**Assignment #5: Advanced Regression**

**Submit through link: eCampus -> Assignments->Assignment 5 Submission**

**Deadline: November 8 (Thursday) @19:00 pm**

**The filename should have this format: LastName-FirstName-hw05.doc**

**For problems 2 and 3 you do not need to use R.**

**Problem 1 (8pt)**

In this question, we will predict the number of applications received (Apps) using the other variables in the College data set (ISLR package).

(a) Perform best subset selection to the data. 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.

Answer: best\_subset<- regsubsets(Apps~.,data = College,nvmax = 19)

summary\_best<- summary(best\_subset)

# adjusted R^2 graph

plot(summary\_best$adjr2,type = 'o', xlab= "No of Predictors", ylab= "Adjusted R square")

which.max(summary\_best$adjr2)

points(13, summary\_best$adjr2[13], col="red",cex=2,pch=20 )

# Cp graph

plot(summary\_best$cp, type = 'o', xlab= "No of Predictors", ylab= "Cp")

which.min(summary\_best$cp)

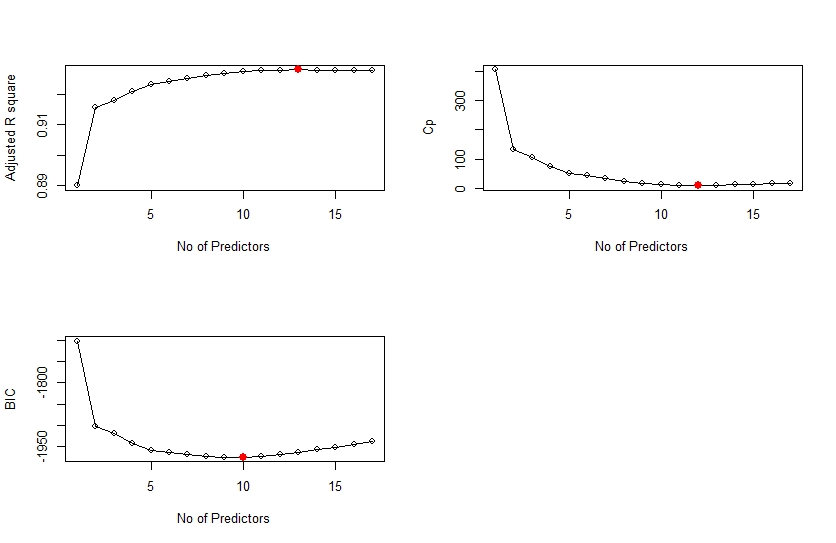
points(12, summary\_best$cp[12], col="red",cex=2,pch=20 )

# BIC graph

plot(summary\_best$bic, type = 'o', xlab= "No of Predictors", ylab= "BIC")

which.min(summary\_best$bic)

points(10, summary\_best$bic[10], col="red",cex=2,pch=20 )



**#Coefficients of Best model using BIC**

coef(best\_subset,id = 10)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | PrivateYes | Accept | Enroll | Top10perc | Top25perc | Outstate | Room. Board | PhD | Expend | Grad.Rate |
| -100.51 | -575.07 | 1.58 | -0.5 | 49.1 | -13.86 | -0.09 | 0.16 | -10.02 | 0.07273776 | 7.33268904 |

**# Coefficients of best model using adjusted R square**

coef(best\_subset,id = 13)

# **Coefficients of best model using Cp**

coef(best\_subset,id = 12)

.

(b) Repeat (a) using forward stepwise selection and backwards stepwise selection. How does your answer compare to the results in (a)?

Answer

# Forward stepwise

fwd\_subset<- regsubsets(Apps~.,data = College,nvmax = 19,method = "forward")

summary\_fwd<- summary(fwd\_subset)

par(mfrow= c(2,2))

# adjusted R^2 graph

plot(summary\_fwd$adjr2,type = 'o', xlab= "No of Predictors", ylab= "Adjusted R square")

which.max(summary\_fwd$adjr2)

points(13, summary\_fwd$adjr2[13], col="red",cex=2,pch=20 )

# Cp graph

plot(summary\_fwd$cp, type = 'o', xlab= "No of Predictors", ylab= "Cp")

which.min(summary\_fwd$cp)

