**Assignment #5: Advanced Regression**

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

**Deadline: November 8 (Friday) @5:00 pm**

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

**Problem 1 (10pt)**

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.

> library(ISLR)

> library(leaps)

> data(College)

> set.seed(1)

> names(College)



> regfit <- regsubsets( Apps ~ .,data = College,nvmax=18)

> summary1 <-summary(regfit)

> par(mfrow=c(1,3))

> plot(summary1$cp,xlab="number of variables",ylab="C\_p")

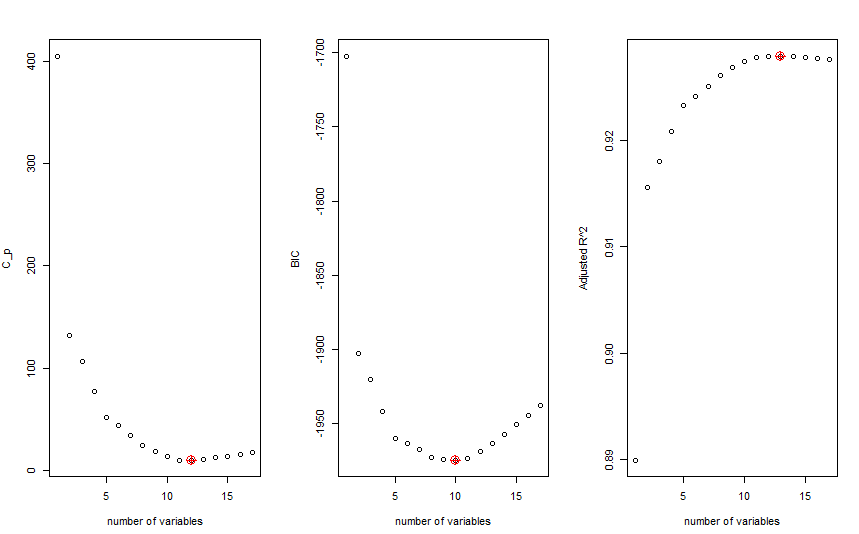
> points(which.min(summary1$cp),summary1$cp[which.min(summary1$cp)],col="red",cex=2,pch=10)

> plot(summary1$bic,xlab="number of variables",ylab="BIC")

> points(which.min(summary1$bic),summary1$bic[which.min(summary1$bic)],col="red",cex=2,pch=10)

> plot(summary1$adjr2,xlab="number of variables",ylab="Adjusted R^2")

> points(which.max(summary1$adjr2),summary1$adjr2[which.max(summary1$adjr2)],col="red",cex=2,pch=10)



*Cp chooses model with 12 predictors, BIC chooses model with 10 predictors, Adjusted Rˆ2 chooses model with 13 predictors. In general, the best model is the one with less number of predictors, in this case it is with 10 predictors which is chosen by BIC.*

**Coefficients of best chosen model:**

> coef(regfit,which.min(summary1$bic))



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

**Forward Selection:**

> forwardsel <-regsubsets( Apps ~ .,data = College,nvmax=17,method="forward")

> summary2 <- summary(forwardsel)

> par(mfrow=c(1,3))

> plot(summary2$cp,xlab="number of variables",ylab="C\_p")

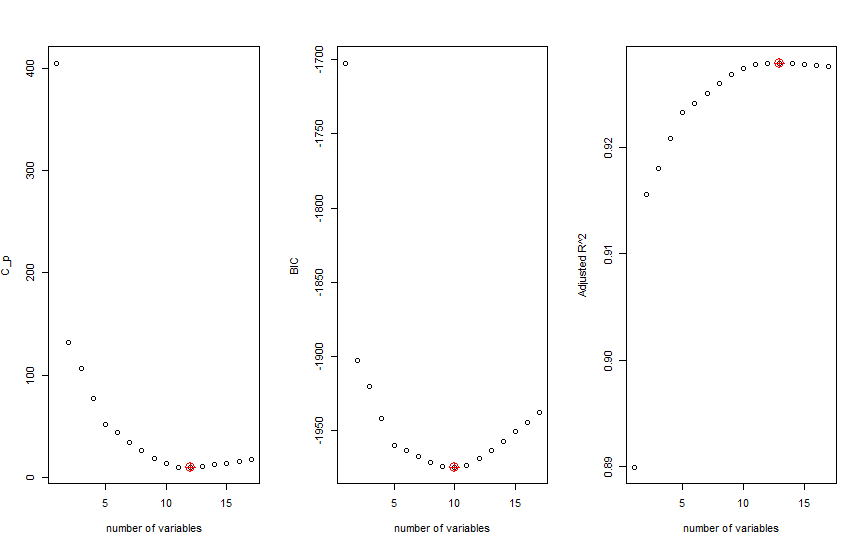
> points(which.min(summary2$cp),summary2$cp[which.min(summary2$cp)],col="red",cex=2,pch=10)

