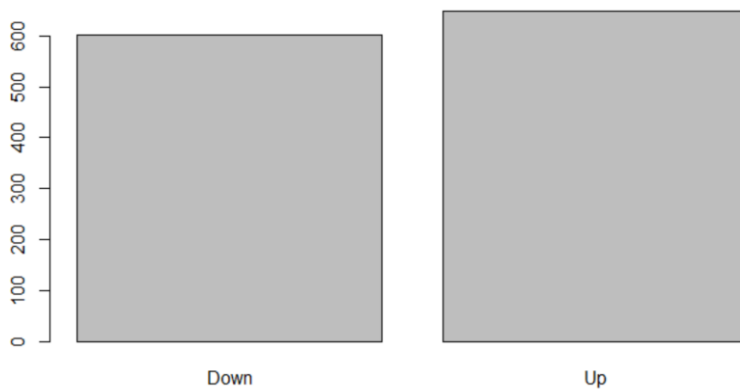


**Assignment 8**

1. Run Lab\_LDA01.R in R

```
> require(ISLR)
> require(MASS)
> require(descr)
> attach(Smarket)
The following objects are masked from Smarket (pos = 3):
    Direction, Lag1, Lag2, Lag3, Lag4, Lag5, Today, Volume, Year
> ## Linear Discriminant Analysis
> freq(Direction)
Direction
  Frequency Percent
Down      602   48.16
Up        648   51.84
Total    1250  100.00
```



```

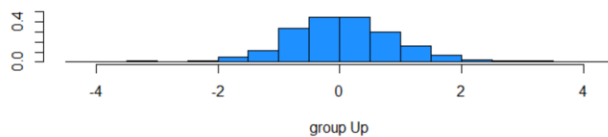
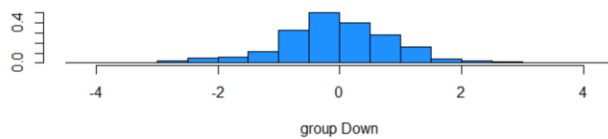
> train = Year<2005
> lda.fit=lda(Direction~Lag1+Lag2,data=Smarket, subset=Year<2005)
> lda.fit
Call:
lda(Direction ~ Lag1 + Lag2, data = Smarket, subset = Year <
  2005)

Prior probabilities of groups:
      Down      Up
0.491984 0.508016

Group means:
      Lag1      Lag2
Down 0.04279022 0.03389409
Up   -0.03954635 -0.03132544

Coefficients of linear discriminants:
      LD1
Lag1 -0.6420190
Lag2 -0.5135293
> plot(lda.fit, col="dodgerblue")

```



```

> Smarket.2005=subset(Smarket,Year==2005) # Creating subset with 2005 data for prediction
> lda.pred=predict(lda.fit,Smarket.2005)
> names(lda.pred)
[1] "class" "posterior" "x"
> lda.class=lda.pred$class
> Direction.2005=Smarket$Direction[!train]
> table(lda.class,Direction.2005)
      Direction.2005
lda.class Down Up
      Down   35  35
      Up    76 106
> data.frame(lda.pred)[1:5,]
  class posterior.Down posterior.Up      LD1
999   Up    0.4901792    0.5098208 0.08293096
1000  Up    0.4792185    0.5207815 0.59114102
1001  Up    0.4668185    0.5331815 1.16723063
1002  Up    0.4740011    0.5259989 0.83335022
1003  Up    0.4927877    0.5072123 -0.03792892
> table(lda.pred$class,Smarket.2005$Direction)
      Down Up
Down   35  35
Up    76 106
> mean(lda.pred$class==Smarket.2005$Direction)
[1] 0.5595238

```

2. Review ISLR Chapters 6 and look up answers for the following questions
3. From the three methods (best subset, forward stepwise, and backward stepwise):

a. Which of the three models with  $k$  predictors has the smallest training RSS?  
Best subset selection has the smallest training RSS. Forward and Backward stepwise models depend on which predictors are selected initially that can result in complications should a poor choice be initially made that later cannot be undone.

b. Which of the three models with  $k$  predictors has the smallest test RSS?  
Best subset may have the smallest test RSS because it is inclusive of more models than the other methods presented. However, one of the other models may fit the data being analyzed better, as they would be less likely to overfit. In this case, the outcome will weigh more towards the selection of test set or validation process than on the selection method.

