Fitting Logistic Regression

Wednesday, April 19, 2017 8:47 PM

$$\ell(y, x; \beta) = -y x^T \beta + \log\left(1 + e^{x^T\beta}\right)$$

$$\frac{\partial}{\partial \beta} \ell(y, x; \beta) = -y x + \frac{e^{x^T\beta} \cdot x}{1 + e^{x^T\beta}} = -y x + \text{logit}^{-1}(x^T\beta) \cdot x$$

$$= (p - y) x \quad \text{if} \quad p = \mathbb{P}\{Y = 1 \mid X = x, \beta\}$$

$$\frac{\partial^2}{\partial \beta \partial \beta^T} \ell(y, x; \beta) = \frac{e^{x^T\beta} x x^T}{1 + e^{x^T\beta}} - \frac{e^{2x^T\beta} x x^T}{(1 + e^{x^T\beta})^2} = \frac{e^{x^T\beta}}{(1 + e^{x^T\beta})^2} x x^T$$

$$= p(1-p) x x^T \succeq 0$$ so ℓ is convex!

Empirical risk: $R_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, x_i; \beta)$

$$\frac{\partial}{\partial \beta} R_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} (p_i - y_i) x_i = \frac{1}{n} X^T r \quad \text{where}$$

$$p_i = \text{logit}^{-1}(x_i^T \beta), \quad r_i = p_i - y_i$$

$$\frac{\partial^2}{\partial \beta \partial \beta^T} R_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} p_i (1 - p_i) x_i x_i^T = \frac{1}{n} X^T \underset{\text{diagonal}}{W} X$$

$$W_{ii} = p_i (1 - p_i)$$

OPT

Newton-Raphson

until convergence criteria

$$\beta_{t+1} \leftarrow \beta_t - H^{-1} g$$

Hessian ↳ gradient at β_t

Idea: Approximate $R_n(\beta)$ by local quadratic

$$R_n(\beta) \approx R_n(\beta_t) + g^T(\beta-\beta_t) + \frac{1}{2}(\beta-\beta_t)^T H(\beta-\beta_t)$$

↳ argmin $= \beta_t - H^{-1}g$.

Logistic: $H^{-1}g = \underbrace{(X^TWX)^{-1}X^Tr}_{\text{weighted least squares}}$

$\boxed{\text{Newton Raphson} \longrightarrow \text{iteratively re-weighted least squares}}$