INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 05 ISSUE: 07 | JULY - 2021 ISSN: 2582-3930

Linear Regression Vs Logistic Regression

MOHAMMED ILIYAS [1], Dr. VIBHA M B [2]
[1] PG Scholar, Department of Master of Computer Applications Dayananda Sagar College of Engineering, Bangalore, India Visvesvaraya Technological University, Karnataka, India

[2] Assistant Professor, Department of Master of Computer Applications Dayananda Sagar College of Engineering, Bangalore, India

ABSTRACT

Machine Learning is used for prediction and machine learning is the part of Artificial Intelligence. There is a lot of algorithms based on the types of Machine Learning. The two widely being used machine learning algorithms which is Linear regression and Logistic regression. Keywords—Linear Regression, Logistic regression, best fit line, use-case

INTRODUCTION

The two widely being used machine learning algorithms which is Linear regression and Logistic regression. Now even though the linear regression and logistic regression are the most widely used algorithms there is still a lot of confusion between them and today will compare the two algorithms and see how they work let's take a look at the agenda is:

1: Various Types of Machine Learning
2: Regression vs Classification
3: What is Linear Regression
4: What is Logistic Regression
5: Linear Regression Use case
6: Logistic Regression Use case
7: Linear Vs Logistic Regression

1. Types of Machine Learning

Machine Learning Definition: Machine Learning is used for prediction and for prediction we need to train a model and to train a model we need lot of data.

Supervised Learning:

Supervised learning model is trained using data which is already tagged with a correct answer it can be further grouped into regression and classification, as input data is feeds into the models it adjusts accordingly through a reinforcement learning process, and can solve a large variety of real-world problems.

Unsupervised Learning:

Unsupervised Learning model is trained using data that is neither classified not labelled and allowing the model to act on that information without any guidance. Clustering, Self-Supervised and auto encoding are all Unsupervised Learning method.

Reinforcement Learning:

Reinforcement Learning is one of the types of machine learning which is being growing from past few years. In this an agent learns to act in this environment by certain actions and either it gets rewards on the action or it gets a punishment and observing the reward which gets from those actions.

Figure 1: Types of Machine Learning

2. Regression Vs Classification

Here we will focus on Supervised Learning as Linear and Logistic Regression are actually Supervised Learning Algorithm.

Classification

It is a technique which is used to arrive at an illustrative that shows the data starting for organization with a outrider or messenger variable.

Regression

It is used to predict continuous values. There are three methods for evaluating the regression: Variance, Bias, error.

Classification	Regression
• Classification is about predicting label, or is the task of predicting discrete class label. • The dependent variables are known as categorical and unordered.	• Regression is about predicting quantity, or is the task of predicting continuous quality. • The dependent variables are known as continuous or numerical.

3. Linear Regression

We can use the linear regression to predict some continuous quality.

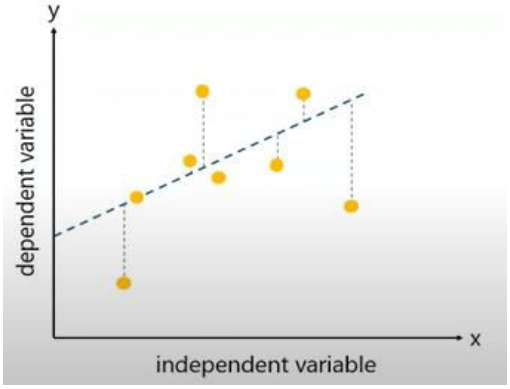

Figure 2: Linear regression best fit line

Training and Testing

1. As we know Machine Learning is used for prediction and for prediction, we need to train a model and to train a model we need lot of data. In linear regression there is only one target feature. There are four things to train a model and that is Train-input, Train-output, Test-input and Test-output.
2. It uses Train-input and train-output to create train model, we use 25%-70% of the data to create a train model, after creation of model we give Text-input to predict the Test-output.