#### 4. Application exercise:

Generate simulated data, and then use this data to perform best subset selection.

1. Use the `rnorm()` function to generate a predictor  $X$  of length  $n = 100$ , as well as a noise vector  $\epsilon$  of length  $n = 100$ .

```
> set.seed(1)
> X=rnorm(100)
> eps=rnorm(100)
```

2. Generate a response vector  $y$  of length  $n = 100$  according to the model:

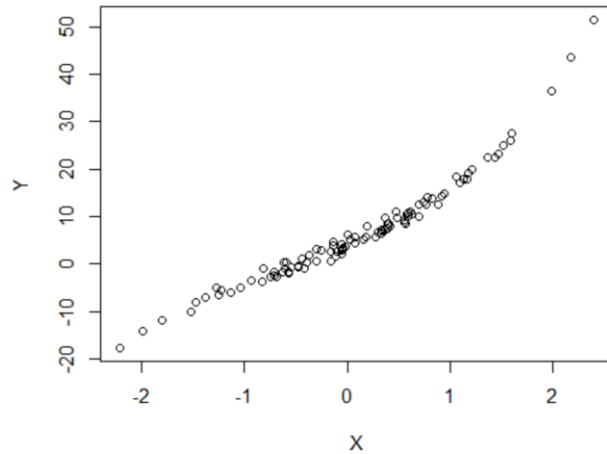
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon,$$

where  $\beta_0, \beta_1, \beta_2$  and  $\beta_3$  are 4, 9, 2, 1 respectively.

```
> beta0=4
> beta1=9
> beta2=2
> beta3=1
> Y=beta0+beta1*X+beta2*X^2+beta3*X^3+eps
```

Plot x and y.

```
> plot(X,Y)
```



3. Use the leaps package: require(leaps)

```
> require(leaps)
Loading required package: leaps
```

4. Use the regsubsets() function from the leaps package to perform best subset selection in order to choose the best model containing the predictors.  $x, x_2, \dots, x_{10}$ .

```
> regsubsets(Y~poly(X,10,row=T), data=data.frame(Y,X), nvmax=10)
Subset selection object
Call: regsubsets.formula(Y ~ poly(X, 10, row = T), data = data.frame(Y,
  X), nvmax = 10)
10 Variables (and intercept)
      Forced in Forced out
poly(X, 10, row = T)1      FALSE      FALSE
poly(X, 10, row = T)2      FALSE      FALSE
poly(X, 10, row = T)3      FALSE      FALSE
poly(X, 10, row = T)4      FALSE      FALSE
poly(X, 10, row = T)5      FALSE      FALSE
poly(X, 10, row = T)6      FALSE      FALSE
poly(X, 10, row = T)7      FALSE      FALSE
poly(X, 10, row = T)8      FALSE      FALSE
poly(X, 10, row = T)9      FALSE      FALSE
poly(X, 10, row = T)10     FALSE      FALSE
1 subsets of each size up to 10
Selection Algorithm: exhaustive
```

What is the best model obtained according to Cp, BIC, and adjusted R2? Show some plots to provide evidence for your answer, and report the coefficients of the best model obtained. Note you will need to use the data.frame() function to create a single data set containing both x and y .

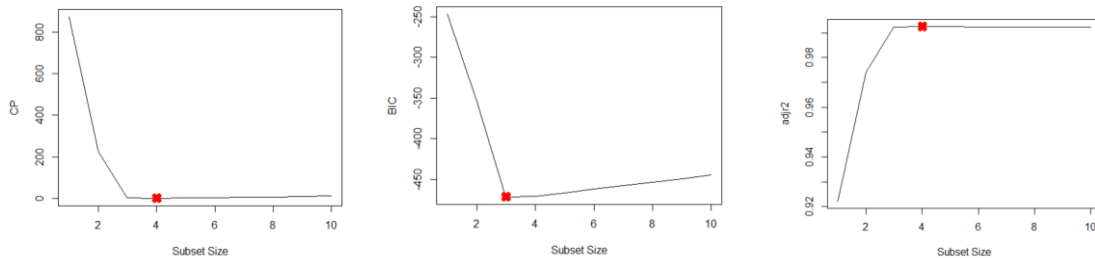
```
> data.all<-data.frame(y=Y, x=X)
> mod.full=regsubsets(y~poly(x, 10, raw=T), data=data.all, nvmax=10)
> mod.summary=summary(mod.full)
```

