.

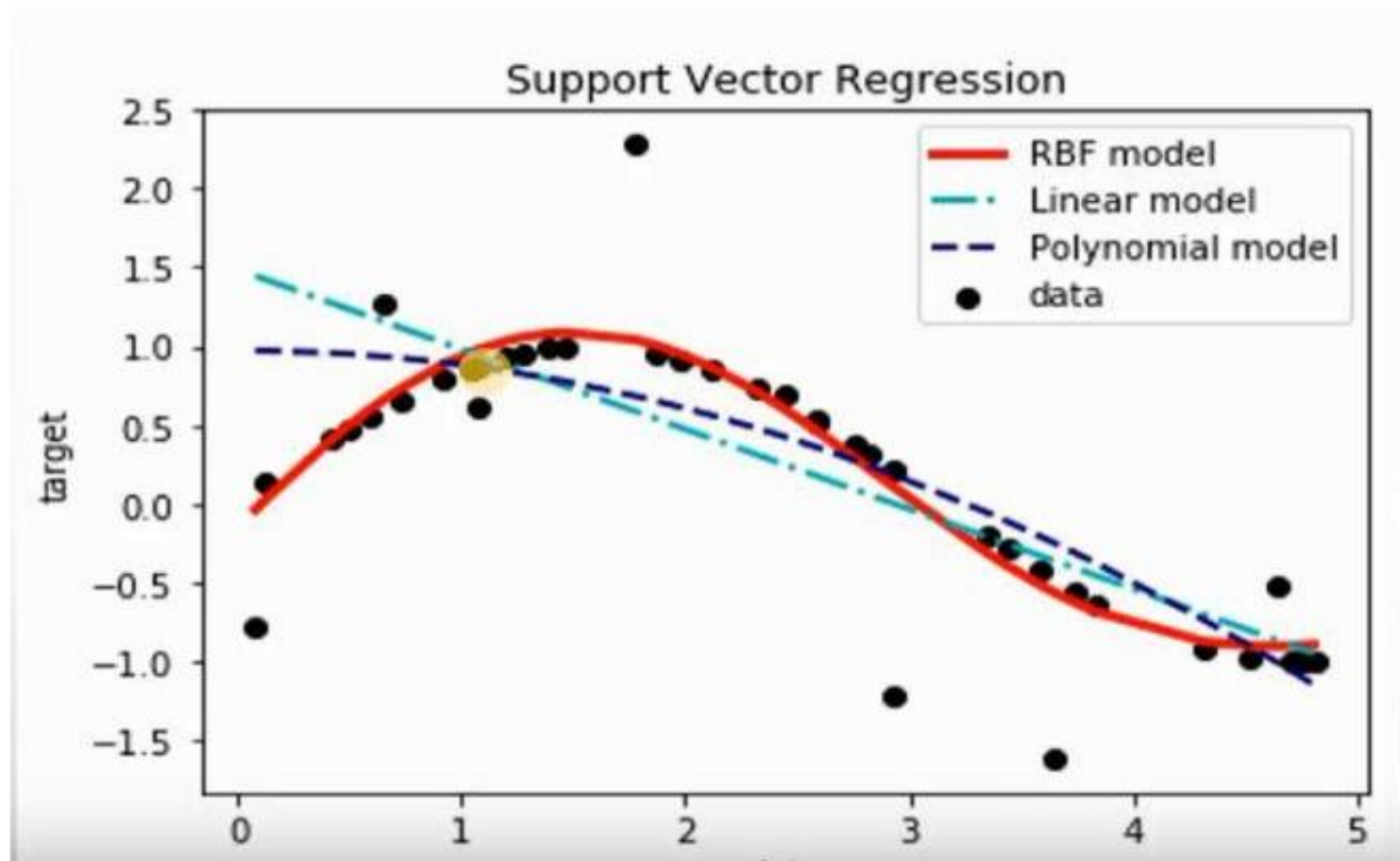
Support vector regression

Kernel: The function used to map a lower dimensional data into a higher dimensional data. **Hyper Plane:** In SVM this is basically the separation line between the data classes. Although in SVR we are going to define it as the line that will help us predict the continuous value or target value **Boundary line:** In SVM there are two lines other than Hyper Plane which creates a margin. The support vectors can be on the Boundary lines or outside it. This boundary line separates the two classes. In SVR the concept is same. **Support vectors:** These are the data points which are closest to the boundary. The distance of the points is minimum or least.

