3. Separation in Logistic Regression

(a)

3/3 points (graded)

We consider a 1-dimensional logistic regression problem, i.e., assume that data $X_i \in \mathbb{R}, i = 1, \ldots, n$ is given and that get independent observations of

$$Y_i | X_i \sim \text{Ber}\left(\frac{\mathbf{e}^{\beta X_i}}{1 + \mathbf{e}^{\beta X_i}}\right),$$

where $\beta \in \mathbb{R}$.

Moreover, recall that the associated log likelihood for β is then given by

$$\ell(\beta) = \sum_{i=1}^{n} \left(Y_i X_i \beta - \ln\left(1 + \exp\left(X_i \beta\right)\right)\right)$$

Calculate the first and second derivate of ℓ. *Instructions: The summation $\sum_{i=1}^{n}$ is already placed to the left of the answer box. Enter the summands in terms of β, X_i (enter "X_i") and Y_i (enter "Y_i").*

$\ell'(\beta) = \sum_{i=1}^{n}$ Y_i*X_i-exp(X_i*beta)*X_i/(1+exp(X_i*beta)) ✔ **Answer:** (Y_i * X_i) - (X_i / (1+exp(-X_i * beta)))

$Y_i \cdot X_i - \frac{\exp(X_i \cdot \beta) \cdot X_i}{1+\exp(X_i \cdot \beta)}$

$\ell''(\beta) = \sum_{i=1}^{n}$ -exp(X_i*beta)*X_i*X_i/(1+exp(X_i*beta))^2 ✔ **Answer:** -X_i^2 * exp(-X_i * beta) / (1+exp(-X_i*beta))^2

$-\frac{\exp(X_i \cdot \beta) \cdot X_i \cdot X_i}{(1+\exp(X_i \cdot \beta))^2}$

What can you conclude about $\ell'(\beta)$?

- ( ) ℓ' is neither increasing nor decreasing on the whole of $\mathbb{R}$.
- (•) ℓ' is strictly decreasing.
- ( ) ℓ' is strictly increasing.

✔

Solution:

The first derivative is given by

$$\begin{aligned}\ell'(\beta) = & \sum_{i=1}^{n}\left(Y_iX_i - X_i\frac{\mathbf{e}^{X_i\beta}}{1+\mathbf{e}^{X_i\beta}}\right)\\ = & \sum_{i=1}^{n}\left(Y_iX_i - X_i\frac{1}{1+\mathbf{e}^{-X_i\beta}}\right)\end{aligned}$$

The second derivative is

$$\ell''(\beta) = -\sum_{i=1}^{n} X_i^2 \frac{\mathbf{e}^{-X_i\beta}}{\left(1+\mathbf{e}^{-X_i\beta}\right)^2}.$$

Since $X_i^2 > 0$ and $\mathbf{e}^{-X_i\beta} > 0$ for all β, $\ell'(\beta)$ is strictly decreasing. Note that this also means that $\ell(\beta)$ is strictly concave.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

(b)

3/3 points (graded)

Imagine we are given the following data ($n = 2$):

$$X_1 = 0 \quad Y_1 = 0$$
$$X_2 = 1 \quad Y_2 = 1$$

In order to give the maximum likelihood estimator, we want to solve

$$\ell'(\beta) = 0$$

for the given data.

First, we rewrite this as

$$\ell'(\beta) = f(\beta) + g,$$

where

$$f(\beta) = -\sum_{i=1}^{n} X_i \frac{1}{1+\mathbf{e}^{-X_i\beta}}.$$

and g is some appropriate value.

What is the range of $f(\beta)$?

- ( ) $\mathbb{R}$
- ( ) $\mathbb{R}_{<0} = \{r \in \mathbb{R} : r < 0\}$
- (•) $(-1, 0)$, the unit open interval
- ( ) $\{-1, 0\}$, the set containing two values, -1 and 0

✔

What is g?

1

✔ **Answer:** 1

What can you conclude about the solution β?

- ( ) $\beta = 1$.
- ( ) $\beta = 0$.
- (•) There is no β that solves $\ell'(\beta) = 0$.

All $\beta \in \mathbb{R}$ solve $\ell'(\beta) = 0$.

Solution:

Given $X_1 = 0$, $X_2 = 1$, we can plug these values into the expression for ℓ':

$$\ell'(\beta) = 1 - \frac{1}{1 + e^{-\beta}}$$

$$f(\beta) = -\frac{1}{1 + \mathbf{e}^{-\beta}},$$

which has range $(-1, 0)$.

