Predictive Analytics
Regression and Classification
Lecture 7 : Part 2

Sourish Das

Chennai Mathematical Institute

Aug-Nov, 2019

Logistic Regression

- Logistic Regression with logit-link

$$\log\left(\frac{p}{1-p}\right) = x^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \cdots + x_p\beta_p$$

- Logistic Regression with probit-link

$$\Phi^{-1}(p) = x^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \cdots + x_p\beta_p$$

- How to estimate $\boldsymbol{\beta}$?

Likelihood Function of Probit Model

- Suppose $\{y_i, x_i\}_{i=1}^n$ contains n independent samples
- For the single observation, the probit model is

$$\begin{aligned} \mathbb{P}(y_i = 1|x_i) &= \Phi(x_i^T \boldsymbol{\beta}) = p_i \\ \mathbb{P}(y_i = 0|x_i) &= 1 - \Phi(x_i^T \boldsymbol{\beta}) = 1 - p_i \end{aligned}$$

- The likelihood of a single observation (y_i, x_i) is $f(y_i, x_i|\boldsymbol{\beta}) = p_i^{y_i}(1-p_i)^{(1-y_i)} = \Phi(x_i^T\boldsymbol{\beta})^{y_i}[1-\Phi(x_i^T\boldsymbol{\beta})]^{(1-y_i)}$
- Since the observations are independent, the joint likelihood of the entire sample is

$$\mathcal{L}(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n} \left(\Phi(x_i^T\boldsymbol{\beta})^{y_i}[1-\Phi(x_i^T\boldsymbol{\beta})]^{(1-y_i)} \right)$$

Log-Likelihood Function of Probit Model

- The joint likelihood of the probit model is

$$\mathcal{L}(\beta; \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n} \left(\Phi(x_i^T \beta)^{y_i} [1 - \Phi(x_i^T \beta)]^{(1-y_i)} \right)$$

- The log-likelihood function of probit model is

$$\ln \mathcal{L}(\beta; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^{n} \left(y_i \ln \Phi(x_i^T \beta) + (1 - y_i) \ln(1 - \Phi(x_i^T \beta)) \right)$$

- The negative log-likelihood function of probit model is

$$-\ln \mathcal{L}(\beta; \mathbf{y}, \mathbf{X}) = -\sum_{i=1}^{n} \left(y_i \ln \Phi(x_i^T \beta) + (1 - y_i) \ln(1 - \Phi(x_i^T \beta)) \right)$$

Likelihood Function of Logit Model

- Suppose $\{y_i, x_i\}_{i=1}^n$ contains n independent samples

- For the single observation, the logit model is

$$
\begin{aligned}
\mathbb{P}(y_i = 1|x_i) &= \frac{\exp(x_i^T\boldsymbol{\beta})}{1+\exp(x_i^T\boldsymbol{\beta})} \\
\mathbb{P}(y_i = 0|x_i) &= 1 - \frac{\exp(x_i^T\boldsymbol{\beta})}{1+\exp(x_i^T\boldsymbol{\beta})} = \frac{1}{1+\exp(x_i^T\boldsymbol{\beta})}
\end{aligned}
$$

- The likelihood of a single observation (y_i, x_i) is $f(y_i, x_i|\boldsymbol{\beta}) = p_i^{y_i}(1-p_i)^{(1-y_i)} = \left(\frac{\exp(x_i^T\boldsymbol{\beta})}{1+\exp(x_i^T\boldsymbol{\beta})}\right)^{y_i}\left[1 - \frac{\exp(x_i^T\boldsymbol{\beta})}{1+\exp(x_i^T\boldsymbol{\beta})}\right]^{(1-y_i)}$

Likelihood Function of Logit Model

- The likelihood of a single observation (y_i, x_i) is

$$
\begin{aligned}
f(y_i, x_i|\boldsymbol{\beta}) &= p_i^{y_i}(1-p_i)^{(1-y_i)} \\
&= \left(\frac{\exp(x_i^T\boldsymbol{\beta})}{1+\exp(x_i^T\boldsymbol{\beta})}\right)^{y_i}\left(\frac{1}{1+\exp(x_i^T\boldsymbol{\beta})}\right)^{(1-y_i)}
\end{aligned}
$$

- Since the observations are independent, the joint likelihood of the entire sample is

$$
\mathcal{L}(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n}\left(p_i^{y_i}[1-p_i]^{(1-y_i)}\right),
$$

where $p_i = \frac{\exp(x_i^T\boldsymbol{\beta})}{1+\exp(x_i^T\boldsymbol{\beta})}$

Maximum Likelihood Estimates of β

- The Maximum Likelihood Estimates (MLE) of β is

$$
\begin{aligned}
\hat{\beta} &= \underset{\boldsymbol{\beta}}{\text{argmax}}\ \mathcal{L}(\beta; \mathbf{y}, \mathbf{X}) \\
&= \underset{\boldsymbol{\beta}}{\text{argmax}}\ \ln \mathcal{L}(\beta; \mathbf{y}, \mathbf{X}) \\
&= \underset{\boldsymbol{\beta}}{\text{argmin}}\ [-\ln \mathcal{L}(\beta; \mathbf{y}, \mathbf{X})]
\end{aligned}
$$

- Gradient Descent Algorithm can be used to minimise negative log-likelihood function

- This is general recepie of any statistical model

- **For simple linear regression one can show that the OLS estimator of β is also MLE**

In practice...

- You don't have to implement any optimization
- There are two Python module which implement logistic regression
  1. `statsmodels`
  2. `sklearn` (aka., scikit learn)
- There is a built-in function called `glm` in `R`-package named `stats`.
- You can use whatever you find suitable to you

In the next video...

- I will discuss the statistical inference with logistic regression...