SNS COLLEGE OF ENGINEERING
Kurumbapalayam(Po), Coimbatore – 641 107
(An autonomous institution)
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Mrs. N.Padmashri/Assistant Professor

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATASCIENCE

What is Regression?

What is regression?

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Regression is the process of predicting a continuous value

Can we Predict?

- We can use the regression methods to predict a continuous value such as co2 emission
- In Regression there are two types of variables:
  - ✓A Dependent variable and
  - ✓One or more Independent variable
- Dependent variables are state, target and try to predict
- Independent variable are also known as explanatory variables can be the causes of those states

X: Independent variable

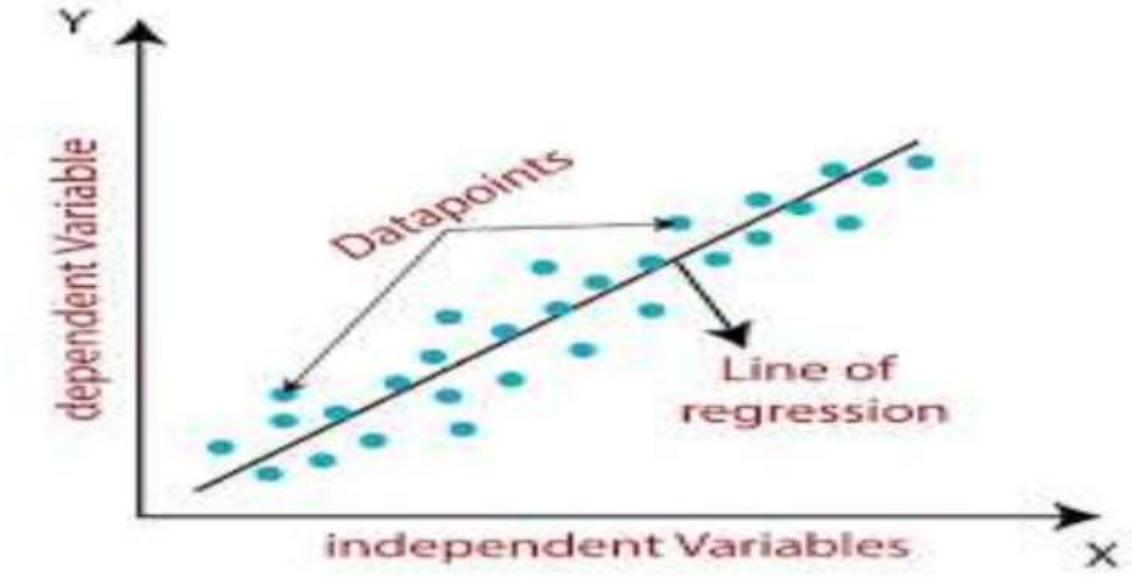
Y: Dependent variable

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.6	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Continuous Values

Regression is the process of predicting a continuous value

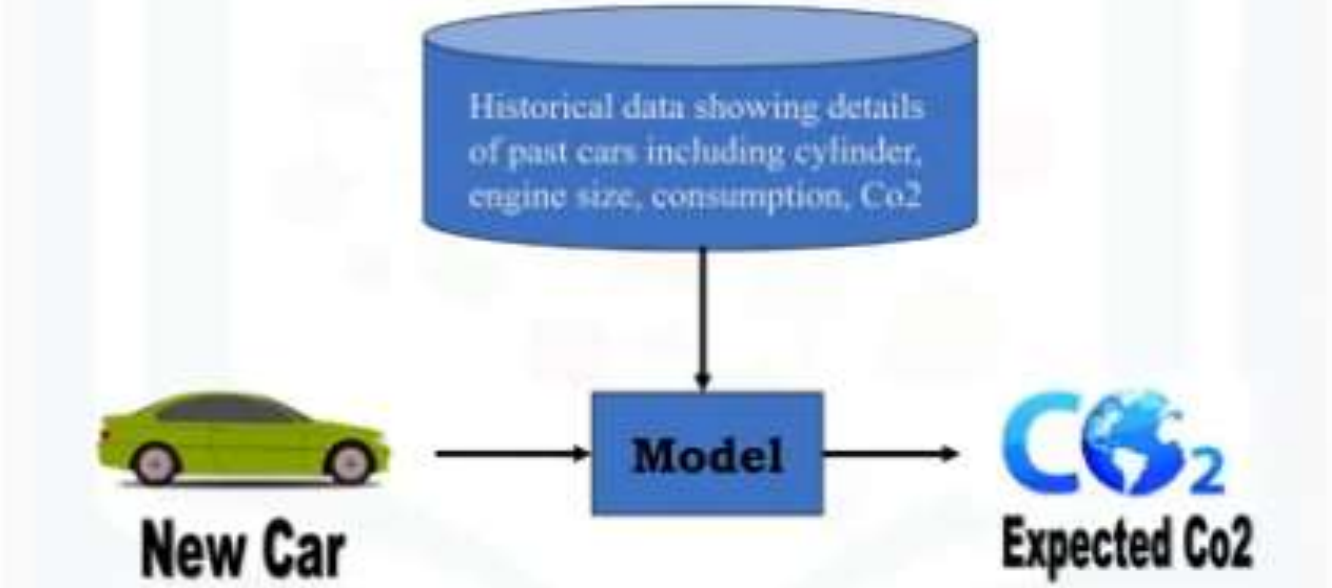
Line of Regression



- **Dependent value should be continuous** and cannot be a discrete value.
- The independent variable or variable can be measured on either a categorical, **or continuous measurement scale.**

REGRESSION MODEL

What is a regression model?



