

Logistic regression

Logistic regression is a statistical method for analysing a dataset in which there are one or more independent variables that determine an outcome (target/independent variable). The outcome is measured with a categorical variable. Based on the type and number of categories, Logistic regression can be classified as:

1. Binary: target variable can have only 2 possible types: "0" or "1" which may represent "win" vs "loss", "pass" vs "fail", "dead" vs "alive", "male" vs "female" etc.
2. Multinomial: target variable can have 3 or more possible types which are not ordered(i.e. types have no quantitative significance) like "disease A" vs "disease B" vs "disease C", "General Class" vs "ST" vs "SC" vs "OBC", etc..
3. Ordinal: it deals with target variables with ordered categories. For example, a test score can be categorized as: "very poor", "poor", "good" and "very good". Here, each category can be given a score like 0, 1, 2, 3. Similarly Educational qualifications "illiterate", "primary", "class VIII", "madhyamik", "HS", "graduate", "post- graduate" can be given a score like 0, 1, 2, 3, 4, 5, 6.

Comparison to linear regression

Given data on time spent studying and exam scores. Linear Regression and logistic regression can predict different things:

- Linear Regression could help us predict the student's test score on a scale of 0 - 100. Linear regression predictions are continuous (numbers in a range).
- Logistic Regression could help use predict whether the student passed or failed. Logistic regression predictions are discrete (only specific values or categories are allowed). We can also view probability scores underlying the model's classifications.