Support vector regression

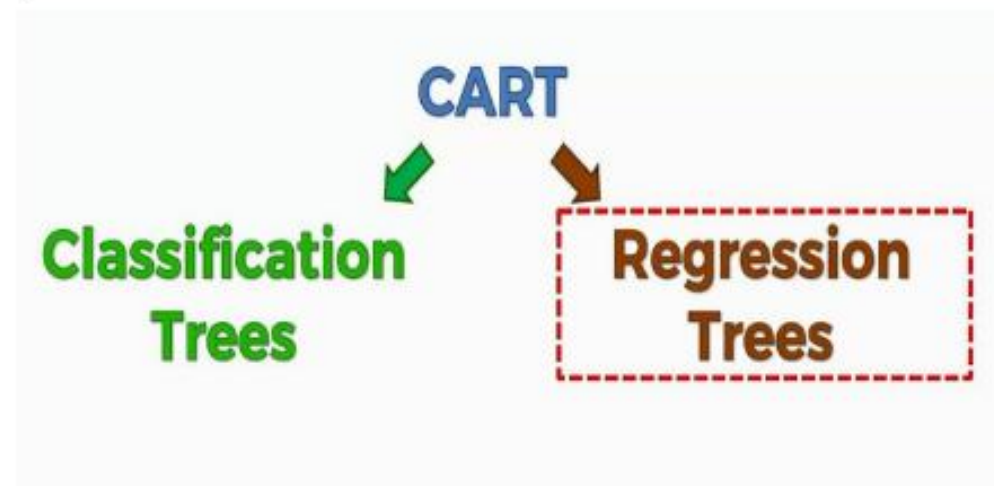


Support vector regression

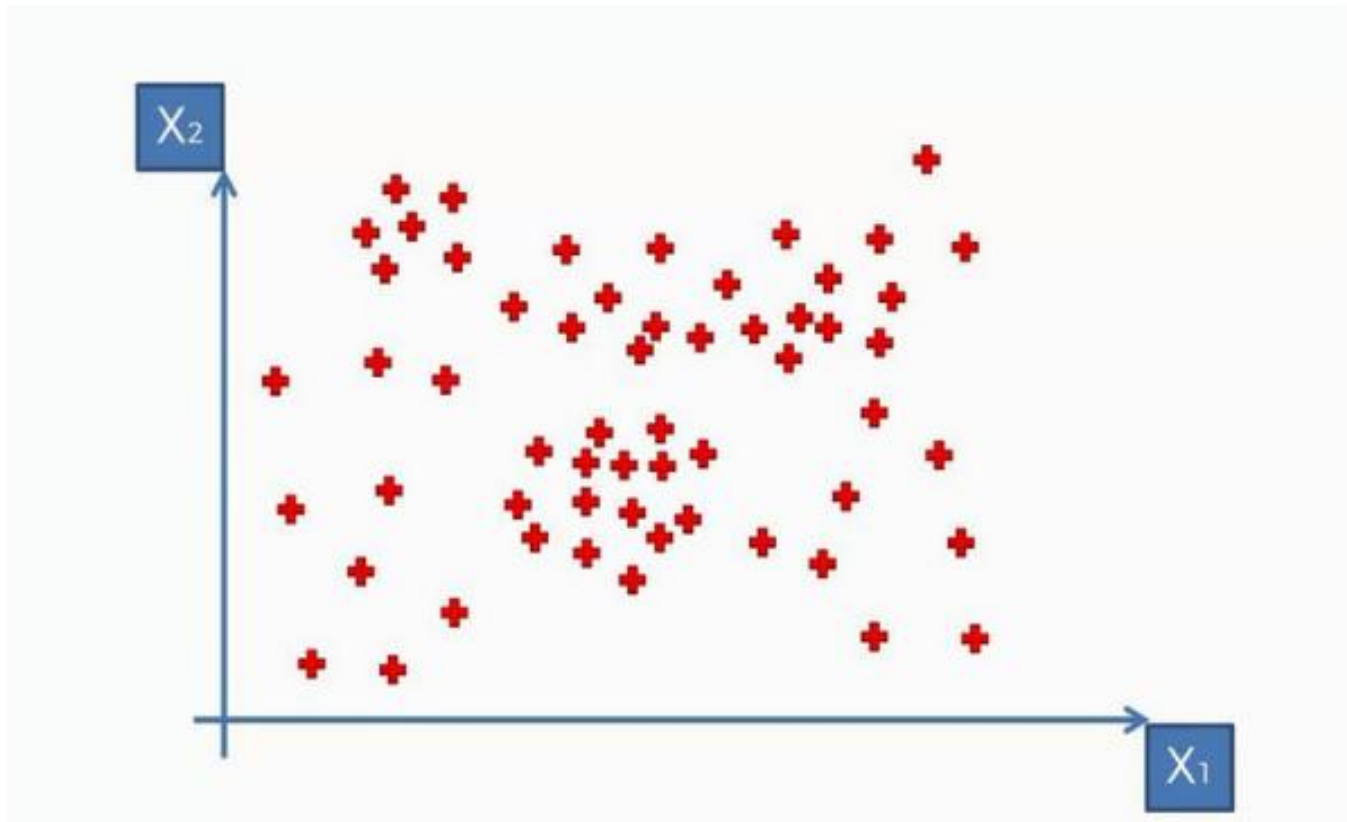


What is Cart

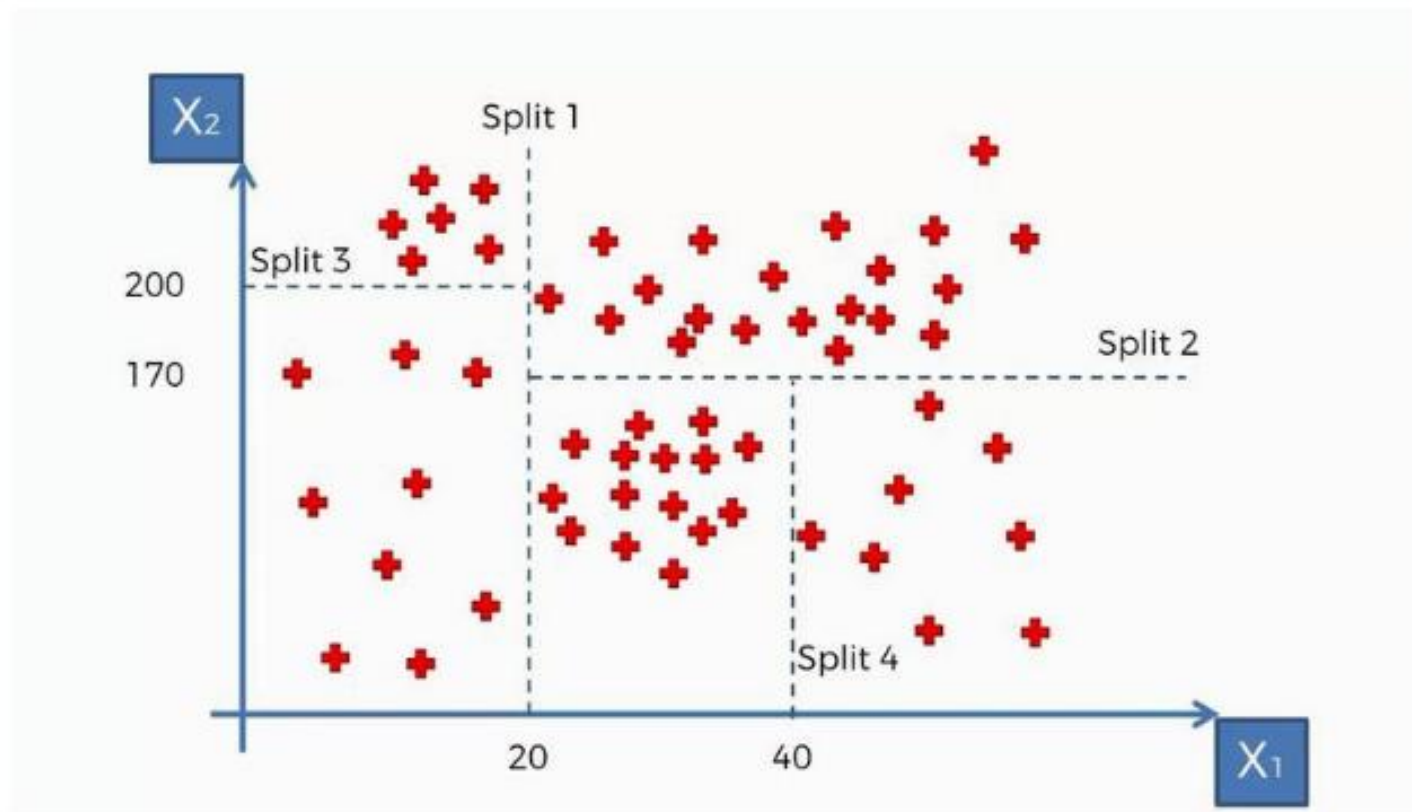
A Classification And Regression Tree (CART), is a predictive model, which explains how an outcome variable's values can be predicted based on other values. A CART output is a decision tree where each fork is a split in a predictor variable and each end node contains a prediction for the outcome variable



