

Individual Assignment 5

Charlie Ling

9/29/2021

Problem 8: In this exercise, we will generate simulated data, and will then use this data to perform best subset selection.

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

```
set.seed(2)
X=rnorm(100)
e=rnorm(100)
```

- (b) Generate a response vector Y of length $n=100$ according to the model

$$Y=B_0+B_1X+B_2X^2+B_3X^3+e$$

Where B_0 , B_1 , B_2 , and B_3 are constants of your choice.

$$Y=0.3+0.5*X+0.2*X^2+0.1*X^3+e$$

- (c) Use the `regsubsets()` function to perform best subset selection in order to choose the best model containing the predictors X, X^2, \dots, X^{10} . What is the best model obtained according to C_p , BIC, and adjusted R^2 ? 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 .

```
df=data.frame(Y,X)
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.0.5
```

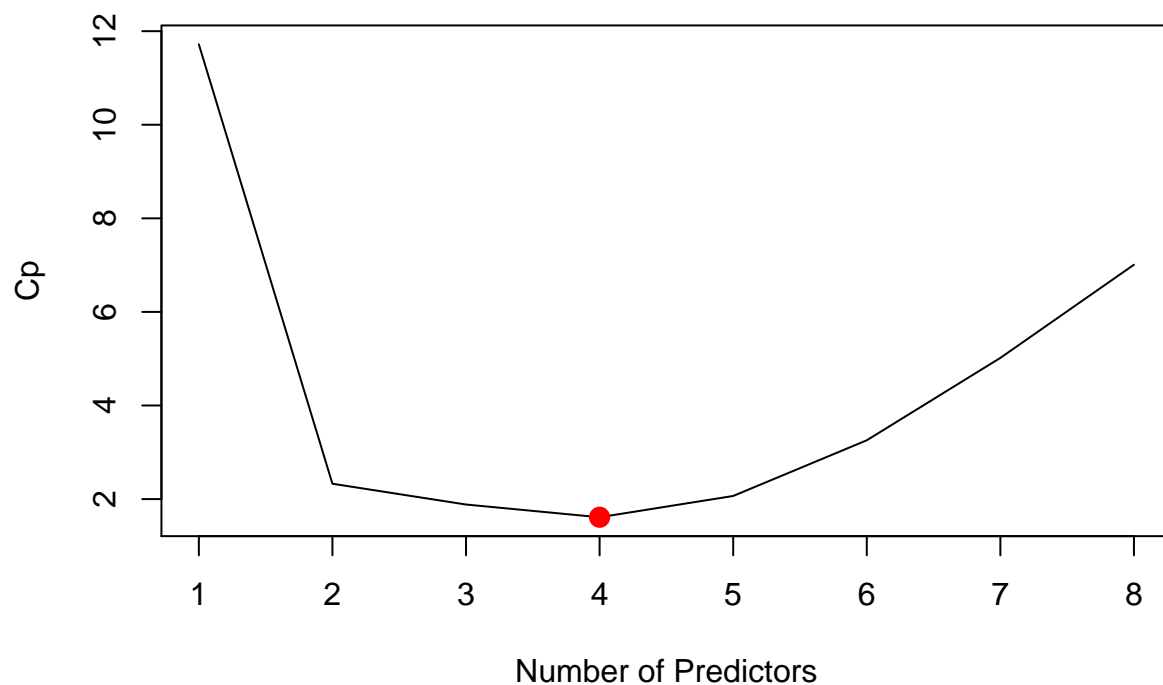
```
regfit.full = regsubsets(Y~poly(X,10),data=df)
reg.summary = summary(regfit.full)
```

```
plot(reg.summary$cp,xlab="Number of Predictors", ylab = "Cp", type = "l")
```

```
which.min(reg.summary$cp)
```

```
## [1] 4
```

```
points(4,reg.summary$cp[4],col="red",cex=2,pch=20)
```