> plot(summary2$bic,xlab="number of variables",ylab="BIC")

> points(which.min(summary2$bic),summary2$bic[which.min(summary2$bic)],col="red",cex=2,pch=10)

> plot(summary2$adjr2,xlab="number of variables",ylab="Adjusted R^2")

> points(which.max(summary2$adjr2),summary2$adjr2[which.max(summary2$adjr2)],col="red",cex=2,pch=10)



*Here also in forward selection method, Cp chooses model with 12 predictors, BIC chooses model with 10 predictors, Adjusted Rˆ2 chooses model with 13 predictors. In general, the best model is the one with less number of predictors, in this case it is with 10 predictors which is chosen by BIC.*

**Coefficients of best chosen model:**

> coef(forwardsel,which.min(summary2$bic))



**Backward Selection:**

> backwardsel <-regsubsets(Apps ~ .,data= College,nvmax=17,method="backward")

> summary3 <- summary(backwardsel)

> par(mfrow=c(1,3))

> plot(summary3$cp,xlab="number of variables",ylab="C\_p")

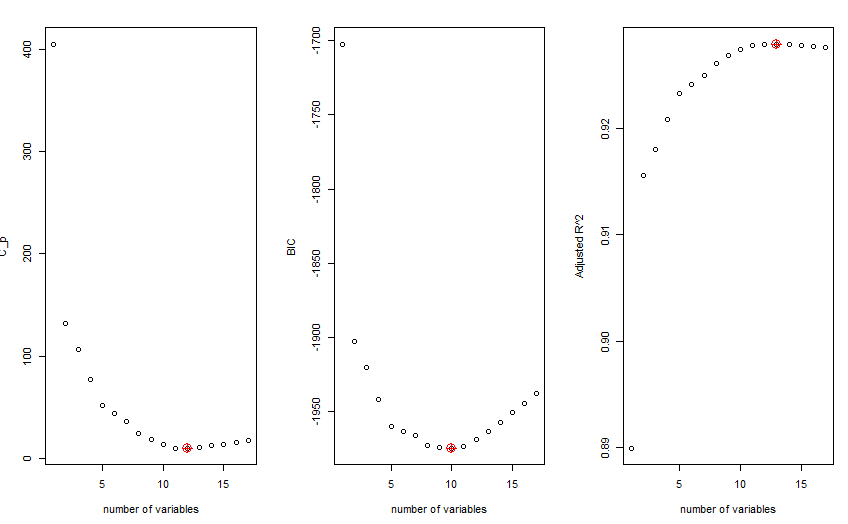
> points(which.min(summary3$cp),summary3$cp[which.min(summary3$cp)],col="red",cex=2,pch=10)

> plot(summary3$bic,xlab="number of variables",ylab="BIC")

> points(which.min(summary3$bic),summary3$bic[which.min(summary3$bic)],col="red",cex=2,pch=10)

> plot(summary3$adjr2,xlab="number of variables",ylab="Adjusted R^2")

> points(which.max(summary3$adjr2),summary3$adjr2[which.max(summary3$adjr2)],col="red",cex=2,pch=10)



*Here also in backward selection method,Cp chooses model with 12 predictors, BIc chooses model with 10 predictors, Adjusted Rˆ2 chooses model with 13 predictors. In general, the best model is the one with less number of predictors, in this case it is with 10 predictors which is chosen by BIC.*

**Coefficients of best chosen model:**

> coef(backwardsel,which.min(summary3$bic))



(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.

