There are four types of machine learning algorithms. Those are,

* **Supervised:** Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.

Under the umbrella of supervised learning fall: Classification, Regression and Forecasting.

1. **Classification**: In classification tasks, the machine learning program must draw a conclusion from observed values and determine to  
    what category new observations belong. For example, when filtering emails as ‘spam’ or ‘not spam’, the program must look at existing observational data and filter the emails accordingly.
2. **Regression**: In regression tasks, the machine learning program must estimate – and understand – the relationships among variables. Regression analysis focuses on one dependent variable and a series of other changing variables – making it particularly useful for prediction and forecasting.
3. **Forecasting**: Forecasting is the process of making predictions about the future based on the past and present data, and is commonly used to analyze trends.

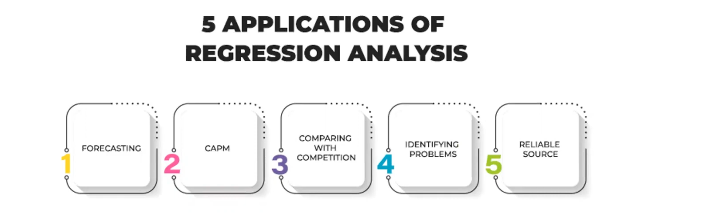
* **Semi-Supervised:** Semi-supervised learning is similar to supervised learning, but instead uses both labeled and unlabelled data. Labeled data is essentially information that has meaningful tags so that the algorithm can understand the data, whilst unlabelled data lacks that information.   
  By using this combination, machine learning algorithms can learn to label unlabelled data.
* **Unsupervised:** Here, the machine learning algorithm studies data to identify patterns. There is no answer key or human operator to provide instruction. Instead, the machine determines the correlations and relationships by analysing available data. In an unsupervised learning process, the machine learning algorithm is left to interpret large data sets and address that data accordingly.

Under the umbrella of unsupervised learning, fall

1. **Clustering**: Clustering involves grouping sets of similar data (based on defined criteria). It’s useful for segmenting data into several groups and performing analysis on each data set to find patterns.
2. **Dimension reduction**: Dimension reduction reduces the number of variables being considered to find the exact information required.

* **Reinforcement:** Reinforcement learning focuses on regimented learning processes, where a machine learning algorithm is provided with a set of actions, parameters and end values. By defining the rules, the machine learning algorithm then tries to explore different options and possibilities, monitoring and evaluating each result to determine which one is optimal. Reinforcement learning teaches the machine trial and error. It learns from past experiences and begins to adapt its approach in response to the situation to achieve the best possible result.

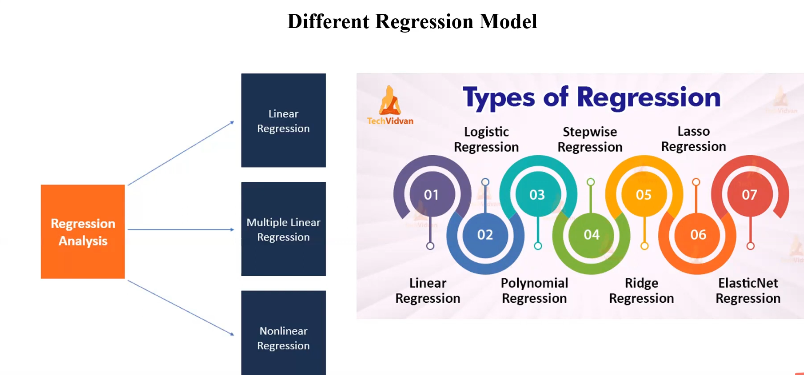
Today, We’re going to discuss a supervised learning **regression** model called **“LINEAR REGRESSION”**.

****

**Linear Regression (Supervised Learning/Regression)**

**Regression:** In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables.

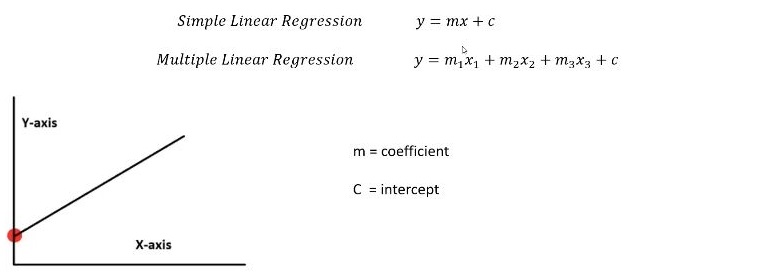
There are more types of regression analysis than those listed here, but these five are probably the most commonly used. Make sure you pick the right one, and it can unlock the full potential of your data, setting you on the path to greater insights.



* **1. Linear regression:** One of the most basic types of regression in machine learning, linear regression comprises a predictor variable and a dependent variable related to each other in a linear fashion. Linear regression involves the use of a best fit line, as described above.  
  You should use linear regression when your variables are related linearly. For example, if you are forecasting the effect of increased advertising spend on sales. However, this analysis is susceptible to outliers, so it should not be used to analyze big data sets.
* **2. Logistic regression:** Does your dependent variable have a discrete value? In other words, can it only have one of two values (either 0 or 1, true or false, black or white, spam or not spam, and so on)? In that case, you might want to use logistic regression to analyze your data.  
  Logistic regression uses a sigmoid curve to show the relationship between the target and independent variables. However, caution should be exercised: logistic regression works best with large data sets that have an almost equal occurrence of values in target variables. The dataset should not contain a high correlation between independent variables (a phenomenon known as multicollinearity), as this will create a problem when ranking the variables.
* **3. Ridge regression:** If, however, you do have a high correlation between independent variables, ridge regression is a more suitable tool. It is known as a regularization technique, and is used to reduce the complexity of the model. It introduces a small amount of bias (known as the ‘ridge regression penalty’) which, using a bias matrix, makes the model less susceptible to overfitting.
* **4. Lasso regression:** Like ridge regression, lasso regression is another regularization technique that reduces the model’s complexity. It does so by prohibiting the absolute size of the regression coefficient. This causes the coefficient value to become closer to zero, which does not happen with ridge regression.  
  The advantage? It can use feature selection, letting you select a set of features from the dataset to build the model. By only using the required features – and setting the rest as zero – lasso regression avoids overfitting.
* **5. Polynomial regression:** Polynomial regression models a non-linear dataset using a linear model. It is the equivalent of making a square peg fit into a round hole. It works in a similar way to multiple linear regression (which is just linear regression but with multiple independent variables), but uses a non-linear curve. It is used when data points are present in a non-linear fashion.  
  The model transforms these data points into polynomial features of a given degree, and models them using a linear model. This involves best fitting them using a polynomial line, which is curved, rather than the straight line seen in linear regression. However, this model can be prone to overfitting, so you are advised to analyze the curve towards the end to avoid odd-looking results.

Linear regression is the most basic type of regression. Simple linear regression allows us to understand the relationships between two continuous variables.

It does this by finding the best fit linear line between the independent and dependent variable.



**Practical Application:** <https://colab.research.google.com/drive/1M-hTg5elSNvIxZg2hsFXiVFkAfcDtK7l#scrollTo=H56FiIAJFXzS>