#Model 4 is the best model obtained according to Cp

```
coef(regfit.full,4)
```

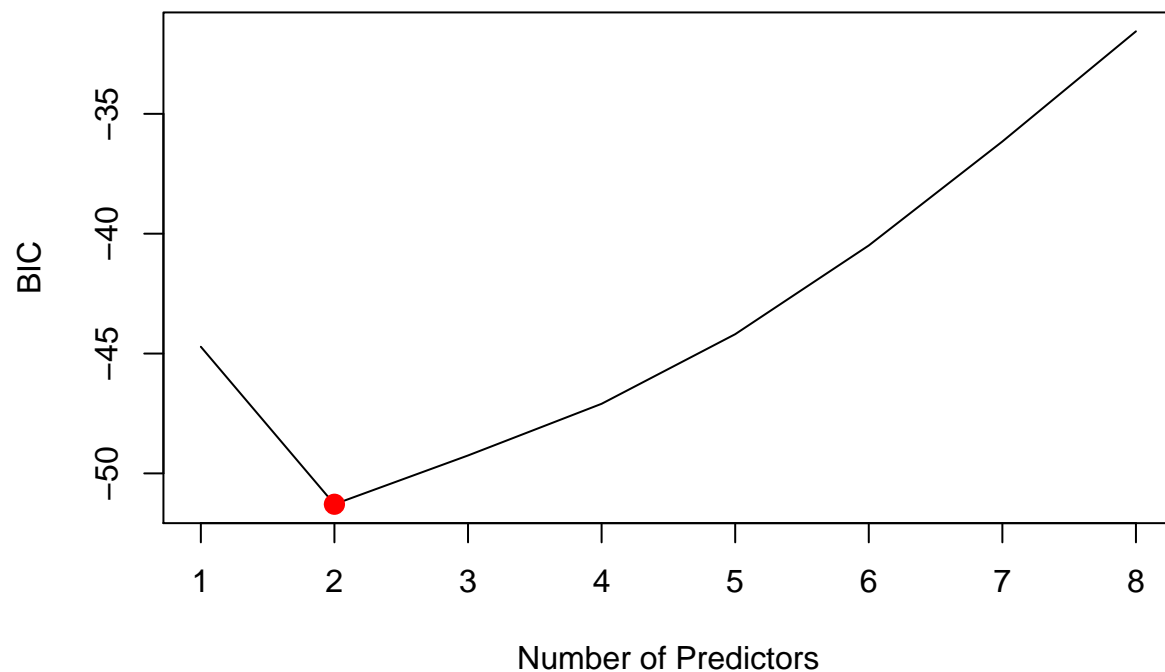
```
## (Intercept) poly(X, 10)1 poly(X, 10)2 poly(X, 10)3 poly(X, 10)10
## 0.5854608 8.8254685 3.3945159 1.5159303 1.5726414
```

```
plot(reg.summary$bic,xlab="Number of Predictors", ylab = "BIC", type = "l")
```

```
which.min(reg.summary$bic)
```

```
## [1] 2
```

```
points(2,reg.summary$bic[2],col="red",cex=2,pch=20)
```



```
#Model 2 is the best model obtained according to bic
coef(regfit.full,2)
```

