Lecture 21   22 Nov

Midterm 2
- In-class   Dec 02
- Take home due on 29 Nov in class

Final — optional

Midterm 2 Take home   Model/Variable Selection

Candidate Predictors $(X_1, X_2, \ldots, X_8)$

Select the "best" model

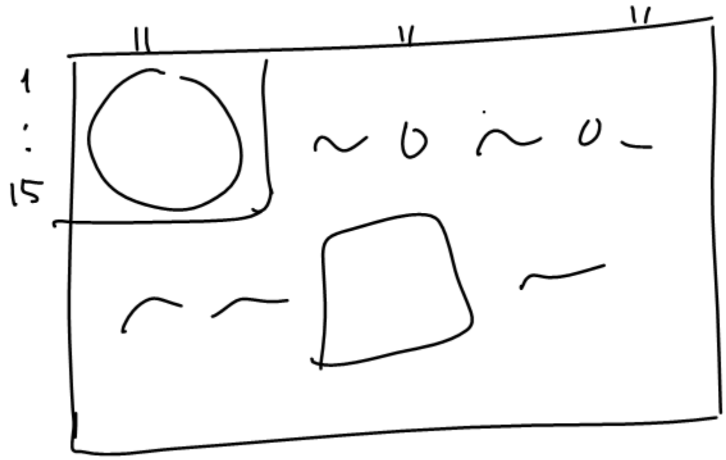
(1) All subsets regression : ALL $2^8 = 256$ models

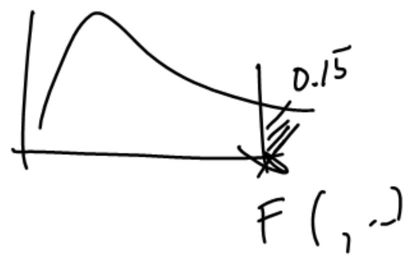
#	X_1	X_2		X_8	AIC	BIC
1	0	0	..	0		
2	1	0		0		
3	0	1		0		
⋮						
256	1	1		1		

If # Predictors p is very large

Group predictors

Correlation Matrix for $X_1, \ldots, X_p$.

