

To print higher-resolution math symbols, click the  
**Hi-Res Fonts for Printing** button on the jsMath control panel.

# CMSC 478 — Spring 2017 — C. S. Marron

## Lab 6: Subset Selection

### Data Description

In this lab, you will work with the **College** dataset of college admissions data. The dataset is part of the **ISLR** package; if you have not already installed the package, you may [download the CSV file](#) of the **College** dataset. Use `?ISLR::College` in R to see a description of the dataset.

### Exercises

**Exercise 1:** Create a new variable **PACcept** that contains the percentage of applicants accepted to each College (so **PACcept** is just `Accept / Apps`). You will evaluate models to predict **PACcept** using all other variables *except* **Accept** and **Apps**.

1. Use `regsubsets()` (part of the **leaps** library) to perform best subset selection.
2. What are the best models obtained according to  $C_p$ , BIC, and adjusted  $R^2$ ?
3. What are the coefficients of the best models?

**Exercise 2:** Determine the best models using both forward and backward stepwise selection.

1. For both forward and backward stepwise selection, what is the best model obtained according to  $C_p$ , BIC, and adjusted  $R^2$ ?
2. Use the `plot()` function to display the variable selection plots for each selection method (full, forward, and backward) ordered by  $R^2$ . How do the models selected by the methods differ?
3. Display the variable selection plots for each selection method ordered by BIC. How do the models selected by the methods differ?

**Exercise 3:** Use five-fold cross-validation to determine the best model. How does the model selected using cross-validation compare to the best models using Adjusted  $R^2$  and BIC? You will need the following prediction function to predict **PACcept** from the object returned by `regsubsets()`.

```
predict.regsubsets = function(object, newdata, id, ...) {
  form = as.formula(object$call[[2]])
  mat = model.matrix(form, newdata)
  coefi = coef(object, id=id)
```

```
xvars= names(coefi)
mat[,xvars] %*% coefi
}
```

*Hint:* See the example on page 250 of the textbook.