Homework6

Jose Chavarria

4/29/2020

```
library(ISLR)
data('College')
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

```
## Loaded glmnet 2.0-18
```

```
library(pls)
```

```
##
## Attaching package: 'pls'
```

```
## The following object is masked from 'package:stats':
##
##     loadings
```

A.

```
set.seed(1)
subset <- sample(nrow(College), nrow(College)*.8, replace = F)
train <- College[subset,]
test <- College[-subset,]
```

B.

```
model1 <- lm(Apps~., data = train)
pred.lm <- predict(model1, test)
lm.er <- mean((pred.lm - test$Apps)^2)
lm.er
```

```
## [1] 1075064
```

C.

```
train.mat <- model.matrix(Apps~., data = train)
test.mat <- model.matrix(Apps~., data = test)
grid <- 10^seq(4,-2,length=100)

fit.ridge <- glmnet(train.mat, train$Apps, alpha=0, lambda = grid, thresh = 1e-12)
cv.ridge <- cv.glmnet(train.mat, train$Apps, alpha=0, lambda = grid, thresh = 1e-12)

bestlam.ridge <- cv.ridge$lambda.min
bestlam.ridge
```

```
## [1] 0.01
```

```
pred.ridge <- predict(fit.ridge, s = bestlam.ridge, newx = test.mat)
mean((pred.ridge - test$Apps)^2)
```

```
## [1] 1075062
```

MSE = 769,103.1

D.

```
fit.lasso <- glmnet(train.mat, train$Apps, alpha = 1, lambda = grid, thresh = 1e-12)
cv.lasso <- cv.glmnet(train.mat, train$Apps, alpha = 1, lambda = grid, thresh = 1e-12)
bestlam.lasso <- cv.lasso$lambda.min
bestlam.lasso
```

```
## [1] 16.29751
```

```
pred.lasso <- predict(fit.lasso, s = bestlam.lasso, newx = test.mat)
mean((pred.lasso - test$Apps)^2)
```

```
## [1] 1106467
```

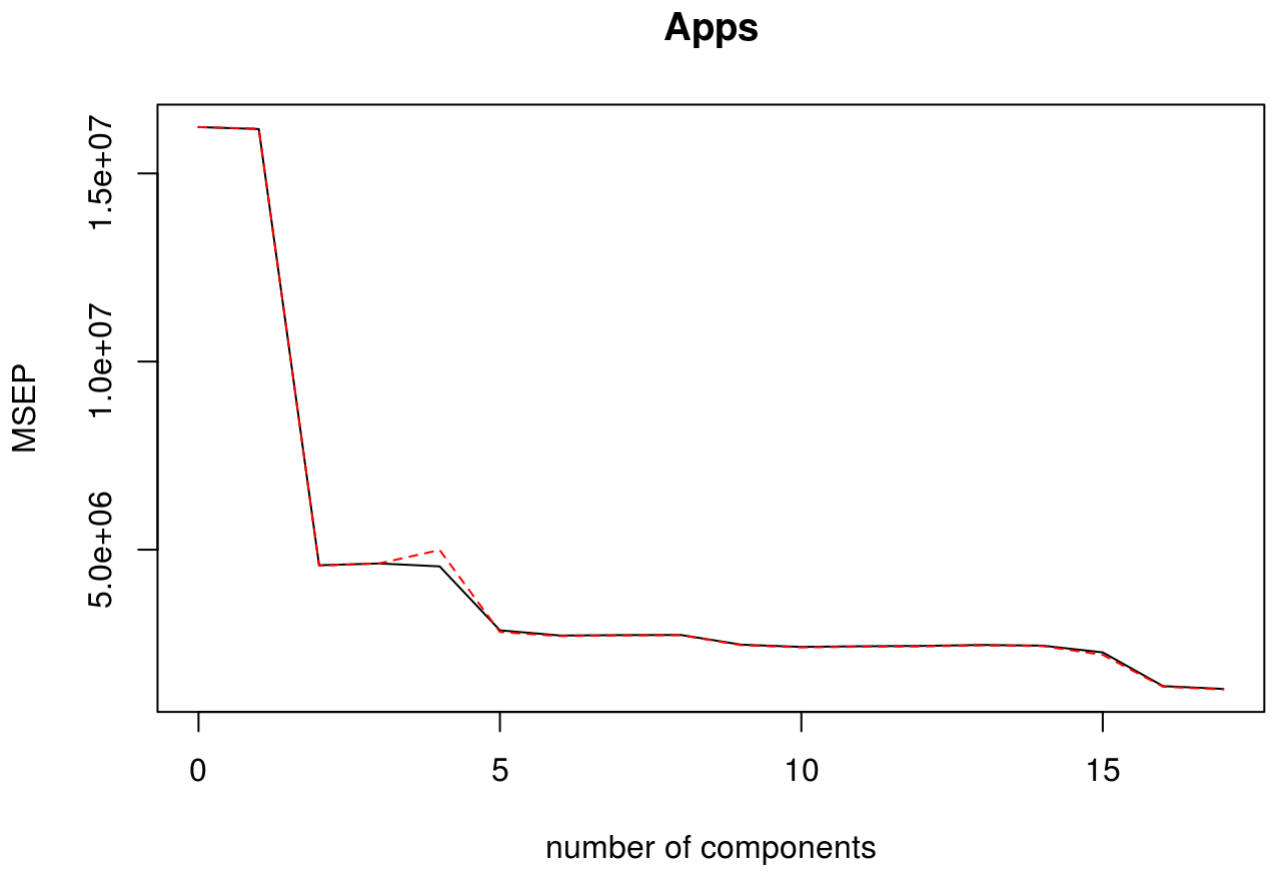
```
predict(fit.lasso, s = bestlam.lasso, type = "coefficients")
```

