Lecture 37

Logistic Regression - the Theory

In this lecture, we will discuss the theory behind logistic regression and the parallels to and distinctions from ordinary least squares regression. We have already seen that regression is a powerful framework in which you can address many other questions for continuous response data and this remains true for binary responses.

Theory Behind Logistic Regression

For logistic regression, the responses Y_i are 0 and 1 where 1 represents some event occurring (such as survival, death, cancer, etc...) and 0 represents that event not occurring. You want to predict Y_i from some predictor variables measured on individual i: $X_{1,i}, X_{2,i}, \ldots, X_{k,i}$.

It isn't reasonable to predict only 0's and 1's. What is reasonable, is to predict the probability of the event for an individual. In fact, for reasons described before, the model is

$$\text{logit}(\pi)1-\pi = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

and the probability π can be recovered by

$$\pi = \frac{e^{\text{logit}(\pi)}}{1+e^{\text{logit}(\pi)}}$$

The estimates of the coefficients do not come from least squares but from maximum likelihood:

$$\Pr(\text{data}|\beta_0, \beta_1, \ldots, \beta_k) = \prod_{i=1}^{n} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}$$

where the terms in the product are a cheap way of getting π_i if $Y_i = 1$ and $(1-\pi_i)$ if $Y_i = 0$ and the π_i's depend on the coefficients through the link function as above.

The coefficients that make the above probability the greatest are the maximum likelihood estimators and they are found numerically in statistical software packages such as Minitab. Statisticians like maximum likelihood estimators for the following reasons:

1. the estimates are asymptotically unbiased
2. the estimates are asymptotically efficient (that is, they have the smallest variance possible)

3. the estimates are asymptotically normally distributed

The problem is that all of these are "asymptotically" - as you collect more and more data, the results hold. But you don't in fact have any guarantee that the estimates are good with a data set of size 10 or of size 100 or even of size 1000 necessarily. All you know is that if you keep collecting data, eventually the properties of unbiasedness, efficiency, and normality will hold.

What is the interpretation of the coefficients of a logistic regression? Suppose that you have a model

$$\text{logit}(\pi)1 - \pi = 1 - 2x.$$

The coefficient -2 gives the additive decrease in log odds if x is increased by 1 unit. That is

$$\text{logit}(\pi)(x+1) - \text{logit}(\pi)(x) = 1 - 2(x+1) - (1 - 2x) = -2$$

Thus the odds decrease by a factor of e^{-2}:

$$\frac{\pi(x+1)}{1-\pi(x+1)} = e^{-2}\frac{\pi(x)}{1-\pi(x)} \approx .135\frac{\pi(x)}{1-\pi(x)}$$

Thus the odds ratio of the event decreases by 86.5% for every increase of one unit in x.

References and Readings

Exercises for Lecture 37

1. –
2. –