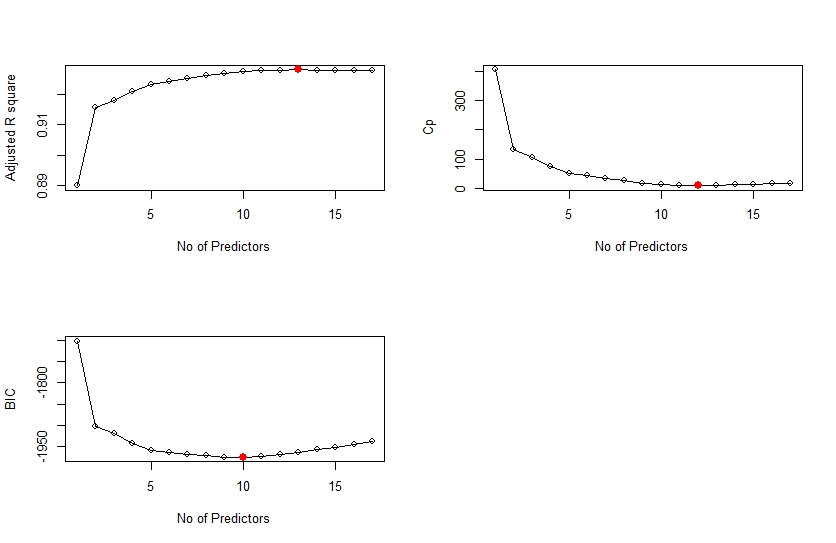
points(12, summary\_fwd$cp[12], col="red",cex=2,pch=20 )

# BIC graph

plot(summary\_fwd$bic, type = 'o', xlab= "No of Predictors", ylab= "BIC")

which.min(summary\_fwd$bic)

points(10, summary\_fwd$bic[10], col="red",cex=2,pch=20 )



**#Coefficients of Best forward stepwise model using BIC**

coef(fwd\_subset,id = 10)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | PrivateYes | Accept | Enroll | Top10perc | Top25perc | Outstate | Room. Board | PhD | Expend | Grad.Rate |
| -100.51 | -575.07 | 1.58 | -0.5 | 49.1 | -13.86 | -0.09 | 0.16 | -10.02 | 0.07273776 | 7.33268904 |

**# Coefficients of best forward stepwise model using adjusted R square**

coef(fwd\_subset,id = 13)

# **Coefficients of best forward stepwise model using Cp**

coef(fwd\_subset,id = 12)

#Backward Stepwise

bck\_subset<- regsubsets(Apps~.,data = College,nvmax = 19,method = "backward")

summary\_bck<- summary(fwd\_subset)

par(mfrow= c(2,2))

# adjusted R^2 graph

plot(summary\_bck$adjr2,type = 'o', xlab= "No of Predictors", ylab= "Adjusted R square")

which.max(summary\_bck$adjr2)

points(13, summary\_bck$adjr2[13], col="red",cex=2,pch=20 )

# Cp graph

plot(summary\_bck$cp, type = 'o', xlab= "No of Predictors", ylab= "Cp")

which.min(summary\_bck$cp)

points(12, summary\_bck$cp[12], col="red",cex=2,pch=20 )

# BIC graph

plot(summary\_bck$bic, type = 'o', xlab= "No of Predictors", ylab= "BIC")

which.min(summary\_bck$bic)

points(10, summary\_bck$bic[10], col="red",cex=2,pch=20 )

**# Coefficients of best backward stepwise model using BIC.**

coef(bck\_subset,id = 10)