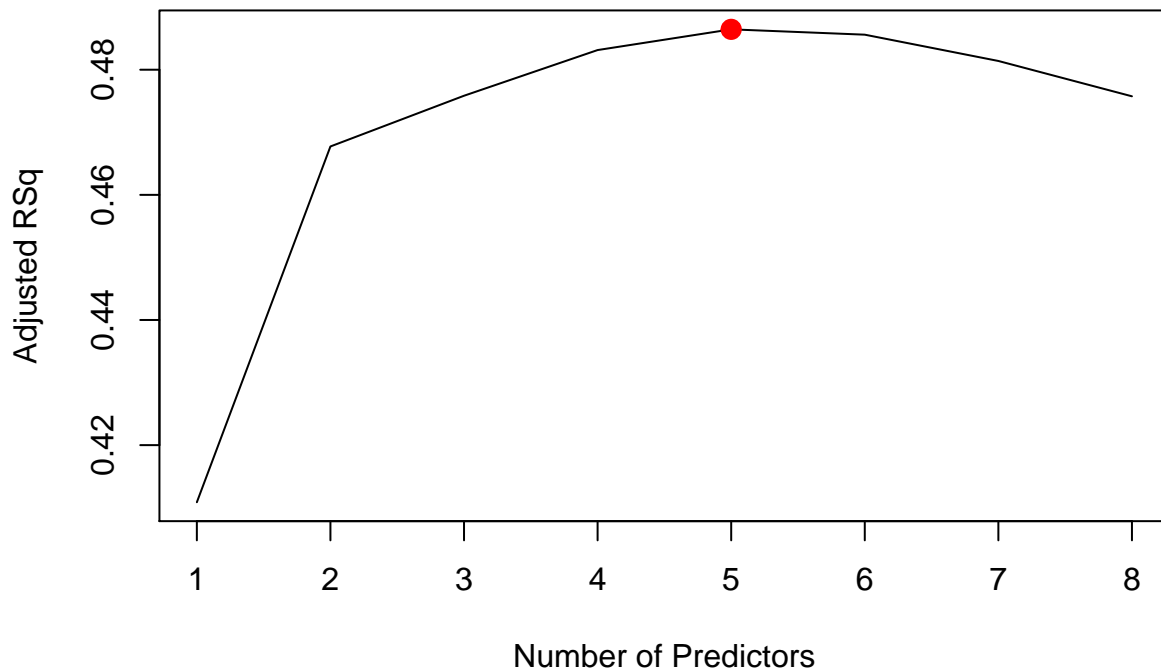
```
## (Intercept) poly(X, 10)1 poly(X, 10)2
## 0.5854608 8.8254685 3.3945159
```

```
plot(reg.summary$adjr2,xlab="Number of Predictors", ylab = "Adjusted RSq", type = "l")
```

```
which.max(reg.summary$adjr2)
```

```
## [1] 5
```

```
points(5,reg.summary$adjr2[5],col="red",cex=2,pch=20)
```



```
#Model 5 is the best model obtained according to adjr2
coef(regfit.full,5)
```

```
##      (Intercept)  poly(X, 10)1  poly(X, 10)2  poly(X, 10)3  poly(X, 10)7
##      0.5854608    8.8254685    3.3945159    1.5159303    1.2502713
## poly(X, 10)10
##      1.5726414
```

(d) Repeat (c), using forward stepwise selection and also using backwards stepwise selection. How does your answer compare to the results in (c)?

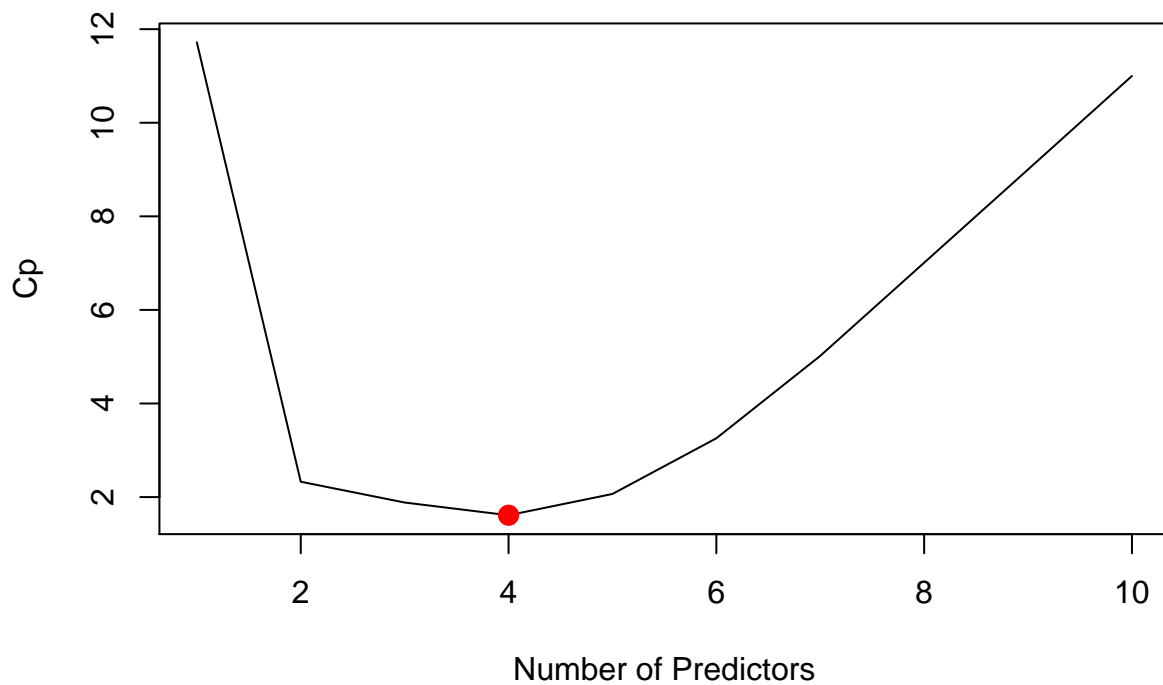
```
regfit.bwd = regsubsets(Y~poly(X,10),data=df, nvmax = 10, method = "backward")
reg.summary = summary(regfit.bwd)
```

```
plot(reg.summary$cp,xlab="Number of Predictors", ylab = "Cp", type = "l")
```

```
which.min(reg.summary$cp)
```

```
## [1] 4
```

```
points(4,reg.summary$cp[4],col="red",cex=2,pch=20)
```

