📝LOGISTIC REGRESSION

* It is supervised learning techniques
* It is used for predicting the categorical dependent variable using a given set of independent variables.
* The outcome must be 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 value which lie between 0 and 1.
* 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 can be created by logistic function or sigmoid function
* It is a significant machine learning algorithm because it can provide probabilities and classify new data using continuous and discrete datasets.
* It uses the concept of predictive modelling as regression; therefore, it is called logistic regression, but is used to classify samples; therefore, it falls under the classification algorithm.

✍️LOGISTIC FUNCTION (SIGMOID FUCNTION)

* It is a mathematical function used to map the predicted values to the probabilities.
* It maps any real value into another value within a range of 0 and 1.
* We use the concept of threshold value, which defined the probabilities of either 0 or 1. Such as values above the threshold values tends to 1, and the value below the threshold values tends to 0.

👉ASSUMPTION FOR LOGISTIC REGRESSION

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

📝TYPES OF LOGISTIC REGRESSION

1. Binomial:

There can be only two possible types of the dependent variables, such as 0 or 1, pass or fail, etc.

1. Multinomial:

There can be more than 2 possible unordered types of the dependent variable, such as “cat”, “dog” or “sheep”.

1. Ordinal:

There can be more than 2 passible ordered typed of the dependent variable, such as “low”, “medium”, “high”.

✍️EXAMPLE:

|  |  |
| --- | --- |
| Hours studies | Result (1=pass, 0=fail) |
| 29 | 0 |
| 15 | 0 |
| 33 | 1 |
| 28 | 1 |
| 39 | 1 |

The model optimized by a method:

Log(odds) = -64 +2\*hours

Q1). How to calculate the probability of pass for the student who studies 33 hours?

Z = -64+2\*hours

= -64 + 2\*33

= -64 +66

= 2

P = 1/(1+ez)

=1/(1+e2)

= 0.88

So, the student who studies 33 hours have the probability of 88% to pass, and 12% to fail.

Q2). At least how many hours the student should study that makes sure will pass the course with probability of more than 95%?

* We use sigmoid function for probability

P = 1/(1+ez)

0.95 = 1/(1+ez)

0.95(1+ez) = 1

0.95 + 0.95(ez) = 1

0.95(ez) = 1-0.95

0.95(ez) = 0.05

ez = 0.05/0.95

ln(ez) = ln(0.05/0.95)

z = ln(0.05/0.95)

z = 2.94

Now, find the hours

z = -64 + 2\*hours

hours = (z+64)/2

hours = (2.94+64)/2

hours = 33.47