> library(glmnet)

Loading required package: Matrix

Loading required package: foreach

Loaded glmnet 2.0-18

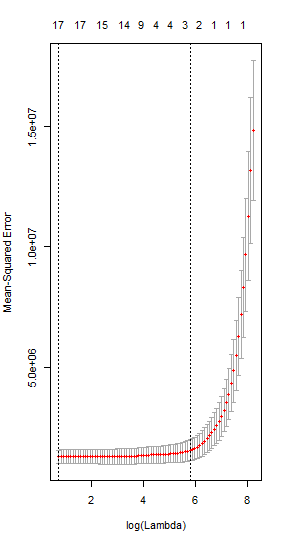
> data(College)

> set.seed(1)

> matrix1<- model.matrix(College$Apps ~ .,data=College)

> cross <- cv.glmnet(matrix1,College$Apps,alpha=1)

> plot(cross)



> optimallambda <- cross$lambda.min

> optimallambda

[1] 2.137223

**The Optimum value of Lambda is 2.137**

> lasso1 <- glmnet(matrix1,College$Apps,alpha=1)

> predict(lasso1,s=optimallambda,type="coefficients")

19 x 1 sparse Matrix of class "dgCMatrix"

1

(Intercept) -471.39372069

(Intercept) .

PrivateYes -491.04485135

Accept 1.57033288

Enroll -0.75961467

Top10perc 48.14698891

Top25perc -12.84690694

F.Undergrad 0.04149116

P.Undergrad 0.04438973

Outstate -0.08328388

Room.Board 0.14943472

Books 0.01532293

Personal 0.02909954

PhD -8.39597537

Terminal -3.26800340

S.F.Ratio 14.59298267

perc.alumni -0.04404771

Expend 0.07712632

Grad.Rate 8.28950241

(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.

> data(College)

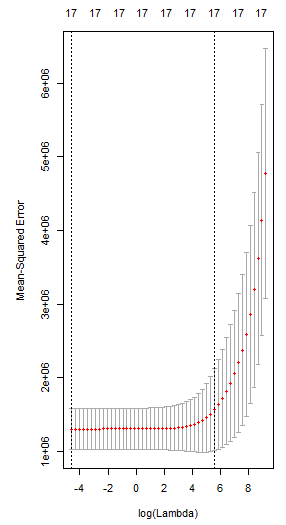
> set.seed(1)

> matrix2<- model.matrix(College$Apps ~ .,data=College)

> grid <- 10^seq(4,-2,length=50)

> cross1 <- cv.glmnet(matrix2,College$Apps,alpha=0,lambda=grid)

> plot(cross1)



> optimallambda2 <- cross1$lambda.min

> optimallambda2

[1] 0.01

**The optimum value of lambda is 0.01**

> ridgeco <- glmnet(matrix1,College$Apps,alpha=0)

> predict(ridgeco,s=optimallambda,type="coefficients")

19 x 1 sparse Matrix of class "dgCMatrix"

1

(Intercept) -1.468326e+03

(Intercept) .

PrivateYes -5.278781e+02

Accept 1.004588e+00

Enroll 4.313442e-01

Top10perc 2.580619e+01

Top25perc 5.501092e-01

F.Undergrad 7.258520e-02

P.Undergrad 2.420595e-02

Outstate -2.407454e-02

Room.Board 1.987732e-01

Books 1.285477e-01

Personal -8.146131e-03

PhD -4.028284e+00

Terminal -4.811071e+00

S.F.Ratio 1.302180e+01

perc.alumni -8.544783e+00

Expend 7.589013e-02

Grad.Rate 1.126699e+01

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

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.

> set.seed(1)

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

> y<- College$Apps

> training<-sample(1:nrow(x),nrow(x)/2)

> testing<-(-training)

> y.testing<-y[testing]

> x.training<-x[training,]

> x.testing<-x[testing,]

> y.training<-y[training]

> data.training<-data.frame(y=y.training,x=x.training)

> data.testing<-data.frame(y=y.testing,x=x.testing)

> reg<-regsubsets(y~.,data=data.training,nvmax = 17)

> test.mat<-model.matrix(y~.,data = data.testing,nvmax=17)

> errors<-rep(NA,3)

> coeff<-coef(reg,id=10)

> prediction<-test.mat[,names(coeff)]%\*%coeff

> errors[1]<-mean((prediction-y.testing)^2)

> coeff<-coef(reg,id=12)