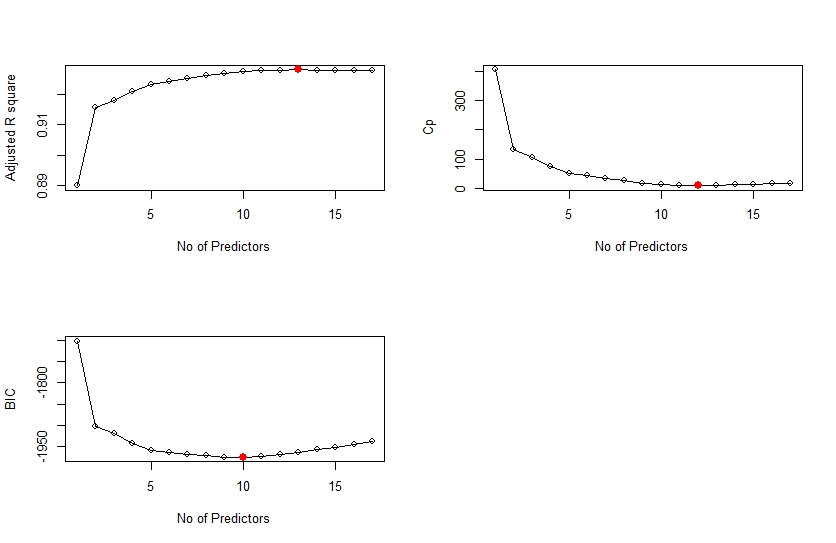
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | PrivateYes | Accept | Enroll | Top10perc | Top25perc | Outstate | Room. Board | PhD | Expend | Grad.Rate |
| -100.51 | -575.07 | 1.58 | -0.5 | 49.1 | -13.86 | -0.09 | 0.16 | -10.02 | 0.07273776 | 7.33268904 |

**# Coefficients of best backward stepwise model using adjusted R square**

coef(bck\_subset,id = 13)

# **Coefficients of best backward stepwise model using Cp**

coef(bck\_subset,id = 12)



All the three subset selection techniques are giving the same subset as the best model for adjusted r square, Cp and BIC respectively.

(c) Fit a lasso model on the data. 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.

Answer:

# fitting a lasso model

library(glmnet)

x<- model.matrix(Apps~., College)[,-1]

y<- College$Apps

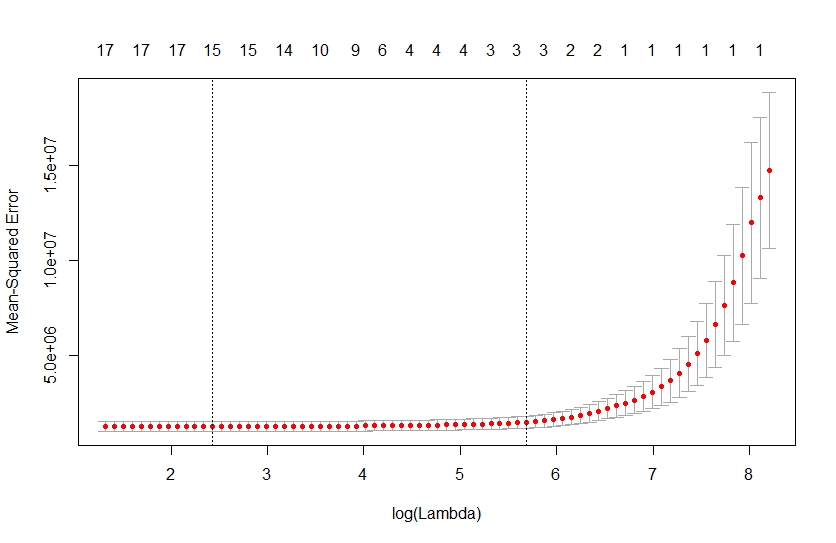
grid<- 10^seq(10,-2,length=100)

reg\_lasso<- glmnet(x, y, alpha=1, lambda= grid )

set.seed(100)

cv.lasso<- cv.glmnet(x,y, alpha=1)

plot(cv.lasso)



lambda\_best<-cv.lasso$lambda.min

lambda\_best

3.40

predict(reg\_lasso, s= lambda\_best, type = "coefficients")

|  |
| --- |
| (Intercept) -480.13227170  PrivateYes -489.67603364  Accept 1.56289346  Enroll -0.70168936  Top10perc 47.19755475  Top25perc -12.12761643  F.Undergrad 0.03409185  P.Undergrad 0.04380077  Outstate -0.08179465  Room.Board 0.14814139  Books 0.01187453  Personal 0.02773581  PhD -8.24548722  Terminal -3.21178041  S.F.Ratio 13.98403377  perc.alumni -0.12644001  Expend 0.07661310  Grad.Rate 8.06493716 |
|  |
| |  | | --- | | > | |
|  |
| |  | | --- | |  | |

(d) Fit a ridge regression model on the data. 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.

