Homework4

Xingyuan Chen

2019/4/10

1. We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain $p + 1$ models, containing $0, 1, 2, ..., p$ predictors. Explain your answers:

(a) Which of the three models with k predictors has the smallest training RSS?

Answer:
Best subset selection.

(b) Which of the three models with k predictors has the smallest test RSS?

Answer:
All models may have the smallest test RSS, depend which is closest to the real true model.

(c) True or False:

i. The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the (k+1)-variable model identified by forward stepwise selection.
True.

ii. The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the (k+1)-variable model identified by backward stepwise selection.
True.

iii. The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the (k+1)-variable model identified by forward stepwise selection.
False.

iv. The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the (k+1)-variable model identified by backward stepwise selection.
False.

v. The predictors in the k-variable model identified by best subset are a subset of the predictors in the (k+1)-variable model identified by best subset selection.
False.

2. For parts (a) through (c), indicate which of i. through iv. is correct. Justify your answer.

(a) The lasso, relative to least squares, is:

i. More flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.

ii. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.

iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.

iv. Less flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.
Answer:iii.

(b) Repeat (a) for ridge regression relative to least squares.
Answer:iii.

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 ϵ of length $n = 100$.

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

(b) 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 $\beta_0, \beta_1, \beta_2, \beta_3$ are constants of your choice.

```
b0 = 1
b1 = 2
b2 = 3
b3 = 4
Y = b0 + b1 * X + b2 * X^2 + b3 * X^3 + eps
```

(c) Use the regsubsets() function to perform best subset selection in order to choose the best model containing the predictors $X, X^2, ..., X^{10}$. 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.

```
library(leaps)
data = data.frame(y = Y, x = X)
model = regsubsets(y ~ poly(x, 10, raw = T), data = data, nvmax = 10)
summary = summary(model)
which.min(summary$cp)
```

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

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

```
## [1] 3
```

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

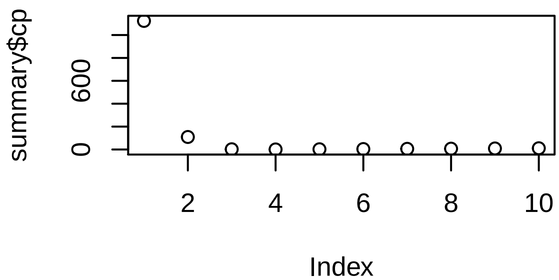
```
## [1] 4
```

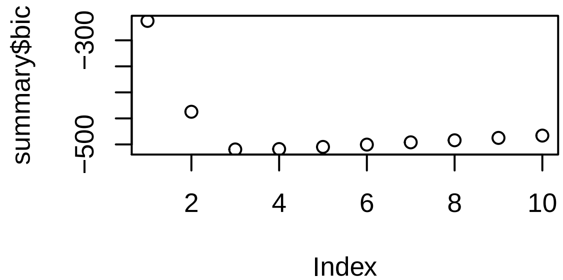
```
par(mfrow = c(2, 2))
plot(summary$cp)
plot(summary$bic)
plot(summary$adjr2)
coefficients(model, id=3)
```

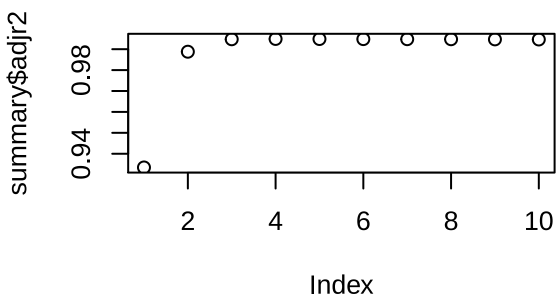
```
##           (Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)2
##              1.061507              1.975280              2.876209
## poly(x, 10, raw = T)3
##              4.017639
```

```r
coefficients(model, id=4)
```

```
##           (Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)2
##            1.07200775            2.38745596            2.84575641
## poly(x, 10, raw = T)3 poly(x, 10, raw = T)5
##            3.55797426            0.08072292
```