> prediction<-test.mat[,names(coeff)]%\*%coeff

> errors[2]<-mean((prediction-y.testing)^2)

> coefff<-coef(reg,id=13)

> prediction<-test.mat[,names(coeff)]%\*%coeff

> errors[3]<-mean((prediction-y.testing)^2)

> errors

[1] 1122431 1118948 1118948

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

> set.seed(1)

> training <-sample(1:nrow(x),nrow(x)/2)

> testing<-(-training)

> y.testing<-y[testing]

> lasso <-glmnet(x[training,],y[training],alpha=1,lambda=optimallambda)

> prd <- predict(lasso,s=optimallambda,newx=x[testing,])

> mean((prd-y.testing)^2)

[1] 1115208

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

> set.seed(1)

> training <-sample(1:nrow(x),nrow(x)/2)

> testing<-(-training)

> y.testing<-y[testing]

> ridge <-glmnet(x[training,],y[training],alpha=0,lambda=optimallambda2)

> prd <- predict(ridge,s=optimallambda,newx=x[testing,])

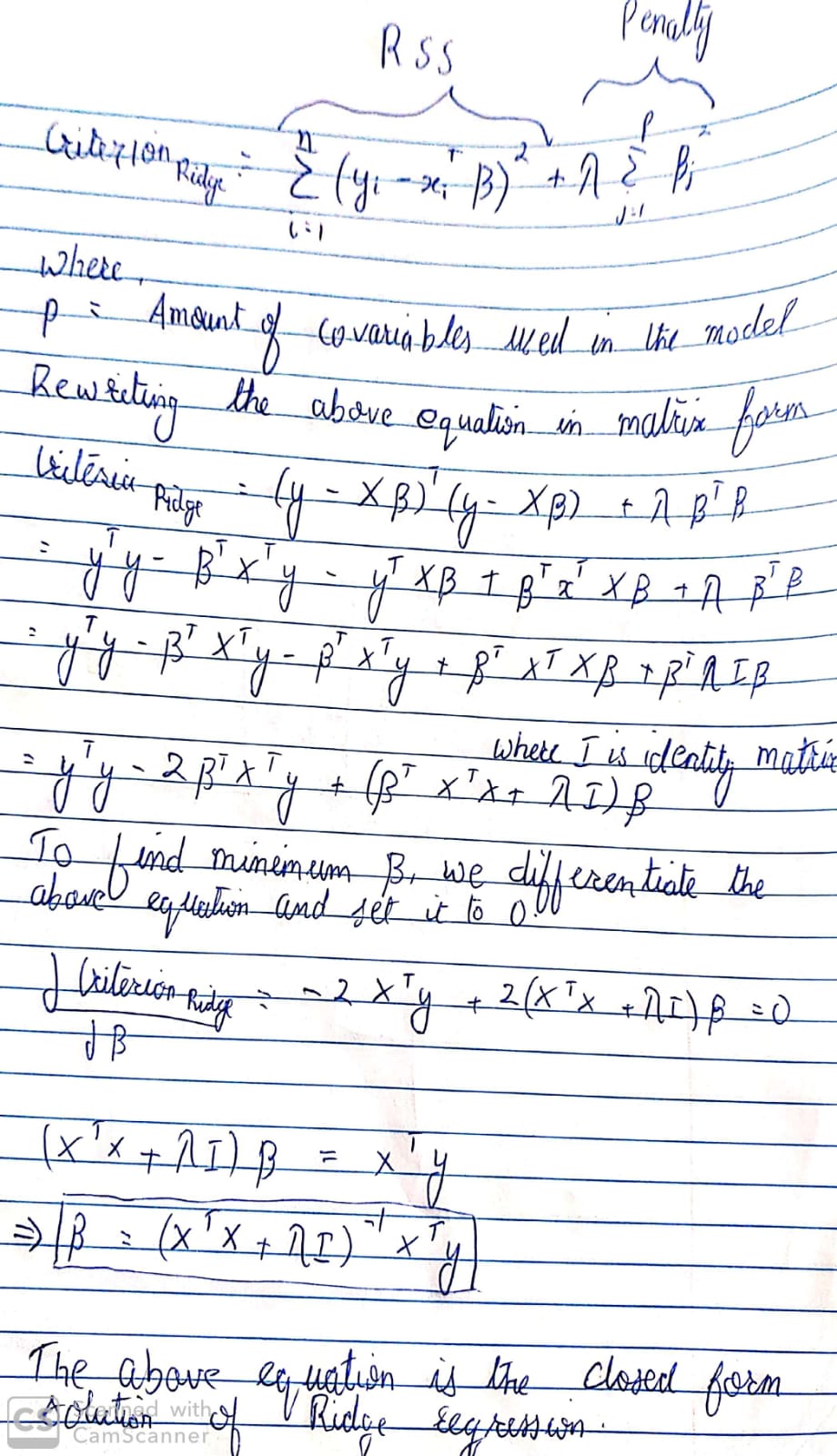
> mean((prd-y.testing)^2)