Find the model size for best cp, BIC and adjusted R2

```
> min.cp=which.min(mod.summary$cp)
> min.bic=which.min(mod.summary$bic)
> max.adj2=which.max(mod.summary$adj2)
```

Plot cp, BIC, and adjusted R2

```
> plot(mod.summary$cp, xlab="Subset Size", ylab="CP", pch=20, type="l")
> points(min.cp, mod.summary$cp[min.cp], pch=4, col="red", lwd=7)
> plot(mod.summary$bic, xlab="Subset Size", ylab="BIC", pch=20, type="l")
> points(min.bic, mod.summary$bic[min.bic], pch=4, col="red", lwd=7)
> plot(mod.summary$adj2, xlab="Subset Size", ylab="adjr2", pch=20, type="l")
> points(max.adj2, mod.summary$adj2[max.adj2], pch=4, col="red", lwd=7)
```



All three criteria (CP, BIC, and adjusted R2) select 3-variable models. The coefficients of the best 3-variable model:

```
> coefficients(mod.full, id=3)
      (Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)2 poly(x, 10, raw = T)3
      4.061507      8.975280      1.876209      1.017639
```

5. Repeat 3, using forward stepwise selection and using backwards stepwise selection. How does your answer compare to the results in 3?

```
> mod.fwd=regsubsets(y~poly(x, 10, raw=T), data=data.all, nvmax=10, method="forward")
> mod.bwd=regsubsets(y~poly(x, 10, raw=T), data=data.all, nvmax=10, method="backward")
> fwd.summary=summary(mod.fwd)
> bwd.summary=summary(mod.bwd)
```

### Find best forward-selected model size

```
> min.cp.f=which.min(fwd.summary$cp)
> min.bic.f=which.min(fwd.summary$bic)
> max.adj2.f=which.max(fwd.summary$adj2)
```

### Find best backward-selected model size

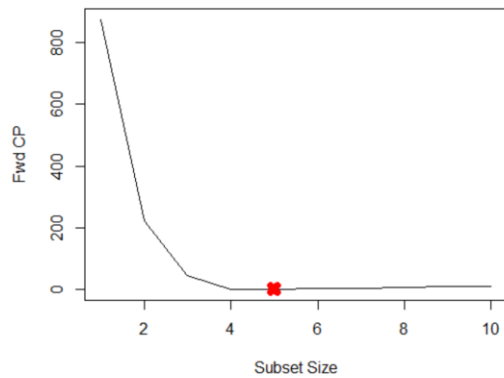
```
> min.cp.b=which.min(bwd.summary$cp)
> min.bic.b=which.min(bwd.summary$bic)
> max.adj2.b=which.max(bwd.summary$adj2)
```

### Plot the statistics

```
> par(mfrow=c(3,2))
```

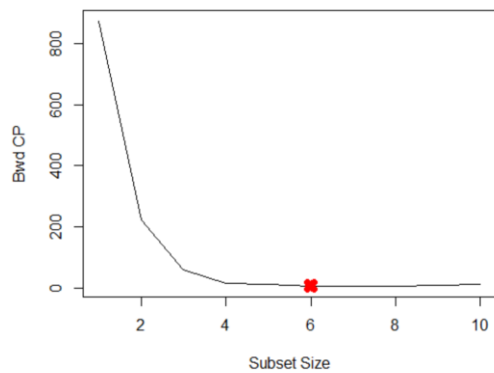
### Forward CP

```
> plot(fwd.summary$cp, xlab="Subset Size", ylab="Fwd CP", pch=20, type="l")
> points(min.cp.f, fwd.summary$cp[min.cp.f], pch=4, col="red", lwd=7)
```



