lecture 13: 7

$$\lambda I_p X^T = \lambda X^T I_n$$

$$X^T X X^T + \lambda I_p X^T = X^T X X^T + \lambda X^T I_n$$

$$(X^T X + \lambda I_p) X^T = X^T (X X^T + \lambda I_n)$$

$$\Rightarrow \cancel{(X^T X + \lambda I_p)^{-1} (X^T X + \lambda I_p)} X^T = (X^T X + \lambda I_p)^{-1} X^T (X X^T + \lambda I_n)$$

$$\Rightarrow X^T (X X^T + \lambda I_n)^{-1} = (X^T X + \lambda I_p)^{-1} X^T \cancel{(X X^T + \lambda I_n)(X X^T + \lambda I_n)^{-1}}$$

$$\Rightarrow X^T (X X^T + \lambda I_n)^{-1} = (X^T X + \lambda I_p)^{-1} X^T$$

$$\therefore X^T (X X^T + \lambda I_n)^{-1} Y = (X^T X + \lambda I_p)^{-1} X^T Y$$

lecture 14: 1

$$\arg\min_{b \in \mathbb{R}} \tfrac{1}{2}(b - b_0)^2 + \lambda |b| = \arg_{b \in \mathbb{R}} \tfrac{1}{2} b^2 - b_0 b + \lambda |b|$$

$b \geq 0$, $\tfrac{1}{2} b^2 - b_0 b + \lambda |b| = \tfrac{1}{2} b^2 - (b_0 - \lambda) b$

If $b_0 - \lambda \geq 0$, $\arg\min_{b \geq 0} \tfrac{1}{2} b^2 - (b_0 - \lambda) b = b_0 - \lambda$

If $b_0 - \lambda \leq 0$, $\arg\min_{b \geq 0} \tfrac{1}{2} b^2 - (b_0 - \lambda) b = 0$

$b < 0$, $\tfrac{1}{2} b^2 - b_0 b + \lambda |b| = \tfrac{1}{2} b^2 - (b_0 + \lambda) b$

If $b_0 + \lambda \geq 0$, $\arg\min_{b < 0} \tfrac{1}{2} b^2 - (b_0 - \lambda) b = 0$

If $b_0 + \lambda \leq 0$, $\arg\min_{b < 0} \tfrac{1}{2} b^2 - (b_0 - \lambda) b = b_0 + \lambda$

Therefore,
$$\arg\min_{b \in \mathbb{R}} \tfrac{1}{2}(b - b_0)^2 + \lambda |b| = \begin{cases} b_0 - \lambda & \text{if } b_0 \geq \lambda \\ 0 & \text{if } -\lambda \leq b_0 \leq \lambda \\ b_0 + \lambda & \text{if } b_0 \leq -\lambda \end{cases}$$

$$= \mathrm{sign}(b_0)(|b_0| - \lambda)_+$$

lecture 15: 1

$$y_i = \hat\beta_0 + \hat\beta_1 f_{1i} + \hat\beta_2 f_{2i} + \hat\beta_{12} f_{1i} \cdot f_{2i} + \hat\varepsilon_i$$

① $f_{1i} = f_{2i} = 1$

$y_{11,i} = \hat\beta_0 + \hat\beta_1 + \hat\beta_2 + \hat\beta_{12} + \hat\varepsilon_{11,i}$, where $\hat\varepsilon_{11,i}$ is the residual

based on $f_{1i} = f_{2i} = 1$, we let the # of observation: $\{f_{1i}=1, f_{2i}=1\} := n_{11}$

$\sum_{i: f_{1i}=f_{2i}=1} y_{11,i} = n_{11}(\hat\beta_0 + \hat\beta_1 + \hat\beta_2 + \hat\beta_{12}) + \sum_{i: f_{1i}=f_{2i}=1} \hat\varepsilon_{11,i}$, where $\sum_{i: f_{1i}=f_{2i}=1} \hat\varepsilon_{11,i} = 0$

