Lab work 3

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

library(MASS)
```

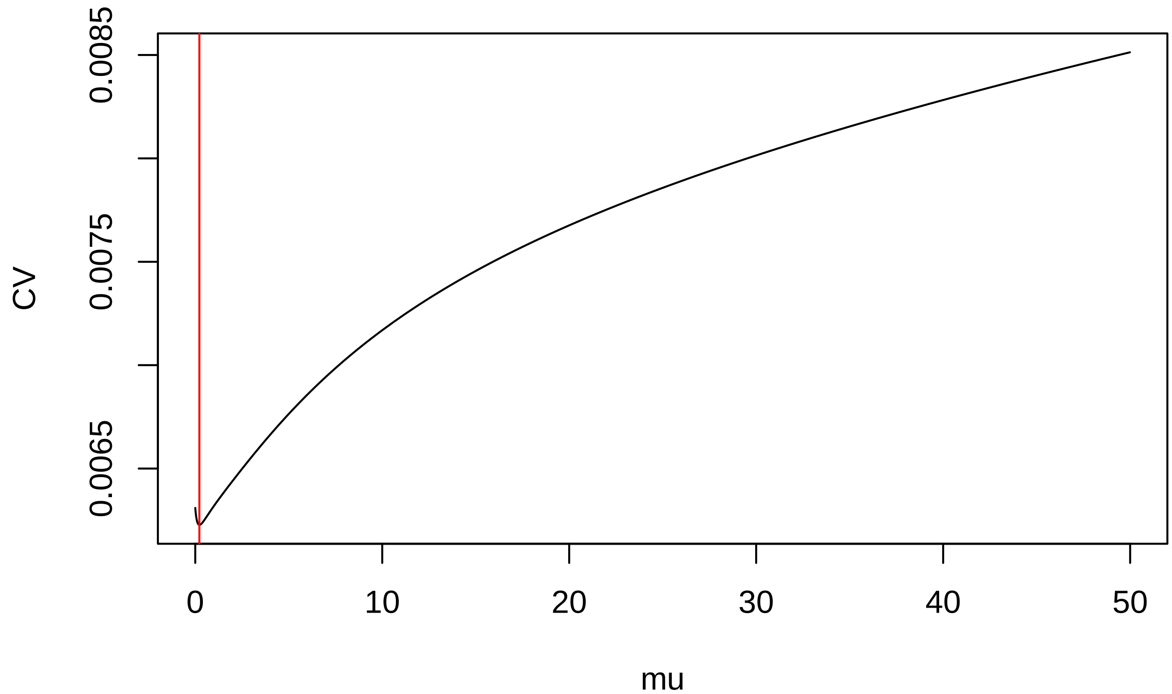
RIDGE

Using ridge regression, lambda - set of values of the regularization parameter:

```
fit<-lm.ridge(adi~.-adi, data = Data[(nn-69):(nn-20),], lambda = seq(0.001,50,.01))
```

GRAPH OF CROSS-VALIDATION VALUES

```
plot(fit$lambda,fit$GCV,type = 'l', xlab = 'mu', ylab = 'CV')
i<-which.min(fit$GCV)
abline(v=fit$lambda[i],col='red')
```



Determining the number of the CV minimum value:

```
fit$lambda[i]
```

```
## [1] 0.221
```

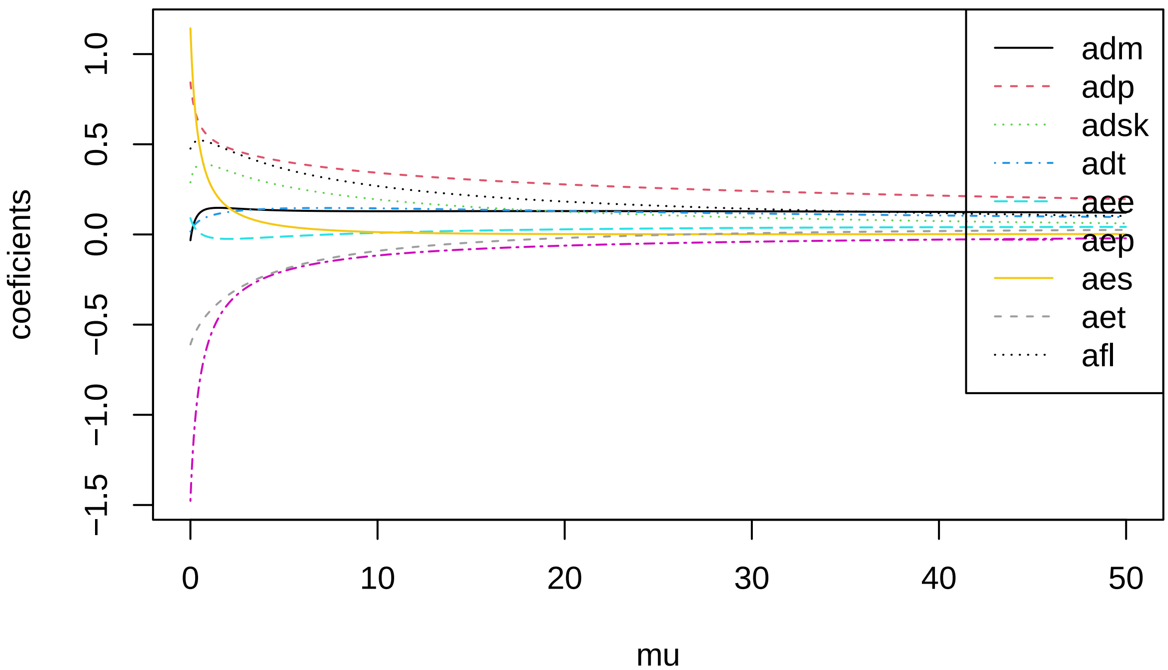
```
fit$GCV[i]
```

```
##       0.221
## 0.006227653
```

The optimal value is achieved when the regularization parameter equals 0.221.

GRAPH OF THE DEPENDENCE OF COEFFICIENTS ON THE REGULARIZATION PARAMETERS

```
matplot(fit$lambda,t(fit$coef), type = 'l', col=1:9, lty=1:9, xlab = 'mu',ylab = 'coeficients')
legend('topright',col = 1:9,legend = colnames(Data)[2:10],lty=1:9)
```

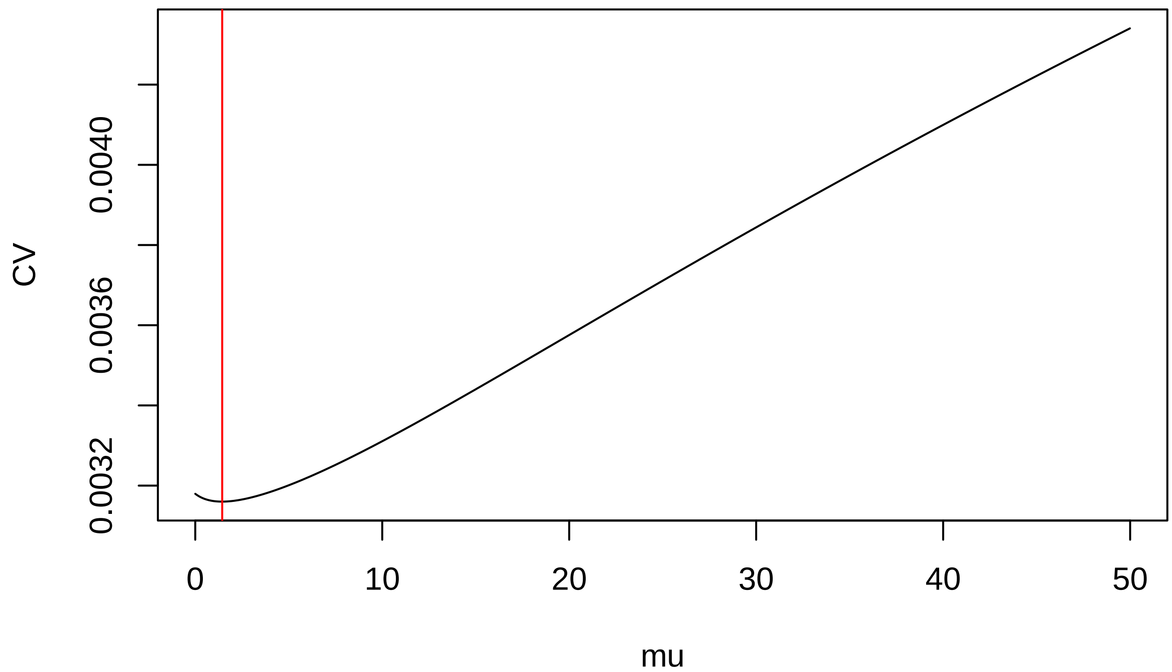


TEST DATA

Lets consider our model on test data and build a model on all data except last 20 sessions:

