LECTURE FIVE

Multiple Linear Regression

Data: y_i $\underbrace{x_{i1} \quad x_{i2} \cdots x_{im}}$, $i = 1, \ldots, n$

Model: $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \underbrace{\varepsilon_i}$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Inference: **Prior:** $\beta_0, \ldots, \beta_m, \log \sigma \overset{iid}{\sim} \text{Unif}(-C, C)$

Posterior:

$$f_{\beta_0, \ldots, \beta_m | \text{data}}(\beta_0 \ldots, \beta_m) \propto \left[\frac{S(\beta_0, \ldots, \hat{\beta}_m)}{S(\beta_0, \ldots, \beta_m)}\right]^{\frac{n}{2}} \left.\right\} \text{peaked at the least squares estimator}$$

$$S(\beta_0, \ldots, \beta_m) = \underbrace{\sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_m x_{im}\right)^2}$$

$\hat{\beta}_0, \ldots, \hat{\beta}_m$: least squares estimators

$S(\hat{\beta}_0, \ldots, \hat{\beta}_m)$

Multivariate t - density

$x_1 \ldots x_p$

$$f(x_1 \ldots x_p) \propto \left[\frac{1}{1 + \frac{1}{\nu}\underbrace{(x-\mu)^T \Sigma^{-1} (x-\mu)}_{\text{Quadratic Form}}}\right]^{\frac{\nu+p}{2}}$$

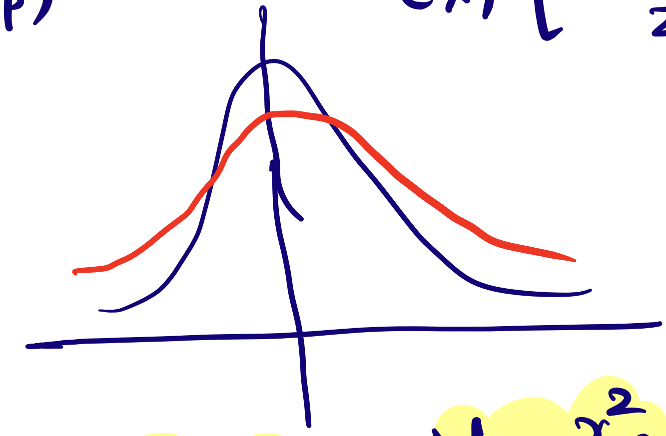
$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$$

$$t_p(\mu, \underset{p \times p \text{ matrix}}{\underset{\downarrow}{\Sigma}}, \nu) \xrightarrow{\nu} \text{scalar degrees of freedom}$$

① Connection to Multivariate Normal

$$N_p(\mu, \Sigma) \longrightarrow p \times p : \text{Covariance Matrix}$$

$$f(x_1 \ldots x_p) = \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$$



μ : mean

Σ : covariance matrix.

Suppose: $X \sim N_p(\mu, \Sigma)$, $V \sim \chi^2_\nu$ independent.

Fact:

$$\boxed{T = \mu + \frac{X-\mu}{\sqrt{V/\nu}} \hookrightarrow t_p(\mu, \Sigma, \nu)}$$

$V \sim \chi^2_\nu$

$V = Z_1^2 + \cdots + Z_\nu^2$

$Z_i \overset{iid}{\sim} N(0,1)$

② $$\boxed{T_j = \mu_j + \frac{X_j - \mu_j}{\sqrt{V/\nu}}}$$

$X_j \sim N(\mu_j, \Sigma(j,j))$

$$T_j \sim t_1(\mu_j, \Sigma(j,j), \nu)$$

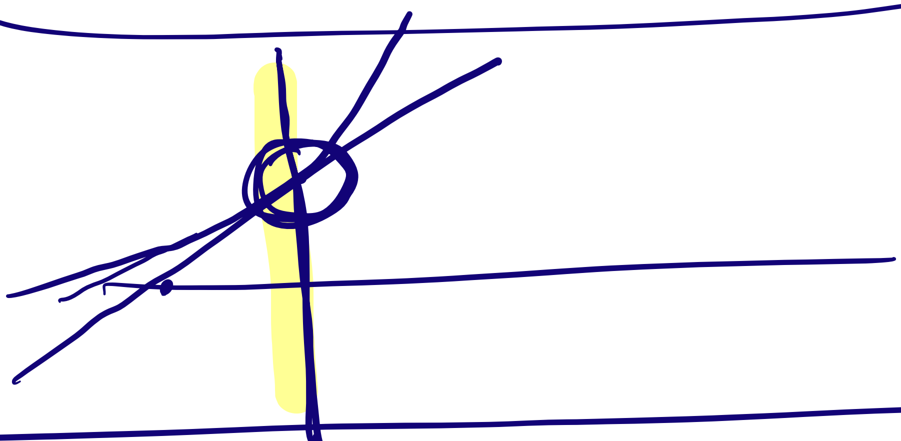
③ If the degrees of freedom ν is large, then the t-density is the same as the Normal density.