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

```r
fwd = regsubsets(y ~ poly(x, 10, raw = T), data = data, nvmax = 10,
    method = "forward")
bwd = regsubsets(y ~ poly(x, 10, raw = T), data = data, nvmax = 10,
    method = "backward")
fwd.summary = summary(fwd)
bwd.summary = summary(bwd)
which.min(fwd.summary$cp)
```

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

```r
which.min(bwd.summary$cp)
```

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

```r
which.min(fwd.summary$bic)
```

```
## [1] 3
```

```r
which.min(bwd.summary$bic)
```

```
## [1] 3
```

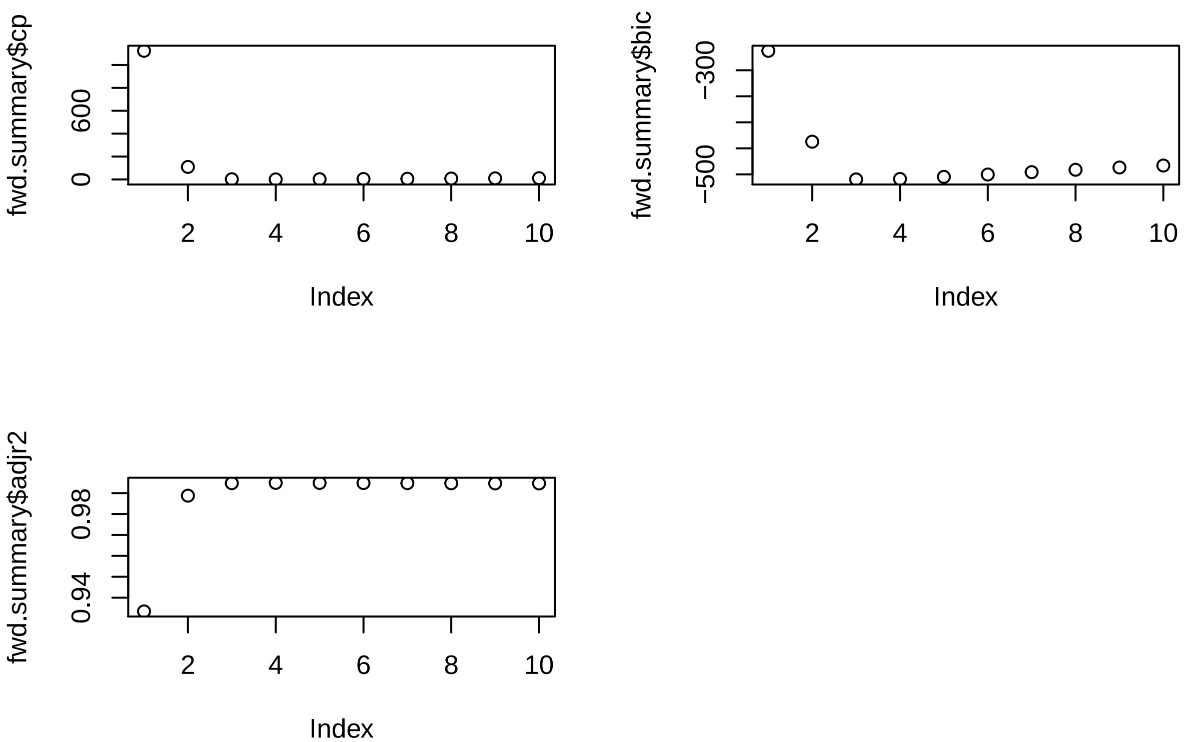
```r
which.max(fwd.summary$adjr2)
```

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

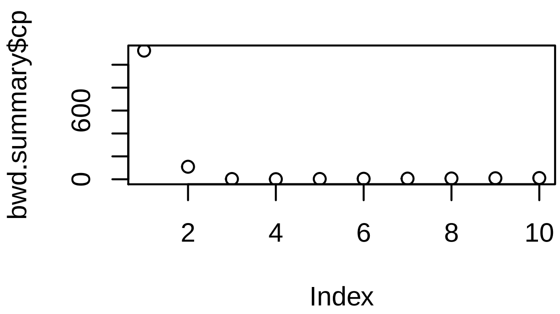
```r
which.max(bwd.summary$adjr2)
```

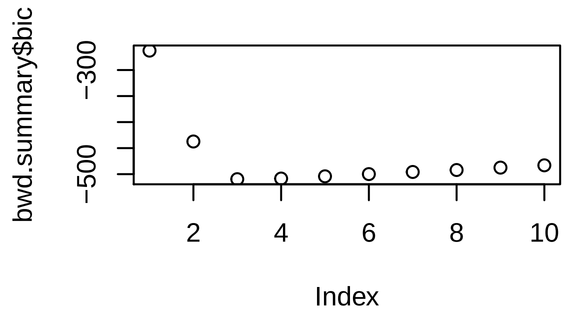
```
## [1] 4
```

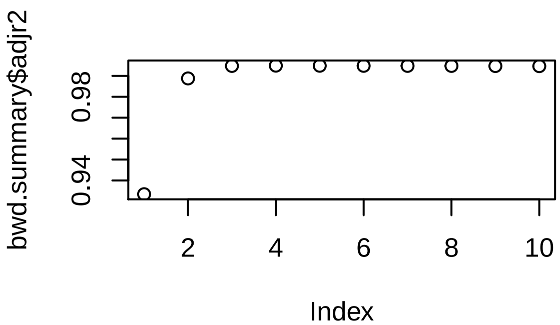
```r
par(mfrow = c(2, 2))
plot(fwd.summary$cp)
plot(fwd.summary$bic)
plot(fwd.summary$adjr2)
par(mfrow = c(2, 2))
```




```r
plot(bwd.summary$cp)
plot(bwd.summary$bic)
plot(bwd.summary$adjr2)
```







Answer is the same.