$\therefore \bar y_{11} = \hat\beta_0 + \hat\beta_1 + \hat\beta_2 + \hat\beta_{12}$

② $f_{1i} = 1, f_{2i} = 0$

$y_{10,i} = \hat\beta_0 + \hat\beta_1 + \hat\varepsilon_{10,i}$, let the # of observation: $\{f_{1i}=1, f_{2i}=0\} := n_{10}$

And we know $\sum_{i: f_{1i}=1, f_{2i}=0} \hat\varepsilon_{10,i} = 0$

Thus, we can get: $\bar y_{10} = \hat\beta_0 + \hat\beta_1$

③ $f_{1i} = 0, f_{2i} = 1$, we get: $\bar y_{01} = \hat\beta_0 + \hat\beta_2$

④ $f_{1i} = f_{2i} = 0$, we get: $\bar y_{00} = \hat\beta_0$

$$\begin{cases} \bar y_{11} = \hat\beta_0 + \hat\beta_1 + \hat\beta_2 + \hat\beta_{12} \\ \bar y_{10} = \hat\beta_0 + \hat\beta_1 \\ \bar y_{01} = \hat\beta_0 + \hat\beta_2 \\ \bar y_{00} = \hat\beta_0 \end{cases} \Rightarrow \hat\beta_{12} = (\bar y_{11} - \bar y_{10}) - (\bar y_{01} - \bar y_{00})$$

lecture 16 : 6

$$Y = X_1 \hat{\beta}_{w,1} + X_2 \hat{\beta}_{w,2} + \hat{\varepsilon}_w$$

$$\Sigma^{-1/2} Y = \Sigma^{-1/2} X_1 \hat{\beta}_{w,1} + \Sigma^{-1/2} X_2 \hat{\beta}_{w,2} + \Sigma^{-1/2} \hat{\varepsilon}_w$$

$\tilde{Y} = \Sigma^{-1/2} Y$, $\tilde{X}_1 = \Sigma^{-1/2} X_1$, $\tilde{X}_2 = \Sigma^{-1/2} X_2$, $\operatorname{cov}(\Sigma^{-1/2} \hat{\varepsilon}_w) = \sigma^2 I_n$

$\therefore$ $\hat{\beta}_{w,1}$, $\hat{\beta}_{w,2}$ is the ols fit of $\tilde{Y}$ on $\tilde{X}_1$, $\tilde{X}_2$

By FWL Theorem, we know:

$$\hat{\beta}_{w,2} = (\hat{X}_2' \hat{X}_2)^{-1} \hat{X}_2' \hat{Y}$$

Where $\hat{X}_2$ is the residual of $\tilde{X}_2$ ols on $\tilde{X}_1$,

$\hat{Y}$ is the residual of $\tilde{Y}$ ols on $\tilde{X}_1$.

$$\hat{X}_2 = \tilde{X}_2 - \tilde{X}_1 (\tilde{X}_1' \tilde{X}_1)^{-1} \tilde{X}_1' \tilde{X}_2$$

$$= \Sigma^{-1/2} X_2 - \Sigma^{-1/2} X_1 (X_1' \Sigma^{-1} X_1)^{-1} X_1' \Sigma^{-1} X_2$$

$\Rightarrow$ $\Sigma^{1/2} \hat{X}_2$ = residual of X_2 ols fit on $X_1 := \tilde{X}_{2,w}$

$\Rightarrow$ $\hat{X}_2 = \Sigma^{-1/2} \tilde{X}_{2,w}$

Similarly, $\hat{Y} = \Sigma^{-1/2} Y - \Sigma^{-1/2} X_1 (X_1' \Sigma^{-1} X_1)^{-1} X_1' \Sigma^{-1} Y$

$\Rightarrow$ $\Sigma^{1/2} \hat{Y}$ = residual of Y wls fit on $X_1 := \tilde{Y}_w$

