Homework 7

Due: 11:59pm, Thursday, April 14

Instruction: Please scan or typeset your solutions and upload them as a single pdf file to Canvas. Do not just take a picture of your solutions.

0. Readings (recommended) Section 4.3 in *An Introduction to Statistical Learning with Applications in R*, 2nd Edition, `https://www.statlearning.com`

We consider the logistic regression model for the `Default` data set. (see Notes 8)

1. Let y_i be the default, x_{i1} be the student factor, x_{i2} be the balance, and x_{i3} be the income, assume $Y_i \sim \text{bernoulli}(p_i)$ with

$$p_i = P(Y_i = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}}, \quad i = 1, \ldots, n$$

Write down the pmf $f(y_i)$.

2. Write down the joint distribution $f(y_1, \ldots, y_n)$.

3. What is the likelihood function $L(\beta_0, \beta_1, \beta_2, \beta_3)$.

4. Write down the log likelihood function, $l(\beta_0, \beta_1, \beta_2, \beta_3) = \log L(\beta_0, \beta_1, \beta_2, \beta_3)$, and negative log likelihood function $-l(\beta_0, \beta_1, \beta_2, \beta_3)$.

5. The maximum likelihood estimators of $\beta_0, \beta_1, \beta_2, \beta_3$ are

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = \text{argmax}\, L(\beta_0, \beta_1, \beta_2, \beta_3)$$

explain that it is equivalent to the followings

$$\begin{aligned}(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) &= \text{argmax}\, l(\beta_0, \beta_1, \beta_2, \beta_3) \\ &= \text{argmin}\, -l(\beta_0, \beta_1, \beta_2, \beta_3)\end{aligned}$$

6. In class, we use Newton-Raphson iteration to obtain the estimates

$$\beta^{(t+1)} = \beta^{(t)} + (X^T D(\beta^{(t)}) X)^{-1} X^T (y - p(\beta^{(t)})), \quad t = 0, 1, \ldots,$$

carry out this computation in `R`, what is your $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$?

7. How do you interpret $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$?

8. Given your estimates, what is your default prediction of a person who is not a student, and has balance of 900 and income of 20,000?

9. Use `glm` command, what is your $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$? Are they the same as your answers in Question 6?

10. (Optional) Derive the Newton-Raphson iteration.