(Regression: $\nu = n-m-1$)

$$\left[\frac{1}{1+\frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)}\right]^{\frac{\nu+p}{2}}$$

$$1+\frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu) \approx \exp\left(\frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$\boxed{1+a \approx e^{a} \text{ when } a \text{ is small}}$$



Back to Regression

$$\text{posterior} \propto \left[\frac{S(\hat{\beta})}{S(\beta)}\right]^{\frac{n}{2}}$$

$$\underset{n\times 1}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \underset{n\times(m+1)}{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & x_{21} & & x_{2m} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & & x_{nm} \end{bmatrix}$$

$$\underset{(m+1)\times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_m \end{pmatrix}$$

$$S(\beta) = \|y - X\beta\|^2$$

$$= \|y - X\hat{\beta}\|^2 + \|X\hat{\beta} - X\beta\|^2$$

$$= S(\hat{\beta}) + (\hat{\beta}-\beta)^T X^T X (\hat{\beta}-\beta)$$

posterior: $\left(\frac{S(\hat{\beta})}{S(\beta)}\right)^{\frac{n}{2}}$

$$= \left(\frac{S(\hat{\beta})}{S(\hat{\beta}) + (\hat{\beta}-\beta)^T X^T X (\hat{\beta}-\beta)}\right)^{\frac{n}{2}}$$

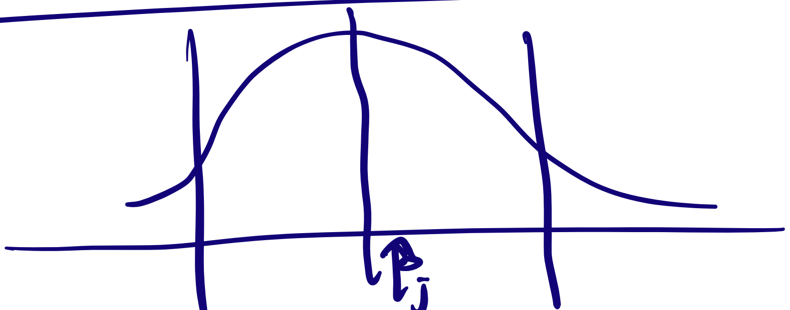
$$= \left[\frac{1}{1 + \frac{1}{S(\hat{\beta})} \underset{(\beta-\hat{\beta})^T X^T X (\beta-\hat{\beta})}{(\hat{\beta}-\beta)^T X^T X (\hat{\beta}-\beta)}}\right]^{\frac{n}{2}}$$

$$\longrightarrow \boxed{t_{m+1}\left(\hat{\beta}, \frac{S(\hat{\beta})}{n-m-1}(X^T X)^{-1}, n-m-1\right)}$$

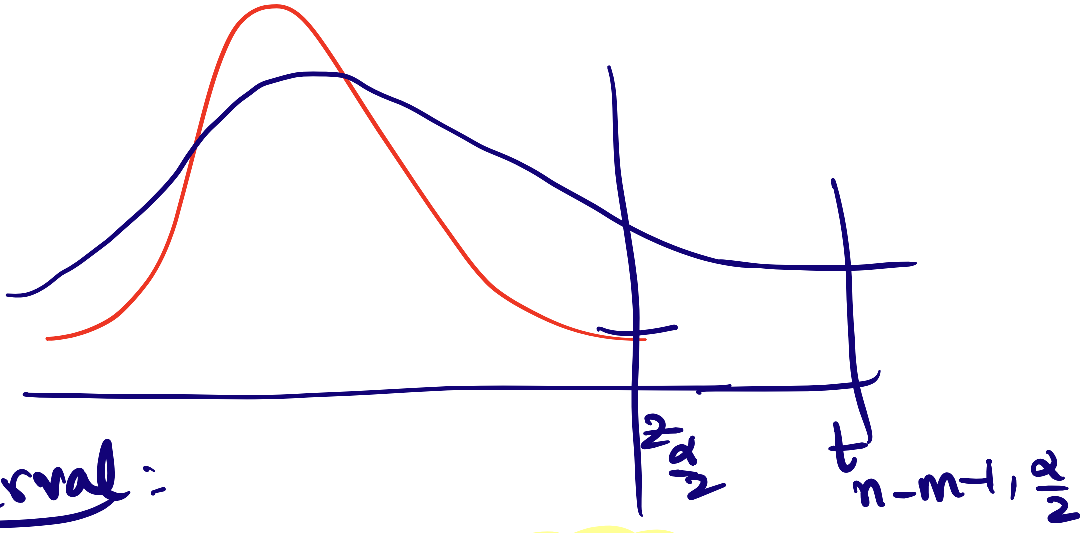
$$\boxed{\beta_j} \quad \beta_j \,|\, \text{data} \sim t_1\left(\hat{\beta}_j, \frac{S(\hat{\beta})}{n-m-1}(X^T X)^{j+1,j+1}, n-m-1\right)$$

