Simple Logistic Regression

While linear regression is used for predicting a continuous response variable, logistic regression is used for predicting a categorical response variable. It's particularly useful when the response variable is binary (i.e., it takes on only two possible outcomes, usually coded as 0 and 1).

The primary idea behind logistic regression is to find the probability of the response variable being true (1) given the values of the predictor variables.

In simple logistic regression, we have only one predictor variable. The form of the logistic regression model is:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \tag{1.1}$$

where:

- p is the probability of the positive class (i.e., the outcome $y = 1$).

- β_0 and β_1 are the parameters of the model.

- x is the predictor variable.

On the left-hand side, we have the natural log of the odds ratio (also called the logit), rather than just p itself. This is done to ensure that the predicted probabilities lie between 0 and 1. The function $\frac{p}{1-p}$ is called the odds, and can take any value between 0 and ∞.

In contrast to linear regression, where the parameters are estimated using least squares, the parameters in logistic regression are usually estimated using maximum likelihood estimation.

Maximum likelihood estimation finds the parameter values that make the observed data most likely under the model.

In a simple logistic regression model, the probability that $Y = 1$ given x is:

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \tag{1.2}$$

And the probability that $Y = 0$ given x is:

$$1 - p(x) = \frac{1}{1 + e^{\beta_0 + \beta_1 x}} \tag{1.3}$$

This is a draft chapter from the Kontinua Project. Please see our website (https://kontinua. org/) for more details.

Answers to Exercises