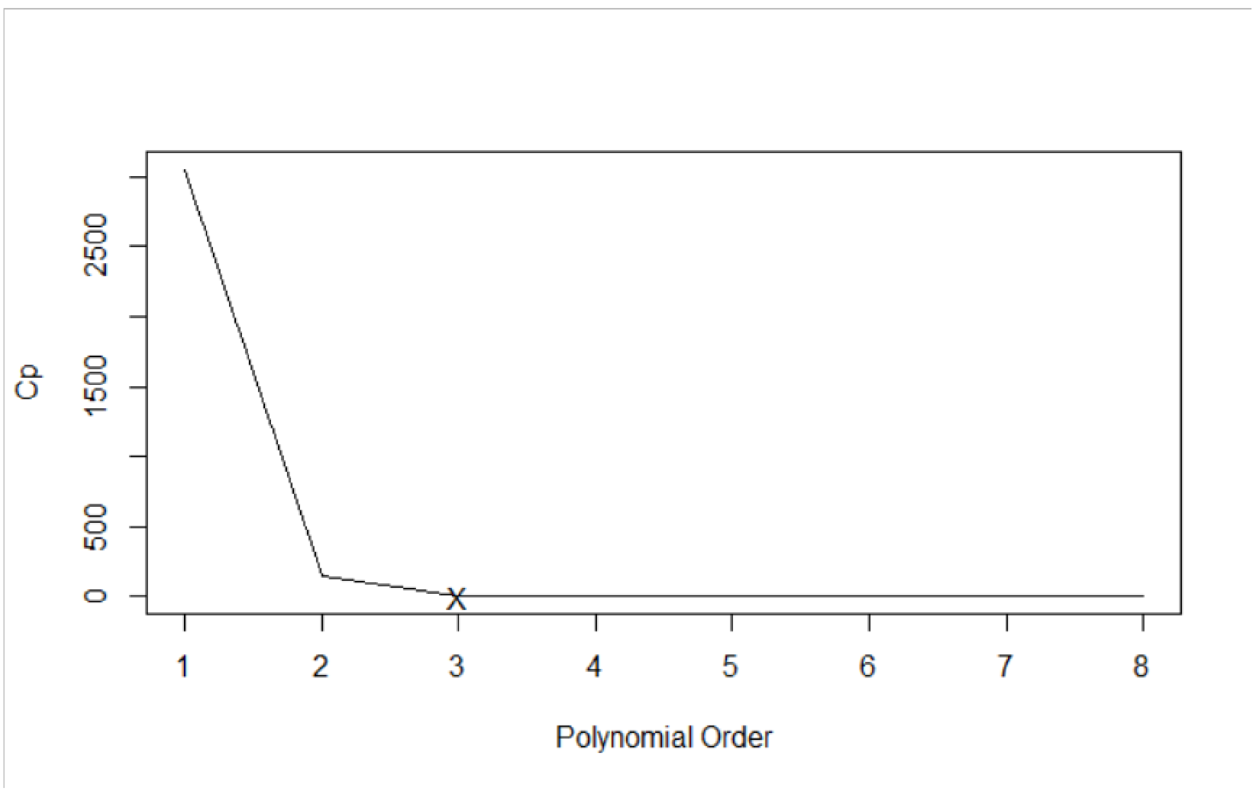
Answers to Review Week Five

1. Chapter 6, question 3.
   (a) The parameter s will increase as λ is increased. This will result in in a steady increase in the training RSS.
   (b) The test RSS should show a pattern similar to the test MSE, initial decrease followed by an increase.
   (c) The variance will decrease as the s gets larger.
   (d) The squared bias will increase as s increases.
   (e) The irreducible error stays constant.
