Lab8

Setup

```r
library(MPV)      # Dataset
```

```
Loading required package: lattice

Loading required package: KernSmooth

KernSmooth 2.23 loaded
Copyright M. P. Wand 1997-2009

Loading required package: randomForest

randomForest 4.7-1.2

Type rfNews() to see new features/changes/bug fixes.
```

```r
library(MASS)      # For stepwise selection
```

```

Attaching package: 'MASS'

The following object is masked from 'package:MPV':

    cement
```

```r
library(leaps)      # For subset selection
library(car)        # For PRESS statistic calculation
```

```
Loading required package: carData
```

```r
library(stats)      # For AIC and BIC calculation
```

Problem 1

P 1. Consider the Hald Cement data set given in Table 10.1 of Montgomery Book.

Load the Dataset

```r
data(cement)
data <- cement
```

Define the full model

```r
# Full model with all predictors
full_model <- lm(y ~ x1 + x2 + x3 + x4, data=data)
```

Part (i) : Subset Models Selection Based on Different Criteria

(i) Find at least two subset models based on:

- (a) R^2
- (b) R_p^2
- (c) Mallow C_p statistics
- (d) Forward selection
- (e) Backward elimination
- (f) Step wise selection

$(a)R^2$

```r
# (a) & (b): R-squared and Adjusted R-squared based models using regsubsets
subset_models <- regsubsets(y ~ x1 + x2 + x3 + x4, data=data, nbest=1)
subset_summary <- summary(subset_models)

# Select models based on R² and Adjusted R²
best_r2_model <- which.max(subset_summary$rsq)

best_r2_formula <- paste("y ~", paste(names(coef(subset_models, best_r2_model))[-1], collapse

best_r2_formula
```

```
[1] "y ~ x1 + x2 + x3 + x4"
```

$(b) {R_p}^2$

```r
best_adj_r2_model <- which.max(subset_summary$adjr2)
best_adj_r2_formula <- paste("y ~", paste(names(coef(subset_models, best_adj_r2_model))[-1],

best_adj_r2_formula
```

```
[1] "y ~ x1 + x2 + x4"
```

(c) Mallow C_p Statistics

```r
# (c): Mallow's Cp statistic
best_cp_model <- which.min(subset_summary$cp)

best_cp_formula <- paste("y ~", paste(names(coef(subset_models, best_cp_model))[-1], collapse

best_cp_formula
```

```
[1] "y ~ x1 + x2"
```

(d) Forward selection

```r
null_model <- lm(y ~ 1, data=data)

# Forward Selection
forward_model <- step(null_model, direction="forward", scope=formula(full_model), trace=FALSE

forward_model
```

```
Call:
lm(formula = y ~ x4 + x1 + x2, data = data)

Coefficients:
(Intercept)           x4           x1           x2
    71.6483      -0.2365       1.4519       0.4161
```

(e) Backward Elimination

```r
# Backward Elimination
backward_model <- step(full_model, direction="backward", trace=FALSE)

backward_model
```

```
Call:
lm(formula = y ~ x1 + x2 + x4, data = data)

Coefficients:
(Intercept)           x1           x2           x4
    71.6483       1.4519       0.4161      -0.2365
```

(f) Stepwise Selection

```r
# Stepwise Selection
stepwise_model <- step(null_model, direction="both", scope=formula(full_model), trace=FALSE)
stepwise_model
```

```
Call:
lm(formula = y ~ x4 + x1 + x2, data = data)

Coefficients:
(Intercept)           x4           x1           x2
    71.6483      -0.2365       1.4519       0.4161
```

Part (ii) : Compute PRESS, AIC, and BIC for selected subset models

(ii) For the selected subset models, find

  (a) Value of the PRESS statistics
  (b) AIC
  (c) BIC

(a) PRESS Statistics

```
# Define function for PRESS statistic
PRESS <- function(model) {
  pr <- residuals(model)/(1 - lm.influence(model)$hat)
  sum(pr^2)
}

# Fit the selected models
model_r2 <- lm(best_r2_formula, data=data)
model_adj_r2 <- lm(best_adj_r2_formula, data=data)
model_cp <- lm(best_cp_formula, data=data)

# PRESS statistic
press_r2 <- PRESS(model_r2)
press_adj_r2 <- PRESS(model_adj_r2)
press_cp <- PRESS(model_cp)
```

