Department of AI&DS

MACHINE LEARNING
22AD2203R

Topic:

LINEAR MODELS

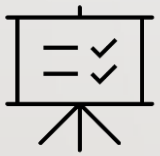
Session - 06

Dr. NARENDRA BABU TATINI
Associate Professor
Department of IoT

AIM OF THE SESSION

To know students about the Linear and types of Linear techniques.

INSTRUCTIONAL OBJECTIVES



This session is designed to:

1. Understand the Linear Models.
2. Identify the types of Linear Models.

LEARNING OUTCOMES

At the end of this session, you should be able to:

1. Define Linear Model, and
2. Describe the Linear Model Techniques.

TOPICS

- Linear Regression
- Logistic Regression
- Support Vector Machine

INTRODUCTION LINEAR MODEL

- **Linear Models** :Machine learning, a linear model refers to a model that assumes a linear relationship between the input features and the target variable.
- The key characteristic of a linear model is that it represents the relationship between the independent variables (input features) and the dependent variable (target) as a linear combination of the features, possibly with the addition of a constant term .
- The general form of a linear model can be represented as
- $$y=\beta 0+\beta 1x1+\beta 2x2+\ldots+\beta nxn$$

TYPES OF LINEAR MODEL

1.Linear Regression: Used for predicting a continuous target variable, it fits a linear relationship between the input features and the target variable.

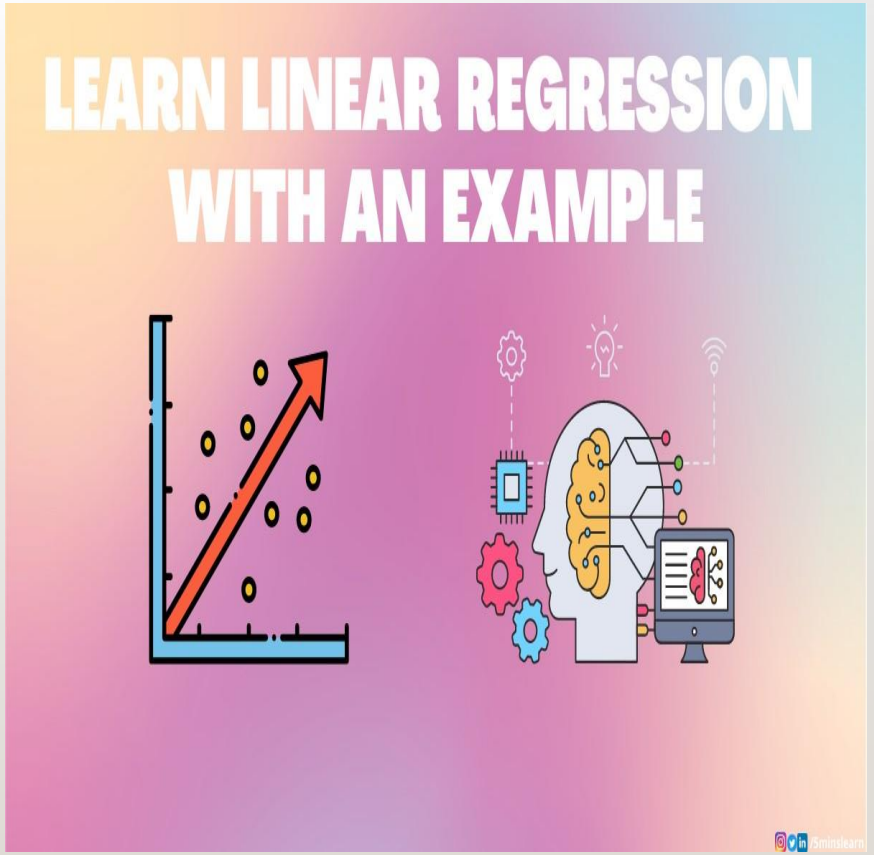
2.Logistic Regression: Employed for binary classification tasks, it models the probability of an instance belonging to a particular class given the input features.

3.Linear Support Vector Machines (SVM): Used for both regression and classification tasks, SVMs attempt to find the optimal hyperplane that separates data points of different classes or predicts the target variable based on input features.

