



# SNS COLLEGE OF TECHNOLOGY

Coimbatore-37.

An Autonomous Institution



**COURSE NAME : 19CSE301-INTRODUCTION TO DATA SCIENCE**

**III YEAR/ V SEMESTER**

**UNIT – IV CLASSIFICATION**

**Topic: Regression on Probabilities**

Mrs.G.Swathi

Assistant Professor

Department of Computer Science and Engineering



# Introduction

- Regression is defined as a statistical method that helps us to analyze and understand the relationship between two or more variables of interest.
- In regression, we normally have one dependent variable and one or more independent variables.
- we try to “regress” the value of the dependent variable “Y” with the help of the independent variables. In other words, we are trying to understand, how the value of ‘Y’ changes w.r.t change in ‘X’.



# Types of variable

- **Dependent Variable:** This is the variable that we are trying to understand or forecast.
- **Independent Variable:** These are factors that influence the analysis or target variable and provide us with information regarding the relationship of the variables with the target variable.



# Types of Regression

- Linear Regression
- Polynomial Regression
- Logistic Regression

## Linear Regression

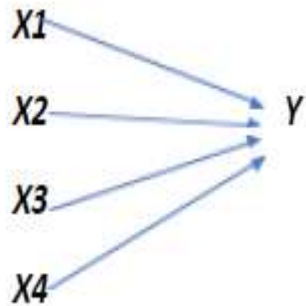
Linear Regression is a predictive model used for finding the *linear* relationship between a dependent variable and one or more independent variables.



# Simple & Multiple Linear Regression

$X \longrightarrow Y$

- If the relationship between Independent and dependent variables is multiple in number, then it is called Multiple Linear Regression





# Types of Regression

As the model is used to predict the dependent variable, the relationship between the variables can be written in the below format.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Where,

$Y_i$  – Dependent variable

$\beta_0$  -- Intercept

$\beta_1$  – Slope Coefficient

$X_i$  – Independent Variable

$\epsilon_i$  – Random Error Term

$$\begin{array}{ccccc} SST & = & SSR & + & SSE \\ \text{Total Sum of} & & \text{Regression Sum} & & \text{Error Sum} \\ \text{Squares} & & \text{of Squares} & & \text{of Squares} \end{array}$$

The main factor that is considered as part of Regression analysis is understanding the variance between the variables. For understanding the variance, we need to understand the measures of variation.



# Assumption

- Linear Relationship
- Normality
- No or Little Multicollinearity
- No Autocorrelation in errors
- Homoscedasticity

## Coefficient of determination $r^2$

The coefficient of determination is the portion of the total variation in the dependent variable that is explained by variation in the independent variable. A higher value of  $r^2$  better is than the model with the independent variables being considered for the model.

$r^2 = SSR$  Note: The value of  $r^2$  is the range of  $0 \leq r^2 \leq 1$



# References

- 1 Tom M. Mitchell, “Machine Learning”, McGraw-Hill Education (India) Private Limited, 2013.
- 2 Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, “An Introduction to Statistical Learning: with Applications in R”, Springer; First Edition 2013.
- 3 P. Flach, —Machine Learning: The art and science of algorithms that make sense of data, Cambridge University Press, 2012.





Thank  
You