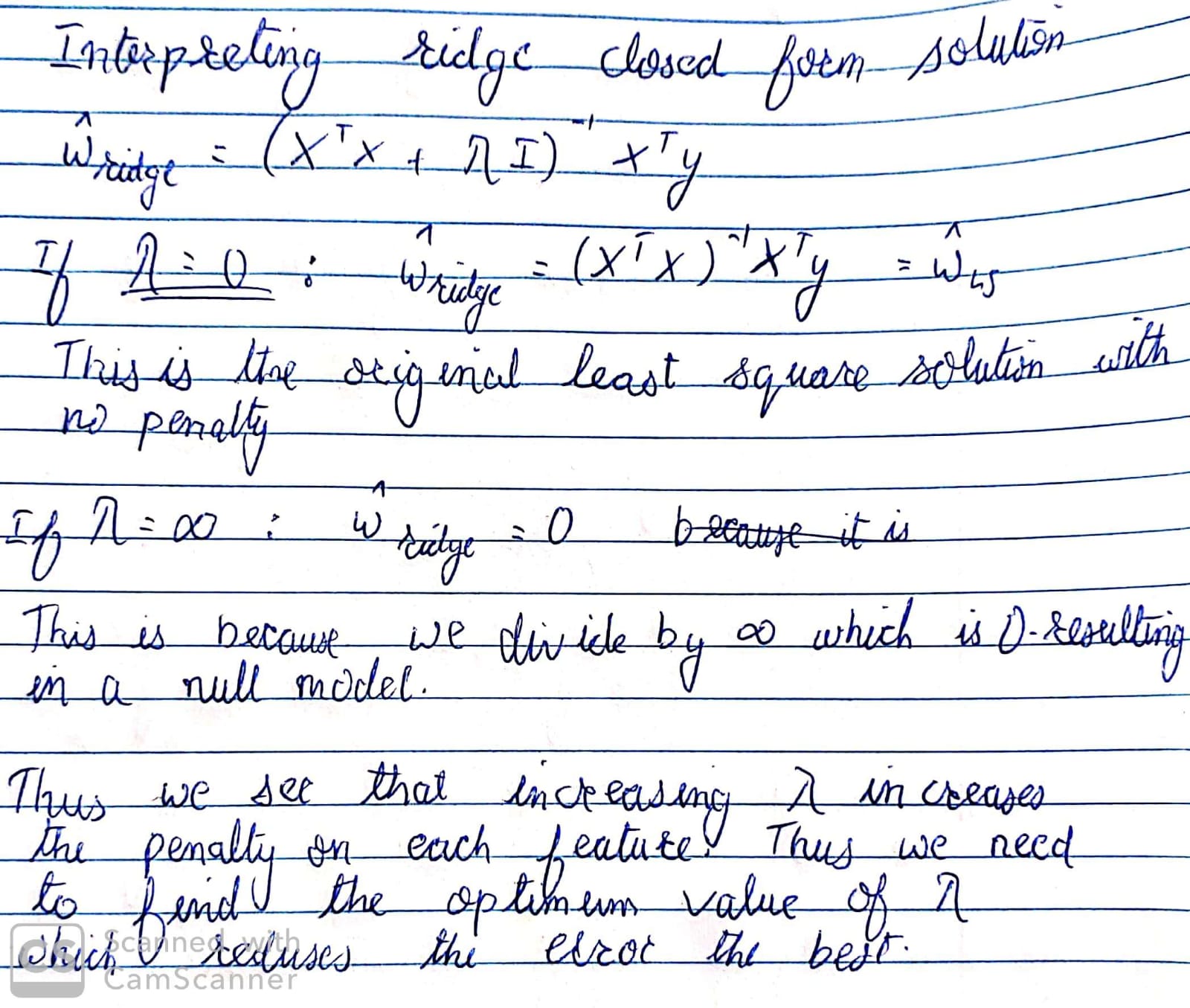
[1] 1132539

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

**Comparing above three models best model is Lasso model with the optimum Lambda obtained by cross validation.**

**Problem 2 (5pt)** 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 .)