1.LINEAR REGRESSION

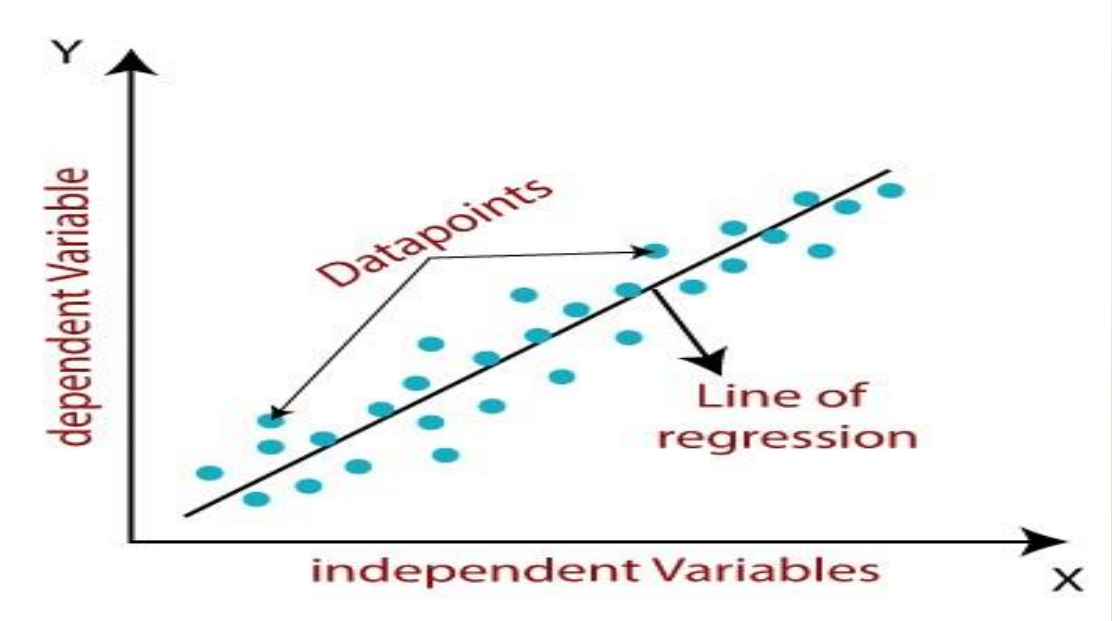
- Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as **sales, salary, age, product price,** etc.

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



LINEAR REGRESSION

- The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:



LINEAR REGRESSION

Types of Linear Regression

Linear regression can be further divided into two types of the algorithm:

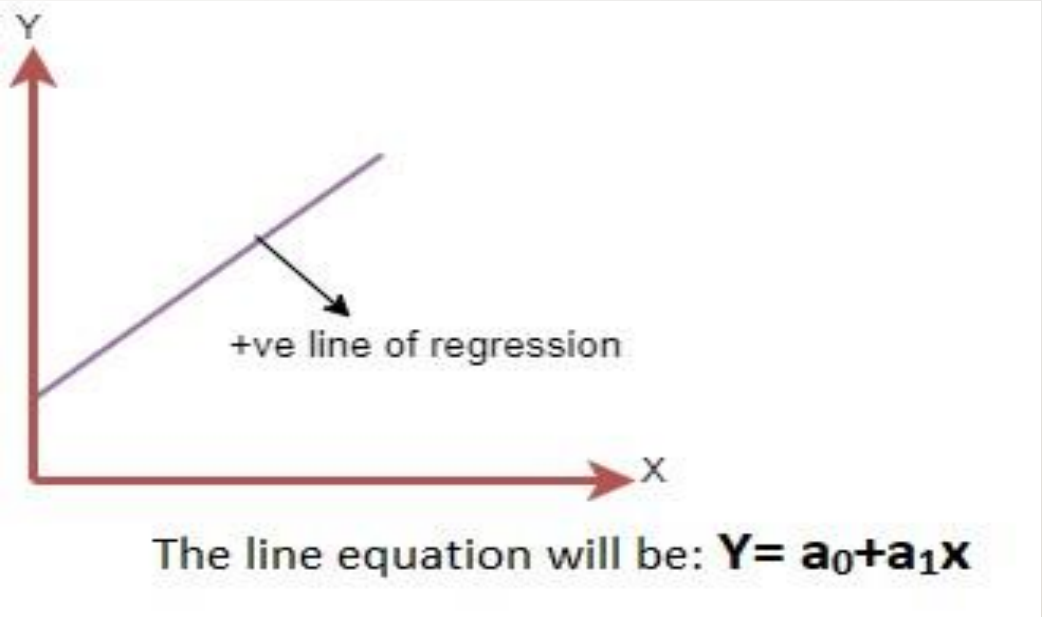
- **Simple Linear Regression:**
  If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.
- **Multiple Linear regression:**
  If more than one independent variable is used to predict the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

