

# Lab #9 - Logistic Regression Part I

Econ 224

September 25th, 2018

## Notation

In this lab we'll be looking at logistic regression, a commonly-used binary classification method. To make things simpler, we'll use some slightly different notation and terminology than ISL. First we'll define the *column vectors*  $X$  and  $\beta$  as follows:

$$X = \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

Notice that the first element of  $X$  is *not*  $X_1$ : it is simply the number 1. There's an important reason for this that you'll see in a moment. From the reading, we know that logistic regression is a *linear model* for the *log odds*, namely

$$\log \left[ \frac{P(X)}{1 - P(X)} \right] = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

where  $P(X)$  is shorthand for  $\mathbb{P}(Y = 1|X)$ . Note that when I write  $\log$  I **always** mean the natural logarithm. Also note that when I write  $\exp(z)$  I mean  $e^z$ . This comes in handy if  $z$  is a complicated expression.

Using the vector notation introduced above, we can express this more compactly as

$$\log \left[ \frac{P(X)}{1 - P(X)} \right] = X' \beta$$

since

$$X' \beta = \begin{bmatrix} 1 & X_1 & X_2 & \cdots & X_p \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

I will call  $X' \beta$  the *linear predictor* since it is the linear function of  $X$  that we use to predict  $Y$ . By exponentiating both sides of the log-odds expression from above and re-arranging, obtain the following:

$$\begin{aligned} \frac{P(X)}{1 - P(X)} &= \exp(X' \beta) \\ P(X) &= [1 - P(X)] \exp(X' \beta) \\ P(X) + P(X) \exp(X' \beta) &= \exp(X' \beta) \\ P(X)[1 + \exp(X' \beta)] &= \exp(X' \beta) \\ P(X) &= \frac{\exp(X' \beta)}{1 + \exp(X' \beta)} \\ P(X) &= \Lambda(X' \beta) \end{aligned}$$

where the function  $\Lambda$  is defined as follows

$$\Lambda(z) = \frac{e^z}{1 + e^z}$$

## Exercise #1

- (a) Verify that  $\Lambda(z) = \frac{1}{1 + e^{-z}}$ .
- (b) Using (b), write an alternative expression for  $P(X)$ .

## Solution to Exercise #1

- (a) Dividing the numerator and denominator by  $e^z$ , which cannot result in division by zero since  $e^z$  is always positive, we have

$$\Lambda(z) = \frac{e^z}{1 + e^z} = \frac{1}{1/e^z + 1} = \frac{1}{1 + e^{-z}}$$

- (b)  $P(X) = \frac{1}{1 + \exp(-X'\beta)}$

## Interpreting $\beta$ in a Logistic Regression

From the expression above, we see that  $\beta_j$  is the partial derivative of the log-odds with respect to  $X_j$ . But it's difficult to think in terms of log-odds. By doing some calculus (see the exercises below), we can work out the partial derivative of  $p(X)$  with respect to  $X_j$ , but this will *not* turn out to equal  $\beta_j$ . Because  $P(X)$  is not a linear function of  $X$ , the derivative varies with  $X$ , which makes things fairly complicated. There are two main approaches for dealing with this problem. One is to evaluate the derivative at a “typical” value of  $X$  such as the sample mean. Another is to use the “divide by 4 rule.” This rule says that if we increase  $X_j$  by one unit,  $P(X)$  will change by *no more than*  $\beta_j/4$ . In the following exercise, you'll derive this rule.

## Exercise #2

- (a) Analyze the function  $\Lambda(z)$ : calculate its derivative, and its limits as  $z \rightarrow -\infty$  and  $+\infty$ . What values can this function take? Is it increasing? Decreasing? Explain.
- (b) Use the chain rule and your answer to (a) to find the partial derivative of  $\Lambda(X'\beta)$  with respect to  $X_j$ .
- (c) What is the maximum value of the *derivative* of  $\Lambda(z)$ ? At what value of  $z$  does it occur?
- (d) Use your answers to parts (a), (b) and (c) to justify the “divide by 4 rule.”
- (e) The “divide by 4 rule” provides an upper bound on the effect of  $X_j$  on  $P(X)$ . When is this upper bound close to the derivative you calculated in part (c)?

