

LINEAR REGRESSION Vs LOGISTIC REGRESSION

Linear regression- uses supervised machine learning algorithm to predict continuous values. It assumes a relationship between the dependent and independent variables and finds a plane/line that best fits describing the variables given.

Logistic regression- uses supervised machine learning algorithm that helps in binary classification (classifying tasks in one or two labels) separating discrete values.

Steps of linear regression

Step 1: we have a dataset with dependent variable Y which is a function of x i.e. $Y = f(X)$ where x is an independent variable. The data provided can form an equation of line that best fits the data.

$Y = mx + c$ where m is the slope of the line and c is the intercept of line

Step 2: putting random values of m and c to predict the best fit. Thus, we get an output value Y.

Step 3: in logistic regression, after predicting the value of y, we check whether it's accurate or not by the actual value given (labels already provided because it's supervised learning model). We name Y as \hat{y} .

In linear regression, we calculate the error by mean square error method (MSE), and also call error as loss function where,

$$L = \frac{1}{n} \sum ((y - \hat{y})^2)$$

, where L is the loss function and n is the number of observation and y is the actual value and \hat{y} is the predicted value.

Step 4: to get the best fitted line, the value of loss function should be minimum. To minimize its value, we use a method called Gradient Descent Method.

Gradient Descent Method

To calculate the loss function, the error is represented in second order i.e., have 2 weights m and c (parabolic).

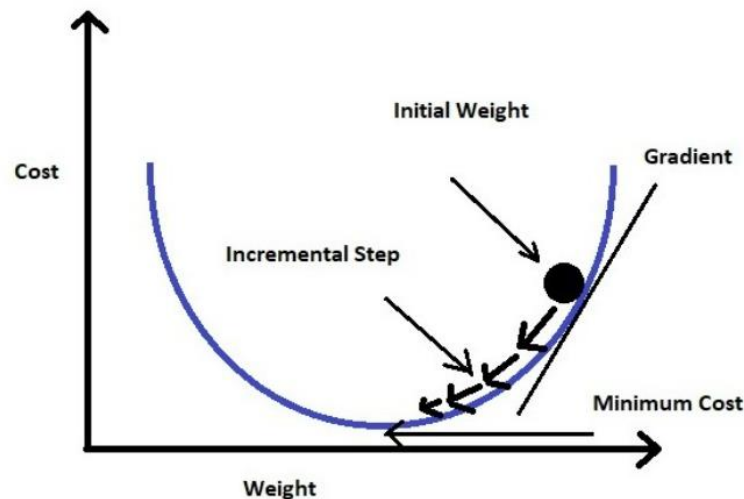


Fig 1: Gradient Descent

To minimize the error, we need to reach the bottom of curve.

1. Calculate the first order derivative of loss function for m and c
2. Subtract the result by initial weights, multiplying by learning rate (α)
3. Keep repeating until you get the minimum value (consider threshold as very small value otherwise it can take forever to reach minimum value as 0).

Step 5: after minimizing the loss function, we get a best fit and thus then we can predict any value of Y for given x .

Logistic regression

In this approach, we classify elements into two groups and calculate probability of each element of the set.

Steps of logistic regression

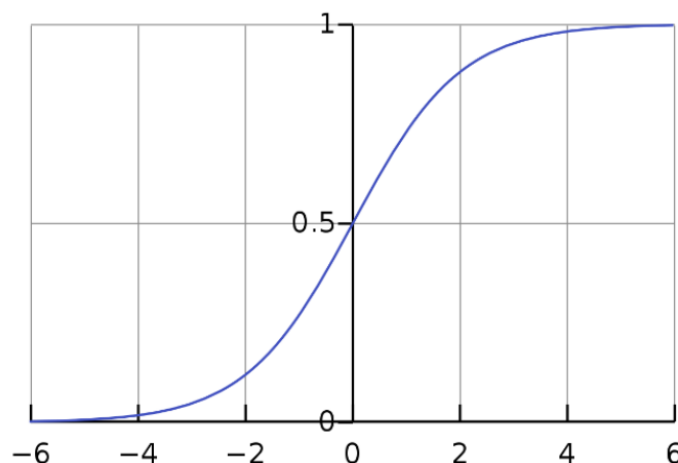
We decide a probability threshold and if the element is greater than the threshold, we put the element in one group and vice versa.

Step 1: find the best fitted line by linear regressions' steps

Step 2: it might be possible that the regression line so obtained might contain outliers, thus, we might not able to classify the data into two classes.