(b) AIC

```
aic_r2 <- AIC(model_r2)
aic_adj_r2 <- AIC(model_adj_r2)
aic_cp <- AIC(model_cp)
```

(c) BIC

```
bic_r2 <- BIC(model_r2)
bic_adj_r2 <- BIC(model_adj_r2)
bic_cp <- BIC(model_cp)
```

Ouput the Results

```r
cat("Model based on R-squared:\n")
```

```
Model based on R-squared:
```

```r
cat("Formula:", best_r2_formula, "\n")
```

```
Formula: y ~ x1 + x2 + x3 + x4
```

```r
cat("PRESS:", press_r2, "\n")
```

```
PRESS: 110.3466
```

```r
cat("AIC:", aic_r2, "\n")
```

```
AIC: 65.83669
```

```r
cat("BIC:", bic_r2, "\n\n")
```

```
BIC: 69.22639
```

```r
cat("Model based on Adjusted R-squared:\n")
```

```
Model based on Adjusted R-squared:
```

```r
cat("Formula:", best_adj_r2_formula, "\n")
```

```
Formula: y ~ x1 + x2 + x4
```

```r
cat("PRESS:", press_adj_r2, "\n")
```

```
PRESS: 85.35112
```

```r
cat("AIC:", aic_adj_r2, "\n")
```

```
AIC: 63.86629
```

```r
cat("BIC:", bic_adj_r2, "\n\n")
```

```
BIC: 66.69103
```

```r
cat("Model based on Mallow's Cp:\n")
```

```
Model based on Mallow's Cp:
```

```r
cat("Formula:", best_cp_formula, "\n")
```

```
Formula: y ~ x1 + x2
```

```r
cat("PRESS:", press_cp, "\n")
```

```
PRESS: 93.88255
```

```r
cat("AIC:", aic_cp, "\n")
```

```
AIC: 64.31239
```

```r
cat("BIC:", bic_cp, "\n\n")
```

```
BIC: 66.57219
```

```r
cat("Forward Selection Model:\n")
```

```
Forward Selection Model:
```

```r
cat("Formula:", as.character(forward_model$call$formula), "\n")
```

```
Formula: ~ y x4 + x1 + x2
```

```r
cat("PRESS:", PRESS(forward_model), "\n")
```

```
PRESS: 85.35112
```

```r
cat("AIC:", AIC(forward_model), "\n")
```

```
AIC: 63.86629
```

```r
cat("BIC:", BIC(forward_model), "\n\n")
```

```
BIC: 66.69103
```

```r
cat("Backward Elimination Model:\n")
```

```
Backward Elimination Model:
```

```r
cat("Formula:", as.character(backward_model$call$formula), "\n")
```

```
Formula: ~ y x1 + x2 + x4
```

```r
cat("PRESS:", PRESS(backward_model), "\n")
```

```
PRESS: 85.35112
```

```r
cat("AIC:", AIC(backward_model), "\n")
```

```
AIC: 63.86629
```

```r
cat("BIC:", BIC(backward_model), "\n\n")
```

```
BIC: 66.69103
```

```r
cat("Stepwise Selection Model:\n")
```

```
Stepwise Selection Model:
```

```r
cat("Formula:", as.character(stepwise_model$call$formula), "\n")
```

```
Formula: ~ y x4 + x1 + x2
```

```r
cat("PRESS:", PRESS(stepwise_model), "\n")
```

```
PRESS: 85.35112
```

```r
cat("AIC:", AIC(stepwise_model), "\n")
```

```
AIC: 63.86629
```

```r
cat("BIC:", BIC(stepwise_model), "\n")
```

```
BIC: 66.69103
```