```
## 19 x 1 sparse Matrix of class "dgCMatrix"
##                          1
## (Intercept) -5.179713e+02
## (Intercept)  .
## PrivateYes  -4.602369e+02
## Accept       1.503720e+00
## Enroll      -3.504466e-01
## Top10perc    3.553232e+01
## Top25perc   -3.274772e+00
## F.Undergrad  .
## P.Undergrad  .
## Outstate    -6.420967e-02
## Room.Board   1.521474e-01
## Books        .
## Personal     3.613238e-03
## PhD         -5.173327e+00
## Terminal    -4.564563e+00
## S.F.Ratio    7.569006e+00
## perc.alumni -3.004123e+00
## Expend       6.566683e-02
## Grad.Rate    5.421194e+00
```

The MSE is larger than both ridge regression and least squares

> E.

```
fit.pcr <- pcr(Apps~., data = train, scale = TRUE, validation = "CV")
validationplot(fit.pcr, val.type = "MSEP")
```



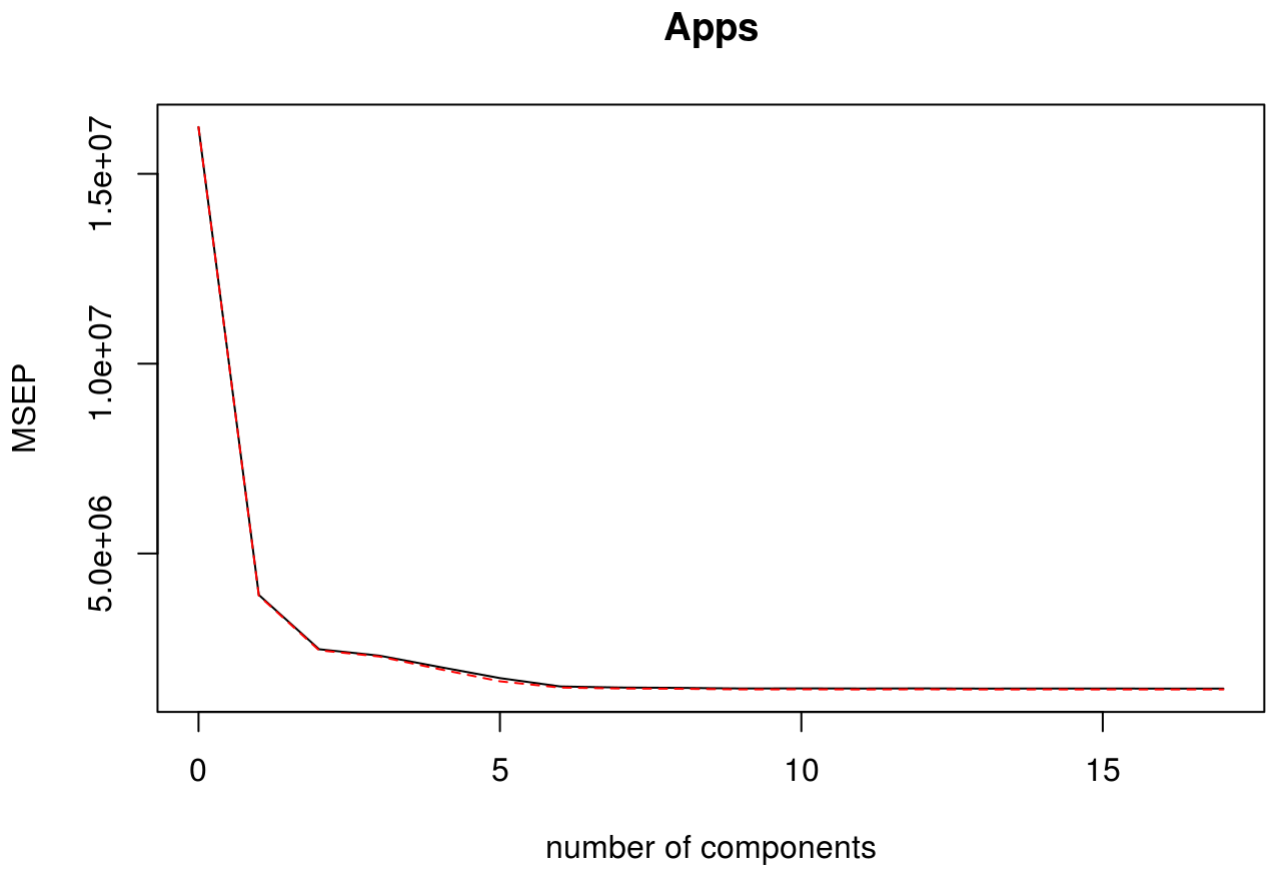

```
pred.pcr <- predict(fit.pcr, test, ncomp = 10)
mean((pred.pcr - test$Apps)^2)
```

```
## [1] 1553814
```

The MSE is higher than the the previous models

> F.

```
fit.pls <- plsr(Apps ~ ., data = train, scale = TRUE, validation = "CV")