LINEAR REGRESSION

- **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:

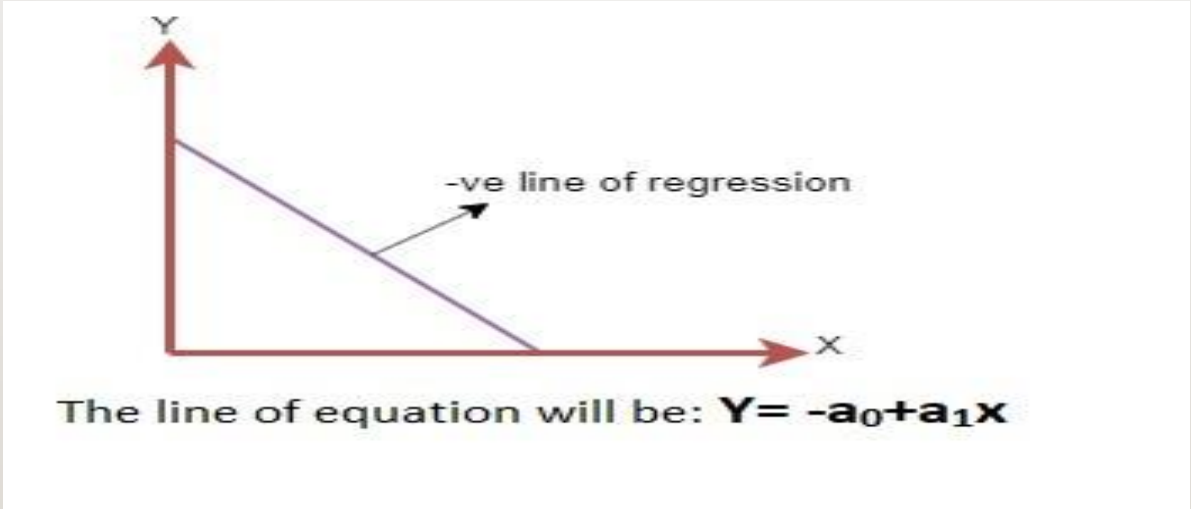
Positive Linear Relationship:

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.



LINEAR REGRESSION

- **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.



2.LOGISTIC REGRESSION

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, **it gives the probabilistic values which lie between 0 and 1**.
- Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas **Logistic regression is used for solving the classification problems.**

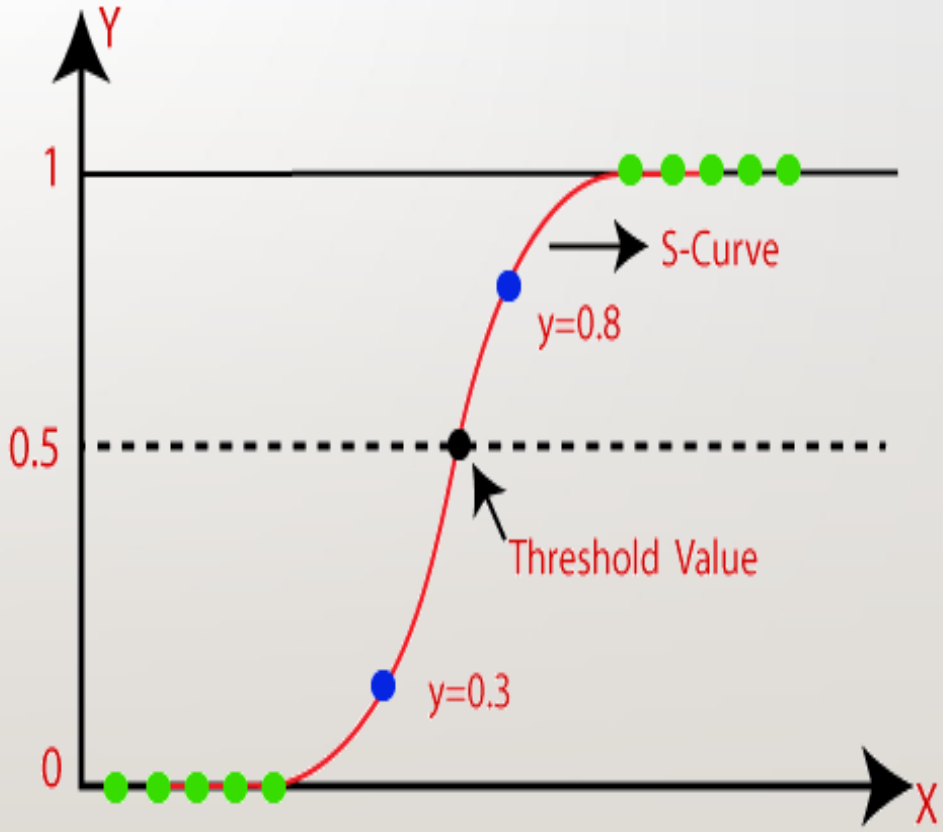
LOGISTIC REGRESSION

- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).
- The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification.

