Introduction of logistic model

Yongshi JIE

Definition

In statistics, logistic regression, or logit regression, is a mathematical model used in statistics to estimate the probability of an event occuring having been given some previous data. Logistic Regression works with binary data, where either the event happens (1) or the event does not happen (0). So given some feature x it tries to find out whether some event y happens or not.

For example, if y represents whether a sports team wins a match, then y will be 1 if they win the match or y will be 0 if they do not. This is known as Binomial Logistic Regression. There is also another form of Logistic Regression which uses multiple values for the variable y. This form of Logistic Regression is known as Multinomial Logistic Regression.

Definition

Logistic regression does not look at the relationship between the two variables as a straight line. Instead, Logistic regression uses the natural logarithm function to find the relationship between the variables and uses test data to find the coefficients. The function can then predict the future results using these coefficients in the logistic equation.

Logistic regression uses the concept of odds ratios to calculate the probability. This is defined as the ratio of the odds of an event happening to its not happening.

$$Odds = \frac{P(y=1|x)}{1-P(y=1|x)}$$

The natural logarithm of the odds ratio is then taken in order to create the logistic equation is know as the logit:

$$Logit(P(x)) = \ln\left(\frac{P(y=1|x)}{1-P(y=1|x)}\right)$$

In Logistic regression the Logit of the probability is said to be linear with respect to x, so the logit becomes:

$$Logit(P(x)) = a + bx$$

Definition

Using the two equations together then gives the following:

$$\ln\left(\frac{P(y=1|x)}{1-P(y=1|x)}\right) = a + bx$$

Using the two equations together then gives the following:

$$\frac{P(y=1|x)}{1-P(y=1|x)} = e^{a+bx}$$

This then leads to the probability:

$$P(y=1|x) = \frac{e^{a+bx}}{1+e^{a+bx}} = \frac{1}{1+e^{-(a+bx)}}$$

The final equation is the logistic regression.

We can use the vector form and the logistic equation can be changed to:

$$P(y=1|x) = \frac{1}{1+e^{-(w^T x)}}$$