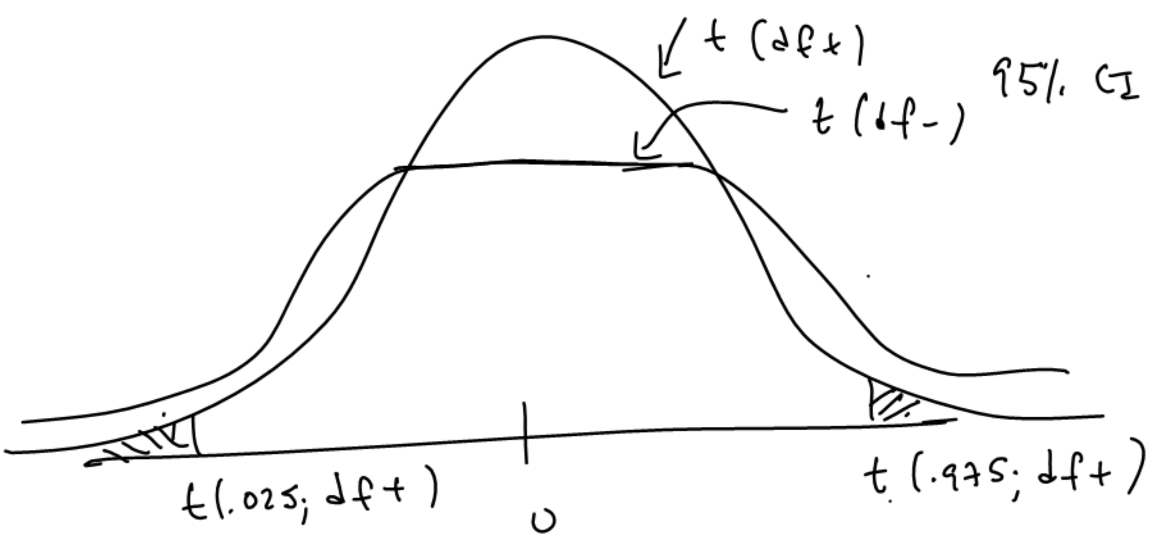




(2) Stepwise Methods — Forward

— Backward

— F/B

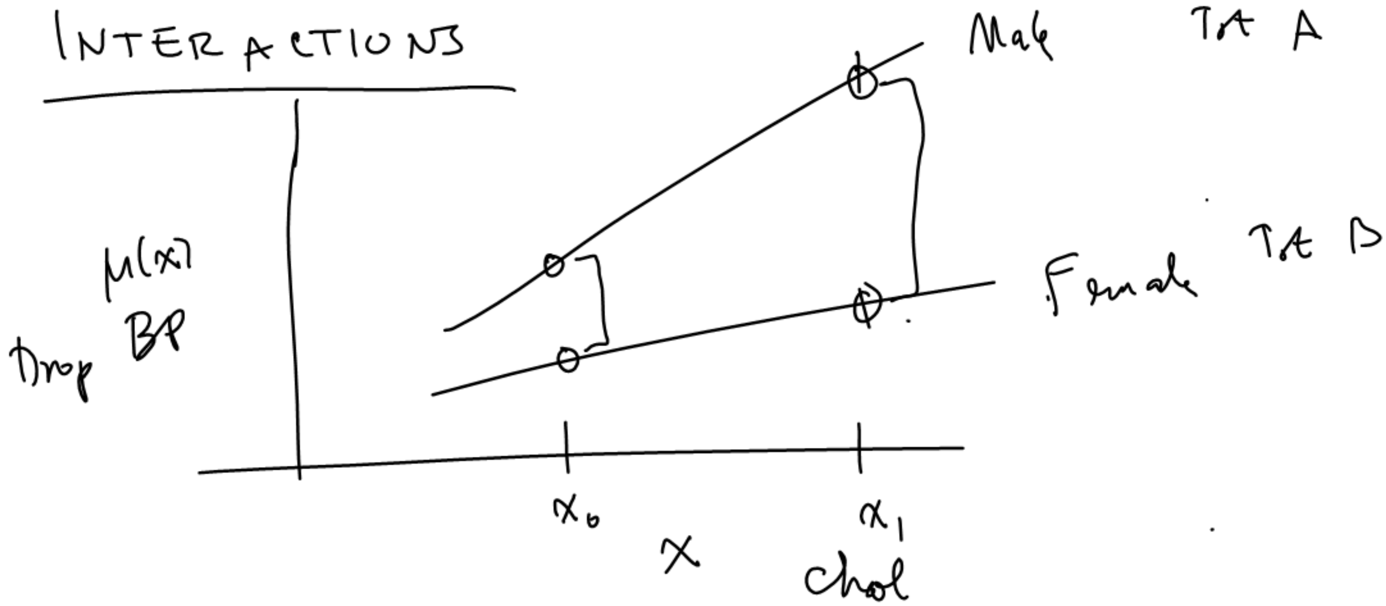


(3) Build in interactions between $\{X_1, \dots X_Q\}$

$X_1 \cdot X_2$, $X_1 \cdot X_3$, $X_2 \cdot X_3$, ...

(4) Interpret the results

INTERACTIONS



There is an interaction between gender and x if the difference between males & females differs [varies] across values of x, e.g.

$$\mu^M(x_0) - \mu^F(x_0) \neq \mu^M(x_1) - \mu^F(x_1)$$

Formally:

$$Y_i = \mu(x_i, G_{1i}, G_{2i}) + \varepsilon_i \quad \text{where}$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \quad \text{and} \quad G_{1i} = \begin{cases} 1, F \\ 0, M \end{cases} \quad G_{2i} = \begin{cases} 1, M \\ 0, F \end{cases}$$

$$\mu(x_i, G_{1i}, G_{2i}) = (\beta_0^F + \beta_1^F x_i) G_{1i} + \underbrace{(\beta_0^M + \beta_1^M x_i) G_{2i}}$$

The mean function for females & males, respectively, are:

$$\mu(x_i, G_{1i} = 1, G_{2i} = 0) = \beta_0^F + \beta_1^F x_i$$

$$\mu(x_i, G_{1i} = 0, G_{2i} = 1) = \beta_0^M + \beta_1^M x_i$$

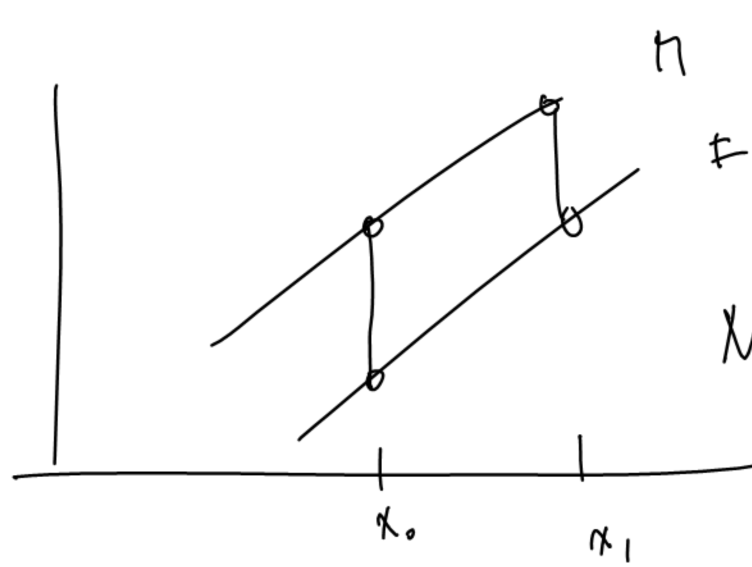
WLOG let $i = 1, \dots, n_1$ females
$i = n_1+1 \dots n_1+n_2$ males

