Multinomial Logistic Regression Tutorial

Barbara Flores

Table of contents

1 Overview

1.1 Generalized linear model

Generalized Linear Models (GLMs) are a class of statistical models in which the relationship between a dependent variable and one or more independent variables follows a specific probability distribution. The general formula of a Generalized Linear Model (GLM) is expressed as follows:

$$g(\mu) = X\beta$$

where
- $g(\mu)$: The link function connecting the mean μ of the response variable distribution with the linear combination of predictor variables $X\beta$
- X:The design matrix containing the values of the predictor variables.
- β: The vector of coefficients associated with each predictor variable.

The choice of probability distribution and the link function allows the adaptation of the model to different types of data and relationships between variables."

1.1.1 Link function

- The link function connects the distribution of the outcome to Y X and indicates how the expected value of the response relates to the linear combination of explanatory variables.
- GLM assumptions include independent distribution of data, a specified distribution for the dependent variable (e.g., Bernoulli for logistic regression), and a linear relationship between the transformed expected response and the explanatory variables.
- Logistic regression and link functions
- Motivation for using logistic regression
- Outcome variable type
- Model example data on pumpkin seeds
- GLM terms: generalized linear models and link functions
- GLM assumptions: independent data, specified distribution for dependent variable, linear relationship between transformed expected response and explanatory variables
- Explanation of why there is no error term
- Motivation for logistic regression and link functions
- Logistic regression is used for binary outcome variables
- Generalized linear models (GLMs) assume the response variable follows a distribution
- Link function connects the distribution of the outcome to the linear combination of explanatory variables
- GLM assumptions include independent distribution of data, specified distribution for the dependent variable, and linear relationship between the transformed expected response and explanatory variables

1.1.2 Link function

1.1.3 GLM Assumptions

1.2 Multinomial Logistic Regression

$Pr(\theta|y) \propto Pr(y|\theta)Pr(\theta)$
$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$

2 Probability Distribution

3 Data Example:

4 Probability Distribution