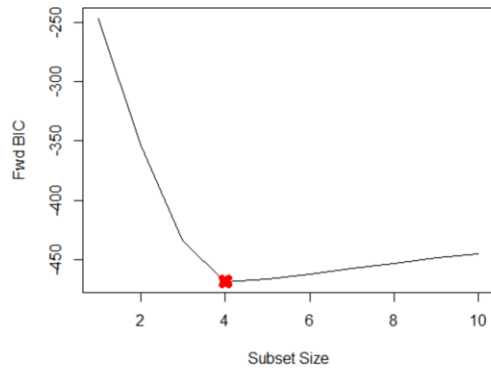
### Backward CP

```
> plot(bwd.summary$cp, xlab="Subset Size", ylab="Bwd CP", pch=20, type="l")
> points(min.cp.b, bwd.summary$cp[min.cp.b], pch=4, col="red", lwd=7)
```



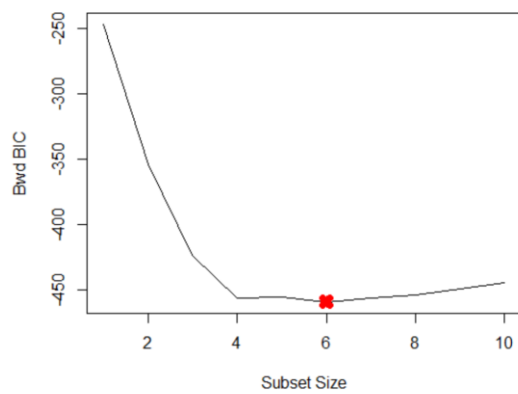
### Forward BIC

```
> plot(fwd.summary$bic, xlab="Subset Size", ylab="Fwd BIC", pch=20, type="l")  
> points(min.bic.f, fwd.summary$bic[min.bic.f], pch=4, col="red", lwd=7)
```



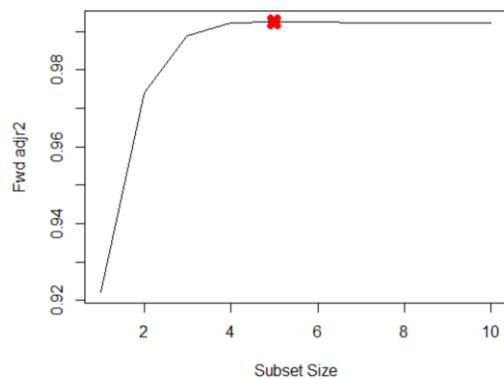
### Backward BIC

```
> plot(bwd.summary$bic, xlab="Subset Size", ylab="Bwd BIC", pch=20, type="l")  
> points(min.bic.b, bwd.summary$bic[min.bic.b], pch=4, col="red", lwd=7)
```



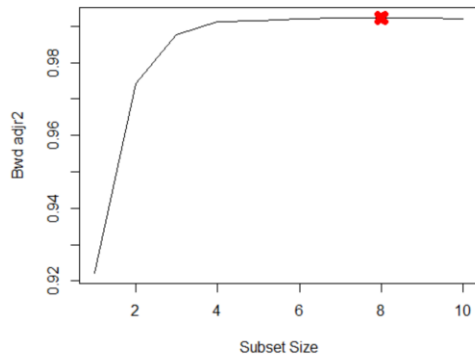
### Forward Adj R2

```
> plot(fwd.summary$adjr2, xlab="Subset Size", ylab="Fwd adjr2", pch=20, type="l")  
> points(max.adjr2.f, fwd.summary$adjr2[max.adjr2.f], pch=4, col="red", lwd=7)
```



## Backward Adj R2

```
> plot(bwd.summary$adjr2, xlab="Subset Size", ylab="Bwd adjr2", pch=20, type="l")
> points(max.adj2.b, bwd.summary$adjr2[max.adj2.b], pch=4, col="red", lwd=7)
```



All statistics select 3-variable models except for Adjusted R2 backward selection – these are the coefficients:

```
> coefficients(mod.fwd, id=3)
(Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)3 poly(x, 10, raw = T)4
 4.7873698      9.2517551      0.8949995      0.4048019

> coefficients(mod.bwd, id=3)
(Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)4 poly(x, 10, raw = T)5
 4.7875054      10.1410271      0.3898228      0.1474814

> coefficients(mod.bwd, id=4)
(Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)4 poly(x, 10, raw = T)5 poly(x, 10, raw = T)6
 4.4954810      9.9317370      1.0171097      0.1829648      -0.1332826
```

The performance of both forward and backward stepwise selection methods still presents the same recommended model as the results in 3.