

Fitting Logistic Regression

Wednesday, April 19, 2017 8:47 PM

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

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

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

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

$$= p(1-p) x x^T \geq 0 \text{ so } \mathcal{L} \text{ is convex!}$$

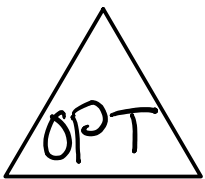
$$\text{Empirical risk: } R_n(\beta) = \frac{1}{n} \sum_{i=1}^n \ell(y_i, x_i; \beta)$$

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

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

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

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



Newton-Raphson

until convergence criteria

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

Hessian $\hat{\tau}$ gradient at β_*

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

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

$$\hookrightarrow \text{argmin} = \beta_* + H^{-1}g.$$

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

Newton Raphson \rightarrow iteratively re-weighted least squares