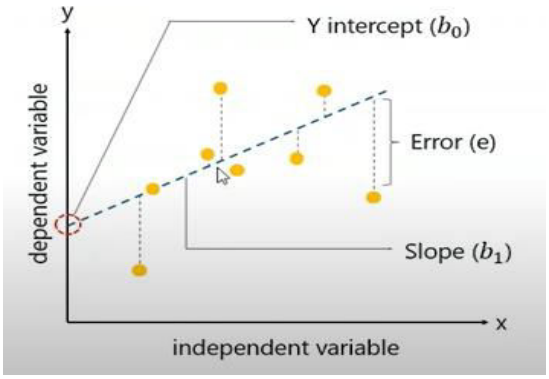

Figure 3(a): Linear Regression equation Explanation

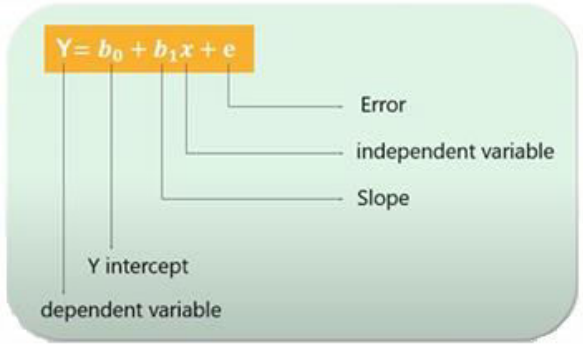

Figure 3(b): Linear Regression equation Explanation

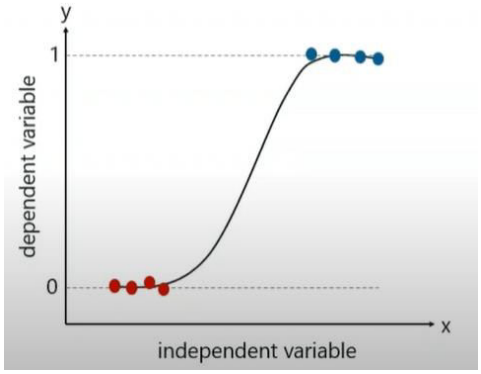

Figure 4: Logistic Regression Sigmoid Curve

Where, b0(beta0) term which we're going to be calling the intercept or the constant.
b1(beta1) which is the term that multiplies to X and we going to be calling the coefficient of x or the slope of x
e is going to be our error term which is going to minimize.
And we call these linear equations because they will appear as a straight line if we plot them in the dimensional plot.

4. Logistic Regression

In logistic regression the dependent variable is categorical and when we discuss liner regression the dependent variable is always continues so this is one major difference between logistic and linear regression, now what do we mean by categorial variable, so when we see categorical variable, that the variable can hold values like 1 or 0, Yes or No and so on. so basically a binary type of variable in logistic regression the outcome is always categorical so when the resulting outcome has only two possible values it is obvious desirable to have a model that predicts the value as either 0 or 1or is probabilities or the ranges between 0 and 1 liner regression does not has this capability because if we use linear regression to model a binary response variable the resulting model will not predict the values in the range 0 and 1 because linear regression works on continues dependent variables and not on categorical variables. We can't expect categorical results in linear regression model because we feeding it only continues variables.

Let's understand from the above diagram, consider the graph it's not a linear line, instead it is a curve, now this curve basically called as S curve or sigmoid curve. now we know that we mention logistic regression only calculates the outcome as 0 or 1. So why do we have a curve from 0 to 1? It should be a straight line at 1 and a straight line at 0 because obviously its outputting values as only 0 and 1. Now this is because logistic regression calculates the probability. Let say that the final probability of an event occurring is 0.6 what does it mean does my output belong to class no 1 or does belon to class no 0. When such situation occurs, we set a threshold value, let say that my threshold value is 0.5 this means that any value in the range of o.5 to 1 is classified as 1 and any value below 0.5 is considered as 0. This way u can easily get a binary result in variable which is exactly what we need when it comes to logistic regression.

$$\log\left(\frac{Y}{1-Y}\right) = C + B1X1 + B2X2 + \ldots.$$

Now let's try to understand the equation, 'y' over here ranges between 0 and 1, represents the probability of an event to happen so the y value can be set 0 and 1 it can b 0.2, 0.3, 0.9 and all, that's when you set a threshold value so that you can categories the output is either 0 or 1.

Next, we have X1 X2 X3 these are basically the independent variables that we use to calculate the dependent variable now see obviously here is a numerical const and b1 and b2 basically represent the respecting coefficients 0f X1, X2 and so on
C is the constant term which will be the probability of the event happening no other factors are considered.

5. Linear Regression use case

So, we need to understand the relationship between the monthly sales and the cost of the online Advertisement for some xyz company in order to predict this sale in the upcoming months, so basically, we are going to predict sales

based on the amount that they spending on advertising so our dependent variable becomes number of sales and the independent variable is obesely the advertising cost. Here we have represented the advertising cost in terms of 1000$ so with-it data provided in this table.