LOGISTIC REGRESSION

The below image is showing the logistic function:

- **Note:** Logistic regression uses the concept of predictive modeling as regression.
- Therefore, it is called logistic regression.
- But is used to classify samples; Therefore, it falls under the classification algorithm.



LOGISTIC REGRESSION

Logistic Function (Sigmoid Function):

- The sigmoid function is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1.
- The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

LOGISTIC REGRESSION

Assumptions for Logistic Regression:

- The dependent variable must be categorical in nature.
- The independent variable should not have multi-collinearity.

Logistic Regression Equation:

- The Logistic regression equation can be obtained from the Linear Regression equation. The mathematical steps to get Logistic Regression equations are given below:
- We know the equation of the straight line can be written as:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

LOGISTIC REGRESSION

Type of Logistic Regression:

- On the basis of the categories, Logistic Regression can be classified into three types:
- **Binomial:** In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.
- **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"
- **Ordinal:** In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

3.LINEAR SUPPORT VECTOR MACHINES (SVM)

- Support Vector Machines (SVM) are a type of supervised learning algorithm that can be used for classification, regression, or outlier detection.
- In the context of classification, SVMs are particularly useful for binary classification tasks, where the goal is to separate data points into two classes by finding the best possible boundary (hyperplane) that maximizes the margin between the two classes.
- Linear Support Vector Machines (SVM) specifically work with linearly separable data. They aim to find the best hyperplane that separates the data points of one class from those of another class in an N-dimensional space.
- This hyperplane is selected to have the maximum distance to the nearest data point of any class, also known as the margin.

LINEAR SUPPORT VECTOR MACHINES (SVM)

- The basic idea behind SVMs is to transform the input data into a higher-dimensional feature space, where the data points can be separated by a hyperplane.
- However, in the case of linear SVMs, the feature space transformation is linear, which means the hyperplane is a linear combination of the input features.
- The optimization problem in linear SVM involves finding the hyperplane that maximizes the margin while minimizing the classification error.
- This is often formulated as a constrained optimization problem, with the objective of maximizing the margin subject to the constraint that all data points are correctly classified.

LINEAR SUPPORT VECTOR MACHINES (SVM)

Mathematically, the decision function for a linear SVM can be expressed as:

- $f(x) = w^T * x + b$

where:

