Maximum likelihood estimation for regression models

Ciaran Evans

Warmup

Work on the warmup activity (handout), then we will discuss as a class.

Warmup

Suppose that we have independent observations $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ from the model

$$Y_i | \mathbf{x}_i \sim N(\mathbf{x}_i^T \beta, \sigma_i^2).$$

Suppose that the variances $\sigma_1^2, \ldots, \sigma_n^2$ are known. Show that the maximum likelihood estimator of β minimizes the weighted sum of squares

$$WSS(\beta) = \sum_{i=1}^{n} w_i (Y_i - \mathbf{x}_i^T \beta)^2 = (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{W} (\mathbf{y} - \mathbf{X}\beta)$$

Maximum likelihood estimation and logistic regression

Let $(\mathbf{x}_1, Y_1), ..., (\mathbf{x}_n, Y_n)$ be iid samples from the model

$$Y_i | \mathbf{x}_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i^T \beta$$

where the distribution of $\mathbf{x}_i$ does not depend on β.

$$\ell(\beta | \mathbf{y}, \mathbf{X}) \overset{\text{(up to a constant)}}{=} \sum_{i=1}^{n} \left\{ Y_i \mathbf{x}_i^T \beta - \log(1 + e^{\mathbf{x}_i^T \beta}) \right\}$$

$$U(\beta) = \frac{\partial \ell}{\partial \beta} = \mathbf{X}^T(\mathbf{y} - \mathbf{p})$$

Newton's method

- Want β^* such that $U(\beta^*) = \mathbf{0}$
- Begin with initial estimate $\beta^{(0)}$
- Iterative updates:

$$\beta^{(r+1)} = \beta^{(r)} - \left(\mathbf{H}(\beta^{(r)})\right)^{-1} U(\beta^{(r)})$$

The Hessian

$$U(\beta) = \frac{\partial}{\partial \beta} \ell(\beta | \mathbf{y}, \mathbf{X}) = \mathbf{X}^T (\mathbf{y} - \mathbf{p})$$

$$\mathbf{H}(\beta) = \frac{\partial}{\partial \beta} U(\beta) = \frac{\partial}{\partial \beta} \mathbf{X}^T (\mathbf{y} - \mathbf{p}) = \left(-\frac{\partial \mathbf{p}}{\partial \beta} \right) \mathbf{X}$$

$$\frac{\partial \mathbf{p}}{\partial \beta} = \begin{bmatrix} \frac{\partial p_1}{\partial \beta} & \frac{\partial p_2}{\partial \beta} & \cdots & \frac{\partial p_n}{\partial \beta} \end{bmatrix} \in \mathbb{R}^{(k+1) \times n}$$

Putting everything together

Want to maximize the log likelihood $\ell(\beta|\mathbf{y}, \mathbf{X})$.

Class activity

Work on the class activity:

https://sta711-s26.github.io/class_activities/ca_07_2.html

Submit your work on Canvas.