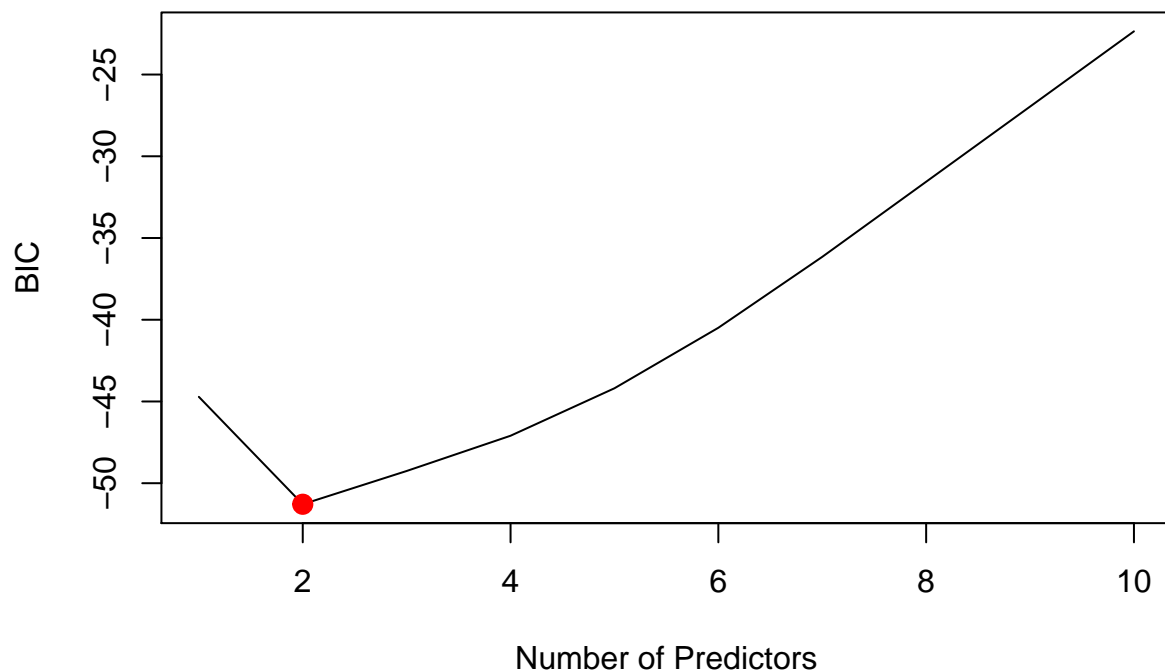


```
#Model 4 is the best model obtained according to Cp
coef(regfit.full,4)
```

```
## (Intercept) poly(X, 10)1 poly(X, 10)2 poly(X, 10)3 poly(X, 10)10
## 0.5854608 8.8254685 3.3945159 1.5159303 1.5726414
plot(reg.summary$bic,xlab="Number of Predictors", ylab = "BIC", type = "l")

which.min(reg.summary$bic)

## [1] 2
points(2,reg.summary$bic[2],col="red",cex=2,pch=20)
```



```
#Model 2 is the best model obtained according to bic
coef(regfit.full,2)
```