Answer:

# fitting a ridge model

library(glmnet)

x<- model.matrix(Apps~., College)[,-1]

y<- College$Apps

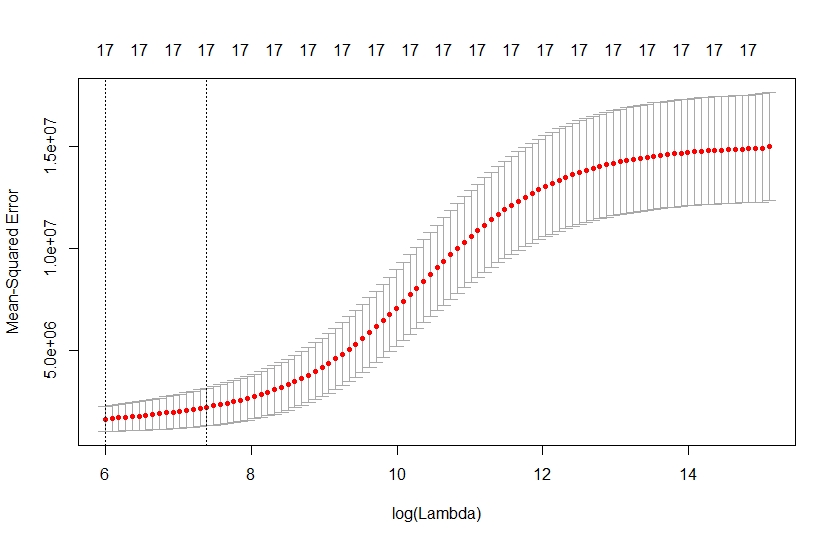
grid<- 10^seq(10,-2,length=100)

reg\_ridge<- glmnet(x, y, alpha=0, lambda= grid )

set.seed(1)

cv.ridge<- cv.glmnet(x,y, alpha=0)

plot(cv.ridge)



lambda\_best<-cv.ridge$lambda.min

lambda\_best

400.4766

predict(reg\_ridge, s= lambda\_best, type = "coefficients")

(Intercept) -1.512577e+03

PrivateYes -5.297867e+02

Accept 9.789533e-01

Enroll 4.663570e-01

Top10perc 2.500038e+01

Top25perc 1.048650e+00

F.Undergrad 7.633057e-02

P.Undergrad 2.446417e-02

Outstate -2.134950e-02

Room.Board 1.997989e-01

Books 1.351305e-01

Personal -8.933673e-03

PhD -3.779999e+00

Terminal -4.713986e+00

S.F.Ratio 1.278148e+01

perc.alumni -8.809281e+00

Expend 7.520720e-02

Grad.Rate 1.134540e+01

(e) Now split the data set into a training set and a test set.

set.seed((1))

train<- sample(nrow(x),size = nrow(x)/2)

test<- -train

i. Fit the best models obtained in the best subset selection (according to C*p*, BIC or adjusted *R*2) to the training set, and report the test error obtained.

Answer: using BIC given model subset for calculating the MSE.

# fitting linear model

lm.best<- lm(Apps~Private+Accept+Enroll+Top10perc+Top25perc+Outstate+Room.Board+PhD+Expend+Grad.Rate,data = College,subset = train)

y\_best<-predict(lm.best, newdata = College[test,])

y\_test<-College$Apps[test]

mean((y\_best - y\_test)^2)

= 1078371

ii. Fit a lasso model to the training set, with *λ* chosen by cross validation. Report the test error obtained.

# predicting through lasso

y\_lasso<-predict(reg\_lasso,newx = x[test,],s = 3.40 )

mean((y\_lasso - y\_test)^2)

= 1062972

iii. Fit a ridge regression model to the training set, with *λ* chosen by cross validation. Report the test error obtained.

