Maximum likelihood estimation for regression models

Ciaran Evans

Maximum likelihood estimation and linear regression

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

$$Y_i|\mathbf{x}_i \sim N(\mu_i, \sigma^2)$$
$$\mu_i = \mathbf{x}_i^T \beta$$

where the distribution of $\mathbf{x}_i$ does not depend on β or σ^2.

$$L(\beta, \sigma^2|\mathbf{y}, \mathbf{X}) =$$

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.

$$L(\beta|\mathbf{y}, \mathbf{X}) \propto \prod_{i=1}^{n} f(Y_i|\mathbf{x}_i, \beta) =$$

Maximizing

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

$$\frac{\partial \ell}{\partial \beta} =$$

Score

Definition (score): Let $\mathbf{y} = (Y_1, ..., Y_n)$ be a sample of n observations from some distribution with parameter vector $\boldsymbol{\theta}$. Let $L(\boldsymbol{\theta}|\mathbf{y})$ be the likelihood function, and $\ell(\boldsymbol{\theta}|\mathbf{y}) = \log L(\boldsymbol{\theta}|\mathbf{y})$ the log-likelihood.

The **score**, which we will denote $U(\boldsymbol{\theta})$, is the gradient of the log-likelihood with respect to $\boldsymbol{\theta}$:

$$U(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta}|\mathbf{y}).$$

Example: For logistic regression: $U(\beta) = \mathbf{X}^T(\mathbf{y} - \mathbf{p})$

Question: How would I solve $\mathbf{X}^T(\mathbf{y} - \mathbf{p}) = \mathbf{0}$?

Newton's method

We want to find β^* such that $U(\beta^*) = \mathbf{0}$. Issue: no closed form solution!

Idea: Approximate $U(\beta^*)$ with a first-order Taylor expansion:

$$U(\beta^*) \approx$$

Newton's method

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

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

The Hessian

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

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

Putting everything together

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