

MATH 680: EXERCISE 3  
A BINARY REGRESSION PROBLEM

# The Binary Regression Model

The binary regression model is a model for binary outcome data formed using predictor values. Suppose that, for  $i = 1, \dots, n$ ,

$$Y_i \sim \text{Bernoulli}(\mu_i(\beta))$$

are an independent sample of binary outcomes, where

$$\mu_i(\beta) \equiv \mu(\beta; \mathbf{x}_i) = \mathbb{E}[Y_i | \mathbf{x}_i]$$

and  $\beta$  is a  $(p \times 1)$  parameter vector. The log-likelihood is then

$$\begin{aligned} \ell(\beta) &= \sum_{i=1}^n [y_i \log \mu_i(\beta) + (1 - y_i) \log(1 - \mu_i(\beta))] \\ &= \sum_{i=1}^n \left[ y_i \log \left( \frac{\mu_i(\beta)}{1 - \mu_i(\beta)} \right) + \log(1 - \mu_i(\beta)) \right] \end{aligned}$$

# The Binary Regression Model

$$\begin{aligned}\dot{\ell}(\beta) &= \sum_{i=1}^n \left[ y_i \frac{\dot{\mu}_i(\beta)}{\mu_i(\beta)(1 - \mu_i(\beta))} - \frac{\dot{\mu}_i(\beta)}{1 - \mu_i(\beta)} \right] \\ &= \sum_{i=1}^n \left[ \frac{(y_i - \mu_i(\beta))}{V(\mu_i(\beta))} \right] \dot{\mu}_i(\beta)\end{aligned}$$

say, where

$$V(\mu_i(\beta)) = \mu_i(\beta)(1 - \mu_i(\beta))$$

is the variance of  $Y_i$ . In  $\dot{\ell}(\beta)$ ,

$$\dot{\mu}_i(\beta)$$

is a  $(p \times 1)$  vector.

# The Binary Regression Model

$$\begin{aligned} \ddot{\ell}(\beta) = \sum_{i=1}^n & \left[ \frac{(y_i - \mu_i(\beta))}{V(\mu_i(\beta))} \ddot{\mu}_i(\beta) \right. \\ & \left. - \left( \frac{(y_i - \mu_i(\beta))(1 - 2\mu_i(\beta)) + V(\mu_i(\beta))}{\{V(\mu_i(\beta))\}^2} \right) \dot{\mu}_i(\beta) \dot{\mu}_i(\beta)^\top \right] \end{aligned}$$

In this expression

$$\ddot{\mu}_i(\beta)$$

is a  $(p \times p)$  matrix.

# The Binary Regression Model

Note that

$$\mathbb{E}[Y_i|\mathbf{x}_i] = \mu_i(\beta)$$

so therefore<sup>†</sup>

$$\mathcal{I}_n(\beta) = -\mathbb{E}[\ddot{\ell}(\beta)] = \sum_{i=1}^n \frac{\dot{\mu}_i(\beta)\dot{\mu}_i(\beta)^\top}{V(\mu_i(\beta))}$$

so therefore Fisher scoring can proceed using the update

$$\beta^{(t+1)} = \beta^{(t)} + \{\mathcal{I}_n(\beta^{(t)})\}^{-1} \dot{\ell}(\beta^{(t)}).$$

---

<sup>†</sup>Recall: need the expectation over  $Y_i$  with the  $\mathbf{x}_i$  fixed constants

# The Binary Regression Model

It is common to use a *link function*  $l$  to capture the influence of predictors  $\mathbf{x}_i$  on response  $Y_i$ . We write

$$l(\mu_i(\beta)) = \mathbf{x}_i\beta$$

for  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ . Hence

$$\mu_i(\beta) = l^{-1}(\mathbf{x}_i\beta) = h(\mathbf{x}_i\beta)$$

say, and

$$\dot{\mu}_i(\beta) = \dot{h}(\mathbf{x}_i\beta)\mathbf{x}_i^\top \quad (p \times 1)$$

$$\ddot{\mu}_i(\beta) = \ddot{h}(\mathbf{x}_i\beta)\mathbf{x}_i^\top \mathbf{x}_i \quad (p \times p)$$

Here,  $h(\cdot)$  is regarded as a function of a **scalar** argument, so  $\dot{h}(\cdot)$  and  $\ddot{h}(\cdot)$  are also scalar quantities

$$\dot{h}(t) = \frac{dh(t)}{dt} \qquad \ddot{h}(t) = \frac{d^2h(t)}{dt^2}.$$

# The Binary Regression Model

## Task

1. Write code to implement Newton's method to find the maximum likelihood estimator of  $\beta$  in a binary regression model.
2. Test your code on the data set

`BinaryData.csv`

using the **logistic** link function

$$l(t) = \log \left( \frac{t}{1-t} \right)$$

In the data set, the columns are

`Y X1 X2 X3`

where `X1` is used to represent an intercept term.

3. Develop a Gauss-Newton or Quasi-Newton strategy for analyzing these data.