

Lecture 11

Model Selection

STAT 8020 Statistical Methods II
September 13, 2019

Whitney Huang
Clemson University



Notes

Agenda

- 1 Variable Selection Criteria
- 2 Automatic Search Procedures



Notes

Variable Selection

- What is the appropriate subset size?
- What is the best model for a fixed size?



Notes

Mallows' C_p Criterion

$$\begin{aligned} (\hat{Y}_i - \mu_i)^2 &= (\hat{Y}_i - E(\hat{Y}_i) + E(\hat{Y}_i) - \mu_i)^2 \\ &= \underbrace{(\hat{Y}_i - E(\hat{Y}_i))^2}_{\text{Variance}} + \underbrace{(E(\hat{Y}_i) - \mu_i)^2}_{\text{Bias}^2}, \end{aligned}$$

where $\mu_i = E(Y_i | X_i = x_i)$

- Mean squared prediction error (MSPE):

$$\sum_{i=1}^n \sigma_{\hat{Y}_i}^2 + \sum_{i=1}^n (E(\hat{Y}_i) - \mu_i)^2$$

- C_p criterion measure:

$$\begin{aligned} \Gamma_p &= \frac{\sum_{i=1}^n \sigma_{\hat{Y}_i}^2 + \sum_{i=1}^n (E(\hat{Y}_i) - \mu_i)^2}{\sigma^2} \\ &= \frac{\sum \text{Var}_{\text{pred}} + \sum \text{Bias}^2}{\text{Var}_{\text{error}}} \end{aligned}$$



Notes

C_p Criterion

- Do not know σ^2 nor numerator
- Use $\text{MSE}_{X_1, \dots, X_{p-1}} = \text{MSE}_F$ as the estimate for σ

- For numerator:

- Can show $\sum_{i=1}^n \sigma_{\hat{Y}_i}^2 = p\sigma^2$

- Can also show $\sum_{i=1}^n (E(\hat{Y}_i) - \mu_i)^2 = E(\text{SSE}_F) - (n-p)\sigma^2$

$$\Rightarrow C_p = \frac{\text{SSE} - (n-p)\text{MSE}_F + p\text{MSE}_F}{\text{MSE}_F}$$



Notes

C_p Criterion Cont'd

Recall

$$\Gamma_p = \frac{\sum_{i=1}^n \sigma_{\hat{Y}_i}^2 + \sum_{i=1}^n (E(\hat{Y}_i) - \mu_i)^2}{\sigma^2}$$

- When model is correct $E(C_p) \approx p$
- When plotting models against p
 - Biased models will fall above $C_p = p$
 - Unbiased models will fall around line $C_p = p$
 - By definition: C_p for full model equals p



Notes

Adjusted R^2 Criterion

Adjusted R^2 , denoted by R^2_{adj} , attempts to take account of the phenomenon of the R^2 automatically and spuriously increasing when extra explanatory variables are added to the model.

$$R^2_{\text{adj}} = 1 - \frac{\text{SSE}/(n - p - 1)}{\text{SST}/(n - 1)}$$

- Choose model which maximizes R^2_{adj}
- Same approach as choosing model with smallest MSE



Notes

Predicted Residual Sum of Squares *PRESS* Criterion

- For each observation i , predict Y_i using model generated from other $n - 1$ observations
- $PRESS = \sum_{i=1}^n (Y_i - \hat{Y}_{i(i)})^2$
- Want to select model with small $PRESS$



Notes

Other Approaches

- Akaike's information criterion (AIC)

$$n \log\left(\frac{\text{SSE}_k}{n}\right) + 2k$$

- Bayesian information criterion (BIC)

$$n \log\left(\frac{\text{SSE}_k}{n}\right) + k \log(n)$$

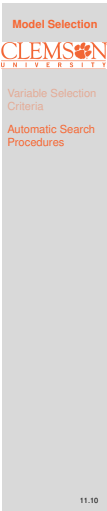
- Can be used to compare **non-nested** models



Notes

Automatic Search Procedures

- Forward Selection
- Backward Elimination
- Stepwise Search
- All Subset Selection



Notes

Notes

Notes
