Discrete Response Model
Lecture 2

datascience@berkeley

Statistical Assumption of Binomial Logistic Regression Models

Statistical Assumptions

Recall from Lecture 1 the five assumptions underlying the binomial probability model. Some of them have their counterparts in the context of conditional response probability.

1. The dependent variable needs to be binary (and not ordinal); specifically, the conditional distribution of y given follows a *Bernoulli distribution*.
2. Observations are independent of each other. In fact, the error term of the model needs to follow an independent and identically distributed random variable.
3. No perfect collinearity.
4. Linearity assumption: linearity of independent variables and log-odds ratio.

The Requirement of No Complete Separation

Complete Separation:

There is a complete separation of data points if there exists a vector $\mathbf{b}$ that correctly allocates all observations to their response groups; that is,

$$\begin{cases} \mathbf{b}'\mathbf{x}_i > 0 & Y_i = 1 \\ \mathbf{b}'\mathbf{x}_i < 0 & Y_i = 2 \end{cases}$$

In this case, the maximum likelihood estimate does not exist; the log-likehood goes to 0 as iteration increases.

Berkeley SCHOOL OF INFORMATION