$(X^T X)^{j+1,j+1}$ = $(j+1, j+1)$th diagonal entry of $(X^T X)^{-1}$

Uncertainty Interval for β_j:



$t_{n-m-1,\frac{\alpha}{2}}$: $(1-\frac{\alpha}{2})$-quantile of univariate standard t with $n-m-1$ d.f



Interval:

$$\left[\hat{\beta}_j - \sqrt{\frac{S(\hat{\beta})}{n-m-1}}\sqrt{(X^TX)^{j+1,j+1}}\;t_{n-m-1,\frac{\alpha}{2}},\right.$$

$$\left.\hat{\beta}_j + \sqrt{\frac{S(\hat{\beta})}{n-m-1}}\sqrt{(X^TX)^{j+1,j+1}}\;t_{n-m-1,\frac{\alpha}{2}}\right]$$

Bayesian Inference ($\beta_0, \ldots, \beta_m, \log\sigma \sim \text{Unif}(-C,C)$ C large)

||

Frequentist Inference

$$(X^TX)^{j+1,j+1} = \left((X^TX)^{-1}\right)(j+1, j+1)$$

① $S(\hat{\beta}) = \sum\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_m x_{im}\right)^2$

↓ smallest possible sum of squares

RSS (Residual Sum of Squares)

Residual: $y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_m x_{im}$

$\longrightarrow$ i^{th} Residual

② $n - m - 1$: Residual Degrees of Freedom

③ $$\frac{S(\hat{\beta})}{n-m-1} = \frac{RSS}{\text{Residual DF}} = \hat{\sigma}^2$$

$$\left.\hat{\sigma} = \sqrt{\frac{RSS}{\text{Residual DF}}}\;\right\} \begin{array}{l}\text{Residual}\\ \text{Standard}\\ \text{Error}\end{array}$$

④ $$\sqrt{\frac{S(\hat{\beta})}{n-m-1}}\sqrt{\left(\left(X^T X\right)^{-1}\right)_{(j+1,j+1)}}$$

$\longrightarrow$ Standard Error for β_j

Nonlinear Regression

CA population data

$$\underset{\substack{\uparrow \\ \text{log population}}}{y_t} = \beta_0 + \beta_1 t + \varepsilon_t$$

$$y_t = \beta_0 + \beta_1 t + \beta_2 \, \text{ReLU}(t-c) + \varepsilon_t$$

$$\text{ReLU}(z) = \max(z, 0) = z_+$$

$$y_t = \beta_0 + \beta_1 t + \beta_2 (t-c)_+ + \varepsilon_t$$

$$t \le c \rightarrow \beta_0 + \beta_1 t \rightarrow \text{slope: } \beta_1$$

$$t > c \rightarrow \beta_0 + \beta_1 t + \beta_2 (t-c) \rightarrow \text{slope: } \beta_1 + \beta_2$$

① What is c?

② β_1 & $\beta_1 + \beta_2$

$$\boxed{\begin{array}{l}\text{Compare to Cubic:}\\ y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon_t\end{array}}$$

Know c:

$$X_c = \begin{bmatrix} 1 & 1 & (1-c)_+ \\ \vdots & 2 & (2-c)_+ \\ \vdots & \vdots & \vdots \\ 1 & n & (n-c)_+ \end{bmatrix}$$

sm.OLS(y, X_c)

sm.OLS(y, X)

$$X = \begin{bmatrix} 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & n & n^2 & n^3 \end{bmatrix}$$

$$\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

Parameter Estimation

$$\beta_0, \beta_1, \beta_2, c, \sigma$$

Least Squares:

$$\underbrace{\sum_{t=1}^{n} \left[y_t - \beta_0 - \beta_1 t - \beta_2 \,\mathrm{ReLU}(t-c) \right]^2}_{S(\beta, c)}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

minimize over β, c

Strategy

① Fix c & $\underbrace{\text{minimize } S(\beta, c)}_{\text{Same as doing linear regression}}$ over β

with

$$X_c = \begin{bmatrix} 1 & 1 & \text{ReLU}(1-c) \\ \vdots & 2 & \vdots \\ \vdots & \vdots & \vdots \\ 1 & n & \text{ReLU}(n-c) \end{bmatrix}$$

Do regression of y on X_c.

Calculate least squares estimate $\hat{\beta}_c$,

& $RSS(c)$

② Minimize $RSS(c)$ over c.

$$\boxed{c \in \{1, \ldots, n\}}$$