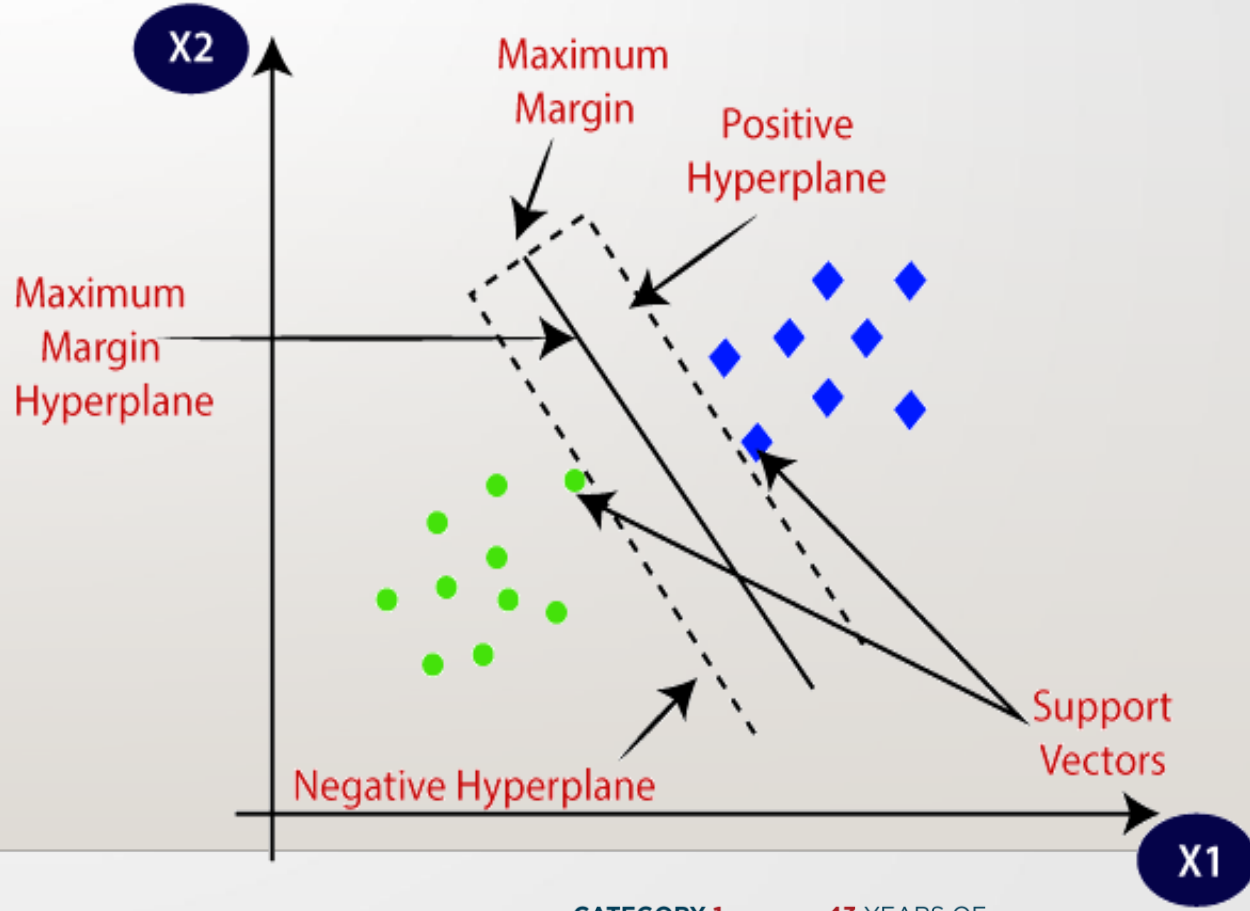
- f(x) is the decision function that predicts the class label of a data point x.
- w is the weight vector perpendicular to the hyperplane.
- b is the bias term.
- The parameters w and b are determined during the training phase of the SVM algorithm, using techniques such as gradient descent or convex optimization methods

LINEAR SUPPORT VECTOR MACHINES (SVM)

- One of the key advantages of linear SVMs is their ability to handle high-dimensional data efficiently and effectively.
- However, a limitation of linear SVMs is that they can only separate data points that are linearly separable.
- which means they may not perform well on datasets that are not linearly separable without the use of more complex kernels or other techniques.

LINEAR SUPPORT VECTOR MACHINES (SVM)

- Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms.
- which is used for Classification as well as Regression problems.
- However, primarily, it is used for Classification problems in Machine Learning.



Self-Assessment Questions

1. What is the primary assumption of a linear model?

a) Homoscedasticity
b) Heteroscedasticity
c) Multicollinearity
d) Autocorrelation

2. Which of the following is a method for dealing with multicollinearity in linear models?

a) Lasso regression
b) Ridge regression
c) Principal Component Analysis (PCA)
d) All of the above

Self-Assessment Questions

3. What is the purpose of the R-squared value in linear regression?

a) To measure the goodness of fit of the model

b) To detect multicollinearity

c) To identify outliers in the data

d) To assess the normality of residuals

4. In linear regression, what does the p-value of a coefficient indicate?

a) The strength of the relationship between the dependent and independent variables

b) The significance of the coefficient in the model

c) The variance inflation factor

d) The correlation between the variables

REFERENCES FOR FURTHER LEARNING OF THE SESSION

Text Books:

1. Mitchell, Tom. Machine Learning. New York, NY: McGraw-Hill, 1997. ISBN: 9780070428072.
2. MacKay, David. Information Theory, Inference, and Learning Algorithms. Cambridge, UK: Cambridge University Press, 2003. ISBN: 9780521642989.

Reference Books:

1. EthemAlpaydin "Introduction to Machine Learning ", The MIT Press (2010).
2. Stephen Marsland, "Machine Learning an Algorithmic Perspective" CRC Press, (2009).

Sites and Web links:

1. Data Science and Machine Learning: https://www.edx.org/course/data-science-machinelearning.
2. Machine Learning: https://www.ocw.mit.edu/courses/6-867-machine-learning-fall-2006/.

THANK YOU

Team – MACHINE LEARNING USING PYTHON