Fitting logistic regression models

Announcements

- Office hour times:
  - Monday 3 - 4 (sign up for 15-minute slots)
  - Wednesday 11 - 12 (15-minute slots)
  - Wednesday 12 - 12:45 (drop-in)
  - Thursday 1 - 2 (drop-in)
- Homework 1 and Challenge Assignment 1 released on course website

Course components

- Regular homework assignments
  - Practice material from class
- Challenge assignments
  - Learn additional material related to course
- 2 take-home exams
  - Demonstrate knowledge of theory and methodology
  - No final exam!
- 2 projects
  - Apply material to real data and real research questions

Assigning grades: specifications grading

To get a **B** in the course:

- Receive credit for at least 5 homework assignments
- Master one project
- Master at least 80% of the questions on both exams

To get an **A** in the course:

- Receive credit for at least 5 homework assignments
- Master both projects
- Master at least 80% of the questions on both exams
- Master at least 2 challenge assignments

Late work and resubmissions

- You get a bank of **5** extension days. You can use 1--2 days on any assignment, exam, or project.
- No other late work will be accepted (except in extenuating circumstances!)
- "Not yet mastered" challenge questions, exams, and projects may be resubmitted once

Recap: three ways of fitting linear regression models

- Minimize SSE, via derivatives of $\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_{i,1} - \cdots - \beta_k X_{i,k})^2$
- Minimize $||Y - \widehat{Y}||$ (equivalent to minimizing SSE)
- Maximize likelihood (for *normal* data, equivalent to minimizing SSE)

> Which of these three methods, if any, is appropriate for fitting a logistic regression model? Do any changes need to be made for the logistic regression setting?
>
> Discuss with your neighbor for 2--3 minutes.

Maximum likelihood for logistic regression

$$Y_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_k X_{i,k}$$

Suppose we observe independent samples $(X_1, Y_1), \ldots, (X_n, Y_n)$. Write down the likelihood function

$$L(\beta) = \prod_{i=1}^{n} f(Y_i; \beta)$$

for the logistic regression problem. Take 2--3 minutes, then we will discuss as a class.

Maximum likelihood for logistic regression

$L(\beta) =$

> I want to choose β to maximize $L(\beta)$. What are the usual steps to take?

Initial attempt at maximizing likelihood

$$L(\beta) = \prod_{i=1}^{n} p_i^{Y_i}(1 - p_i)^{1-Y_i}$$

$$\ell(\beta) =$$

Iterative methods for maximizing likelihood

Fisher scoring

Fisher scoring for logistic regression

Practice question: Fisher scoring

Suppose that $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i$, and we have

$$\beta^{(r)} = \begin{bmatrix} -3.1 \\ 0.9 \end{bmatrix}, \qquad U(\beta^{(r)}) = \begin{bmatrix} 9.16 \\ 31.91 \end{bmatrix},$$

$$\mathcal{I}(\beta^{(r)}) = \begin{bmatrix} 17.834 & 53.218 \\ 53.218 & 180.718 \end{bmatrix}$$

Use the Fisher scoring algorithm to calculate $\beta^{(r+1)}$ (you may use R or a calculator, you do not need to do the matrix arithmetic by hand). Take ~ 5 minutes, then we will discuss.

Alternative to Fisher scoring: gradient ascent

$$Y_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_k X_{i,k}$$

Choose $\beta = (\beta_0, \ldots, \beta_k)^T$ to maximize $L(\beta)$.

Gradient ascent:

Motivation for gradient ascent: walking uphill

Practice question: gradient ascent

Suppose that $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i$, and we have

$$\beta^{(r)} = \begin{bmatrix} -3.1 \\ 0.9 \end{bmatrix}, \qquad U(\beta^{(r)}) = \begin{bmatrix} 9.16 \\ 31.91 \end{bmatrix}$$

- Use gradient ascent with a learning rate (aka step size) of $\gamma = 0.01$ to calculate $\beta^{(r+1)}$.
- The actual maximum likelihood estimate is $\widehat{\beta} = (-3.360, 1.174)$. Does one iteration of gradient ascent or Fisher scoring get us closer to the optimal $\widehat{\beta}$?
- Discuss in pairs for 2--3 minutes.