What you have never been told about GLM

Cristian Pachón

19/04/2018

Linear Model

Goal:
To find out if there exists a linear relationship between the random variable (rv) Y and the covariables X_1, X_2, X_3, ..., X_p.
Model:
$\forall i \;\; Y_i = \beta_0 + x_{i1}\beta_1 + +x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \epsilon_i = \mu_i + \epsilon_i$
Hipothesis:

- $\forall i \;\; \epsilon_i \sim N(0, \sigma_i^2)$.
- $\forall i \;\; \sigma_i^2 = \sigma^2$ (homoscedasticity).
- $\forall i, j, \; i \neq j \;\; \epsilon_i$ independent of ϵ_j.
- X values are fixed or if random, they are independent of errors.

Linear Model

In matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

So, $Y = X\beta + \epsilon$

Generalized Linear Model

From the previous expression, it is easy to see that $Y|X \sim N(X\beta, \sigma^2 \cdot Id_n)$ But what happens if Y is not normal distributed and we still want to find a linear relationship? In such situations we can use GLM, since they allow to assume that:

- Y follows a more general probability distribution, no necessarily Normal.
- The linearity is between a transformation of $\mu = E(Y|X)$ and the covariables (and not Y directly).

GLM components

A GLM has three components:

- Random component.
- Deterministic component.
- Link function.

Random component

Given a random vector $Y = (Y_1, Y_2, Y_3, ..., Y_n)^t$ and given that $\mu = E(Y) = (\mu_1, \mu_2, \mu_3, ..., \mu_n)^t$, each Y_i should be written as:

$$Y_i \sim exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right)$$

Comments:

- $h(\mu_i) = \theta_i$.
- ϕ is known as dispersion parameter (also detonted as σ^2).

Deterministic component

$X_{n\times p}\beta_{p\times 1}$ where $p < n$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_p \end{bmatrix}$$

Link function

Any monotone differentiable function η:

$$\eta = g(\mu) = X\beta$$

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

This can be written as:

$$exp\left(\frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}log(2\pi\sigma^2)\right)$$

- $\theta = \mu$,
- h is the identity function.
- $a(\phi) = \phi = \sigma^2$.
- $b(\theta) = \frac{\theta^2}{2}$.
- $c(y;\phi) = \frac{-y^2}{2\sigma^2} - log(\sqrt{2\pi\sigma^2})$

$$Y_i \sim exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

- $h(\mu) = \theta$.

$$f(y;m,p) = \binom{m}{y} p^y (1-p)^{m-y}$$

This can be written as:

$$exp\left[ylog\left(\frac{\mu}{m-\mu}\right) - mlog\left(\frac{m}{m-\mu}\right) + log\binom{m}{y}\right]$$

- $\theta = log\left(\frac{\mu}{m-\mu}\right) = h(\mu)$,
- $a(\phi) = 1$.
- $b(\theta) = log(1+e^{\theta})$.
- $c(y;\phi) = log\binom{m}{y}$

$$Y_i \sim exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

- $h(\mu) = \theta$.

Link function

It is the link between $E(Y)$ and the covariables X_1, ..., X_p. Remember:

- $g(\mu) = X\beta$.
- Any monotone differentiable function.

Link function

- In the normal distribution, we have seen that $\theta = \mu$. So h is the identity function.
- In the binomial distribution, we have seen that $\theta = log\left(\frac{\mu}{m-\mu}\right) = h(\mu)$.
- When g is choosen so that g = h, then it is said to use the canonical link function. In this case, $\theta = g(\mu) = X\beta$

Link function

In the normal case, when canonical link is used:

- $\mu = \beta_0 + X_1\beta_1 + \cdots + X_p\beta_p$

In the binomail case, when canonical link is used:

- $log\left(\frac{\mu}{m-\mu}\right) = \beta_0 + X_1\beta_1 + \cdots + X_p\beta_p.$
- $\mu = \frac{e^{X\beta}}{1+e^{x\beta}}.$
- $p = \frac{1}{m} \cdot \frac{e^{X\beta}}{1+e^{x\beta}}.$

Variance function

Given that

$$Y \sim exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi)\right)$$

We can obtain the following expressions:

- $E(Y) = b'(\theta)$.
- $Var(Y) = a(\phi)b''(\theta)$.
- $V(\mu) = b''(\theta)$ is known as the **Variance function**.
- In case of binomial distribution, $V(\mu) = \frac{\mu(1-\mu)}{m}$.
- In case of normal distribution, $V(\mu) = 1$. So, **homoscedasticity** must hold.

Parameters estimation

To find the maximum likelihood estimator of β, the following iterative equation has to be solved:

$$X^t W X b^{m+1} = X^t W Z$$

where:

- W depends on the variance function and Z depends.
- Z depends on x, b^m, y, μ and η.

Parameters estimation

Given b^m, we proceed as follows:

- Get $\eta = X\beta^m$.
- Get $\mu = g^{-1}(X\beta^m)$.
- Get $\theta = h(\mu)$.
- Get Var(Y).
- Get W and Z.
- Get b^{m+1} from $X^tWXb^{m+1} = X^tWZ$.

Predictions

Given $\hat{\beta}$ (estimation of β), the predicted mean value is equal to:

$$\hat{y} = \hat{\mu} = g^{-1}(X\beta)$$

Goodness of fit

- The **null model** is defined as the model with just one variable (the intercept).
- The **full model** is defined as the model with as many variables as observations. It is the perfect model.
- We want a model **as close as possible to the full model** but with **less variables**. We will test our model with a hypothesis test.

Goodness of fit

- Let $l(\hat{\mu}, \phi, y)$ be the value of the log-likelihood corresponding to our model.
- Let $l(y, \phi, y)$ be the value of the log-likelihood corresponding to the perfect model.

The scaled deviance is defined as:

$$D^*(y, \hat{\mu}) = 2(l(y, \phi, y) - l(\hat{\mu}, \phi, y))$$

Goodness of fit

In order to check if our model is as good as the perfect model, the following test is done:

$$\begin{cases} H_0\text{: Our model is as good as the perfect one} \\ H_1\text{: Our model is not as good the perfect one} \end{cases}$$

Under H_0, asymptotically $D^*(y, \hat{\mu}) \sim \chi^2_{n-p}$, where p is the number of parameters of our model.