Chapter 13 : Generalized linear model

Table of Contents

1. Introduction

The purpose of this chapter is to show that several seemingly unrelated models are actually all special cases of the generalized linear model.

The Generalized Linear Model can be written as:

$$g\big(E(Y_i)\big) = \alpha + \sum_{j=1}^{p} \beta_j X_{ji} + \epsilon_i$$

- Where g(E(Yi)) is some function of the expected value of Yi (this function g is called the link function).

- and $\varepsilon_i \sim F$ (i.e. the error term has some sort of distribution, e.g. normal in case of a linear regression model ,….) . This is called the distributional family.

2. Linear regression as a generalized linear model

In linear regression, we have following model formulation:

$$\left(E(Y_i)\right) = \alpha + \sum_{j=1}^{p} \beta_j X_{ji} + \epsilon_i$$

and $\varepsilon_i \sim N(0, \sigma^2)$.

- That is, the distributional "family" is normal.
- We predict E(Y). Hence , g(E(Y)) = E(Y). In this case G(.) is the *identity* link function.

3. Logistic regression as a generalized linear model

The *logistic regression model* can then be written as

$$ln \frac{P[Y_i = 1]}{1 - P[Y_i = 1]} = \alpha + \sum_{j=1}^{p} \beta_j X_{ji} + \epsilon_i$$

Note that

- When Y_i is a binary variable (can take the values 0 or 1), it does not have a normal distribution; rather it has a *binomial* distribution. The distributional family is binomial.
- The left hand side is not E(Y). The left hand side is expressed in log odds. We predict g(E(Y)), where g is the *logit* link function.

4. Poisson regression as a generalized linear model

The Poisson regression model is formulated as:

$$\log\big(E(Y_i)\big) = \alpha + \sum_{j=1}^{p} \beta_j X_{ji} + \varepsilon_i$$

Note that

- Here we assume a Poisson distribution for Y_i. The distributional family is Poisson.
- The left hand side is log(E(Y)), where g is the *log* link function.

5. How to use in R

We can use the **glm** function in R.

glm(formula, family = gaussian, data)

- The formula looks like

$$Y \sim X1 + X2$$

where X1 and X2 are the names of

- ✓ Continuous variables
- ✓ Categorical variables

- The family argument specifies
- ✓ the link function
- ✓ the variance function

E.g.

	Family argument
Linear Regression model	gaussian(link = "identity")
Logistic regression model	binomial(link = "logit")
Poisson regression model	poisson(link = "log")