On the other hand,

$$g = 1.$$

This means that the equation $\ell'(\beta) = 0$ will not have a solution on $\mathbb{R}$. In fact, if we were to run an iterative maximization algorithm, β would converge to $+\infty$, which is also what would achieve

$$\lim_{\beta \to \infty} \ell'(\beta) = 0.$$

Submit

You have used 2 of 3 attempts

Answers are displayed within the problem

(c)

5/5 points (graded)

The problem you encountered in part (b) is called **separation** . It occurs when the Y_i can be perfectly recovered by a linear classifier, i.e., when there is a β such that

$$\begin{aligned} X_i\beta > 0 &\implies Y_i = 1, \\ X_i\beta < 0 &\implies Y_i = 0. \end{aligned}$$

In order to avoid this behavior, one option is to use a prior on β. Let us investigate what happens if we assume that β is drawn from a $N(0,1)$ distribution, i.e.,

$$P(\beta, Y|X) = P(\beta) \prod_{i=1}^{n} P(Y_i|X_i, \beta)$$

What is the joint log likelihood $\tilde{\ell}(\beta)$ of this Bayesian model? Again, for simplicity, let's plug in $(X_1, Y_1) = (0,0)$ and $(X_2, Y_2) = (1,1)$. *(Try to work out the general formula on your own. It will also be provided in the solution.)*

$\tilde{\ell}(\beta) =$

-ln(2*pi)/2-ln(2)+beta-beta^2/2-ln(1+e^beta)

Answer: ln(1/sqrt(2*pi)) - (beta^2)/2 - ln(2) + beta - ln(1+exp(beta))

$$-\frac{\ln(2\cdot\pi)}{2} - \ln(2) + \beta - \frac{\beta^2}{2} - \ln\left(1 + e^{\beta}\right)$$

Now, we want to find the maximum a posteriori probability estimate, which is obtained by finding β such that $\tilde{\ell}(\beta) = 0$. To this end, calculate the first and second derivative $\tilde{\ell}'(\beta)$ and $\tilde{\ell}"(\beta)$.

$\ell'(\beta) =$

1-beta-1/(1+e^(-beta))

$1 - \beta - \frac{1}{1+e^{-\beta}}$

✔ **Answer:** -beta + 1 - (1/(1+exp(-beta)))

$\ell''(\beta) =$

-1-e^(-beta)/(1+e^(-beta))^2

$-1 - \frac{e^{-\beta}}{(1+e^{-\beta})^2}$

✔ **Answer:** -1 - (exp(-beta) / (1+exp(-beta))^2)

What can you conclude about $\tilde{\ell}'(\beta)$?

- ○ $\tilde{\ell}'$ is neither increasing nor decreasing on the whole of $\mathbb{R}$.
- ◉ $\tilde{\ell}'$ is strictly decreasing.
- ○ $\tilde{\ell}'$ is strictly increasing.

✔

Given the same data as in (b), what can you say about the existence of a solution?

- ○ Applying the same arguments as in (b), we see that there is no optimal β.
- ◉ Modyfing the notation of f in (b) accordingly, we see that f now ranges over all of $\mathbb{R}$, hence there is a solution.

✔

Solution:

The joint log likelihood is given by

$$\begin{aligned}\tilde{\ell}(\beta) = &\ \ln(P(\beta)) + \sum_{i=1}^{n} \ln(P(Y_i|X_i,\beta)) \\ = &\ \ln\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\beta^2}{2} + \sum_{i=1}^{n} (Y_i X_i \beta - \ln(1+\exp(X_i\beta)))\end{aligned}$$

We obtain the first and second derivatives as before,

$$\tilde{\ell}'(\beta) = -\beta + \sum_{i=1}^{n} \left(Y_i X_i - X_i \frac{1}{1+\mathbf{e}^{-X_i\beta}}\right)$$

$$\tilde{\ell}''(\beta) = -1 - \sum_{i=1}^{n} X_i^2 \frac{\mathbf{e}^{-X_i\beta}}{(1+\mathbf{e}^{-X_i\beta})^2}.$$

Using the same notation as before, if we define

$$f(\beta) = -\beta - \sum_{i=1}^{n} X_i \frac{1}{1+\mathbf{e}^{-X_i\beta}},$$

plugging in the data yields

$$f(\beta) = -\beta - \frac{1}{1+\mathbf{e}^{-X_i\beta}},$$

which is a strictly decreasing function with

$$\begin{aligned}\lim_{\beta\to-\infty} f(\beta) = &\ +\infty \\ \lim_{\beta\to+\infty} f(\beta) = &\ -\infty,\end{aligned}$$

so its range is $\mathbb{R}$. Hence, $f(\beta) = g$ can always be solved.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 7 Generalized Linear Models:Homework 11 / 3. Separation in Logistic Regression

Add a Post

Show all posts ▼ by recent activity ▼

? [Staff]I think something is wrong with grader of Part-1 of C
Could you please check the grader of part-1 of C. 11

Range of $f(\beta)$ in (b) 4