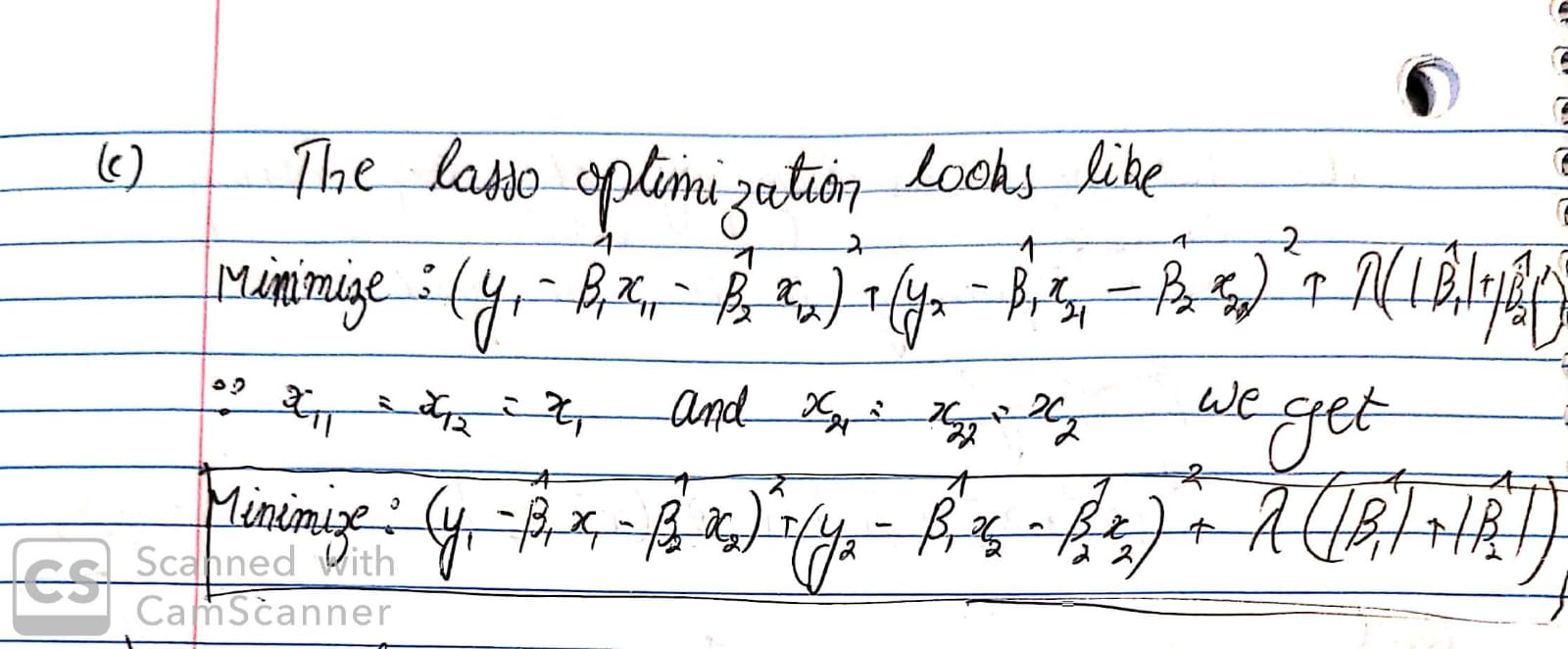




**Problem 3 (Bonus 5pt)** Solve question 5 (c)-(d) in Ch. 6.8 Exercises (hint: In describing the solutions in 5 (d), use equation 6.15).

*(For problems 2 and 3 above, you do not need to use R. )*

5 (c)



5 (d)

