Lecture Notes - 04: Logistic Regression

Dihui Lai

March 9, 2020

Contents

1 Likelihood Function

If a set of random variables Y_1, Y_2 ... Y_n has a joint probability distribution density/mass $f(y_1, y_2, ...y_n; \theta)$, where θ is a set of parameters, the likelihood function is defined as

$$L(\theta) = f(y_1, y_2, ...y_n; \theta) \tag{1}$$

2 Likelihood Function of Logistic Regression

Assuming an event has two possible outcomes $y = 1$ or $y = 0$, with probability p of being 1, i.e. the outcome follows a Bernoulli distribution. As we learned in lecture 2, the probability mass function is

$$\begin{cases} p, & y = 1 \\ 1 - p, & y = 0 \end{cases}$$

Alternatively, we can write this in a more condensed format

$$f(y; p) = p^y (1 - p)^y \tag{2}$$

Assuming we have a set of n events that are independent of each other, the probability mass distribution can be written as

$$f(y_1, y_2, ...y_n; p_1, p_2, ..., p_n) = \prod_{i=1}^{n} p_i^{y^i} (1 - p_i)^{1-y^i}$$

By definition (1), this is also the likelihood function,

$$L(p_1, p_2, ...p_n) = \prod_{i=1}^{n} p_i^{y^i} (1 - p_i)^{1-y^i}$$

To interpreting the likelihood function, let us consider the underlying parameters are the same i.e. $p = p_1 = p_2 ... = p_n$ for all the data entries observed. And we have the likelihood function as

$$L(p) = \prod_{i=1}^{n} p^{y^i}(1-p)^{1-y^i}$$

Let us consider two very special cases where $n = 1$ and $n = 2$

- $n = 1$ i.e. we only have 1 observation, we have $L(p) = p^y(1-p)^{1-y}$

The corresponding log-likelihood is

$$\ell(p_1, p_2, ...p_n) = \sum_{i=1}^{n} \big(y^i \log(p_i) + (1-y^i)\log(1-p_i)\big)$$

the probability of being 1 is modeled as

$$p_i = \frac{1}{1+\exp(-\vec{\beta}\cdot\vec{x}^i)}$$

The log-likelihood function is the defined as the log transformation of the likelihood function

$$\ell = \log(Likelihood) = \sum_{i=1}^{n} y^i \log(p_i) + (1-y^i)\log(1-p_i)$$