2. Chapter 6, question 8.

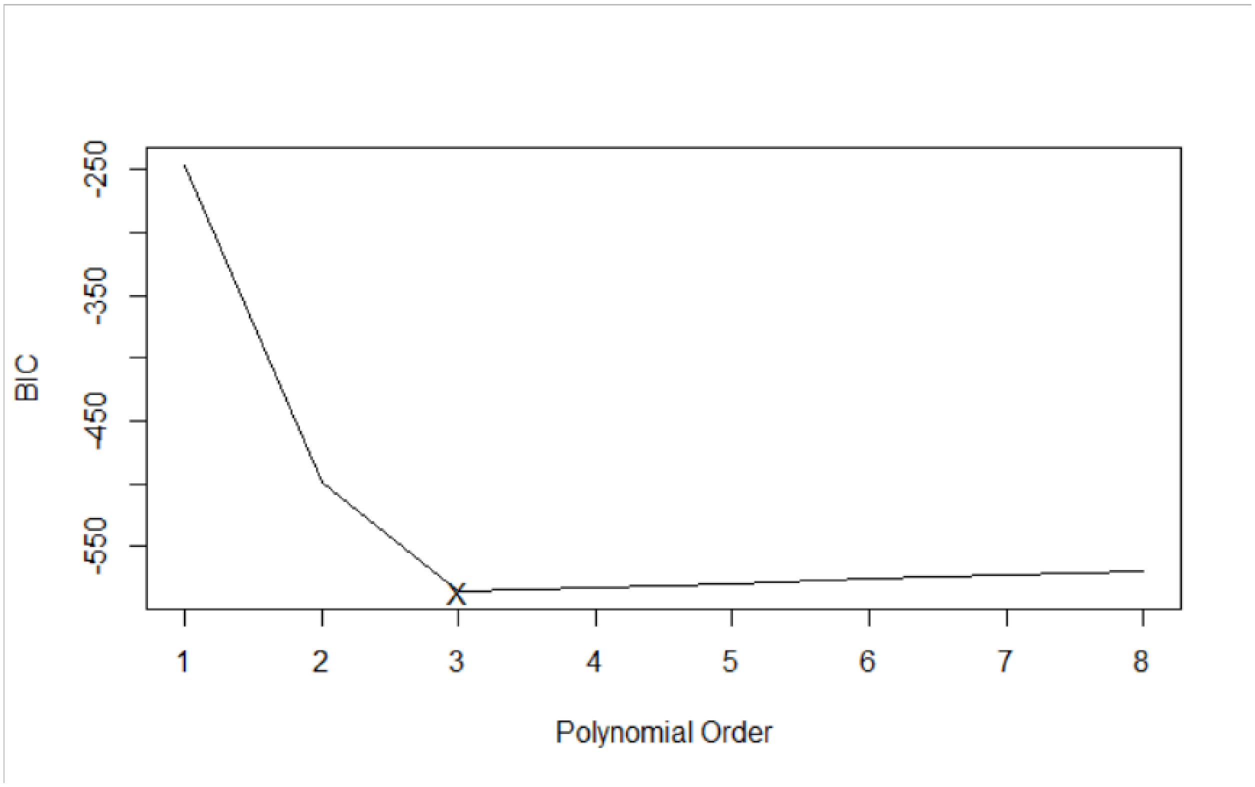
Chapter 6, question 8

```
# (a)
set.seed(100)
x<- rnorm(100)
irr.error<- rnorm(100)
# (b)
b3<- 4
b2<- -3
b1<- 2
b0<- 1
y<- rep(0,100)
y<- b0+b1*x+b2*x^2+b3*x^3+irr.error
sim.data<-
data.frame("Y"=y,"X"=x,"X2"=x^2,"X3"=x^3,"X4"=x^4,"X5"=x^5,"X6"=
x^6,"X7"=x^7,"X8"=x^8,"X9"=x^9,"X10"=x^10)
# (c)
library(leaps)
y.subsets<- regsubsets(Y~., sim.data)
summary.y.subsets<- summary(y.subsets)
summary.y.subsets


...
1 subsets of each size up to 8
Selection Algorithm: exhaustive
         X   X2  X3  X4  X5  X6  X7  X8  X9  X10
1  ( 1 ) " " " " "*" " " " " " " " " " " " " " "
2  ( 1 ) " " "*" "*" " " " " " " " " " " " " " "
