Lecture November 15

Model $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i,$

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

$$\mu(x_{1i}, x_{2i}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

$$\mu(a+1, b) = \beta_0 + \beta_1 (a+1) + \beta_2 (b)$$

$$\mu(a, b) = \beta_0 + \beta_1 a + \beta_2 b$$

$$\Rightarrow \underbrace{\mu(a+1, b) - \mu(a, b)}_{} = \beta_1$$

Change in the mean function
for every unit increase in x_1
but keeping x_2 fixed

$$= \beta_2$$

Set of Potential Predictors / Indep Variables

$$\{ U_1, U_2, \ldots\ldots, U_M \}$$

Response variable Y

(1) Forward Selection

(2) Backward Selection

(3) FB Selection

(4) Regularization

FORWARD ALGORITHM

Step 0. Model $Y_i = \beta_0 + \varepsilon_i$

Step 1. Model $Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$

where x_{1i} is selected from $\{ U_1, \ldots, U_M \}$

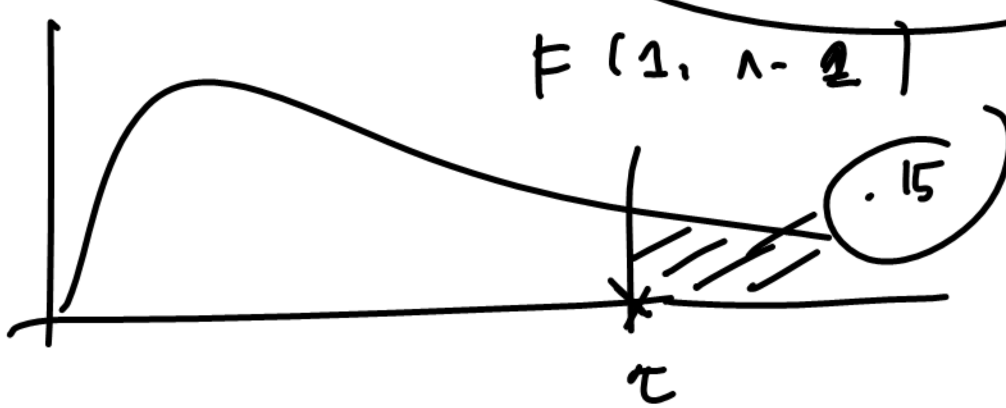
For each candidate predictor U_q, $q = 1 \ldots M$

fit a model $Y_i = \beta_0 + \beta_1 U_{qi} + \varepsilon_i$

$SSE(0)$

$SSE(U_q)$

$$F_q = \frac{(SSE(0) - SSE(U_q))/1}{SSE(U_q)/n-2}$$



Let X_1 be the U_q s.t.

$F_q > \tau$ and $F_q = \arg\max_l F_l$

Step 2. Let $P_2 = \{U_1 \ldots, U_M\} - \{X_1\}$

Model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

For each $U_q \in P_2$, fit the model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 U_{qi} + \varepsilon_i$$

$$SSE(X_1, U_q)$$

$$F_q = \frac{\dfrac{SSE(X_1) - SSE(X_1, U_q)}{1}}{\dfrac{SSE(X_1, U_q)}{n-3}}$$

Choose U_q so that

$$F_q > \tau \equiv F(0.85;\ 1, n-3) \text{ and}$$

$$q = \underset{U_l \in P_2}{\arg\max}\ F_l$$

$$X_2 = F_q$$

<u>CONTINUE</u>...

Subsets

$$\{U_1, U_2, U_3\}$$

$$2^3 = 8.$$

- $Y_i = \beta_0 + \varepsilon_i$
- $Y_i = \beta_0 + \beta_1 U_{1i} + \varepsilon_i$ $\quad U_1$

  $\beta_2 U_{2i}$ $\quad U_2$

  $\beta_3 U_{3i}$ $\quad U_3$

- $Y_i = \beta_0 + \beta_1 U_{1i} + \beta_2 U_{2i} + \varepsilon_i$ $\quad U_1, U_2$

  U_1, U_3

  U_2, U_3

- $Y_i = \beta_0 + \beta_1 U_{1i} + \cdots$

U_1
U_2
$\vdots$
U_{10}

$\xrightarrow{\text{summarize}}$

- Avg: $\bar{U}_{(1:10)}$ WARNING
- PCA

$$S = U_1 - U_2 + \frac{2}{3}U_3 - \dots$$

U_i
$\vdots$

$\xrightarrow{\text{summarize}}$

U_i
$\vdots$

$\longrightarrow$ S

BACKWARD SELECTION

"Complete" model: $\{U_1 \dots U_m\}$

$$Y_i = \beta_0 + \beta_1 U_{1i} + \dots + \beta_m U_{mi} + \varepsilon_i$$

$$SSE(U_1, \dots, U_m)$$

For each q: U_q : $Y_i = \underline{\text{w/o}}\ U_q + \varepsilon_i$