Monthly sales	Advertising cost In 1000s
200	0.5
900	5
450	1.9
680	3.2
490	2.0
300	1.0

Figure 5: Dataset of monthly sales and the online advertising cost

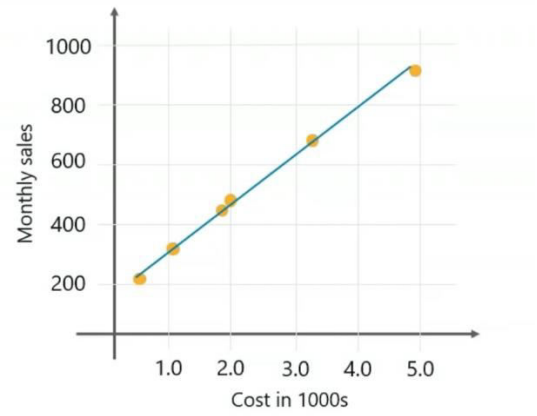

Figure 6: Best fit line of monthly sales and the online advertising cost

Linear regression has a lot of applications in the business setup it is use to forecast sales to perform risk analysis measure profit probability and so on.

6. Logistic Regression use case

Our problem statement here is to predict if any student get enrolled to a school based upon the CGPA. So, consider the dataset it contains two variables the CGPA and the Admission variable.

Admission	CGPA
0	4.2
0	5.1
0	5.5
1	8.2
1	9.0
1	9.1

Figure 7: Dataset of students in CGPA

X indicates CGPA and y indicates weather the student will get admission in collage or not. So, 1 here will mean that you got an admission and 0 means no admit. Now here the logistic regression will help asses a what level of CGPA lead to admission in a college so the logistic model train in such a way that students with the CGPA of 6 and above get admission hence y=1 but if the CGPA is below 6 they don't get an admission that's when y=0. So, the students classified in group y=1 will get an admission where as a student in group y=0 will not get an admission.

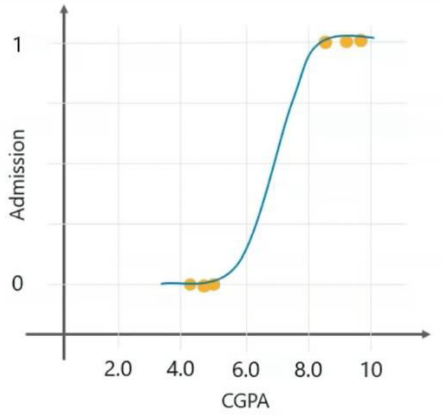

Figure 8: Sigmoid curve of student's admission

Logistic regression is use to solve 'n' number of complex classification problems it is use in the cyber security domain to detect fraudulent customer it is use in image processing classification and the lot other classification problems.

7. Linear Regression Vs Logistic Regression

	L	L o g

		istic Regression
1	I	It is used to nodel a binary d

		ependent variable.

	continuous.	original dependent variable to predict categorica
2	The dependent is always	It is going to use categ

		value.
3	It follows the least square estimation	It follows the exactly opposite of Lin

	, which indicates that it should minimize the	ear Regression. Here, It maximize the probabili

		ty of Y given one value of X.
		square distance of each observed response.

4	Y	log
5		It maps the relationship betw

		en the scalar response and explanatory variabl
		es. it will get a curve which is also known as

		significd curve.
6	L	It's not necessary to have liner relationship.

7	N	The out

put has to be a binary value it should be class

And class B or it should be 0 or 1.

8	L	It's mainly used in classificati
		on when we want to classify dataset into two di

		ffeerreenntt ccllaasssseess tthhaann wwee uussee llooggiissttiicc rreeggrreessssiioonn
		iinn ccyybbeerr sseeccuurriittyy, iinnaaggee pprroocceessssiinngg aanndd ccllaassssi

		fication domain.

CONCLUSION

Linear regression the dependent is always continuous. Linear regression aims at finding the fitting straight line which is also called the regression line. the output of linear regression is always going to be a predicted integer value or basically a continues value. Linear regression is use to predict some continues value.

The Logistic regression used for classification problems. Logistic regression is based on maximum likely-hood estimation. The output has to be a binary value. It maps the relationship between the scalar response and explanatory variables it will get a sigmoid curve. It's not necessary to have liner relationship. It's mainly used in image processing.

REFERENCES

[1]" What is Machine Learning? -Cognizant Softvision" https://www.cognizantsoftvision.com/blog/what-is-machine-learning/

[2] "ML| Linear Regression – GeeksforGeeks" https://www.geeksforgeeks.org/ml-linear-regression-vs-logistic-regression/

[3] "What is Logistic Regression? – Statistics Solutions " https://www.statisticssolutions.com/what-is-logistic-regression/

[4] "Understanding Logistic Regression - Arun Addagatla" https://medium.com/nerd-for-tech/understanding-logistic-regression-782baa868a54