(e) Now fit a lasso model to the simulated data, again using $X, X^2, ..., X^{10}$ as predictors. Use cross-validation to select the optimal value of λ. Create plots of the cross-validation error as a function of λ. Report the resulting coefficient estimates, and discuss the results obtained.

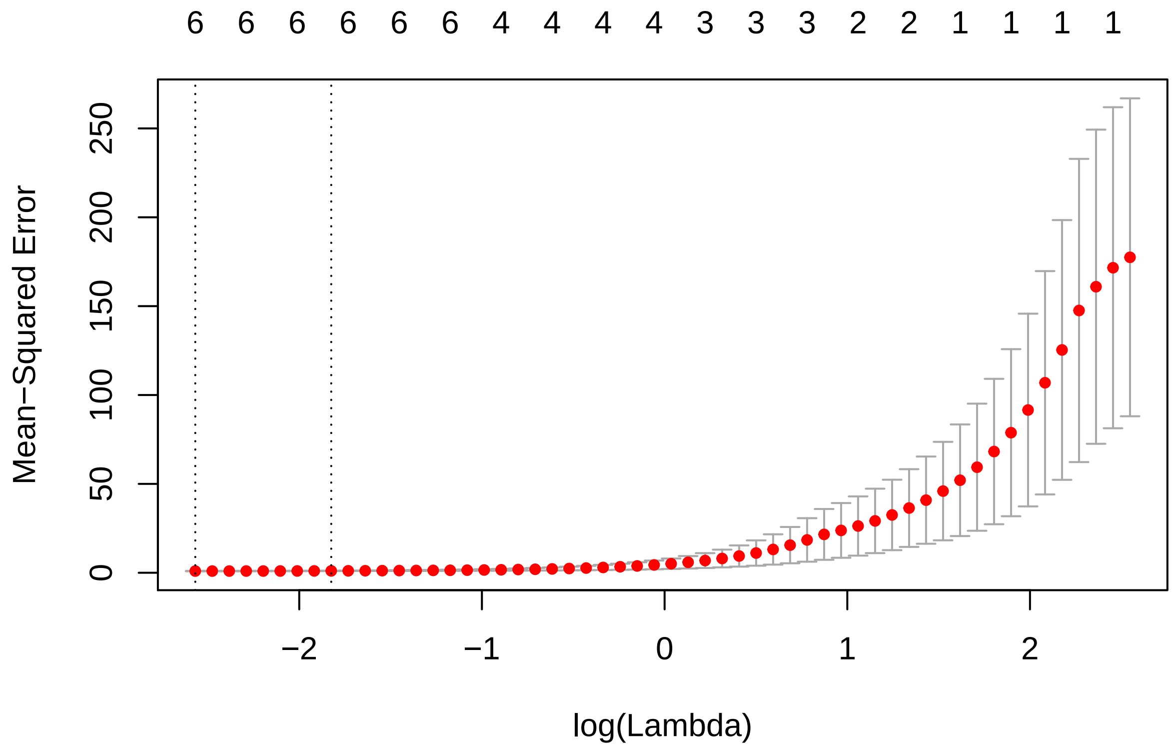
```
library(glmnet)
```

```
## Loading required package: Matrix

## Loading required package: foreach

## Loaded glmnet 2.0-16
```

```
xmatrix = model.matrix(y ~ poly(x, 10, raw = T), data = data)[, -1]
lasso = cv.glmnet(xmatrix, Y, alpha = 1)
minlambda = lasso$lambda.min
minlambda
```

```
## [1] 0.07660225
```

```
plot(lasso)
```




```
bestlasso = glmnet(xmatrix, Y, alpha = 1)
predict(bestlasso,s=minlambda, type = "coefficients")
```

```
## 11 x 1 sparse Matrix of class "dgCMatrix"
##                                1
## (Intercept)           1.182646169
## poly(x, 10, raw = T)1  2.137739131
## poly(x, 10, raw = T)2  2.623547995
## poly(x, 10, raw = T)3  3.813195738
## poly(x, 10, raw = T)4  0.042303133
## poly(x, 10, raw = T)5  0.012404464
## poly(x, 10, raw = T)6  .
## poly(x, 10, raw = T)7  0.003849104
## poly(x, 10, raw = T)8  .
## poly(x, 10, raw = T)9  .
## poly(x, 10, raw = T)10 .
```