validationplot(fit.pls, val.type = "MSEP")
```




```
pred.pls <- predict(fit.pls, test, ncomp = 10)
mean((pred.pls - test$Apps)^2)
```

```
## [1] 1075376
```

G.

```
test.avg <- mean(test$Apps)
lm.r2 <- 1 - mean((pred.lm - test$Apps)^2) / mean((test.avg - test$Apps)^2)
ridge.r2 <- 1 - mean((pred.ridge - test$Apps)^2) / mean((test.avg - test$Apps)^2)
lasso.r2 <- 1 - mean((pred.lasso - test$Apps)^2) / mean((test.avg - test$Apps)^2)
pcr.r2 <- 1 - mean((pred.pcr - test$Apps)^2) / mean((test.avg - test$Apps)^2)
pls.r2 <- 1 - mean((pred.pls - test$Apps)^2) / mean((test.avg - test$Apps)^2)

rsqrd <- matrix(c(lm.r2,ridge.r2,lasso.r2,pcr.r2,pls.r2),ncol=1,byrow=TRUE)
colnames(rsqrd) <- c("R-Square")
rownames(rsqrd) <- c("lm.r2","ridge.r2","lasso.r2","pcr.r2","pls.r2")
rsqrd <- as.table(rsqrd)
rsqrd
```

```
##           R-Square
## lm.r2    0.8926812
## ridge.r2 0.8926814
## lasso.r2 0.8895463
## pcr.r2   0.8448897
## pls.r2   0.8926500
```

Our models appear to predict applications better except for the PCR model when we look at our rsquared