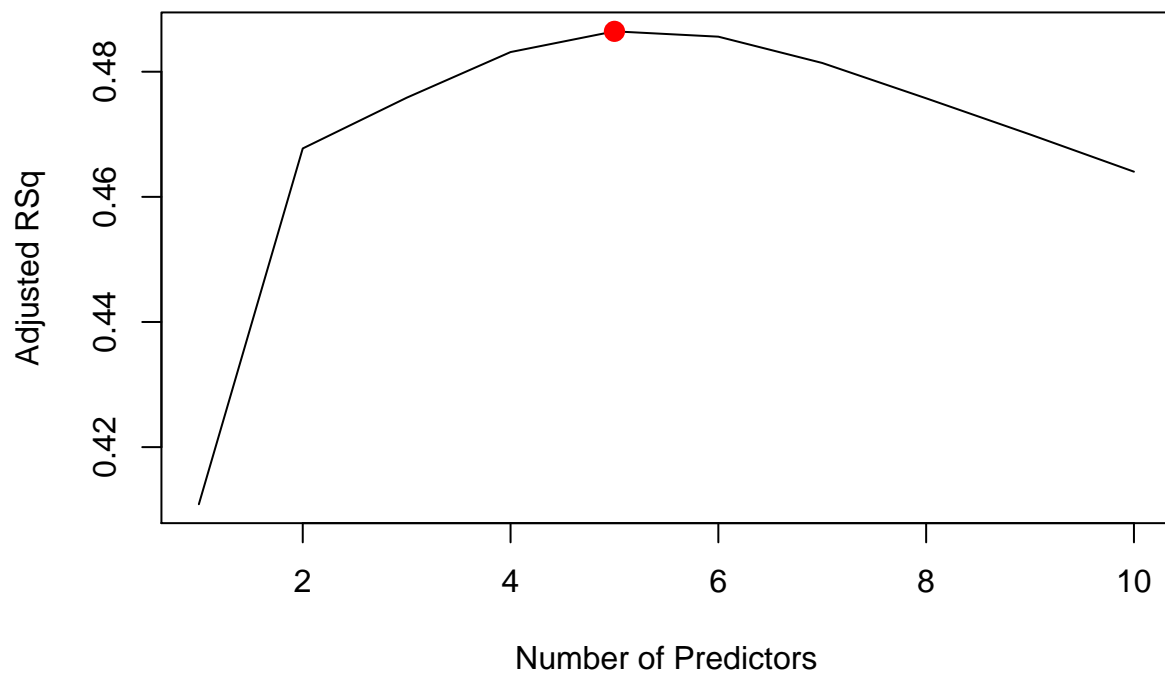
```
## (Intercept) poly(X, 10)1 poly(X, 10)2
## 0.5854608 8.8254685 3.3945159
```

```
plot(reg.summary$adjr2,xlab="Number of Predictors", ylab = "Adjusted RSq", type = "l")
```

```
which.max(reg.summary$adjr2)
```

```
## [1] 5
```

```
points(5,reg.summary$adjr2[5],col="red",cex=2,pch=20)
```



#Model 5 is the best model obtained according to adjr2

```
coef(regfit.full,5)
```

```
## (Intercept) poly(X, 10)1 poly(X, 10)2 poly(X, 10)3 poly(X, 10)7
## 0.5854608 8.8254685 3.3945159 1.5159303 1.2502713
## poly(X, 10)10
## 1.5726414
```

The results are the same to the results in (c)