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ \hline Y_{n_1+1} \\ \vdots \\ Y_{n_1+n_2} \end{pmatrix} = \left(\begin{array}{cc|cc} \underline{\mathbb{1}} & \begin{matrix} x_1 \\ \vdots \\ x_{n_1} \end{matrix} & \underline{0} & \underline{0} \\ \hline \underline{0} & \underline{0} & \underline{\mathbb{1}} & \begin{matrix} x_{n_1+1} \\ \vdots \\ x_{n_1+n_2} \end{matrix} \end{array}\right) \begin{pmatrix} \beta_0^F \\ \beta_1^F \\ \beta_0^M \\ \beta_1^M \end{pmatrix} + \underline{\varepsilon}$$

$$\mu(x_{1i}, G_{1i}, G_{2i}) = \beta_0^F G_{1i} + \beta_1^F \underbrace{G_{1i} x_i} + \beta_0^M G_{2i} + \beta_1^M \underbrace{G_{2i} x_i}$$

Here, ~~the~~ model includes the interaction terms $G_{1i} x_i$ & $G_{2i} x_i$

There is an interaction between "gender & x"



parallel

No interaction

Another parameterization:

$$\mu(x_i, G_{2i}) = \beta_0 + \beta_1 x_i + (\delta_0 + \Delta_1 x_i) G_{2i}$$

The mean function for the female group:

$$\mu(x_i, G_{2i} = 0) = \beta_0 + \beta_1 x_i$$

———————— male group

$$\mu(x_i, G_{2i} = 1) = (\beta_0 + \delta_0) + (\beta_1 + \Delta_1) x_i$$

Consider the model :

$$Y_i = \mu(x_{1i}, x_{2i}) + \varepsilon_i \quad \text{where}$$

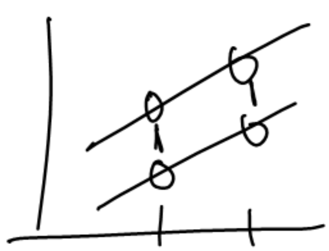
$\varepsilon_i \sim$ iid $N(0, \sigma^2)$

x_{1i} & x_{2i} are continuous-valued

$$\mu(x_{1i}, x_{2i}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \underbrace{\beta_{12}}\, x_{1i} x_{2i}$$

β_{12} captures the interaction between x_1 & x_2

The difference between $\underline{\mu(x_{1i} = x_1^0, x_{2i} = x_2^0)}$ and $\underline{\mu(x_{1i} = x_1^0, \ x_{2i} = x_2^1)}$ is:



$$\underline{\underline{D(x_1^0)}} = \beta_0 + \beta_1 x_1^0 + \beta_2 x_2^1 + \beta_{12} x_1^0 x_2^1$$
$$- \left(\beta_0 + \beta_1 x_1^0 + \beta_2 x_2^0 + \beta_{12} x_1^0 x_2^0\right)$$
$$= \beta_2 (x_2^1 - x_2^0) + \beta_{12} x_1^0 (x_2^1 - x_2^0)$$

The difference between $\mu(x_{1i} = x_1^1 \;\; x_{2i} = x_2^0)$

$\mu(x_{1i} = x_1^1 \;\; x_{2i} = x_2^1)$

$$D(x_1^1) = \beta_0 + \beta_1 x_1^1 + \beta_2 x_2^1 + \beta_{12} x_1^1 x_2^1 - \left(\beta_0 + \beta_1 x_1^1 + \beta_2 x_2^0 + \beta_{12} x_1^1 x_2^0\right)$$

$$= \beta_2 (x_2^1 - x_2^0) + \beta_{12} x_1^1 (x_2^1 - x_2^0)$$

$$D(x_1^1) - D(x_1^0) = \beta_{12} (x_2^1 - x_2^0)(x_1^1 - x_1^0)$$

$$= 0 \quad \forall \text{ pair iff } \beta_{12} = 0.$$