3  ( 1 ) "*" "*" "*" " " " " " " " " " " " " " "
4  ( 1 ) "*" "*" "*" "*" " " " " " " " " " " " "
5  ( 1 ) "*" "*" "*" "*" "*" " " " " " " " " " "
6  ( 1 ) "*" "*" "*" " " "*" " " " " "*" " " "*"
7  ( 1 ) "*" "*" "*" "*" " " "*" " " "*" " " "*"
```

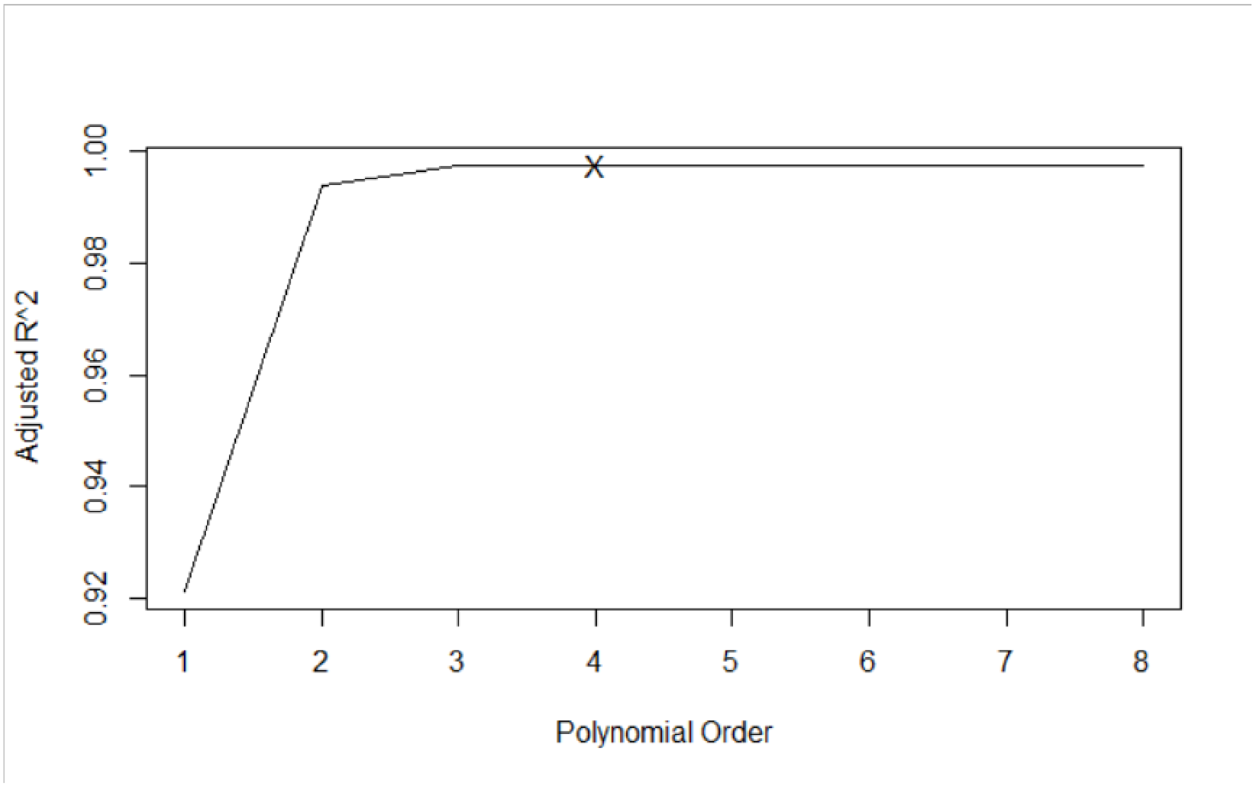
```
8  ( 1 ) "*" "*" "*" "*" "*" "*" " " "*" " " "*"

names(summary.y.subsets)
[1] "which"  "rsq"    "rss"    "adjr2"  "cp"     "bic"
"outmat" "obj"
# Find best model from Cp, BIC and adjusted R^2
summary.y.subsets$cp
[1] 3055.123715  145.957425    3.821967    3.846742    5.309942
6.014323    6.523634
[8]    7.149537
summary.y.subsets$bic
[1] -245.9505 -498.8506 -586.0424 -583.5201 -579.4886 -576.2817