1.We use regression to build such a regression estimation model,
2. Then the model is used to predict the expected CO_2 emission for a new, or unknown car.

REGRESSION TYPES

Types of regression models

- Simple Regression:
  - Simple Linear Regression
  - Simple Non-linear Regression

  Predict co2emission vs EngineSize of all cars
- Multiple Regression:
  - Multiple Linear Regression
  - Multiple Non-linear Regression

  Predict co2emission vs EngineSize and Cylinders of all cars

- Simple regression is when one independent variable is used to estimate a dependent variable.
- It can be either linear, or non-linear.
- For example, predicting CO_2 emission using the variable of engine size.
- When more than one independent variable is present the processes is called multiple linear regression.
- For example, predicting CO_2 emission using engine size and the number of cylinders in any given car.
- Again, depending on the relation between dependent and independent variables it can be either linear or nonlinear regression.

Linear regression

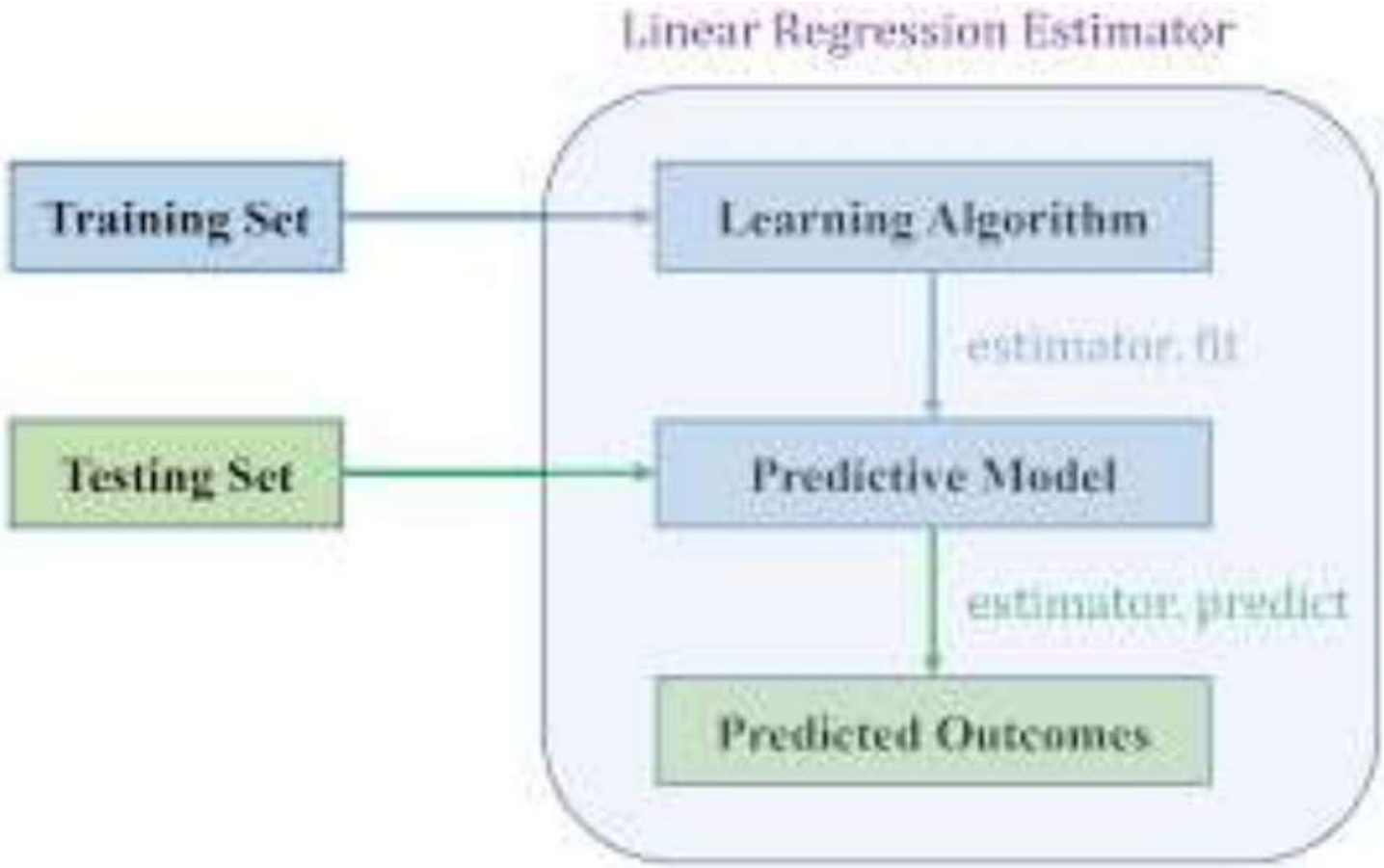
- Linear regression algorithm shows a linear relationship between a dependent (Y) and one or more independent (X) variables, hence called as linear regression.

- Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

Linear models for Regression

- Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task.
- Regression models are a target prediction value based on independent variables.
- Linear regression is one of the easiest and most popular Machine Learning algorithms.
- It is a statistical method that is used for predictive analysis.
- Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc

Linear Regression Estimator



Examples

Applications of regression

- Sales forecasting
- Satisfaction analysis
- Price estimation
- Employment income

Algorithms

Regression algorithms

- Ordinal regression
- Poisson regression
- Fast forest quantile regression
- Linear, Polynomial, Lasso, Stepwise, Ridge regression
- Bayesian linear regression
- Neural network regression
- Decision forest regression
- Boosted decision tree regression
- KNN (K-nearest neighbors)

- Mathematically, we can represent a linear regression as:
  - $y= a0+a1 x+ \varepsilon$ Here,
- Y= Dependent Variable (Target Variable)
- X= Independent Variable (predictor Variable)
- a0= intercept of the line (Gives an additional degree of freedom)
- a1 = Linear regression coefficient (scale factor to each input value).
- ε = random error
- The values for x and y variables are training datasets for Linear Regression model representation

DEFINITIONS

- Simple Linear Regression: If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

- Multiple Linear regression: If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression

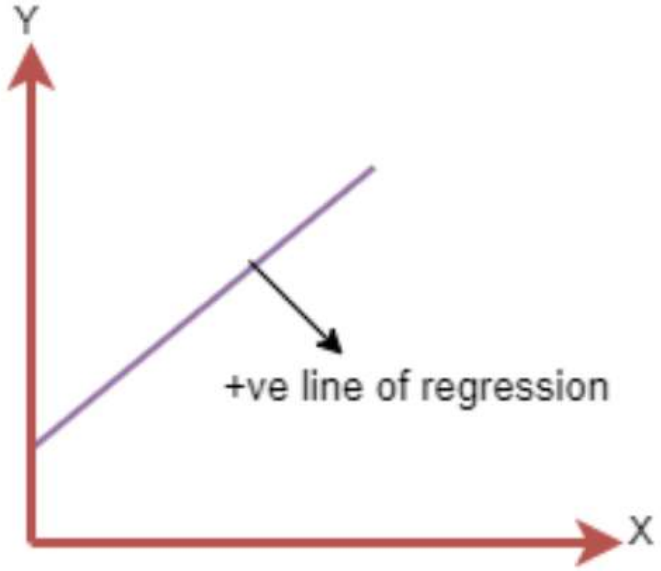
Linear Regression Line

- Linear Regression Line A linear line showing the relationship between the dependent and independent variables is called a regression line.
- A regression line can show two types of relationship
  - Positive Linear Relationship
  - Negative Linear Relationship
- Positive Linear Relationship:

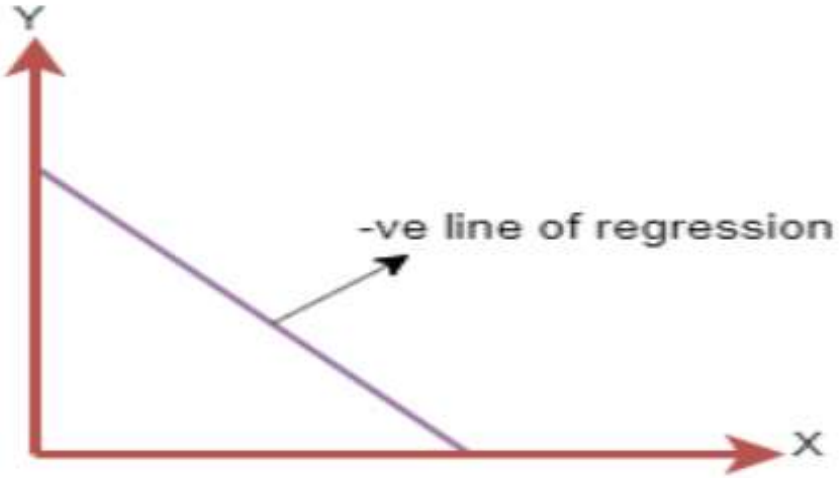
If the dependent variable increases on the Y-axis and independent variable increases on X-axis, then such a relationship is termed as a Positive linear relationship

- Negative Linear Relationship:

If the dependent variable decreases on the Y-axis and independent variable increases on the X-axis, then such a relationship is called a negative linear relationship.



The line equation will be: $Y= a_0+a_1x$



The line of equation will be: $Y= -a_0+a_1x$