# predicting through ridge and then finding MSE

y\_ridge<-predict(reg\_ridge,newx = x[test,],s = 400.4766 )

mean((y\_ridge - y\_test)^2)

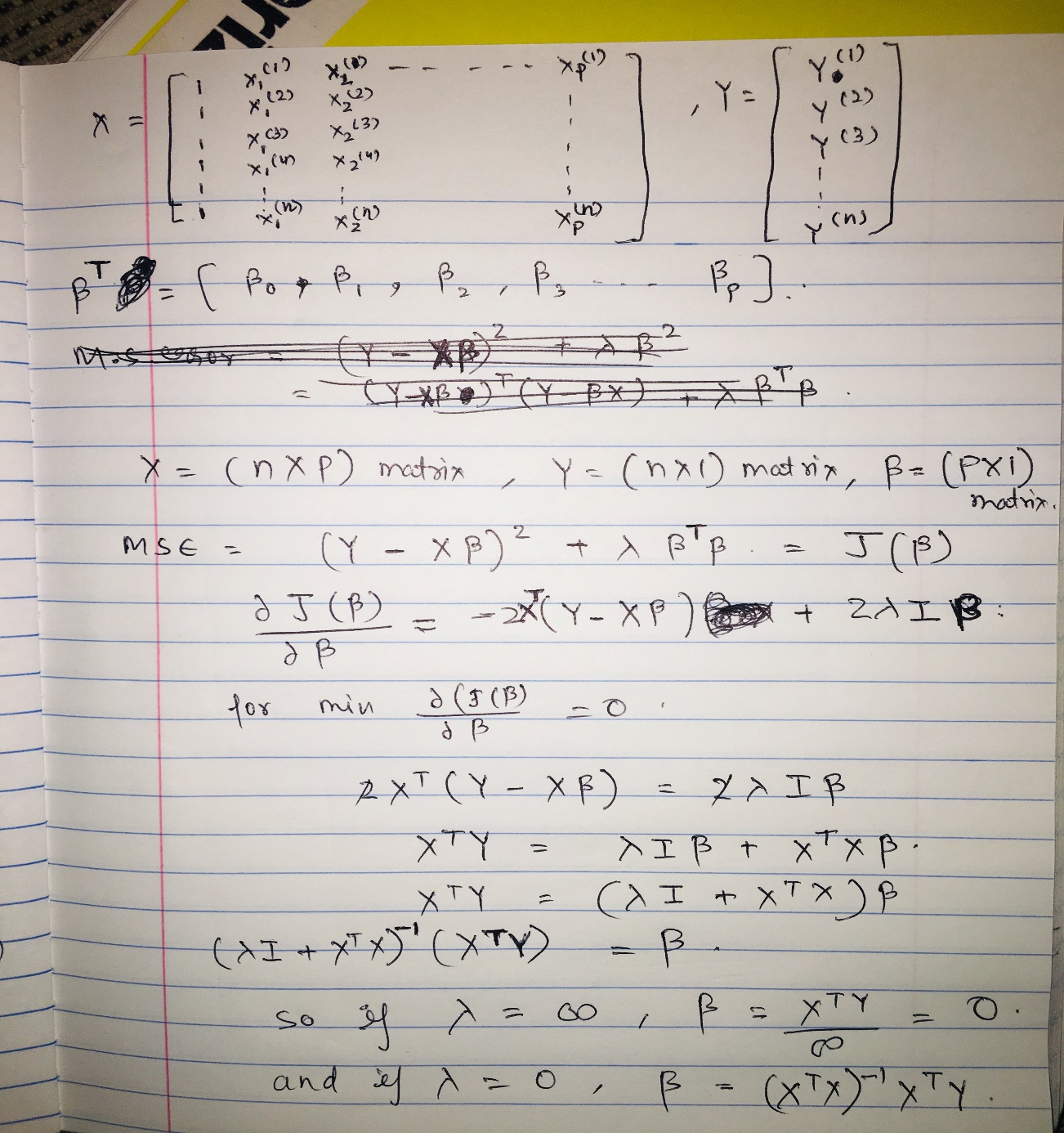
=1010293

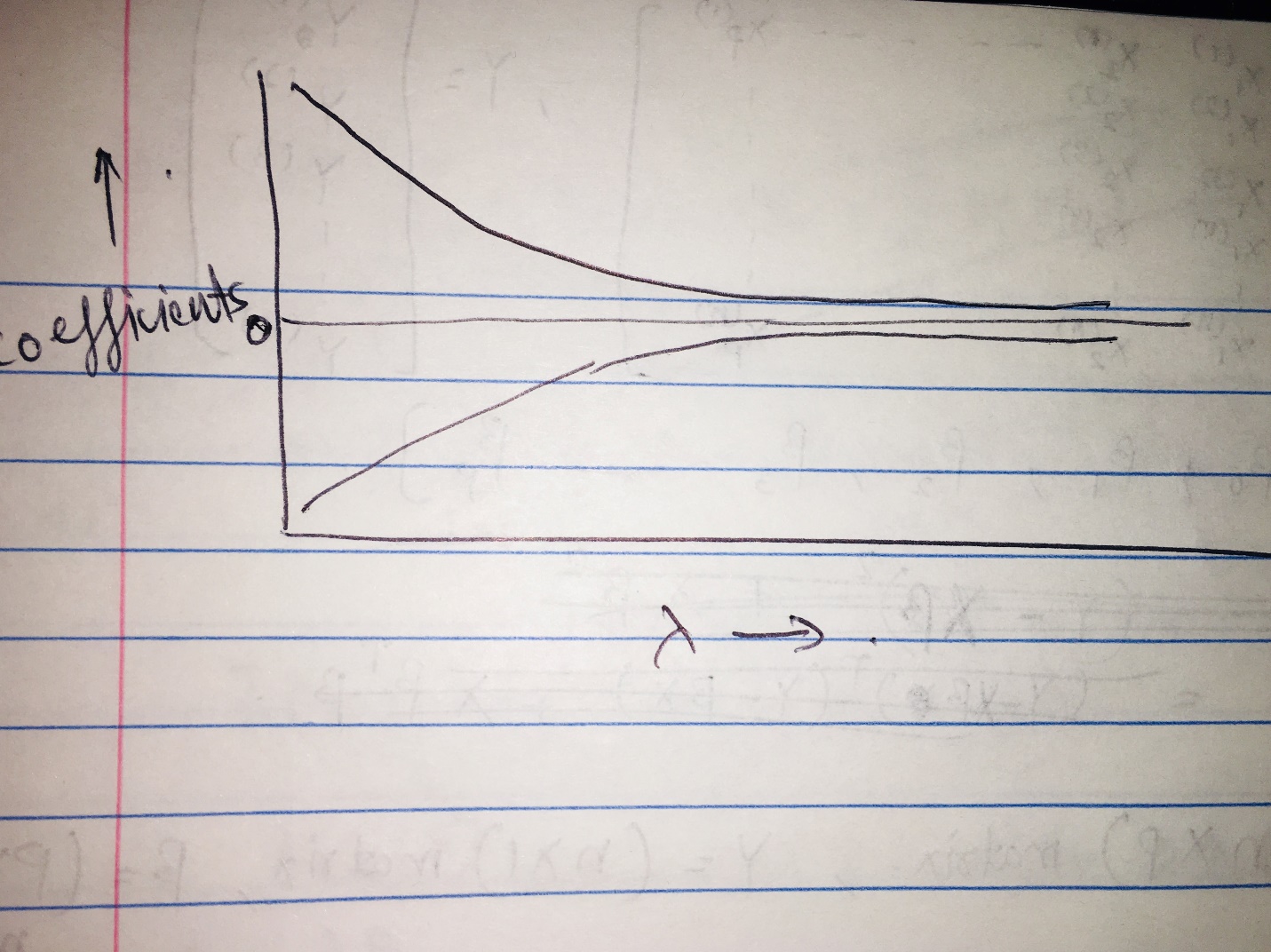
iv. Compare the test errors obtained in the above analysis (i-iii) and determine the optimal model.

Answer : Ridge > Lasso> subset selection models

But again taking different random seeds give a different order with every different seed.