-573.3098 -570.2342
summary.y.subsets$adjr2
[1] 0.9212483 0.9939413 0.9975554 0.9975806 0.9975688 0.9975768
0.9975901 0.9976006
plot(1:8,summary.y.subsets$cp,type="l",xlab="Polynomial Order",
ylab="Cp")
text(3,summary.y.subsets$cp[3],"X")
```




```
plot(1:8,summary.y.subsets$bic,type="l",xlab="Polynomial Order",
ylab="BIC")
text(3,summary.y.subsets$bic[3],"X")
```



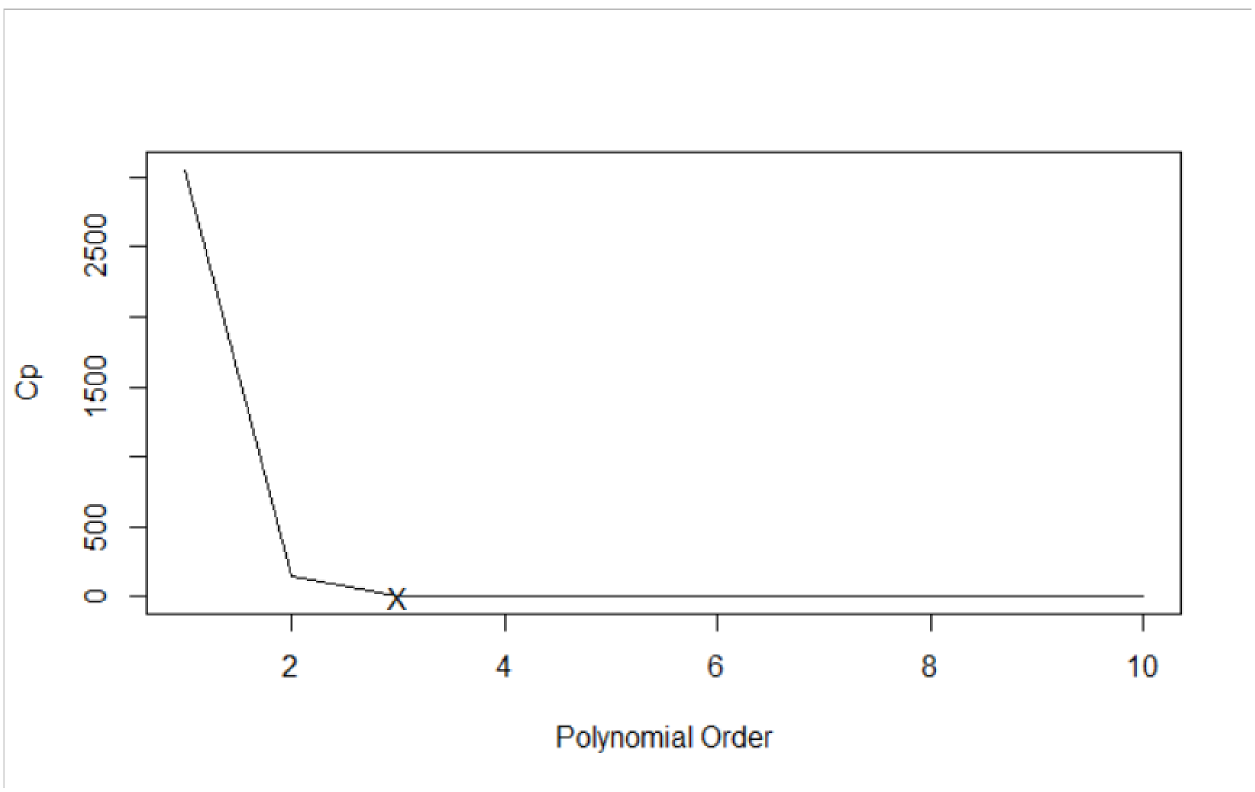

```
plot(1:8,summary.y.subsets$adjr2,type="l",xlab="Polynomial
Order", ylab="Adjusted R^2")
text(4,summary.y.subsets$adjr2[4],"X")
```



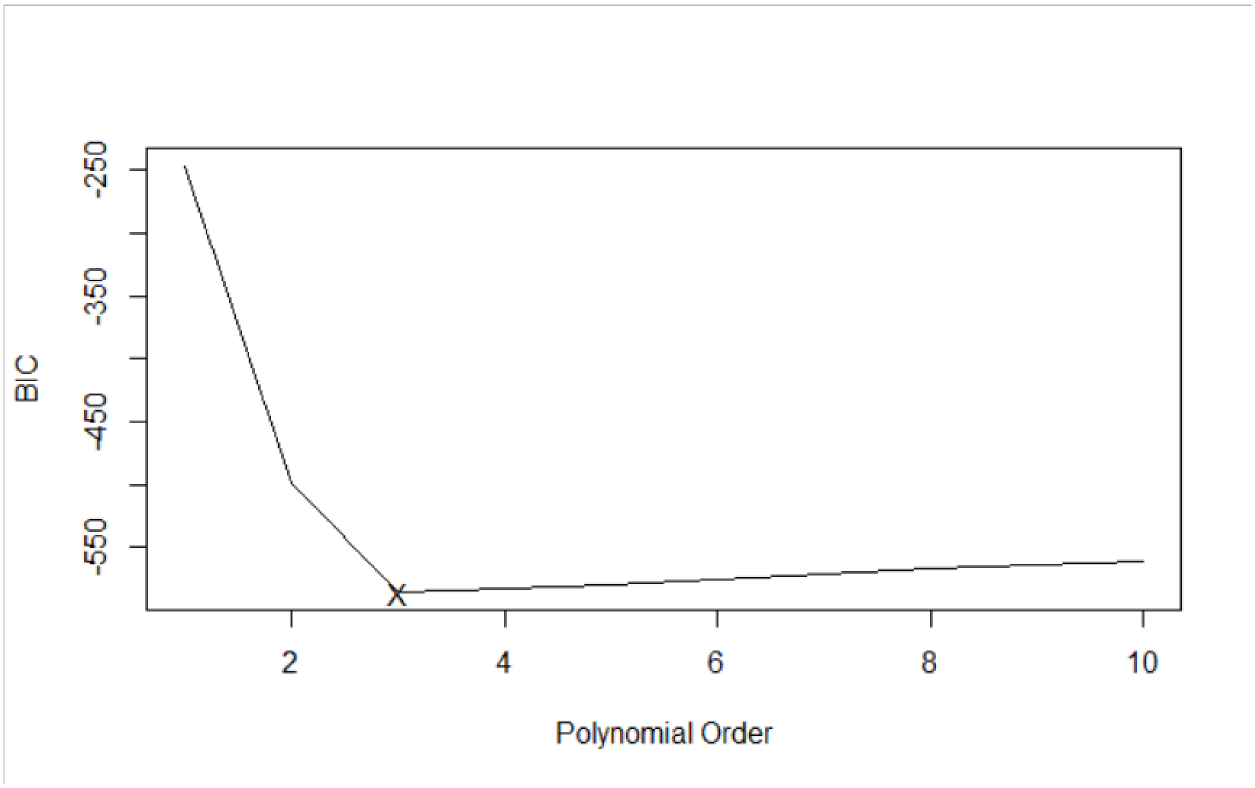

```
# Cp and BIC arrive at the correct model but the adjusted R2
suggests, incorrectly, a quartic model is best.

