GLM

1. Definition of GLM

a. Exponential family with canonical form

GLM: Response variable Y must be a random variable coming from a exponential family with canonical form.

Def. A random variable Y has a distribution in the **Exponential Family** if its pdf (for continous random variables) or its pmf (for discrete random variables) has the following form:

$$\begin{aligned} f(y;\theta) &= s(y)\cdot t(\theta)\cdot \exp\{a(y)\cdot b(\theta)\} \\ &= \exp\{a(y)\cdot \underbrace{b(\theta)}_{\text{natural parameter}} + c(\theta) + d(y)\} \end{aligned}$$

$$*\ \text{Sample space of y does not depend on } \theta$$

If $a(y) = y$, then the distribution of y is in **Canonical** form.

Common exponential family:

$$\begin{aligned} &Y \sim N(\mu,\sigma^2) \\ &Y \sim Bernoulli(\pi) \\ &Y \sim Binomial(n,\pi) \\ &Y \sim Poisson(\lambda) \\ &Y \sim Exp(\lambda) \end{aligned}$$

Ex. $y \sim N(\mu,\sigma^2)$

$$\begin{aligned} f(y;\mu) &= \frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\{-\frac{(y-\mu)^2}{2\sigma^2}\} \\ &= \exp\{-\frac{y^2}{2\sigma^2} + \frac{\mu y}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - lg(\sqrt{2\pi}\cdot\sigma)\} \\ &= \exp\{\underbrace{y}_{a(y)}\underbrace{(\frac{\mu}{\sigma^2})}_{b(\mu)} - \underbrace{(\frac{\mu}{2\sigma^2} + lg(\sqrt{2\pi}\cdot\sigma))}_{c(\mu)} - \underbrace{\frac{y^2}{2\sigma^2}}_{d(y)}\} \end{aligned}$$

$\because a(y) = y,$

$-\infty < y < \infty:$ range of y does not depend on μ

$\therefore y \sim N(\mu,\sigma^2)$ is a distribution in the Exponential Family with Canonical form.

b. Link function

$$g(\mu) = \mathbf{X}\boldsymbol{\beta}$$

Ex.

$g(\mu) = \mu$	$y \sim Normal$
$g(\mu) = lg\frac{\pi}{1-\pi}$	$y \sim Binomial$

c. Linear component

$\beta_0 + \beta_1 X_1 + \beta_2 X_2$	✓
$\beta_0 + \beta_1 X_1^2 + \beta_2 X_2$	✓
$\beta_0 + e^{\beta_1} X_1 + \beta_2 X_2$	✓
$\beta_0 + e^{\beta_1 X_1} + \beta_2 X_2$	✗

2. Multiple regression vs. Multivariate regression

Multiple regression	Multivariate regression
Univariate: one response variable Y	Multivariate: multiple response variables **Y**
$Y_1\ vs.\ X_1, X_2$ $Y_2\ vs.\ X_1, X_2$	$\mathbf{Y}\ vs.\ X_1, X_2$ $\mathbf{Y}$: a vector of response variables

3. Generalized linear model vs. General linear model

	Types of Y
Generalized linear model	$Y \sim N(\mu, \sigma^2) \Rightarrow Y \sim$ Exp. Family with Canonical form
Generalized linear model	Univariate $\Rightarrow$ Multivariate Independent $\Rightarrow$ Dependent