$\Rightarrow$ $\hat{Y} = \Sigma^{-1/2} \tilde{Y}_w$

$$\therefore \hat{\beta}_{w,2} = (\tilde{X}_{2,w}' \Sigma^{-1} \tilde{X}_{2,w})^{-1} \tilde{X}_{2,w} \Sigma^{-1} \tilde{Y}_w$$

$\therefore$ $\hat{\beta}_{w,2}$ is the wls fit of $\tilde{Y}_w$ on $\tilde{X}_{2,w}$

lecture 16: 7

$$\hat{V}_w = n^{-1}\left(n^{-1}\sum_{i=1}^{n} w_i x_i x_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} w_i^2 \hat{\varepsilon}_{w,i}^2 x_i x_i^T\right)\left(n^{-1}\sum_{i=1}^{n} w_i x_i x_i^T\right)^{-1}$$

where $\hat{\varepsilon}_{w,i} = y_i - x_i^T \hat{\beta}_w$

we let $\tilde{y}_i = w_i^{1/2} y_i$, $\tilde{x}_i = w_i^{1/2} x_i$, $\tilde{y}_i = \tilde{x}_i^T \beta + \tilde{\varepsilon}_i$, $\tilde{\varepsilon}_i \sim N(0, \sigma^2)$

$\tilde{\varepsilon}_i = w_i^{1/2} \hat{\varepsilon}_{w,i} = \tilde{y}_i - \tilde{x}_i^T \hat{\beta}_w$

$$\therefore \hat{V}_w = n^{-1}\left(n^{-1}\sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} \frac{\tilde{\varepsilon}_{w,i}^2}{w_i}(w_i^{1/2} x_i)(w_i^{1/2} x_i)^T\right)\left(n^{-1}\sum_{i=1}^{n} (w_i^{1/2} x_i)(w_i^{1/2} x_i)^T\right)^{-1}$$

$$= n^{-1}\left(n^{-1}\sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 \tilde{x}_i \tilde{x}_i^T\right)\left(n^{-1}\sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^T\right)^{-1}$$

lecture 17 : 5

$y_i \sim Bernoulli(q)$, $x_i | y_i = 1 \sim N(\mu_1, \Sigma)$, $x_i | y_i = 0 \sim N(\mu_2, \Sigma)$

$$p(y_i=1|x_i) = \frac{p(x_i, y_i=1)}{f(x_i)} = \frac{p(x_i|y_i=1)\,p(y_i=1)}{p(x_i|y_i=1)\,p(y_i=1) + p(x_i|y_i=0)\,p(y_i=0)}$$

$$= \frac{q \cdot \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\{-\frac{1}{2}(x_i-\mu_1)^T \Sigma^{-1} (x_i-\mu_1)\}}{q \cdot \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\{-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)\} + (1-q) \cdot \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\{-\frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2)\}}$$

$$= \frac{1}{1 + \frac{1-q}{q} \exp\{-\frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 + \mu_1^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu_1 + \mu_2^T\Sigma^{-1}\mu_2 - \mu_2^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_2)}$$

$$= \frac{e^{\alpha + x_i^T\beta}}{1 + e^{\alpha + x_i^T\beta}} = \frac{1}{1 + e^{-(\alpha + x_i^T\beta)}}$$

$$\Rightarrow \log\frac{q}{1-q} - \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2) + \frac{1}{2}\cdot(2\cdot x_i^T\Sigma^{-1}\mu_1 - 2x_i^T\Sigma^{-1}\mu_2) = \alpha + x_i^T\beta$$

$$\Rightarrow \log\frac{q}{1-q} - \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2) + x_i^T\Sigma^{-1}(\mu_1 - \mu_2) = \alpha + x_i^T\beta$$

$$\therefore \alpha = \log\frac{q}{1-q} - \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2)$$

$$\beta = \Sigma^{-1}(\mu_1 - \mu_2)$$