(d)
y.subset.fwd<- regsubsets(Y~.,data =sim.data ,nvmax = 10, method
= "forward")
summary.fwd<-summary(y.subset.fwd)
summary.fwd

1 subsets of each size up to 10
Selection Algorithm: forward
          X   X2  X3  X4  X5  X6  X7  X8  X9  X10
1  ( 1 )  " " " " "*" " " " " " " " " " " " " " "
2  ( 1 )  " " "*" "*" " " " " " " " " " " " " " "
3  ( 1 )  "*" "*" "*" " " " " " " " " " " " " " "
4  ( 1 )  "*" "*" "*" "*" " " " " " " " " " " " "
5  ( 1 )  "*" "*" "*" "*" "*" " " " " " " " " " "
6  ( 1 )  "*" "*" "*" "*" "*" " " " " " " "*" " "
7  ( 1 )  "*" "*" "*" "*" "*" " " "*" " " "*" " "
8  ( 1 )  "*" "*" "*" "*" "*" "*" "*" " " "*" " "
9  ( 1 )  "*" "*" "*" "*" "*" "*" "*" "*" "*" " "
10  ( 1 ) "*" "*" "*" "*" "*" "*" "*" "*" "*" "*"
```

```
y.subset.bwd<- regsubsets(Y~.,data =sim.data ,nvmax = 10, method
= "backward")
summary.bwd<-summary(y.subset.bwd)
summary.bwd
1 subsets of each size up to 10
Selection Algorithm: backward
          X   X2  X3  X4  X5  X6  X7  X8  X9  X10
1  ( 1 )  " " " " "*" " " " " " " " " " " " " " "
2  ( 1 )  " " "*" "*" " " " " " " " " " " " " " "
3  ( 1 )  "*" "*" "*" " " " " " " " " " " " " " "
4  ( 1 )  "*" "*" "*" " " " " "*" " " " " " " " "
5  ( 1 )  "*" "*" "*" " " " " "*" " " "*" " " " "
6  ( 1 )  "*" "*" "*" "*" " " "*" " " "*" " " " "
7  ( 1 )  "*" "*" "*" "*" " " "*" " " "*" " " "*"
8  ( 1 )  "*" "*" "*" "*" " " "*" "*" "*" " " "*"
9  ( 1 )  "*" "*" "*" "*" " " "*" "*" "*" "*" "*"
10  ( 1 ) "*" "*" "*" "*" "*" "*" "*" "*" "*" "*"
```

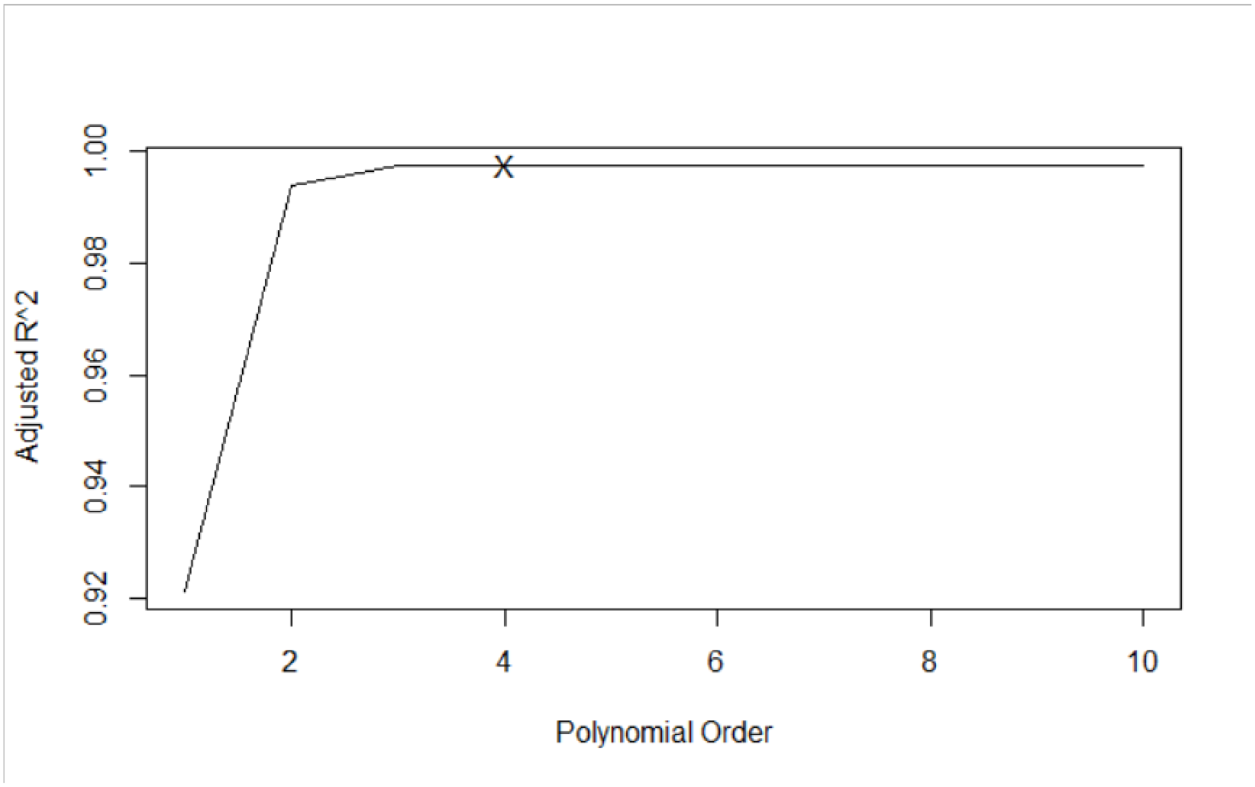
```
# Forward Plots
plot(summary.fwd$cp,type="l",xlab="Polynomial Order", ylab="Cp")
text(which.min(summary.fwd$cp),summary.fwd$cp[which.min(summary.
