

Analytics Assignment 3: GLM - Multinomial Regression

Revanth Chowdary Ganga (rg361)

1. Introduction

1.1 Generalized Linear Models

Generalized Linear Models (GLMs) are a class of statistical models that extend linear regression to handle a broader range of response variable distributions such as binomial, Poisson, and gamma, making them suitable for diverse types of data. Unlike traditional linear regression, GLMs are not constrained by the assumption of normality.

The Primary components of a GLM are:

1. **Response Variable:** with a distribution such as Binomial, Poisson etc.
2. **Linear Predictor:** A Linear combination of the predictor variables (similar to Linear Regression)
3. **Link Function:** to connect the linear predictor to the expected value of the response variable.

The General form of a GLM is:

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Where $g(\mu)$ is the link function, β_i are the coefficients associated with the corresponding predictor variables X_i

1.2 Link Function

A link function is used to connect the linear predictor to the expected value of the response variable i.e. it describes how the mean of the response variable is related to a linear combination of the predictor variables. The link function, denoted as $g(\mu)$, transforms the linear predictor into a scale appropriate for the response variable. The choice of the link function depends on the nature of the response variable and the distribution assumed for it some examples of link functions are **Logit**, **Inverse**, **Log** etc.

1.3 GLM: Multinomial Regression

Multinomial Regression is used when the response variable is categorical in nature and has more than 2 levels (categories). The link function used in Multinomial Regression is a **Logit** function which is represented as follows:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

some sample research questions which can be answered by multinomial include:

1. What Occupation are people most likely to chose based on their parents occupation and their own education
2. What food preferences will an animal have based on its size and habitat

2. Probability Distribution