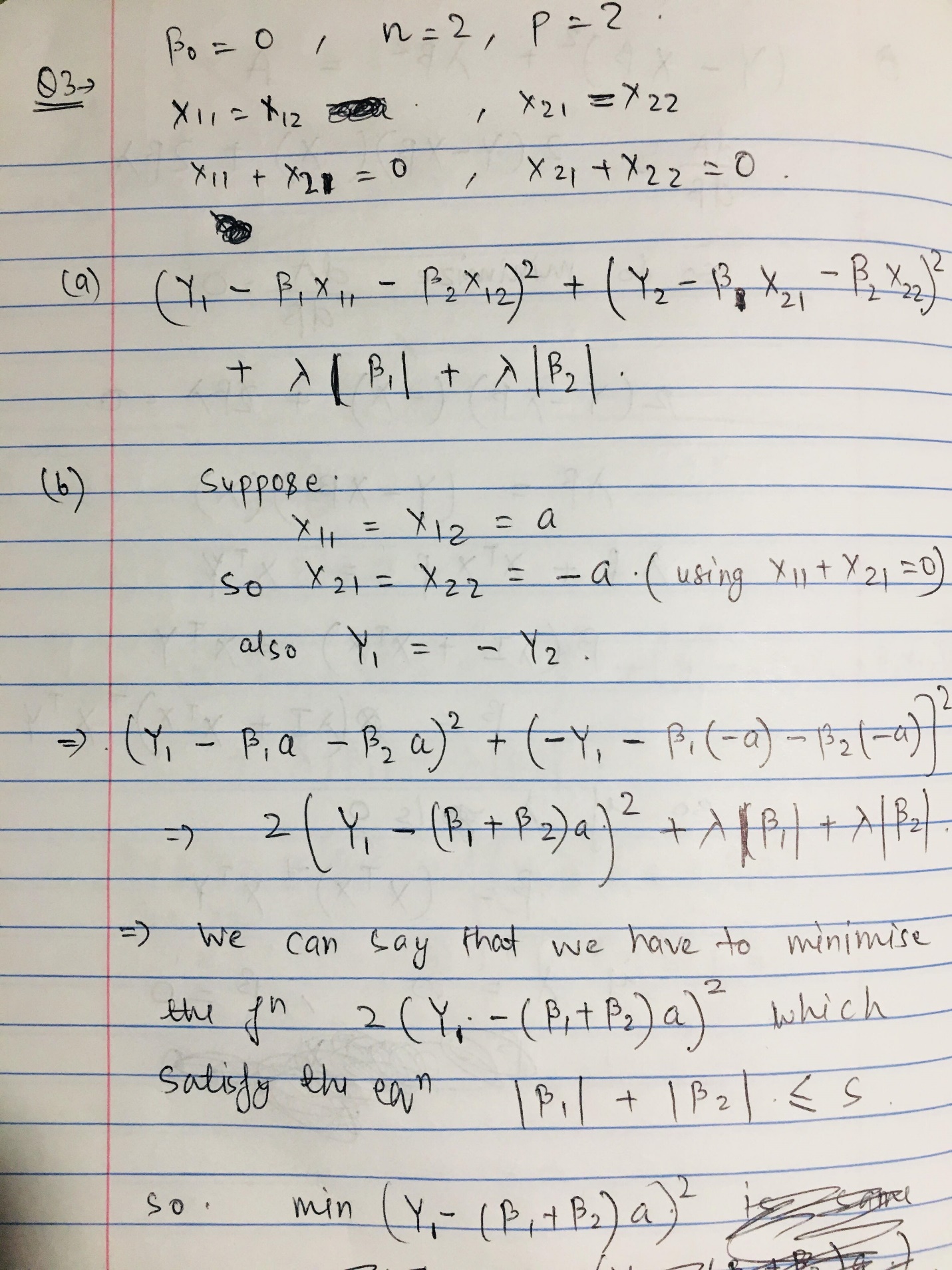
**Problem 2 (7pt)** In the class, we discussed the ridge regression model as one of the shrinkage methods. In this problem, we study the effect of tuning parameter on the model by mathematically calculating the coefficients. To do so, find the optimal value of the objective function given in equation (6.5) in the book (hint: consider as a fixed parameter and differentiate 6.5 with respect to each coefficient. Set the derivative equal to zero to find a closed-form expression for all the coefficients. Then, describe the behavior of the coefficients in terms of . More specifically, discuss the coefficient change when varying from 0 to .)

