Overview of the Generalized Linear Model

Shubham Gupta

July 3, 2019

1 Introduction

- We will apply the concepts of Bayesian analysis(inference, MCMC, etc) to a more complex family of models called **generalized linear models** which consists of models such as t-tests, analysis of variance(ANOVA). multiple regression, logistic regression, log-linear models, etc.

2 Types of Variables

- Two main types of variables: **Predictor** and **Predicted** variables.
- Likelihood functionn expresses probability of values for the **predicted** variable as a function of values of the **predictor** variable.
- Predictor variables are called **depdendant** variables.
- Predicted variables are called **independant** variables.

2.1 Scale types

- Main types are:
  - Metric
  - Ordinal
  - Nominal
  - Count

3 Linear combination of predictors

- GLM expresses influence of predictors as their **weighted sum**.

3.1 Linear function of a single metric predictor

- Linear functions preserver proportinality.

$$y = \beta_0 + \beta_1 x.$$

- This type of equation is called an **affine**.

3.2 Additive combination of metric predictors

- Add predictor variables for combined effect.

$$y = \beta_0 + \sum_{k=1}^{K} \beta_k x_k.$$

3.3 Non additive interaction of metric predictors

- Even if the interactions between two predictors are **not linear**, a new feature(like their product or sum) can help make the dataset linear.

3.4 Nominal Predictors

3.4.1 Linear model for a single nominal predictor

- Also called as **one hot encoding**. Split the nomial variables into multiple columns to model the problem.

$$y = \beta_0 + \beta_{[1]}x_{[1]} + \beta_{[2]}x_{[2]} + \dots.$$

$$y = \beta_0 + \vec{\beta}.\vec{x}.$$

3.4.2 Additive combination of nominal predictors

- Effect of multiple nominal predictors combinations can be represented by:

$$y = \beta_0 + \sum_{n} \beta_{1[j]}x_{1[j]} + \sum_{n} \beta_{2[k]}x_{2[k]} + \dots.$$

3.4.3 Nonadditive interaction of nominal predictors

- $\vec{x}_{1x2}$ refers to a particular combination of values from $\vec{x_1}$ and $\vec{x_2}$.
- Nonadditive interaction is represented by:

$$y = \beta_0 + \beta_{[1]}x_{[1]} + \beta_{[2]}x_{[2]} + \vec{\beta_{1x2}}.\vec{x_{1x2}}.$$

Table 15.1 For the generalized linear model: typical linear functions lin(x) of the predictor variables x, for various scale types of x

		Scale type of predictor x			
		Metric		Nominal	
Single group	Two groups	Single predictor	Multiple predictors	Single factor	Multiple factors
β_0	$\beta_{x=1}$ $\beta_{x=2}$	β_0 $+\beta_1 x$	β_0 $+\sum_k \beta_k x_k$ $+\sum_{j,k} \beta_{j\times k} x_j x_k$ $+\left[\text{higher order interactions}\right]$	β_0 $+\vec{\beta}\cdot\vec{x}$	β_0 $+\sum_k \vec{\beta}_k \cdot \vec{x}_k$ $+\sum_{j,k} \vec{\beta}_{j\times k} \cdot \vec{x}_{j\times k}$ $+\left[\text{higher order interactions}\right]$

The value lin(x) is mapped to the predicted data by functions shown in Table 15.2.

Figure 1: Typical linear functions