SNS COLLEGE OF TECHNOLOGY

Coimbatore-37.
An Autonomous Institution

COURSE NAME : 19CSE301-INTRODUCTION TO DATA SCIENCE

III YEAR/ V SEMESTER

UNIT – IV CLASSIFICATION

Topic: Logistic Model

Mrs.B.Sumathi
Assistant Professor
Department of Computer Science and Engineering

Introduction

- Logistic Regression is also known as Logit, Maximum-Entropy classifier is a supervised learning method for classification.
- It establishes a relation between dependent class variables and independent variables using regression.
- The dependent variable is categorical i.e. it can take only integral values representing different classes.
- The probabilities describing the possible outcomes of a query point are modeled using a logistic function.
- This model is used when we have 2 classes of dependent variables. When there are more than 2 classes, then we have another regression method which helps us to predict the target variable better

Types

- Binary Logistic Regression when the dependent variable is strictly binary
- Multinomial Logistic Regression is when the dependent variable has multiple categories.

There are two types of Multinomial Logistic Regression

- Ordered Multinomial Logistic Regression (dependent variable has ordered values)
- Nominal Multinomial Logistic Regression (dependent variable has unordered categories)

Linear Discriminant Analysis(LDA)

- Discriminant Analysis is used for classifying observations into a class or category based on predictor (independent) variables of the data.

- Discriminant Analysis creates a model to predict future observations where the classes are known.

LDA comes to our rescue in situations when logistic regression is unstable when

1. Classed are well separated

2. Data is small

3. When we have more than 2 classes

Working Process

- The LDA model uses Bayes' Theorem to estimate probabilities.
- They make predictions upon the probability that a new input dataset belongs to each class.
- The class which has the highest probability is considered as the output class and then the LDA makes a prediction.

Polynomial Regression

- Polynomial Regression is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial.

$$y= b0+b1x1+ b2x12+ b2x13+..... bnx1n$$

"In Polynomial regression, the original features are converted into Polynomial features of required degree (2,3,..,n) and then modeled using a linear model."

try to cover it with a linear model, then we can clearly see that it hardly covers any data point. On the other hand, a curve is suitable to cover most of the data points,

Equation and steps

Simple Linear Regression equation: $y = b_0+b_1x$ .....(a)

Multiple Linear Regression equation: $y= b_0+b_1x+ b_2x_2+ b_3x_3+....+ b_nx_n$ .....(b)

Polynomial Regression equation: $y= b_0+b_1x + b_2x^2+ b_3x^3+....+ b_nx^n$ .....(c)

Steps for Polynomial Regression:

- The main steps involved in Polynomial Regression are given below:
  - Data Pre-processing
  - Build a Linear Regression model and fit it to the dataset
  - Build a Polynomial Regression model and fit it to the dataset
  - Visualize the result for Linear Regression and Polynomial Regression model.
  - Predicting the output.

Types and Sigmoid function

A quadratic equation is a general term for a second-degree polynomial equation. This degree, on the other hand, can go up to nth values. Polynomial regression can so be categorized as follows:

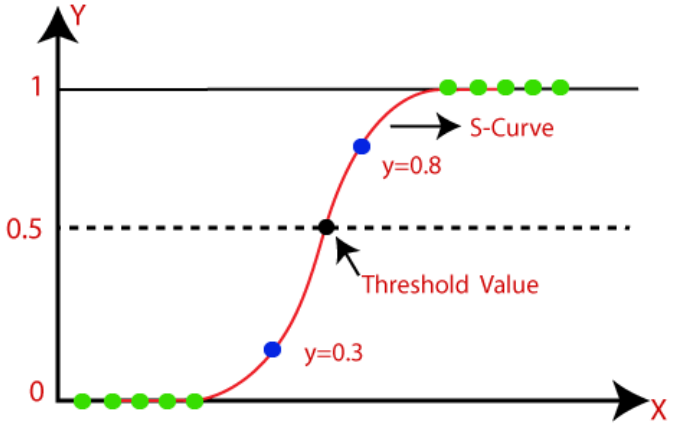
- 1. Linear – if degree as 1
- 2. Quadratic – if degree as 2
- 3. Cubic – if degree as 3 and goes on, on the basis of degree.

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.

- 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**.

Sigmoid function

In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).



Logistic Function(sigmoid)

- 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 Function 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:

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

In Logistic Regression y can be between 0 and 1 only, so for this let's divide the above equation by (1-y):

$\frac{y}{1-y}$; 0 for y= 0, and infinity for y=1

But we need range between -[infinity] to +[infinity], then take logarithm of the equation it will become:

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

Types 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".

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