HW2

$\beta \to p \times i$

1. $\pi_i = P(Y_i = 1 \mid X_i = x_i)$, $Y_i \in (0, 1)$

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \log\left(\frac{P(Y_i=1 \mid X_i)}{1-P(Y_i=1 \mid X_i)}\right) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} = X_i^T \beta$$

$Y_1, Y_2, \dots, Y_n \sim \text{Bern}(\pi)$

PDF is $f(y; \pi) = \pi^{y_i}(1-\pi)^{1-y_i}$

$$E[Y_i \mid X_i] = \pi_i$$

$$\log\left(\frac{E[Y_i \mid X_i]}{1 - E[Y_i \mid X_i]}\right) = X_i^T \beta$$

$$e^{X_i^T \beta} = \frac{E[Y_i \mid X_i]}{1 - E[Y_i \mid X_i]}$$

$$e^{X_i^T \beta} - E[Y_i \mid X_i] e^{X_i^T \beta} = E[Y_i \mid X_i]$$

$$e^{X_i^T \beta} = E[Y_i \mid X_i] + E[Y_i \mid X_i] e^{X_i^T \beta}$$

$$e^{X_i^T \beta} = E[Y_i \mid X_i] (1 + e^{X_i^T \beta})$$

$$\frac{e^{X_i^T \beta}}{1 + e^{X_i^T \beta}} = E[Y_i \mid X_i]$$

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i}(1-\pi_i)^{1-y_i} \longrightarrow \text{sub } \frac{e^{X_i^T\beta}}{1+e^{X_i^T\beta}} \text{ for } \pi_i$$

$$\prod_{i=1}^n \left(\frac{e^{X_i^T\beta}}{1+e^{X_i^T\beta}}\right)^{y_i} \left(1 - \frac{e^{X_i^T\beta}}{1+e^{X_i^T\beta}}\right)^{1-y_i}$$

$$\downarrow \quad \frac{1}{1+e^{X_i^T\beta}}$$

$$(e^{x_i^T\beta})^{y_i - (1+e^{x_i^T\beta})}$$

$$\ell(\beta) = \log L(\beta) = \log \prod_{i=1}^{n} \left(\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right)^{y_i} \left(1 - \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right)^{1-y_i}$$

$$\sum_{i=1}^{n} y_i \left(\log \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right) + (1-y_i)\left[\log\left(1 - \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right)\right]$$

$$\sum_{i=1}^{n} y_i \left[\log e^{x_i^T\beta} - \log(1+e^{x_i^T\beta})\right]$$

$$\sum_{i=1}^{n} y_i \left[x_i^T\beta - \log(1+e^{x_i^T\beta})\right] + (1-y_i)\left[\log\left(\frac{1+e^{x_i^T\beta}}{1+e^{x_i^T\beta}} - \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right)\right]$$

$$\sum_{i=1}^{n} y_i \left[x_i^T\beta - \log(1+e^{x_i^T\beta})\right] + (1-y_i)\log\frac{1}{1+e^{x_i^T\beta}}$$

$$\sum_{i=1}^{n} y_i \left[x_i^T\beta - \log(1+e^{x_i^T\beta})\right] + (1-y_i)\left[\log 1 - \log(1+e^{x_i^T\beta})\right]$$

$$\sum_{i=1}^{n} y_i x_i^T\beta \cancel{- y_i \log(1+e^{x_i^T\beta})} - \log(1+e^{x_i^T\beta}) \cancel{+ y_i \log(1+e^{x_i^T\beta})}$$

$$\ell(\beta) = \sum_{i=1}^{n} y_i x_i^T\beta - \log(1+e^{x_i^T\beta})$$

$\downarrow$

Aside: convex optimization problem?

is $-\log(f)$ convex? $\rightarrow$ is hessian ≥ 0?

$$-\log(f) \rightarrow -\sum_{i=1}^{n} y_i x_i^T\beta - \log(1+e^{x_i^T\beta})$$

$$\sum_{i=1}^{n} \log(1+e^{x_i^T\beta}) - y_i x_i^T\beta$$

gradient: $\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \frac{e^{x_i^T\beta} x_i}{1+e^{x_i^T\beta}} - y_i x_i$

gradient: $\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \frac{e^{X_i^T\beta} X_i}{1+e^{X_i^T\beta}} - y_i X_i$

hessian $\frac{\partial^2 l}{\partial \beta^2} = \sum_{i=1}^{n} \frac{X_i^T X_i e^{X_i^T\beta}(1+e^{X_i^T\beta}) - e^{X_i^T\beta} X_i (X_i e^{X_i^T\beta})}{(1+e^{X_i^T\beta})^2}$

$X_i^T X_i e^{X_i^T\beta} + \cancel{X_i^T X_i e^{2X_i^T\beta}} - \cancel{X_i^T X_i e^{2X_i^T\beta}}$

always positive $\rightarrow$ $\hookrightarrow \sum \frac{X_i^T X_i e^{X_i^T\beta}}{(1+e^{X_i^T\beta})^2} \geq 0$

---

actual gradient: $l(\beta) = \sum_{i=1}^{n} y_i X_i^T\beta - \log(1+e^{X_i^T\beta})$

$\downarrow$

$\frac{\partial l}{\partial b} = \sum_{i=1}^{n} y_i X_i^T - \frac{X_i e^{X_i^T\beta}}{1+e^{X_i^T\beta}}$

$\frac{\partial^2 l}{\partial \beta^2} = \sum_{i=1}^{n} 0 - \frac{X_i^T X_i e^{X_i^T\beta}(1+e^{X_i^T\beta}) - X_i e^{X_i^T\beta}(X_i e^{X_i^T\beta})}{(1+e^{X_i^T\beta})} =$

$- \sum_{i=1}^{n} X_i^T (\pi_i(1-\pi_i)) X_i \leq 0$

Update:

Steepest Descent

$\rightarrow \beta_{t+1} = \beta_t - \alpha \frac{\partial l}{\partial \beta} \Rightarrow \beta_{t+1} = \beta_t - \alpha \left[\sum_{i=1}^{n} y_i X_i^T - \frac{X_i e^{X_i^T\beta}}{1+e^{X_i^T\beta}} \right]$

- Newton

$\rightarrow \beta_{t+1} = \beta_t - \left[\frac{\partial^2 l}{\partial \beta^2}\right]^{-1} \frac{\partial l}{\partial \beta}$