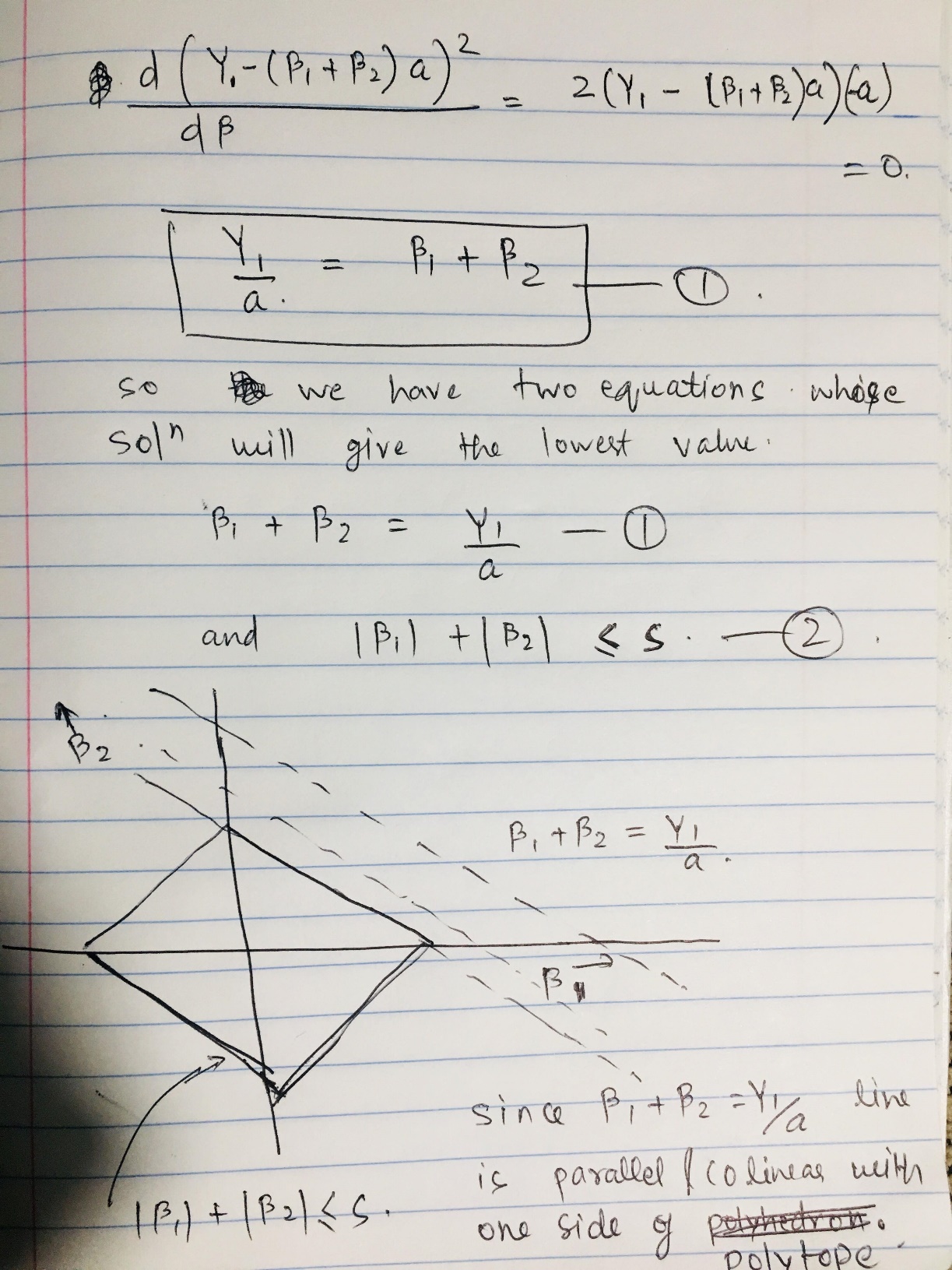
****

****

**We can show through the derived function for parameter coefficients that as the value of lambda increases the Beta coefficients get shrunk to a smaller value which has been shown in the graph above.**

**Problem 3 (Bonus 5pt)** Solve question 5 (c)-(d) in Ch. 6.8 Exercises (Hint: In describing the solutions in 6 (d), use equation 6.15).





Since the two functions are intersecting at a line( not just a single point) if we have enough budget as S. so there will be infinite solutions not just a single one.