What are the key differences between Logistic Regression and Linear Regression?

Authored by

stats writer

May 5, 2024

Logistic Regression and Linear Regression are two commonly used statistical models in data analysis and machine learning. While both models aim to predict a numeric outcome based on a set of independent variables, there are several key differences between them.

Firstly, the type of outcome variable is a major difference between Logistic Regression and Linear Regression. Logistic Regression is used for predicting a categorical outcome, usually with two possible values (e.g. Yes/No, True/False), while Linear Regression is used for predicting a continuous numerical outcome.

Secondly, the relationship between the independent variables and the outcome variable also differs between the two models. Logistic Regression uses a logistic function to model the probability of a certain outcome, while Linear Regression assumes a linear relationship between the independent variables and the outcome.

Furthermore, the assumptions of the two models differ. Linear Regression assumes that the relationship between the independent variables and the outcome is linear, and the residuals (errors) follow a normal distribution. Logistic Regression does not have these assumptions, but it does assume that the outcome variable is binary or ordinal and that the observations are independent.

Another key difference is the output of the two models. Linear Regression produces a continuous numerical output, whereas Logistic Regression produces a probability between 0 and 1, which can be interpreted as the likelihood of a certain outcome.

Finally, the interpretation of the coefficients also differs between the two models. In Linear Regression, the coefficients represent the change in the outcome variable for a unit change in the corresponding independent variable. In Logistic Regression, the coefficients represent the change in the log odds of the outcome variable for a unit change in the corresponding independent variable.

In summary, while both Logistic Regression and Linear Regression aim to predict an outcome variable, they differ in terms of the type of outcome, the relationship between variables, assumptions, output, and interpretation of coefficients. It is important to understand these key differences to choose the appropriate model for a given dataset and research question.

Logistic Regression vs. Linear Regression: The Key Differences

Two of the most commonly used regression models are linear regression and logistic regression.

Both types of regression models are used to quantify the relationship between one or more predictor variables and a , but there are some key differences between the two models:

	Linear Regression	Logistic Regression
Response Variable	Continuous (e.g. price, age, height, distance)	Categorical (yes/no, male/female, win/not win)
Equation Used	$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + ...$	$p(Y) = e^{(\beta0 + \beta1X1 + \beta2X2 + ...)} / (1+ e^{(\beta0 + \beta1X1 + \beta2X2 + ...)})$
Method Used to Fit Equation	Ordinary Least Squares	Maximum Likelihood Estimation
Output to Predict	Continuous value ($150, 40 years, 10 feet, etc.)	Probability (0.741, 0.122, 0.345, etc.)

Here's a summary of the differences:

Difference #1: Type of Response Variable

A linear regression model is used when the response variable takes on a continuous value such as:

PriceHeightAgeDistance

Conversely, a logistic regression model is used when the response variable takes on a categorical value such as:

Yes or NoMale or FemaleWin or Not Win

Difference #2: Equation Used

Linear regression uses the following equation to summarize the relationship between the predictor variable(s) and the response variable:

$Y = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta pXp$

where:

Y: The response variableXj: The jth predictor variableβj: The average effect on Y of a one unit increase in Xj, holding all other predictors fixed

Conversely, logistic regression uses the following equation:

$p(X) = e\beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta pXp / (1 + e\beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta pXp)$

This equation is used to predict the probability that an individual observation falls into a certain category.

Linear regression uses a method known as ordinary least squares to find the best fitting regression equation.

Conversely, logistic regression uses a method known as maximum likelihood estimation to find the best fitting regression equation.

Difference #4: Output to Predict

Linear regression predicts a continuous value as the output. For example:

Price ($150, $199, $400, etc.)Height (14 inches, 2 feet, 94.32 centimeters, etc.)Age (2 months, 6 years, 41.5 years, etc.)Distance (1.23 miles, 4.5 kilometers, etc.)

Conversely, logistic regression predicts probabilities as the output. For example:

40.3% chance of getting accepted to a university.93.2% chance of winning a game.34.2% chance of a law getting passed.

When to Use Logistic vs. Linear Regression

The following practice problems can help you gain a better understanding of when to use logistic regression or linear regression.

Problem #1: Annual Income

Suppose an economist wants to use predictor variables (1) weekly hours worked and (2) years of education to predict the annual income of individuals.

In this scenario, he would use linear regression because the response variable (annual income) is continuous.

Problem #2: University Acceptance

Suppose a college admissions officer wants to use the predictor variables (1) GPA and (2) ACT score to predict the probability that a student will get accepted into a certain university.

In this scenario, she would use logistic regression because the response variable is categorial and can only take on two values - accepted or not accepted.

Problem #3: Home Price

Suppose a real estate agent wants to use the predictor variables (1) square footage, (2) number of bedrooms, and (3) number of bathrooms to predict the selling house of prices.

In this scenario, she would use linear regression because the response variable (price) is continuous.

Problem #4: Spam Detection

Suppose a computer programmer wants to use the predictor variables (1) number of words and (2) country of origin to predict the probability that a given email is spam.

In this scenario, he would use logistic regression because the response variable is categorical and can only take on two values - spam or not spam.

The following tutorials offer more details on linear regression:

The following tutorials offer more details on logistic

regression: