Lab work 4

Artem Storozhuk

IMPORTING DATA

First of all, lets read our data and transform it for further use:

```
filenames = list.files(path='data', full.names=TRUE)
datalist = lapply(filenames,function(x){
  x0 <- read.csv(file = x,header = F)[,c(1,6)];
  colnames(x0) <- c("data", unlist(strsplit(x,"[_.]"))[2]);x0})
y <- Reduce(function(x,y){
  merge(x,y,by="data")
  },datalist)

library(leaps)
```

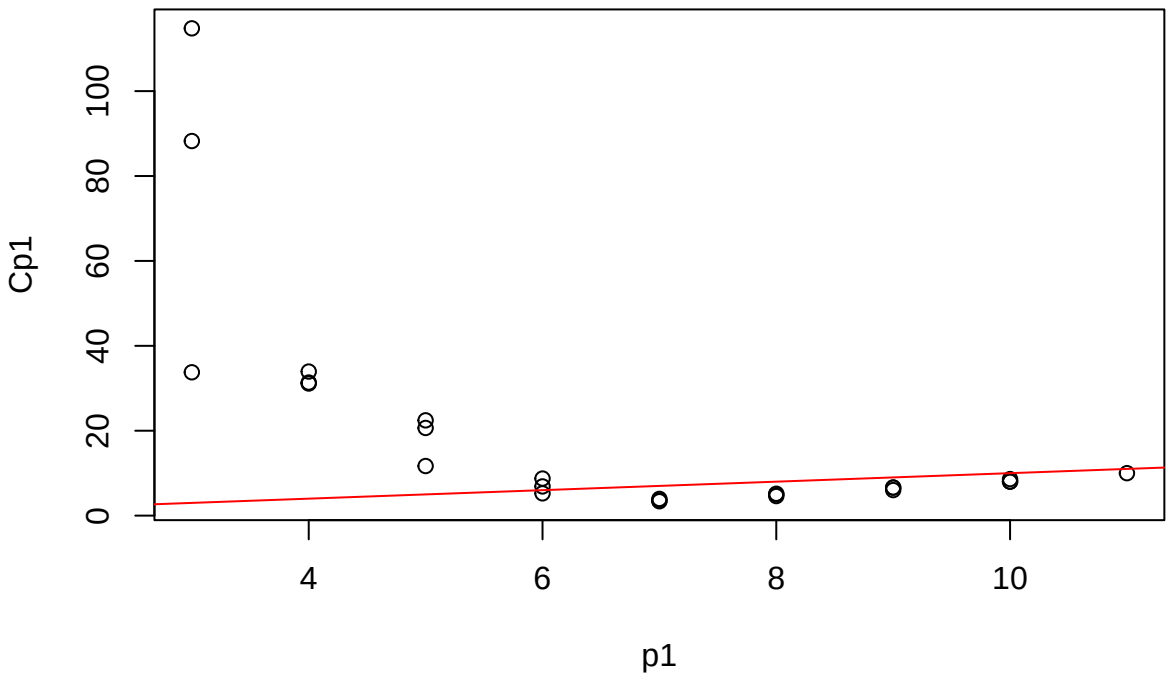
```
## Warning: package 'leaps' was built under R version 4.1.2
```

```
Data <- y[-nrow(y),-1]
nn <- nrow(Data)
```

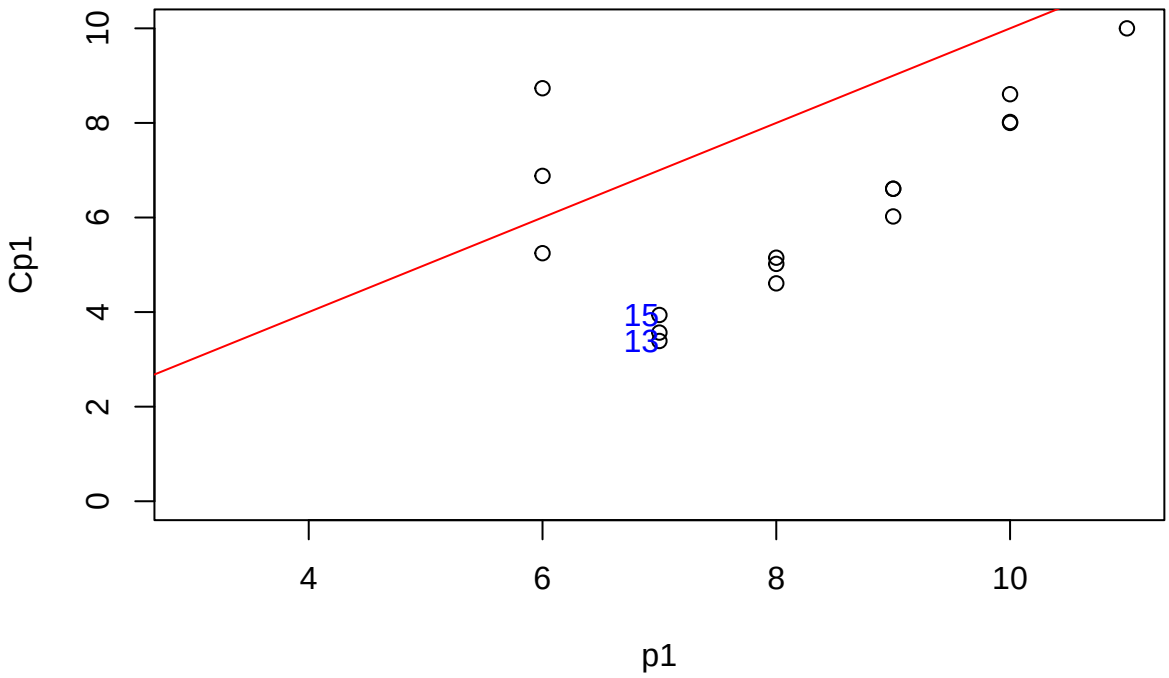
BUILDING A MODEL

Let's create a CpModel1 using 50 sessions before the last 20

```
CpModel1 <- regsubsets(adi~.-adi , data=Data[(nn-70):(nn-20),], nbest = 3, nvmax = 9)
p1 <- apply(summary(CpModel1)$which,1,sum)+1
Cp1 <- summary(CpModel1)$cp
plot(p1,Cp1)
abline(0,1, col='red')
```




```
plot(p1,Cp1, ylim = c(0,10))
abline(0,1,col='red')
index <- c(13,15)
text(p1[index],Cp1[index], labels = index, col = 'blue',adj = 1)
```



As we can see, 13 (Cp= 3.138596) and 15 (Cp= 4.862196) are good choices. Because 13 lies slightly lower, we will choose 15 as our final model

```
summary(CpModel1)$which[15,]
```

```
## (Intercept)         adm         adp        adsk         adt         aee
##        TRUE        TRUE        TRUE        TRUE       FALSE       FALSE
##         aep         aes         aet         afl
##       FALSE       FALSE        TRUE        TRUE
```

```
Cp1[15]
```

```
## [1] 3.938495
```

Selected regressors are adm,adp,adsk,aet,afl and constant term. Coefficients:

```
coef1<-coef(CpModel1,15)
coef1
```

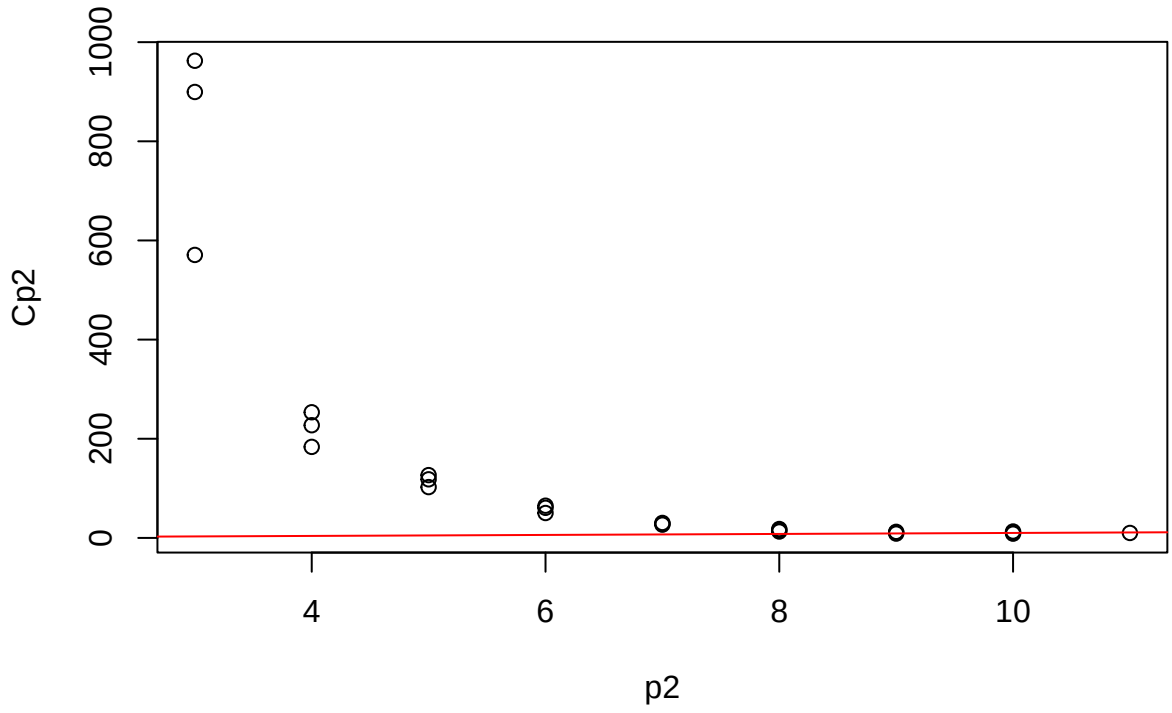
```
## (Intercept)         adm         adp        adsk         aet         afl
## -15.2426295  -0.3391691   0.6970068   0.2151764  -0.1639940   0.4652325
```

Testing model on last 20 sessions:

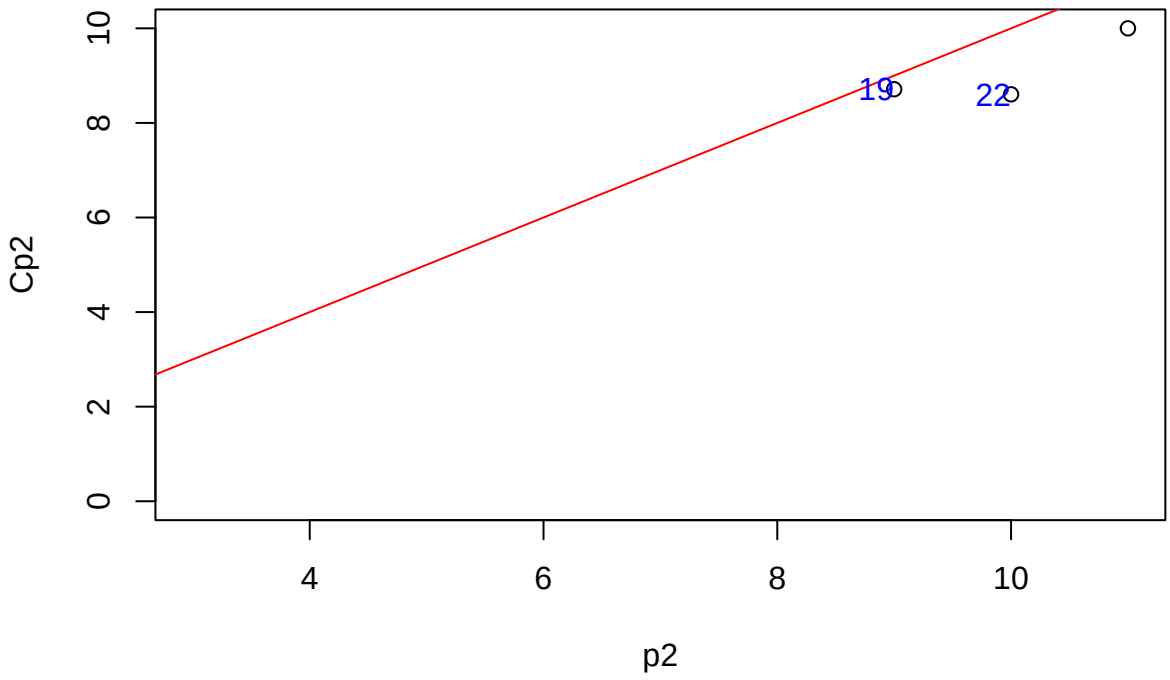
```
test1 <- as.matrix(cbind(const=1, Data[(nn-19):nn, c(2,3,4,9,10)]))
predict1 <- test1%*%coef1
u1 <- Data$adi[(nn-19):(nn)]-predict1
```

Doing the same things as previously (but using 50 sessions before last 20):

```
CpModel2 <- regsubsets(adi~.-adi, data = Data[1:(nn-20),], nbest = 3, nvmax = 9)
p2 <- apply(summary(CpModel2)$which,1,sum)+1
Cp2 <- summary(CpModel2)$cp
plot(p2,Cp2)
abline(0,1, col = 'red')
```




```
plot(p2,Cp2, ylim=c(0,10))
abline(0,1,col='red')
index <- c(19,22)
text(p2[index],Cp2[index], labels = index, col = 'blue',adj = 1)
```



Now 19 is the best candidate.

```
summary(CpModel2)$which[19,]
```

```
## (Intercept)         adm         adp        adsk         adt         aee
##        TRUE        TRUE        TRUE        TRUE        TRUE        TRUE
##         aep         aes         aet         afl
##       FALSE        TRUE        TRUE       FALSE
```

```
Cp2[19]
```

```
## [1] 8.712247
```

Selected regressors are adm,adp,adsk,adt,aee,aes,aet and constant term. Coefficients:

```
coef2 <- coef(CpModel2,19)
coef2
```

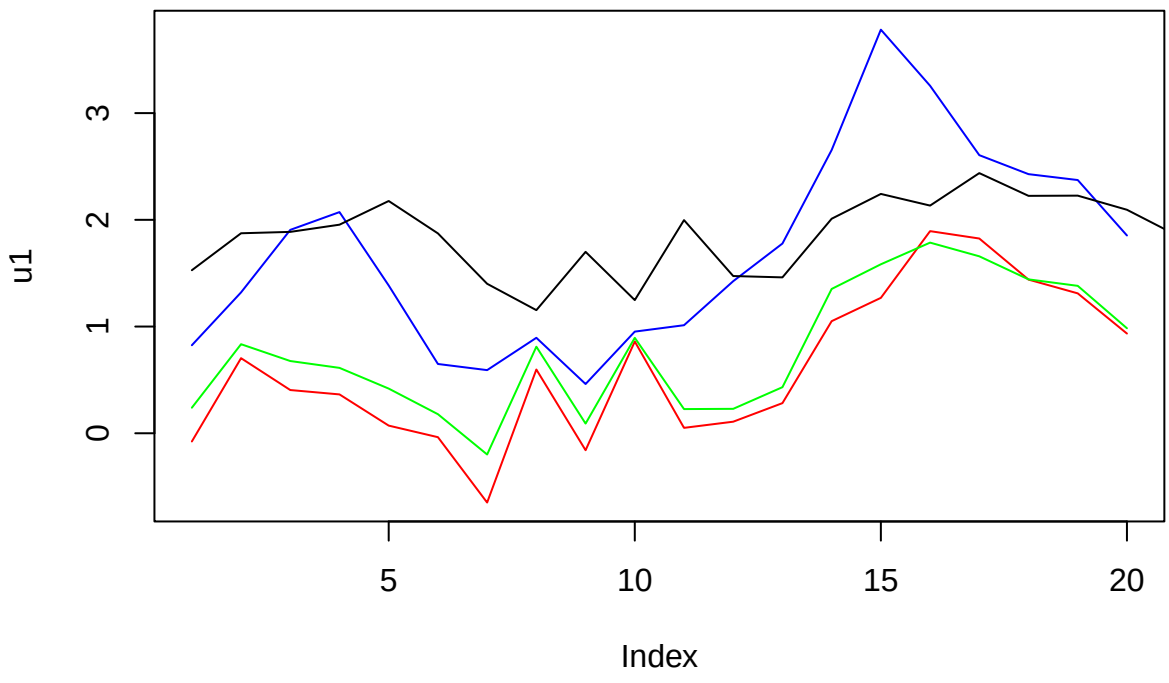
```
## (Intercept)         adm         adp        adsk         adt         aee
##  11.2500784   0.4804021   0.1868610   0.3499128   0.1680017  -0.2734147
##         aes         aet
##  -1.0327890   0.1254802
```

Testing:

```
test2 <- as.matrix(cbind(const=1,Data[(nn-19):nn,c(2,3,4,5,6,8,9)]))
predict2 <- test2%*%coef2
u2 <- Data$adi[(nn-19):nn]-predict2
```

COMPARISON

```
library(MASS)
fit <- lm.ridge(adi~.-adi, data = Data[(nn-69):(nn-20),], lambda = seq(0.001,50,.01))
i <- which.min(fit$GCV)
test3 <- as.matrix(cbind(const=1,Data[(nn-19):nn, 2:10]))
ridge <- lm.ridge(adi~.-adi,data = Data[(nn-69):(nn-20),], lambda =fit$lambda[i] )
coef3 <- coef(ridge)
predict3 <- test3%*%coef3
u3 <- Data$adi[(nn-19):nn]-predict3
model2 <- lm(adi~afl+aep, data = Data[(nn-70):(nn-20),])
u4 <- Data$adi[(nn-20):(nn)]-predict(model2, Data[(nn-20):(nn),])
plot(u1,type = 'l',col='red', ylim=c(min(u1,u2,u3,u4),max(u1,u2,u3,u4)))
lines(u2,col='blue')
lines(u3,col='green')
lines(u4,col='black')
```



We can see that the first model in this work is the best.