```
test <- as.matrix(cbind(const=1,Data[(nn-19):nn, 2:10]))
ridgeFinal <- lm.ridge(adi~.-adi,data = Data[(nn-69):(nn-20),], lambda =fit$lambda[i] )
coefs <- coef(ridgeFinal)
predict <- test%*%coefs
u1<-Data$adi[(nn-19):nn]-predict

fit1 <- lm.ridge(adi~.-adi,data=Data[1:(nn-20),], lambda = seq(0.001,50,.01))
plot(fit1$lambda,fit1$GCV,type = 'l',xlab = 'mu',ylab = 'CV')
i <- which.min(fit1$GCV)
abline(v=fit1$lambda[i],col='red')
```




```
fit1$lambda[i]
```

```
## [1] 1.441
```

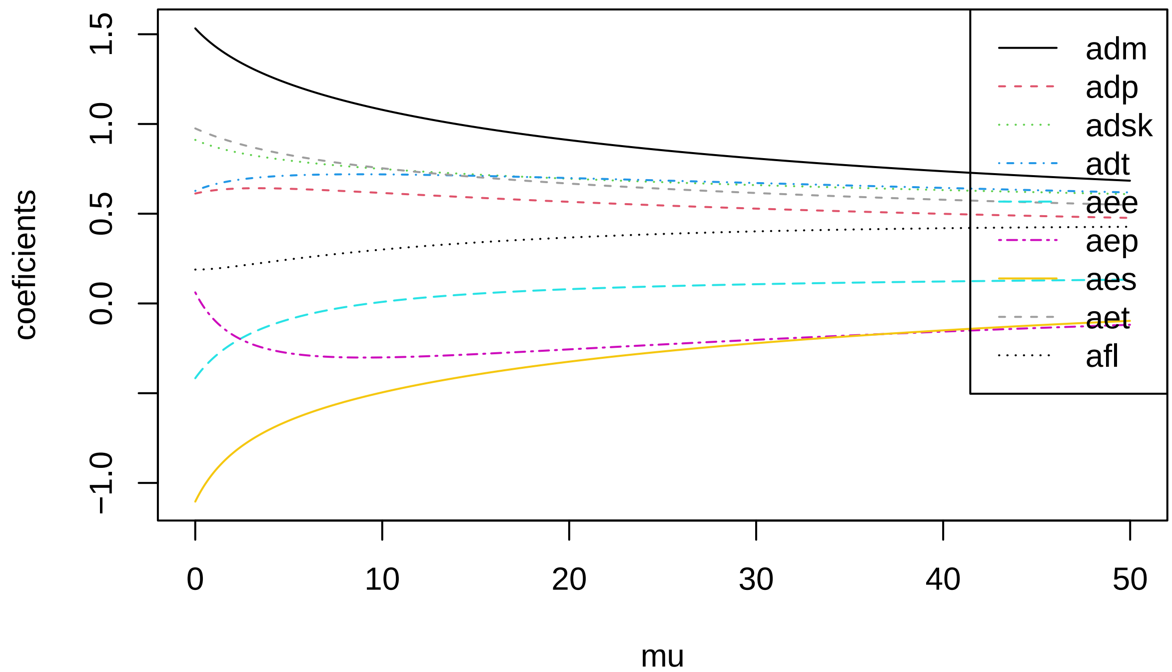
```
fit1$GCV[i]
```