The predicted value gets converted into probability by the help of SIGMOID FUNCTION.

$$S(x) = \frac{1}{1+e^{-x}}$$



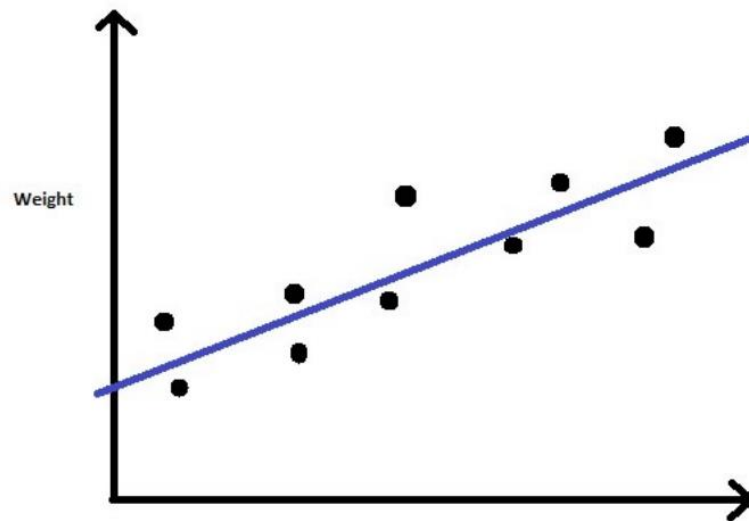
The sigmoid function returns the probability of the output Y in between 0 and 1

Step 3: thus, discrete values are obtained (between 0 and 1) based on threshold value which is usually taken to be 0.5

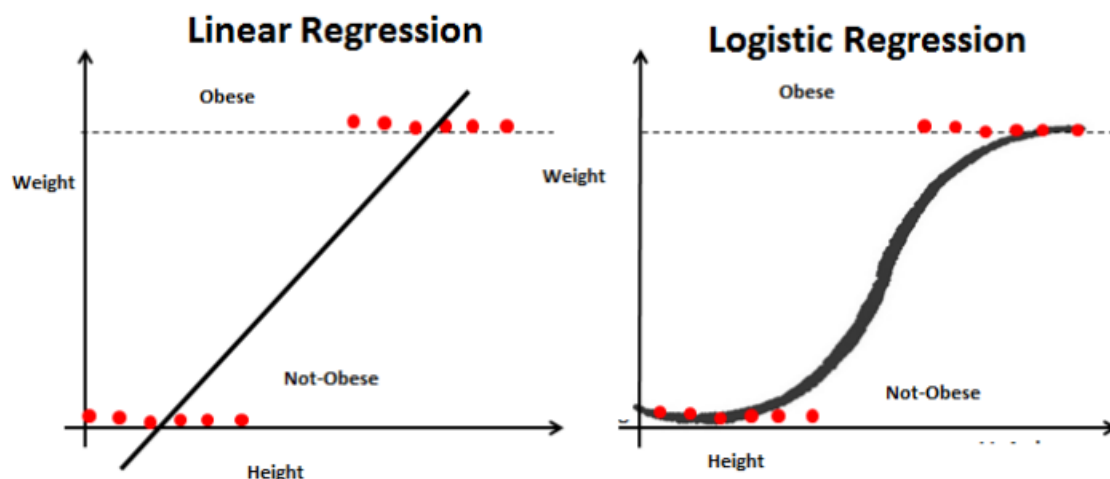
Example to differentiate between linear regression and logistic regression

Suppose we need to predict out values for weight of a student in a course provided unknown values of heights.

If we do this with linear regression,



Supposing we need to make two groups of people as obese and not-obese, so, we have two discrete values for our data also, the data here might contain a lot many outliers, so, sigmoid function is best choice for us because it isn't regression problem and sigmoid function returns us with the probability for each output value thus classifying the output into two classes.



Similarities between linear regression and logistic regression

- Both are supervised ML algos
- Both models are parametric regression i.e. both uses linear equation for prediction.

Difference between linear regression and logistic regression

- Linear regression deals with a regression problem whereas logistic regression deals with a classification problem.
- In linear regression, we have continuous output and in logistic regression we have discrete output
- We have an extra step in logistic regression where we fit the best line that came in linear regression to sigmoid curve
- We calculate loss function with the help of MSR method in linear regression whereas we use estimation in logistic regression.

Reference-

<https://www.analyticsvidhya.com/blog/2020/12/beginners-take-how-logistic-regression-is-related-to-linear-regression/>