

EXAM 2023

Problem 2

d. For college i , $i = 1, \dots, N$:

$$X_i = \begin{pmatrix} 1 & x_{i1} & \dots & x_{ip} \\ 1 & x_{i1} & \dots & x_{ip} \\ \vdots & \vdots & & \vdots \\ 1 & x_{in} & \dots & x_{ip} \end{pmatrix} \quad \left. \vphantom{\begin{pmatrix} 1 & x_{i1} & \dots & x_{ip} \\ 1 & x_{i1} & \dots & x_{ip} \\ \vdots & \vdots & & \vdots \\ 1 & x_{in} & \dots & x_{ip} \end{pmatrix}} \right\} n_i \quad V_i = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \left. \vphantom{\begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}} \right\} \begin{matrix} y_i \\ n_i - y_i \end{matrix}$$

Model matrix for College i :

Response vector for College i :

Since $\hat{\beta}_{ij} = \hat{\beta}_i$, $\hat{z}_{ij}$ will be the same for all j for a given i - Either all $\hat{z}_{ij}$, $j = 1, \dots, n_i$, are 1, and then $\hat{y}_i = \sum_{j=1}^{n_i} \hat{z}_{ij} = n_i$, or all $\hat{z}_{ij}$ are 0, and then $\hat{y}_i = 0$.

$$\hat{y}_i = \sum_{j=1}^{n_i} \hat{\beta}_{ij} = \sum_{j=1}^{n_i} \hat{\beta}_i = n_i \hat{\beta}_i, \text{ as suggested earlier}$$

in the problem.

e) The predicted acceptance probability is
 $\hat{p}_i = 0.7428$ (Top 10 per < 67.5 , Top 25 per < 80.5 ,
Grad Rate < 97.5)

That gives :

$$\hat{y}_i = n \cdot \hat{p}_i = 1660 \cdot 0.7428 = \underline{\underline{1233}}$$

This is very close to the true $y_i = 1232$.

Tree models are easier to interpret.

Problem 3

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i, \quad \varepsilon_i \sim N(0, w_i \sigma^2)$$

The likelihood of this model is:

$$\begin{aligned} L &\stackrel{\text{ind}}{=} \prod_{i=1}^N \frac{1}{\sqrt{2\pi w_i} \sigma} e^{-\frac{1}{2w_i \sigma^2} \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2} \\ &= (2\pi)^{N/2} \sigma^{-N} \prod_{i=1}^N \frac{1}{\sqrt{w_i}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^N \frac{(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2}{w_i}} \end{aligned}$$

The log-likelihood is then:

$$l = -\frac{N}{2} \log(2\pi) - N \log(\sigma) - \frac{1}{2} \sum_{i=1}^N \log(w_i) - \frac{1}{2\sigma^2} \sum_{i=1}^N \frac{(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2}{w_i}$$

Maximising l w.r.t. the β_j s is the same as
minimising $\sum_{i=1}^N \frac{(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2}{w_i}$ w.r.t. the β_j s.

We have:

$$\begin{aligned} W &= \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_N \end{pmatrix} & (Y - X\beta)^T W^{-1} (Y - X\beta) \\ &\rightarrow W^{-1} = \begin{pmatrix} 1/w_1 & & 0 \\ & \ddots & \\ 0 & & 1/w_N \end{pmatrix} & = (y_1 - x_1^T \beta, y_2 - x_2^T \beta, \dots, y_N - x_N^T \beta) \cdot \begin{pmatrix} 1/w_1 & & 0 \\ & \ddots & \\ 0 & & 1/w_N \end{pmatrix} \begin{pmatrix} y_1 - x_1^T \beta \\ \vdots \\ y_N - x_N^T \beta \end{pmatrix} \\ & & = (y_1 - x_1^T \beta \quad \dots \quad y_N - x_N^T \beta) \begin{pmatrix} \frac{y_1 - x_1^T \beta}{w_1} \\ \vdots \\ \frac{y_N - x_N^T \beta}{w_N} \end{pmatrix} \\ & & = \sum_{i=1}^N \frac{(y_i - x_i^T \beta)^2}{w_i} = \sum_{i=1}^N \frac{(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2}{w_i} = \text{RSS} \end{aligned}$$

b) To minimise RSS w.r.t. β , we differentiate and set equal 0:

$$RSS = (Y - X\beta)^T W^{-1} (Y - X\beta)$$

Recall that: $\frac{\partial a^T b}{\partial b} = a$ and $\frac{\partial b^T A b}{\partial b} = (A^T + A)b$

$$\begin{aligned} RSS &= Y^T W^{-1} Y - Y^T W^{-1} X \beta - \underbrace{\beta^T X^T W^{-1} Y}_{= Y^T W^{-1} X \beta} + \beta^T X^T W^{-1} X \beta \\ &= Y^T W^{-1} Y - 2 \underbrace{Y^T W^{-1} X}_{a^T} \underbrace{\beta}_b + \underbrace{\beta^T X^T W^{-1} X}_{A} \underbrace{\beta}_b \end{aligned}$$

$$\begin{aligned} \rightarrow \frac{\partial RSS}{\partial \beta} &= -2 X^T W^{-1} Y + (X^T W^{-1} X + X^T W^{-1} X) \beta \\ &= -2 X^T W^{-1} Y + 2 X^T W^{-1} X \beta \end{aligned}$$

$$\rightarrow -2 X^T W^{-1} Y + 2 X^T W^{-1} X \hat{\beta} = 0$$

$$\rightarrow X^T W^{-1} X \hat{\beta} = X^T W^{-1} Y$$

$$\rightarrow \hat{\beta} = \underline{\underline{(X^T W^{-1} X)^{-1} X^T W^{-1} Y}}$$