```
##       1.441
## 0.003160076
```

The optimal value is achieved when the regularization parameter equals 1.441

GRAPH OF THE DEPENDENCE OF COEFFICIENTS ON THE REGULARIZATION PARAMETERS

```
matplot(fit1$lambda,t(fit1$coef),type = 'l',col = 1:9, lty=1:9, xlab = 'mu',ylab = 'coeficients')
legend('topright', col=1:9,legend = colnames(Data)[2:10],lty = 1:9)
```



Now lets consider our model on last 20 sessions:

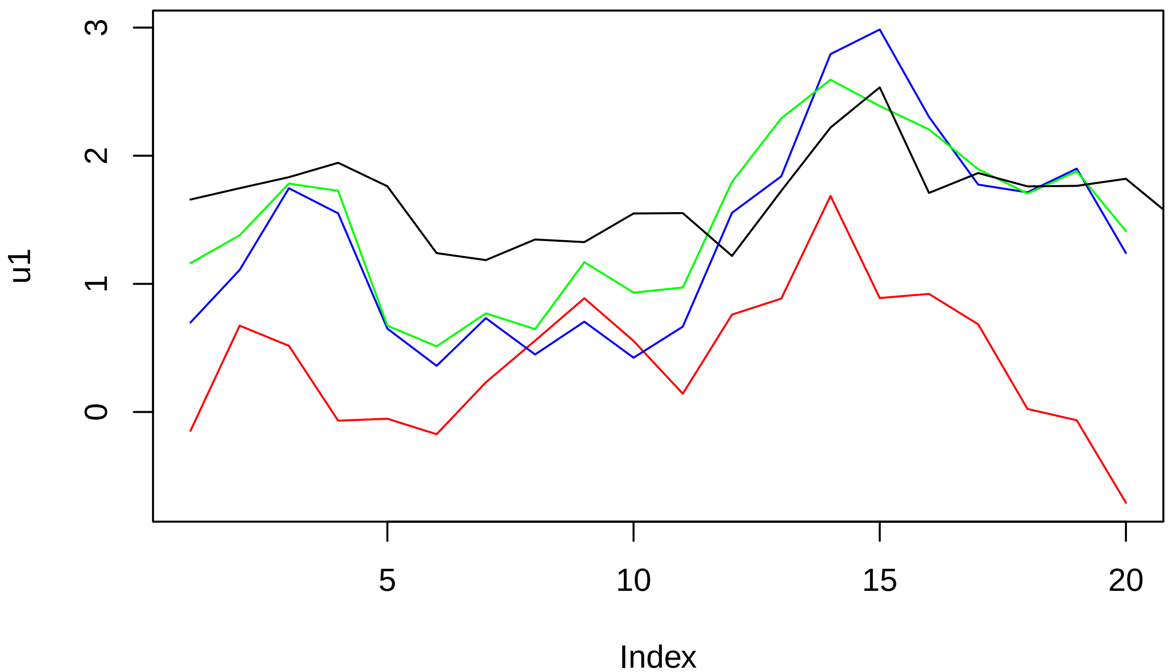
```
test <- as.matrix(cbind(const=1,Data[(nn-19):nn,2:10]))
ridgeFinal1 <- lm.ridge(adi~.-adi, data = Data[1:(nn-20),], lambda = fit1$lambda[i])
coefs1<-coef(ridgeFinal1)
predict1 <- test%*%coefs1
u2 <- Data$adi[(nn-19):nn]-predict1
```

COMPARE

I literally have no idea why almost same parts of code works in one R Markdown and don't work in other, but since it works here we will just compare our models with previous once:

```
X<-Data[-1]
PC<-prcomp(X)
c1<-PC$x[,1]
c2<-PC$x[,2]
c3<-PC$x[,3]
Data1<-data.frame(c1=c1,c2=c2,c3=c3,adi=Data$adi)
model1<-lm(adi~.-adi-c2-c3, data = Data1[(nn-69):(nn-20), ])
base<-as.data.frame(predict(PC, newdata = Data[(nn-19):(nn),2:10]))
predict1<-model1$coefficients[1]+model1$coefficients[2]*base$PC1
u3<-Data$adi[(nn-19):nn]-predict1
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 model **fit** has the best forecast accuracy.