fwd$cp)],"X")
```



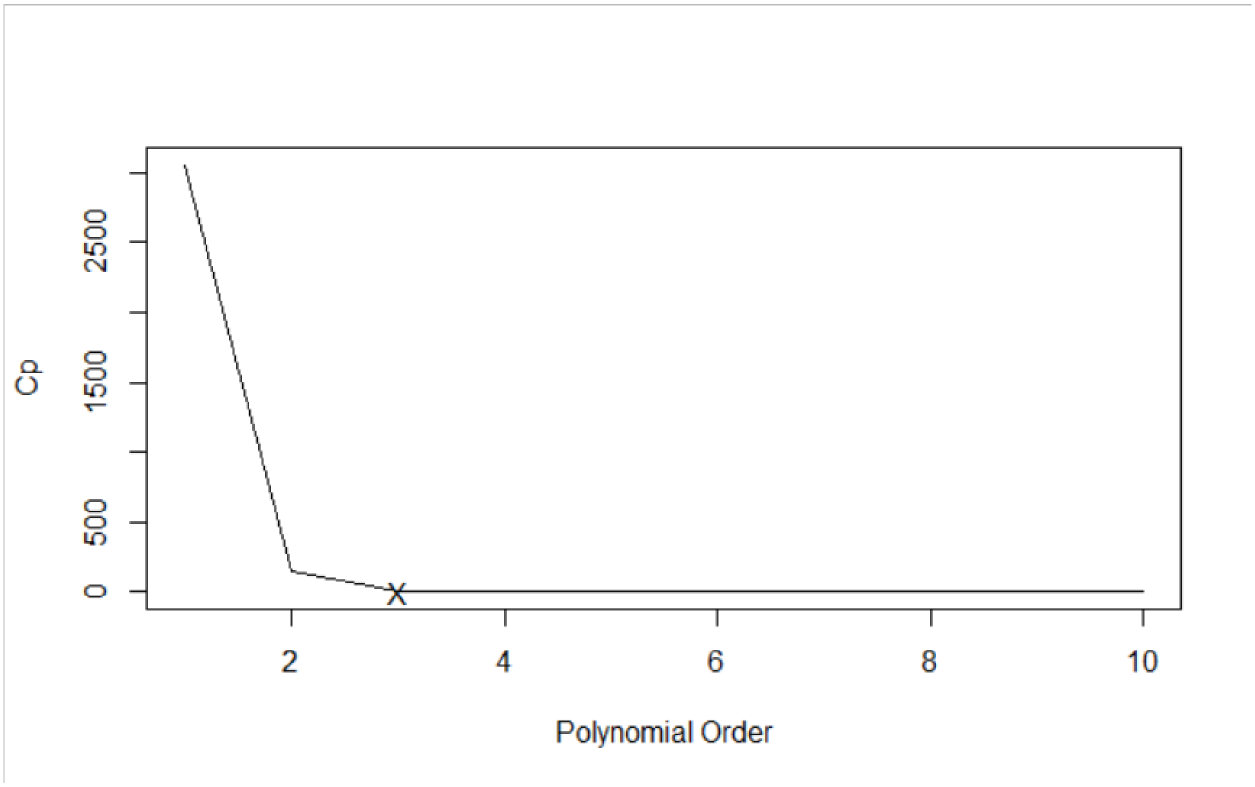
```
plot(summary.fwd$bic,type="l",xlab="Polynomial Order",
ylab="BIC")
text(which.min(summary.fwd$bic),summary.fwd$bic[which.min(summar
y.fwd$bic)],"X")
```



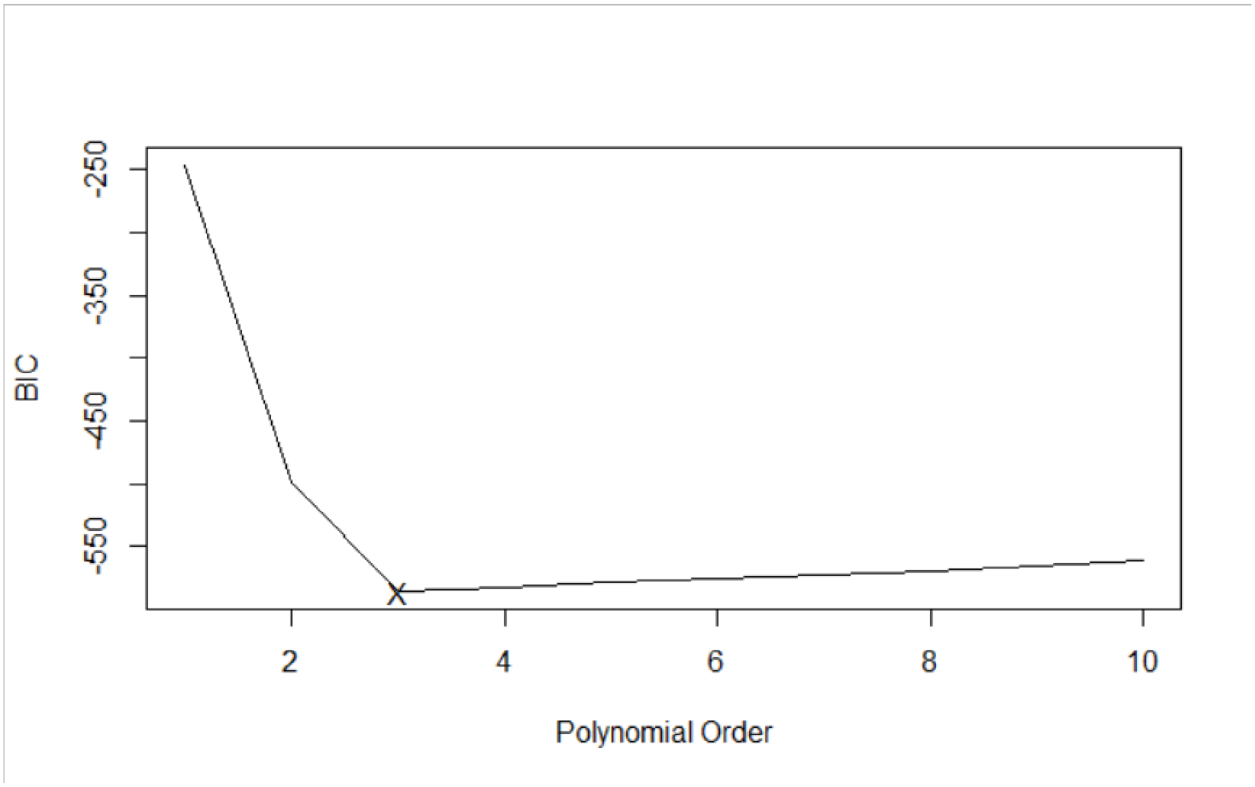

```
plot(summary.fwd$adjr2,type="l",xlab="Polynomial Order",