We have:

$$E(Y) = X\beta, \quad Var(Y) = \sigma^2 W$$

$$\begin{aligned} \rightarrow E(\hat{\beta}) &= E\left((X^T W^{-1} X)^{-1} X^T W^{-1} Y\right) \\ &= (X^T W^{-1} X)^{-1} X^T W^{-1} E(Y) \\ &= \underbrace{(X^T W^{-1} X)^{-1} X^T W^{-1} X}_{\underline{I}} \underbrace{X\beta}_{X\beta} = \underline{\underline{\beta}} \end{aligned}$$

$$\begin{aligned} Var(\hat{\beta}) &= Var\left((X^T W^{-1} X)^{-1} X^T W^{-1} Y\right) \\ &= (X^T W^{-1} X)^{-1} X^T W^{-1} \underbrace{Var(Y)}_{\sigma^2 W} W^{-1} X (X^T W^{-1} X)^{-1} \\ &= \sigma^2 (X^T W^{-1} X)^{-1} X^T \underbrace{W^{-1} W}_{\underline{I}} W^{-1} X (X^T W^{-1} X)^{-1} \\ &= \sigma^2 (X^T W^{-1} X)^{-1} \underbrace{X^T W^{-1} X}_{=\underline{I}} (X^T W^{-1} X)^{-1} \\ &= \underline{\underline{\sigma^2 (X^T W^{-1} X)^{-1}}} \end{aligned}$$

c. We have:

$$\underbrace{\frac{y_i}{\sqrt{w_i}}}_{\tilde{y}_i} = \underbrace{\frac{\beta_0}{\sqrt{w_i}}}_{\tilde{\beta}_0} + \sum_{j=1}^p \beta_j \underbrace{\frac{x_{ij}}{\sqrt{w_i}}}_{\tilde{x}_{ij}} + \underbrace{\frac{\varepsilon_i}{\sqrt{w_i}}}_{\tilde{\varepsilon}_i}$$

$$\rightarrow \tilde{y}_i = \tilde{\beta}_0 + \sum_{j=1}^p \beta_j \tilde{x}_{ij} + \tilde{\varepsilon}_i \quad (**)$$

Since the ε_i 's are normal distributed N , and $\tilde{\varepsilon}_i$ is a linear combination of ε_i , the $\tilde{\varepsilon}_i$'s are also normal distributed and independent.

Further:

$$E(\tilde{\varepsilon}_i) = E\left(\frac{\varepsilon_i}{\sqrt{w_i}}\right) = \frac{1}{\sqrt{w_i}} \underbrace{E(\varepsilon_i)}_{=0} = 0$$

$$\text{Var}(\tilde{\varepsilon}_i) = \text{Var}\left(\frac{\varepsilon_i}{\sqrt{w_i}}\right) = \frac{1}{w_i} \cdot \text{Var}(\varepsilon_i) = \frac{1}{w_i} \cdot w_i \sigma^2 = \sigma^2$$

Hence (**) is a linear regression model in standard form.

Let $\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$, then the least squares estimator

$$\text{of } \tilde{\beta} \quad \hat{\tilde{\beta}} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{y}$$

$$\text{We have: } \tilde{X} = W^{-1/2} X, \quad \tilde{y} = W^{-1/2} y$$

$$\begin{aligned} \rightarrow \hat{\tilde{\beta}} &= (X^T W^{-1/2} W^{-1/2} X)^{-1} X^T W^{-1/2} W^{-1/2} y \\ &= (X^T W^{-1} X)^{-1} X^T W^{-1} y \end{aligned}$$

EXAM 2024

Problem 3

a. $P(Y=1|x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}$, $x_1 = \text{no. of pregnancies}$
 $x_2 = \text{glucose level}$

$$\hat{p} = \frac{e^{-5,75 + 0,123 \cdot 2 + 0,037 \cdot 160}}{1 + e^{-5,75 + 0,123 \cdot 2 + 0,037 \cdot 160}} = \underline{\underline{0,602}}$$

$$\frac{e^{\overbrace{-5,505}^{-5,75 + 0,123 \cdot 2 + 0,037 \cdot x_2}}}{1 + e^{-5,75 + 0,123 \cdot 2 + 0,037 \cdot x_2}} < 0,5$$

$$e^{-5,505 + 0,037 x_2} < 0,5 + 0,5 \cdot e^{-5,505 + 0,037 x_2}$$

$$\cancel{0,5} \cdot e^{-5,505 + 0,037 x_2} < \cancel{0,5} \cdot 1$$

$$-5,505 + 0,037 x_2 < \log(1) = 0$$

$$\rightarrow 0,037 x_2 < 5,505$$

$$\rightarrow x_2 < \frac{5,505}{0,037} \approx 149$$

b- This is due to correlation between pressure and one of the other covariates in Model B.