Lecture 4. Generalized Linear Models

Nan Ye

School of Mathematics and Physics
University of Queensland

Generalized Linear Models

Recall: definition of GLM

- A GLM has the following structure

$$\begin{aligned} &\text{(systematic)} \quad \mathbb{E}(Y \mid \mathbf{x}) = h(\beta^\top \mathbf{x}). \\ &\text{(random)} \quad Y \mid \mathbf{x} \text{ follows an exponential family distribution.} \end{aligned}$$

- This is usually separated into three components
  - The linear predictor $\beta^\top \mathbf{x}$.
  - The response function h.
    People often specify the link function $g = h^{-1}$ instead.
  - The exponential family for the conditional distribution of Y given $\mathbf{x}$.

Recall: remarks on exponential families

- It is common!

  normal, Bernoulli, Poisson, and Gamma distributions are exponential families.

- It gives a well-defined model.

  its parameters are determined by the mean $\mu = \mathbb{E}(Y \mid \mathbf{x})$.

- It leads to a unified treatment of many different models.

  linear regression, logistic regression, ...

In a GLM, we consider exponential families with $T(y) = y$.

Questions

- Given β, how to compute the probability $p(y \mid \mathbf{x}, \beta)$?
- Given β, how to predict the value of y (using mean or mode)?
- Given observed $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, how to find the maximum likelihood estimator (MLE) for β?
- How to find a confidence interval for the MLE?

This Lecture

- Computing $p(y \mid \mathbf{x}, \beta)$
- Fisher scoring method

Evaluating $p(y \mid \mathbf{x}, \beta)$

Example 1. Ordinary linear regression

- Recall: $Y \mid \mathbf{x} \sim N(\mathbf{x}^\top \beta, \sigma^2)$.
- $p(y \mid \mathbf{x}, \beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(y - \mathbf{x}^\top \beta)/2\sigma^2)$.
- σ^2 can be estimated as the variance of the residuals.
- We can predict Y as $\mathbf{x}^\top \beta$, which is both the mean and mode of Y given $\mathbf{x}$.

Example 2. Logistic regression

- Recall: $Y \mid \mathbf{x} \sim B\left(\frac{1}{1+e^{-\mathbf{x}^\top \beta}}\right)$.
- After some calculation: $p(y \mid \mathbf{x}, \beta) = \frac{e^{y\mathbf{x}^\top \beta}}{1+e^{\mathbf{x}^\top \beta}}$.
- We can predict Y as

$$\arg\max_y p(y \mid \mathbf{x}, \beta) = \begin{cases} 1, & \mathbf{x}^\top \beta > 0. \\ 0, & \mathbf{x}^\top \beta \leq 0. \end{cases}$$

A general explicit formula

The idea of going from given β, $\mathbf{x}$ to the distribution of y is shown graphically below

$$\beta, \mathbf{x} \xrightarrow{\quad g^{-1} \quad} \mu \xrightarrow{\quad A'^{-1} \quad} \eta \xrightarrow{\quad \text{exp. fam.} \quad} y$$

- Assume a natural parametrization of the exponential family

$$f(y \mid \eta, \phi) = \exp\left(\frac{\eta T(y) - A(\eta)}{b(\phi)} + c(y, \phi)\right)$$

- Compute the mean $\mu = \mathbb{E}(Y \mid \mathbf{x}) = g^{-1}(\beta^\top \mathbf{x})$.
- Compute the natural parameter $\eta = A'^{-1}(\mu)$.
- Thus the probability of y given $\mathbf{x}$ and β is

$$p(y \mid \mathbf{x}, \beta) = \exp\left(\frac{\eta T(y) - A(\eta)}{b(\phi)} + c(y, \phi)\right),$$

where $\eta = A'^{-1}(g^{-1}(\beta^\top \mathbf{x}))$.

Computing MLE

- We want to choose β to maximize the log-likelihod

$$\ell(\beta) = \sum_{i=1}^{n} \ln p(y_i \mid \mathbf{x}_i, \beta)$$

- We will first cover the Fisher scoring algorithm, a general algorithm for finding MLEs, and then show how it can be applied to GLMs.

Fisher scoring

- An general algorithm for finding an MLE.
- Start with some $\beta^{(0)}$. At iteration $t \geq 0$,

$$\beta^{(t+1)} = \beta^{(t)} + I^{-1}(\beta^{(t)}) \nabla \ell(\beta^{(t)}).$$

where $I(\beta) = -\mathbb{E} \nabla^2 \ell(\beta)$ (known as *Fisher information*).

Notation

- ∇*: the gradient operator* $(\frac{\partial}{\partial \beta_1}, \ldots, \frac{\partial}{\partial \beta_d})$*, as a column vector.*
- $\nabla^\top$ *is the transpose of* ∇*.*
- ∇^2 *denotes the Hessian operator, and is* $\nabla \nabla^\top$*.*

Derivation of Fisher scoring

- Consider the Taylor series expansion of $\ell(\beta')$ around β

$$\ell(\beta') \approx \ell(\beta) + \nabla^\top \ell(\beta) \cdot (\beta' - \beta) + \frac{1}{2}(\beta' - \beta)^\top \nabla^2 \ell(\beta)(\beta' - \beta).$$

  where $\nabla \ell(\beta)$ is the gradient, and $\nabla^2 \ell(\beta)$ is the Hessian.
- The maximizer of the RHS is given by

$$\beta^* = \beta - (\nabla^2 \ell(\beta))^{-1} \nabla \ell(\beta).$$

- This motivates the update (known as Newton-Raphson method)

$$\beta^{(t+1)} = \beta^{(t)} - (\nabla^2 \ell(\beta^{(t)}))^{-1} \nabla \ell(\beta^{(t)}).$$

- Finally, replace the negative Hessian $-\nabla^2 \ell(\beta)$ by its expectation $I(\beta)$.

What You Need to Know

- The explicit form of a GLM model $p(y \mid \mathbf{x}, \beta)$.
- Computing MLE using Fisher scoring.