Homework 2 - Generalized Linear Models

Due on 2021-05-14

Notice

- All homework should be turned in pdf format **using Rmarkdown**.
- Homework can be written in Korean.

1 Exponential family

When a random variable Y belongs to exponential family, then the pdf of the random variable can be written as following form:

$$f_Y(y;\theta,\phi) = exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]$$

where a, b, c are functions, and the parameter θ is often referred to as **canonical parameter**.

1) Express the p.d.f of random variable Y

$$Y \sim Bernulli(p)$$

in the above form, which follows the Bernulli p:

$$f_Y(y;p) = p^y(1-p)^{1-y}\mathbb{1}_{(y\in\{0,1\})}$$

2) Express the p.d.f of random variable Y

$$Y \sim Poisson(\lambda)$$

in the above form, which follows the Poisson λ:

$$f_Y(y;\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}\mathbb{1}_{(y\in\mathbb{Z}^+)}$$

3) We have learned that the link function in GLM is a function g which connects the linear components with the $\mu = \mathbb{E}[Y]$. If link function g satisfies the following condition, then g is called a *canonical link function*.

$$\theta = g(\mu)$$

Verify the logit link for the logistic regression and the log link for the poisson regression are the canonical link functions.

2 Maximum Likelihood Estimator of Poisson Regression

Save Table 1 into `example_data` and do the Poisson regression using `Gender` over `Claims` as follows:

1) Calculate the coefficients by using the following code.

Table 1: Example data for Problem 2

Policy	Policy.Years	Gender	Territory	Claims
1	5	M	East	0
2	5	F	East	0
3	4	M	East	1
4	3	F	West	1
5	4	F	East	0
6	3	F	West	1
7	5	M	West	0
8	5	M	West	2
9	3	M	East	1
10	2	F	East	1
11	4	M	West	1
12	5	F	West	0

```
glm(Claims ~ Gender + Territory,
family = poisson(link=log), data = example_data)
```

2) Find the coefficients of Poisson regression above using an user define `R` function via gradient descent algorithm. Your answer must include the following steps:

- Define Likelihood and Loglikelihood function
- Calculate gradient of Loglikelihood function with repect to β.