ylab="Adjusted R^2")
text(which.max(summary.fwd$adjr2),summary.fwd$adjr2[which.max(su
mmary.fwd$adjr2)],"X")
```



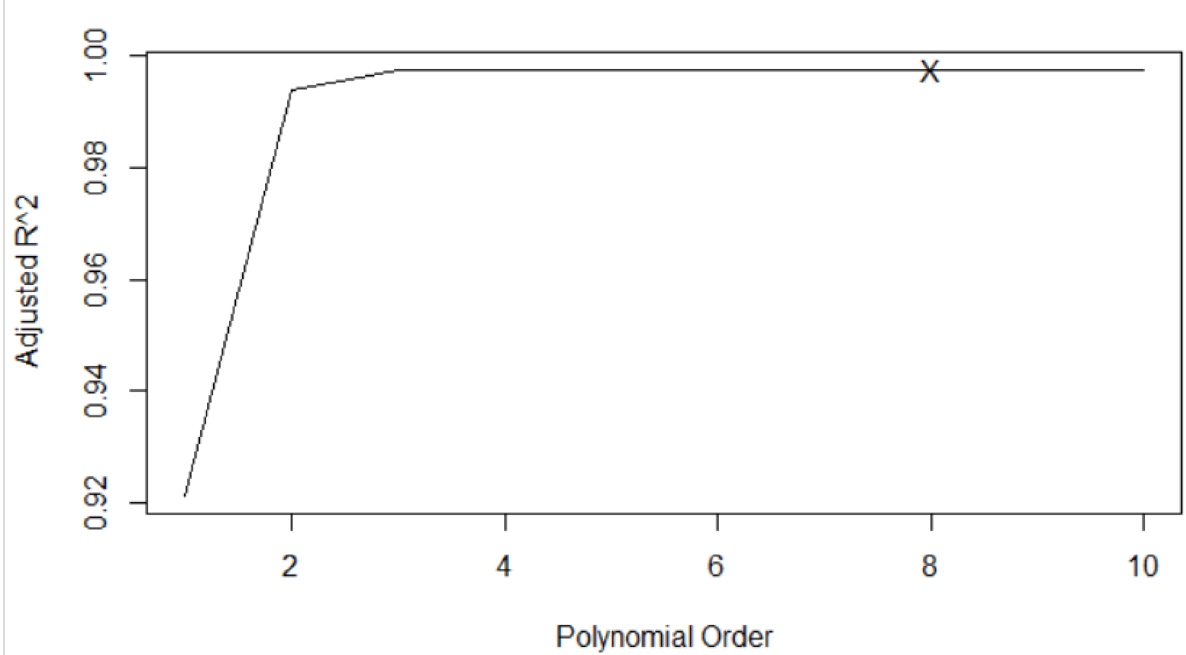

```
# Backward Plots
plot(summary.bwd$cp,type="l",xlab="Polynomial Order", ylab="Cp")
text(which.min(summary.bwd$cp),summary.bwd$cp[which.min(summary.