## Solution to Exercise #2

- (a) The function  $\Lambda$  takes values between 0 and 1. When  $z = 0$ ,  $\Lambda(z) = e^0/(1 + e^0) = 1/2$ . As  $z \rightarrow \infty$ ,  $\Lambda(z) \rightarrow 1$  and as  $z \rightarrow -\infty$ ,  $\Lambda(z) \rightarrow 0$ . We calculate its derivative using the quotient rule as follows

$$\frac{d\Lambda(z)}{dz} = \frac{e^z(1 + e^z) - e^z e^z}{(1 + e^z)^2} = \frac{e^z}{(1 + e^z)^2}$$

Since  $e^z$  is always greater than zero, the derivative is always positive so  $\Lambda(z)$  is strictly increasing.

- (b) The key is to treat the linear predictor  $X'\beta$  as a function of  $X_j$ , namely

$$f(X_j) = X'\beta = \beta_0 + \beta_1 X_1 + \cdots + \beta_j X_j + \beta_{j+1} X_{j+1} + \cdots + \beta_p X_p$$

Now, by the chain rule we have

$$\frac{\partial \Lambda(X'\beta)}{\partial X_j} = \frac{\partial \Lambda(f(X_j))}{\partial X_j} = \frac{\exp(X'\beta)}{[1 + \exp(X'\beta)]^2} \frac{\partial f(X_j)}{\partial X_j} = \frac{\beta_j \exp(X'\beta)}{[1 + \exp(X'\beta)]^2}$$

- (c) To find the value of  $z$  that maximizes the first derivative, we take the *second* derivative of  $\Lambda$  as follows

$$\begin{aligned}\frac{d^2\Lambda(z)}{dz} &= \frac{e^z(1+e^z)^2 - 2e^z(1+e^z)e^z}{(1+e^z)^4} = \frac{e^z(1+2e^z+e^{2z}) - 2e^{2z}(1+e^z)}{(1+e^z)^4} \\ &= \frac{e^z + 2e^{2z} + e^{3z} - 2e^{2z} - 2e^{3z}}{(1+e^z)^4} = \frac{e^z - e^{3z}}{(1+e^z)^4} \\ &= \frac{e^z(1-e^{2z})}{(1+e^z)^4} = \frac{e^z(1+e^z)(1-e^z)}{(1+e^z)^4} = \frac{e^z(1-e^z)}{(1+e^z)^3}\end{aligned}$$

Thus, the first order condition is  $e^z(1-e^z) = 0$ . Since  $e^z$  cannot equal zero for any  $z$ , the only way for this equation to be satisfied is if  $e^z = 1$  which occurs precisely when  $z = 0$ . Substituting into our expression from (a), we find that the derivative of  $\Lambda(z)$  at  $z = 0$  is  $e^0/(1+e^0)^2 = 1/(1+1)^2 = 1/4$ .

- (d) From part (a), we know that the derivative of  $\Lambda(z)$  equals  $e^z/(1+e^z)^2$  which is always positive. From part (c) we know that this derivative is *at most*  $1/4$ . Therefore, the partial derivative of  $\Lambda(X'\beta)$  with respect to  $X_j$  is *at most*  $\beta_j \times 1/4 = \beta_j/4$ .
- (e) When  $X'\beta \approx 0$  it follows that  $\exp(X'\beta)/[1 + \exp(X'\beta)]^2 \approx 1/4$  so the “divide by four” rule gives a good approximation to the actual derivative.

## Latent Variable Interpretation of Logistic Regression

Have them simulate some data from a logistic regression and use it to show all the R commands for running the regression, plotting, etc. Then Thursday’s lab can be about the well-switching example!

## Running a Logistic Regression in R