Simple ~~B~~ Linear Regression

$$y_i \sim \text{Normal}\left[\sum_j x_{ij}\beta_j,\ \sigma^2\right]$$

Normal density: $\dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\dfrac{1}{2}\left(\dfrac{y-\mu}{\sigma}\right)^2\right\}$

For y_i, $\mu_i = \sum_j x_{ij}\beta_j$

$$\ell(\beta) = \frac{1}{2}\sum_i \frac{\left(y_i - \sum_j x_{ij}\beta_j\right)^2}{\sigma^2} + \text{const}$$

$$\cancel{\frac{\partial \ell}{\partial \beta_k} = -\frac{1}{\sigma^2}\sum_i \sum_j x_{ij}\beta_j \cdot x_{ik}}$$

$$\frac{\partial \ell}{\partial \beta_k} = -\frac{1}{\sigma^2}\sum_i x_{ik}\left(y_i - \sum_j x_{ij}\beta_j\right) = 0$$

$$\sum_i x_{ik}y_i = \sum_i \sum_j x_{ik}x_{ij}\beta_j \quad k=0,\dots,p$$

$$X^T y = X^T X \beta$$

$$\frac{\partial^2 \ell}{\partial\beta_k \partial\beta_m} = \frac{1}{\sigma^2}\sum_i x_{ik}x_{im}$$ Matrix: $\dfrac{1}{\sigma^2}X^T X$

$$I = H = \frac{1}{\sigma^2}X^T X \qquad H^{-1} = I^{-1} = \sigma^2 (X^T X)^{-1}$$

$$P(Y_i = 1 \mid x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

$$\cancel{L = \prod_i e}$$

If $y_i = 1$ $\qquad$ If $y_i = 0$:

$$\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\cdots}} \qquad \frac{1}{1 + e^{\cdots}}$$

Combine: $\dfrac{e^{y_i(\beta_0 + \beta_1 x_i)}}{1 + e^{\beta_0 + \beta_1 x_i}}$

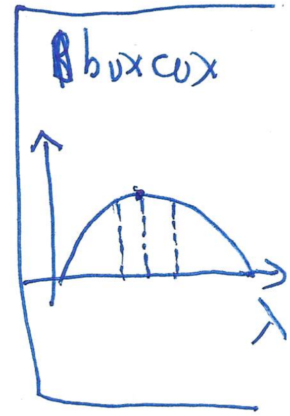
$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} \frac{e^{y_i(\beta_0 + \beta_1 x_i)}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

$$\ell(\beta_0, \beta_1) = -\log L(\beta_0, \beta_1)$$
$$= \sum_{i=1}^{n} \log\left(1 + e^{\beta_0 + \beta_1 x_i}\right) - \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_i)$$

$$\frac{\partial \ell}{\partial \beta_0}, \quad \frac{\partial \ell}{\partial \beta_1}, \quad \frac{\partial^2 \ell}{\partial \beta_0^2}, \ldots$$

$$l = \sum \log\{1+\exp(\beta_0+\beta_1 x_i)\} - \sum y_i(\beta_0+\beta_1 x_i)$$

$$0 = \frac{\partial l}{\partial \beta_0} = \sum \frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} - \sum y_i$$



$$0 = \frac{\partial l}{\partial \beta_1} = \sum \frac{x_i \exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} - \sum y_i x_i$$

$$\frac{\partial l}{\partial \beta_0} = \sum (p_i - y_i) = 0 \qquad \frac{\partial l}{\partial \beta_1} = \sum x_i(p_i - y_i) = 0$$

$p_i \downarrow E(\hat{y}_i)$

$$\sum(\hat{y}_i - y_i) = \sum x_i(\hat{y}_i - y_i) = 0$$

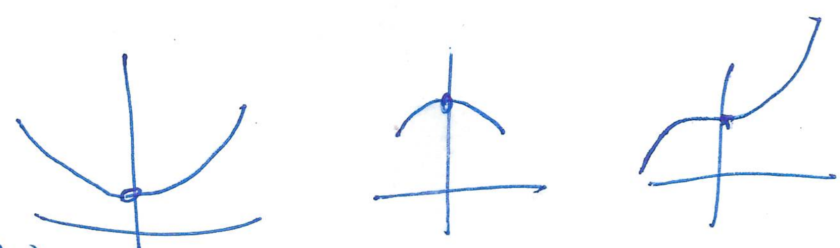
"Fisher scoring algorithm"

$$\frac{\partial^2 l}{\partial \beta_0^2} = \sum p_i(1-p_i) \qquad \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} = \sum x_i p_i(1-p_i) \qquad \frac{\partial^2 l}{\partial \beta_1^2} = \sum x_i^2 p_i(1-p_i)$$

$$H = \begin{pmatrix} \frac{\partial^2 l}{\partial \beta_0^2} & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 l}{\partial \beta_1^2} \end{pmatrix} = \begin{pmatrix} \sum p_i(1-p_i) & \sum x_i p_i(1-p_i) \\ \sum x_i p_i(1-p_i) & \sum x_i^2 p_i(1-p_i) \end{pmatrix}$$

$H \to$ observed information matrix

$\det H \geq 0$ (determinant)



H^{-1}: approx. to covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$