Besides 1,2,3,5, lasso also picked 4 and 7.

(f) Now generate a response vector Y according to the model $Y = \beta_0 + \beta_7 X^7 + \epsilon$, and perform best subset selection and the lasso. Discuss the results obtained.

```
Y = b0 + 7 * X^7 + eps
dataf = data.frame(y = Y, x = X)
modelf = regsubsets(y ~ poly(x, 10, raw = T), data = data, nvmax = 10)
modelfsummary = summary(modelf)
which.min(modelfsummary$cp)
```

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

```
which.min(modelfsummary$bic)
```

```
## [1] 3
```

```r
which.max(modelfsummary$adjr2)
```

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

```r
coefficients(modelf, id = 3)
```

```
##           (Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)2
##              1.061507              1.975280              2.876209
## poly(x, 10, raw = T)3
##              4.017639
```

```r
coefficients(modelf, id = 4)
```

```
##           (Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)2
##            1.07200775            2.38745596            2.84575641
## poly(x, 10, raw = T)3 poly(x, 10, raw = T)5
##            3.55797426            0.08072292
```

```r
xfmatrix = model.matrix(y ~ poly(x, 10, raw = T), data = dataf)[, -1]
lassof = cv.glmnet(xfmatrix, Y, alpha=1)
minlambdaf = lassof$lambda.min
minlambdaf
```

```
## [1] 13.57478
```

```r
bestlassof = glmnet(xfmatrix, Y, alpha = 1)
predict(bestlassof, s = minlambdaf, type = "coefficients")
```

```
## 11 x 1 sparse Matrix of class "dgCMatrix"
##                               1
## (Intercept)            1.904188
## poly(x, 10, raw = T)1  .
## poly(x, 10, raw = T)2  .
## poly(x, 10, raw = T)3  .
## poly(x, 10, raw = T)4  .
## poly(x, 10, raw = T)5  .
## poly(x, 10, raw = T)6  .
## poly(x, 10, raw = T)7  6.776797
## poly(x, 10, raw = T)8  .
## poly(x, 10, raw = T)9  .
## poly(x, 10, raw = T)10 .
```

Best subset picked 1,2,3,(5), while lasso only picked 7.