bwd$cp)],"X")
```



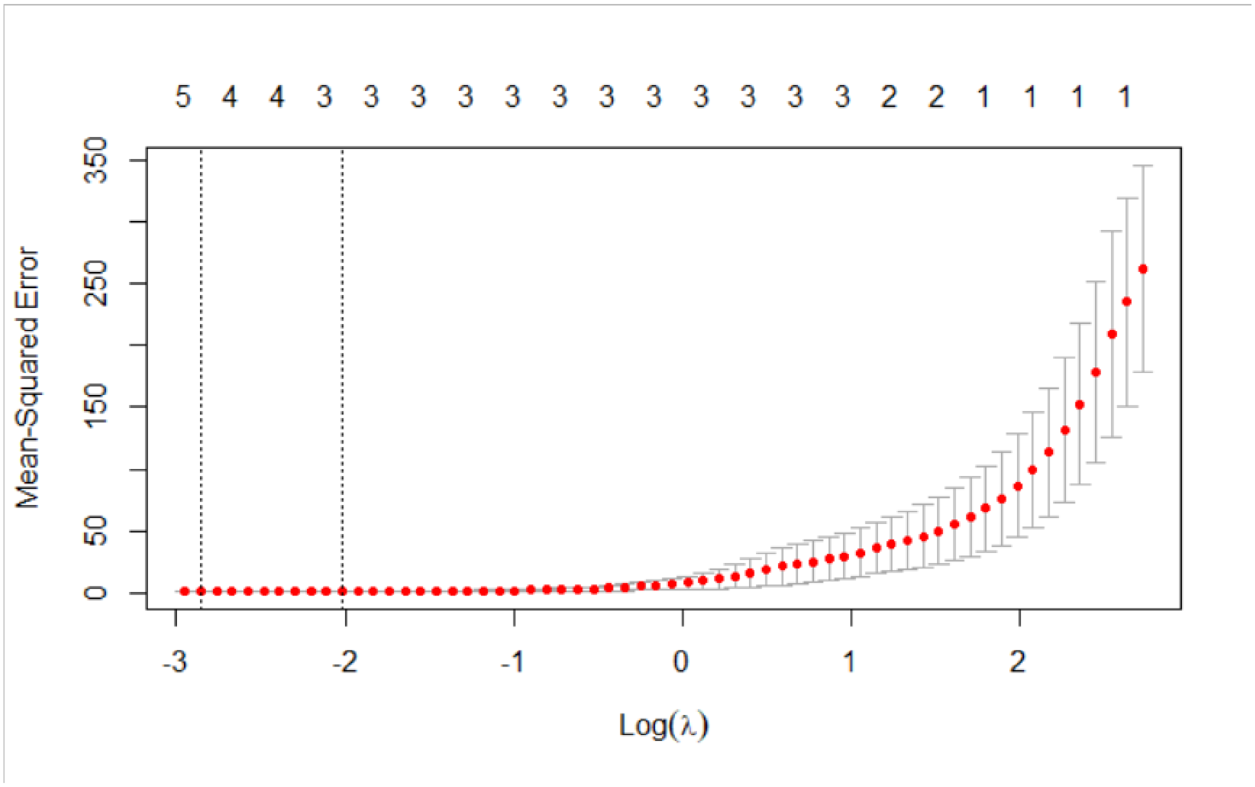

```
plot(summary.bwd$bic,type="l",xlab="Polynomial Order",
ylab="BIC")
text(which.min(summary.bwd$bic),summary.bwd$bic[which.min(summar
y.bwd$bic)],"X")
```




```
plot(summary.bwd$adjr2,type="l",xlab="Polynomial Order",
ylab="Adjusted R^2")
text(which.max(summary.bwd$adjr2),summary.bwd$adjr2[which.max(su
mmary.bwd$adjr2)],"X")
```




```
# As before Cp and BIC identify the correct model with forward
and backward selection but R² does not and is especially wrong
with backward selection.

#(e)
library(glmnet)
set.seed(1)
# sim.data2<- as.matrix(sim.data)
x<- model.matrix(Y~.,data=sim.data)[,-1]
y<- sim.data$Y
grid <- 10^seq(10, -2, length = 100)
sim.lasso<- glmnet(x, y, alpha = 1,lambda = grid)
sim.cv<- cv.glmnet(x, y, alpha = 1,nfolds=10)
plot(sim.cv)
```




```
bestlam <- sim.cv$lambda.min
lasso.coef <- predict(sim.lasso , type = "coefficients",s =
bestlam)[1:11, ]
lasso.coef
(Intercept)             X            X2            X3
X4            X5
 8.941714e-01  1.824323e+00 -2.889171e+00  3.975499e+00
0.000000e+00  5.401150e-03
           X6            X7            X8            X9
 0.000000e+00  9.954679e-07  0.000000e+00  0.000000e+00
> lasso.coef <- predict(sim.lasso , type = "coefficients",s =
bestlam)[1:11, ]
> lasso.coef
  (Intercept)             X            X2            X3
X4            X5
 8.941714e-01  1.824323e+00 -2.889171e+00  3.975499e+00
0.000000e+00  5.401150e-03
           X6            X7            X8            X9
X10
 0.000000e+00  9.954679e-07  0.000000e+00  0.000000e+00
0.000000e+00

# Next look at the + 1 sd deviation λ.
```

```
lasso.coef2 <- predict(sim.lasso , type = "coefficients",s =
sim.cv$lambda[sim.cv$index[2]] )[1:11, ]
lasso.coef2
(Intercept)           X          X2          X3          X4
X5          X6
  0.8255299   1.7795840  -2.8166071   3.9841480   0.0000000
0.0000000   0.0000000
         X7          X8          X9         X10
  0.0000000   0.0000000   0.0000000   0.0000000
# This more conservative result has now reduced the sparse set
to the three relevant features.
```