Variable Selection

HH Chapter 9

Nov 8, 2005

Topics

- Variable Selection
- Stepwise
- Model Selection Criteria
- Model Averaging

1 Variable Selection

Variable Selection

Reasons for reducing the number of variables in the model:

- Philosophical
  - Avoid the use of redundant variables
  - KISS
  - Occam's Razor
- Practical
  - Inclusion of un-necessary terms yields less precise estimates, particularly if explanatory variables are highly correlated with each other

Variable Selection Procedures

- Stepwise Regression: Forward, Stepwise, Backward – add/delete variables until all t-statistics are significant (easy, but not recommended)
- Use a Model Selection Criterion to pick the "best" model
  - R2 (picks largest model)
  - Adjusted R2
  - Mallow's Cp $C_p = (\text{SSE}/\hat{\sigma}^2_{Full}) + 2p_m - n$
  - AIC (Akaike Information Criterion) proportional to Cp for linear models
  - BIC(m) (Bayes Information Criterion) $\hat{\sigma}^2_m + \log(n)p_m$

Trade off model complexity (number of coefficients p_m) with goodness of fit ($\hat{\sigma}^2_m$)

Model Selection

Selection of a single model has the following problems

- When the criteria suggest that several models are equally good, what should we report? Still pick only one model?
- What do we report for our uncertainty after selecting a model?

Typical analysis ignores model uncertainty!

Bayesian Model Averaging

Rather than use a single model, BMA uses all (or potentially a lot) models, but weights model predictions by their posterior probabilities (measure of how much each model is supported by the data)

- Posterior model probabilities

$$p(M_j \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid M_j)p(M_j)}{\sum_j p(\mathbf{Y} \mid M_j)p(M_j)}$$

- Approximate marginals likelihood

$$P(\mathbf{Y} \mid M_j) = \exp\{-.5BIC(M_j)\}$$

- Probability $\beta_j \neq 0$: $\sum_{M_j:\beta_j\neq 0} p(M_j \mid \mathbf{Y})$
- Predictions

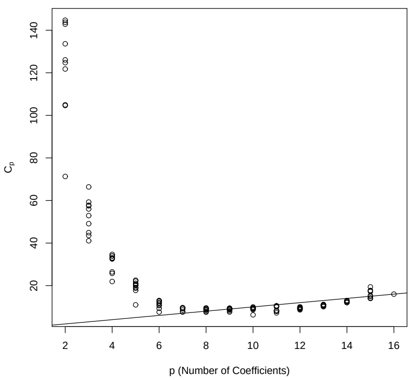
$$\hat{Y^*}|\mathbf{Y} = \sum_j p(M_j|\mathbf{Y})\hat{Y}_{M_j}$$

Example

- Data from Statistical Sleuth 12.17
- 60 cities
- response Mortality
- measures of HC, NOX, SO2
- Is pollution associated with mortality after adjusting for other socio-economic and meteorological factors?
- 15 predictor variables ($2^15 = 32,768$ possible models
- use BMA on a subset of the models

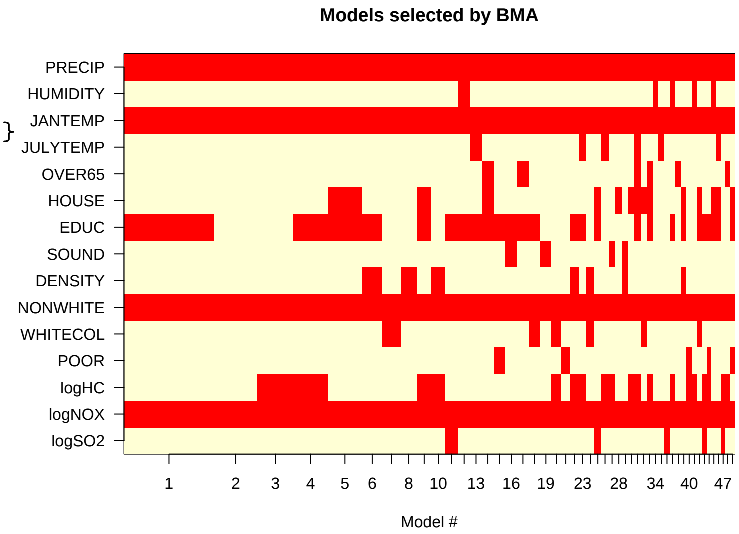
leaps

```
library(leaps)
out = leaps(pollution[,-1],pollution[,1], method="Cp")}
```



bicreg example

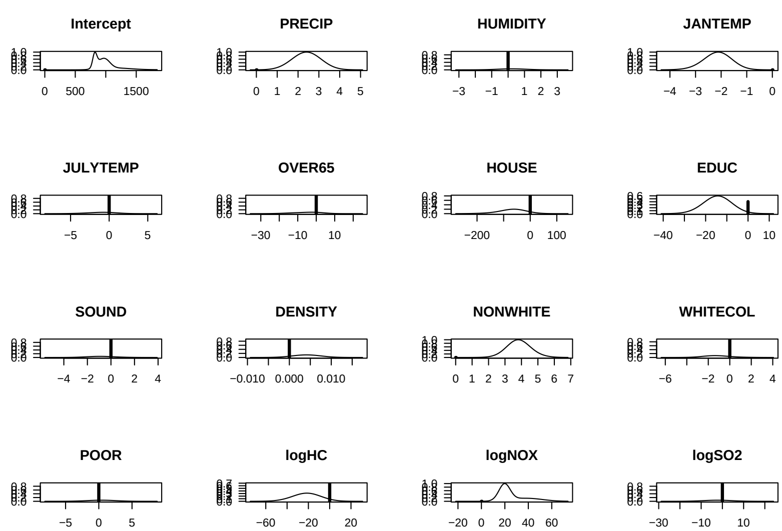
```
library(BMA)
poll.bicreg <- bicreg(pollution[,-1],
                      pollution[,1], nbest=40)
imageplot.bma(poll.bicreg)
plot(poll.bicreg, mfrow=c(4,4))
summary(poll.bicreg)
```



Posterior Probabilities

- What is the probability that there is no pollution effect?
- Sum posterior model probabilities over all models that include no pollution variables
- With the subset of models, posterior probability is 0 (SO2 is in all of the 49 models)
- With enumeration, posterior probability is 0.0038
- Odds that there is an effect $(1 - .0038)/(.0038) = 262.1579$

Posterior Distributions



Model Space