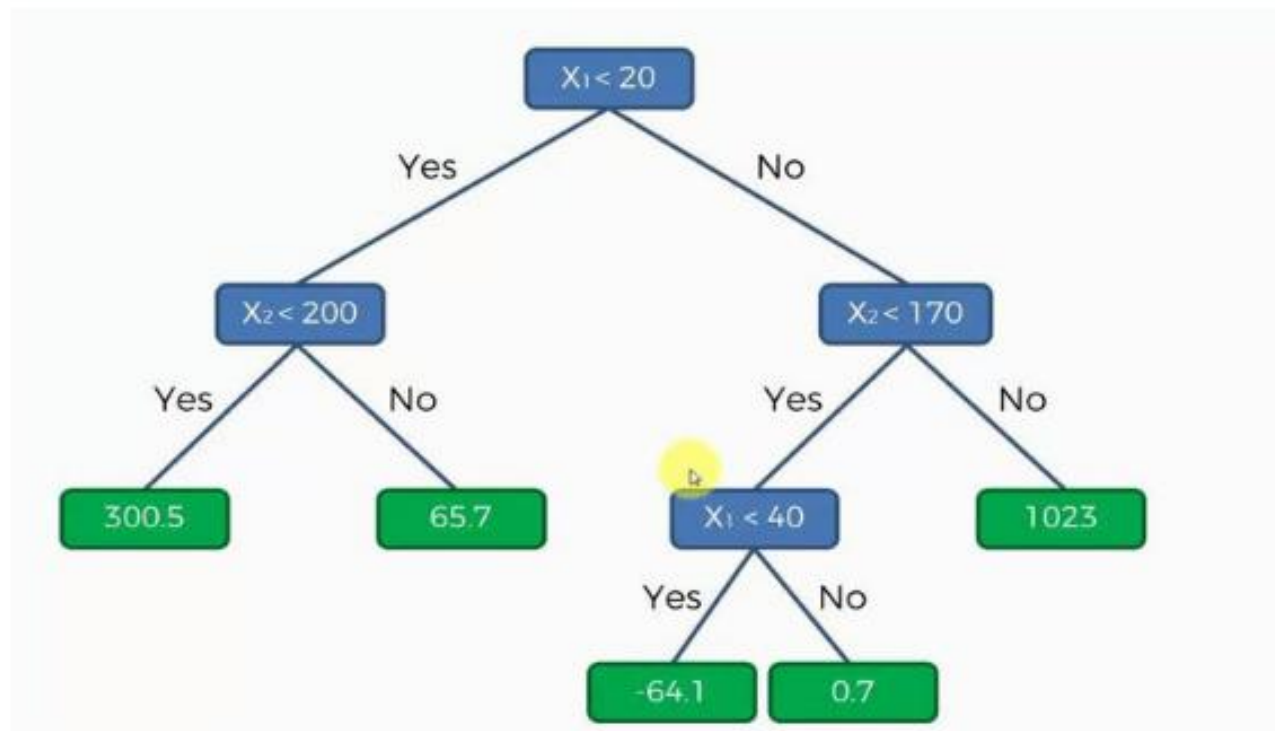
Support vector regression



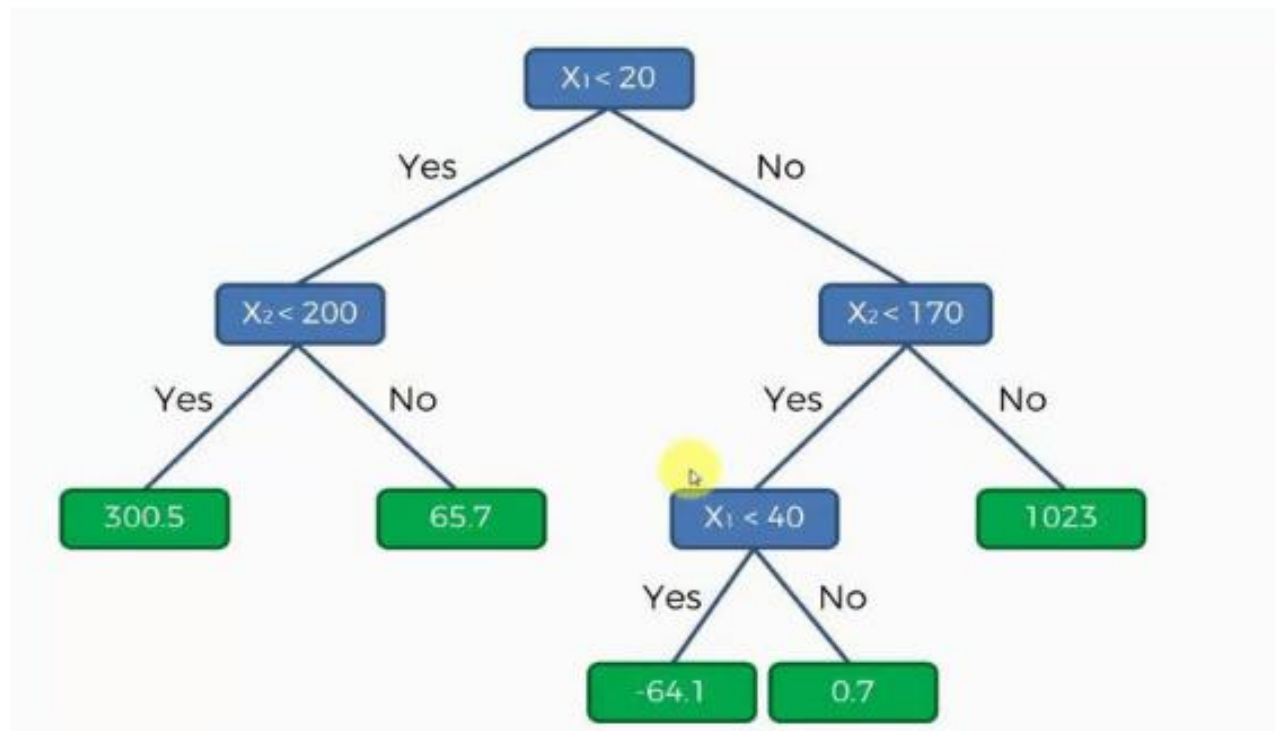
Decision Tree Regression



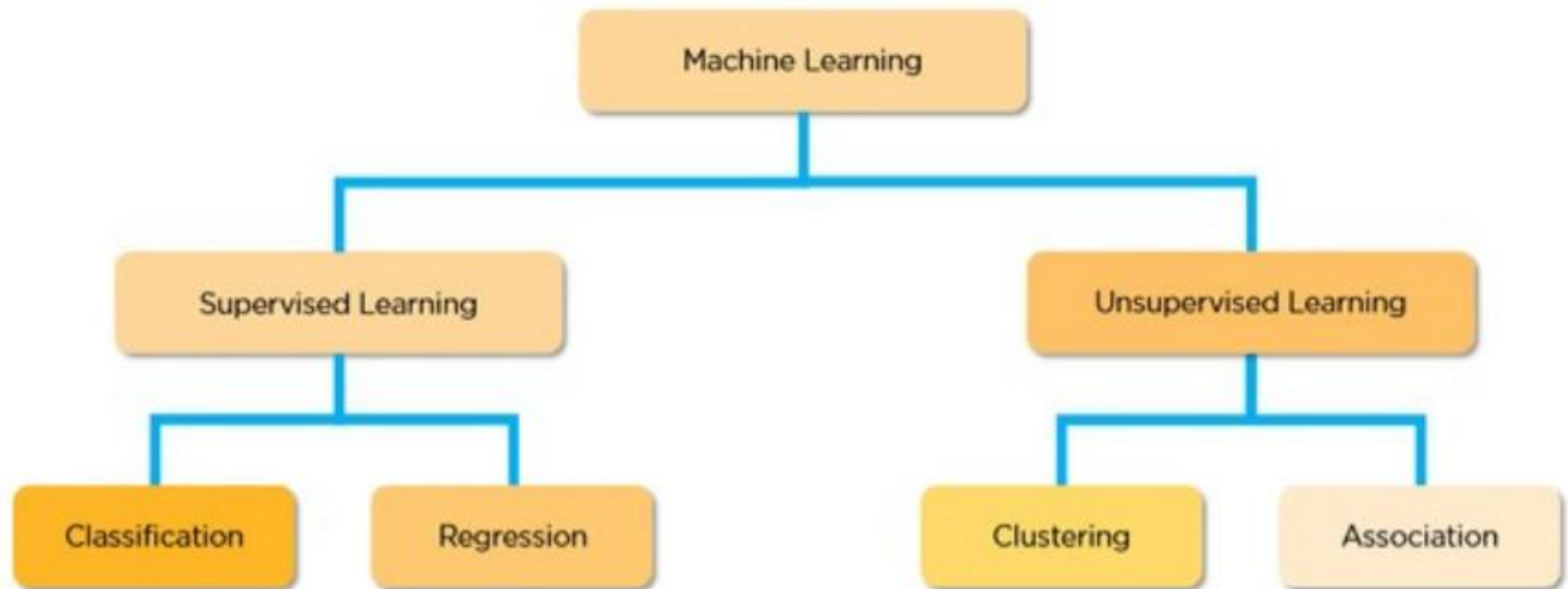
Decision Tree Regression



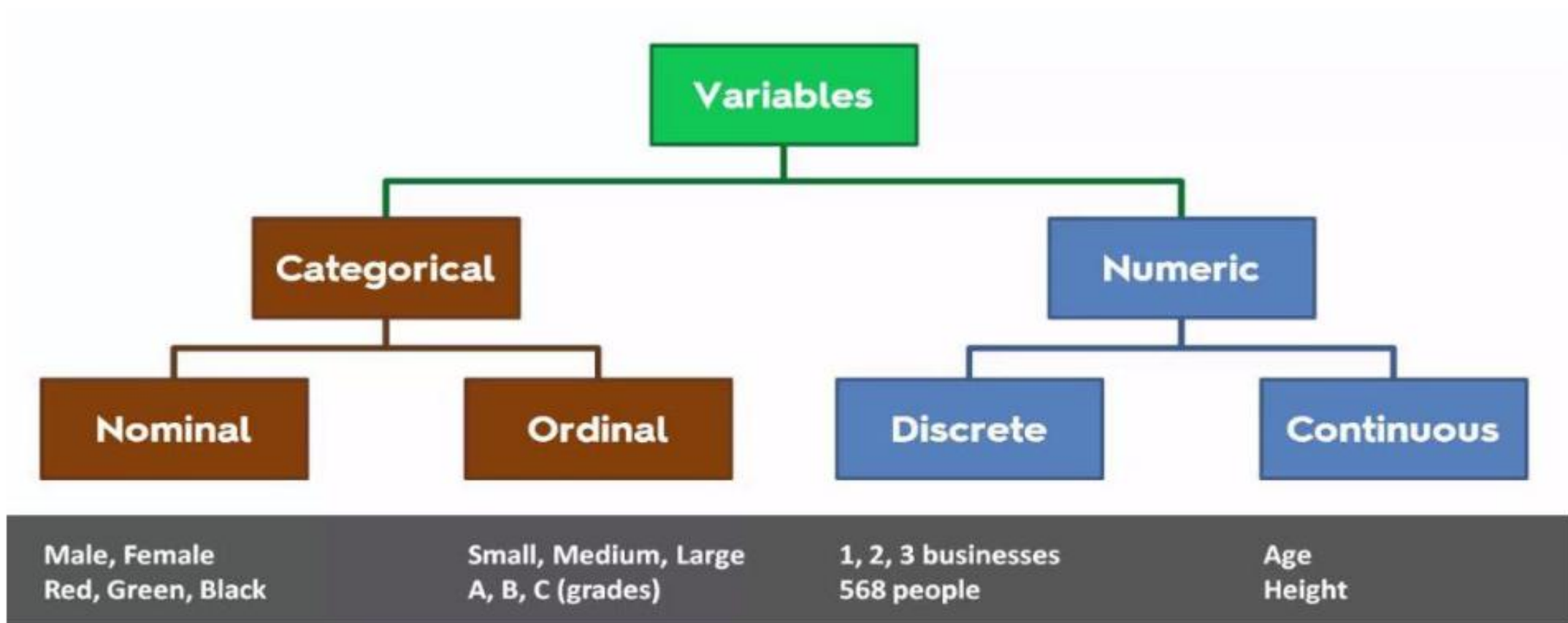
Random Forest Regression



Supervised Learning Of Machine Learning



Variable/Data its types in Machine Learning



Classification In Supervised Learning

Unlike regression where you predict a continuous number, you use classification to predict a category. There is a wide variety of classification applications from medicine to marketing. Classification models include linear models like Logistic Regression, and nonlinear ones like K-NN, Kernel SVM and Random Forests. Learning Classification models: 1.Logistic Regression 2.K-Nearest Neighbors (K-NN) 3.Support Vector Machine (SVM) 4.Kernel SVM 5.Naive Bayes 6.Decision Tree Classification 7.Random Forest Classification

Logistic Regression

Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).

In logistic regression, the dependent variable is binary or dichotomous, i.e. it only contains data coded as 1 (TRUE, success, pregnant, etc.) or 0 (FALSE, failure, non-pregnant, etc.).

Also

Probability is the likelihood or chance of an event occurring. **Probability** = the number of ways of achieving success. the total number of possible outcomes.

Is it classification algo.? So why is it called regression?

Logistic Regression

Logistic regression also called
binary regression

And also it can be Multiple ...

So it become multiple logistic regression