04 Logistic Regression

Friday, 2 April 2021 9:08 AM

Log. Regression is used when the dependent variable is categorical in Nature.

Eg:- Predicting spam or not spam.

Log Reg equation :-

$$\left\{ y = \frac{e^{(b_0 + b_1 x)}}{1 + e^{(b_0 + b_1 x)}} \right\}$$

Types of Log. Reg :-

① Binary Log-Reg {0, 1}

② Multinomial Logistic Regression
{veg, Non-veg, Vegan}

③ Ordinal Logistic Regression
{Rating from 1 to 5}

Assumptions:

① Binary L.R. requires the dependent variable to be binary and so does the ordinal.

② The observations needs to be independent

③ As little multicollinearity as possible among the independent variable.

④ It assums linearity of dependent variable and Log odds. further, it only requires the independent variable to be linearly related to the log odds.

⑤ Logistic regression typically requires a large sample side.

Log-Reg assumes linear Relationship between the target and the input

☆ Logistic Regression models the probability of the default class.

$P(X) = P(Y=1|X)$

(Probability of Y being 1 given x)

Transformed into 0 & 1 in order to create classification basic.

$$P(X) = \frac{e^{(b_0 + b_1 * x)}}{1 + e^{(b_0 + b_1 * X)}}$$

$$\Rightarrow \ln\left(\frac{P(X)}{1 - P(X)}\right) = b_0 + b_1 X$$

(linear eq.)

log of probabilities of default classes.

(odds)

A good Model will give prob close to 1 in case of class A and 0 in case of class B.

Eg. Binary odds can be written as

$$\ln(\text{odds}) = b_0 + b_1 X$$

$$\Rightarrow \text{odds} = e^{b_0 + b_1 x}$$

(b_0 & b_0 is estimated by MLE)

Steps of Preparing data for Log. Reg.

① Binarization of output variable → One hot encode

② Remove noise → dual

③ Gaussian Distribution

④ Remove Correlated Input → VIF (Variance Inflation factor)

⑤ Remove sparsity if the MLE fail to converge.