lab3q5

March 20, 2017

In [1]:
```
## Question 5
library(e1071)
library(class)
#library(knncat)

dt = read.csv("GermanData.csv")
dt$Credit.Risks = as.factor(dt$Credit.Risks)
num.vars <- sapply(dt, is.numeric)
dt[num.vars] <- lapply(dt[num.vars], scale)

######80-20

n.points <- 1000 # number of rows in the dataset
sampling.rate <- 0.8

num.test.set.labels <- n.points * (1 - sampling.rate)
training <- sample(1:n.points, sampling.rate * n.points,
                   replace=FALSE)

train <- subset(dt[training, ], select = c("X.","Age","Job","Credit.amount'

testing <- setdiff(1:n.points, training)
test <- subset(dt[testing, ], select = c("X.","Age","Job","Credit.amount","'


cl = train$Credit.Risks


knn_model.1 = knn(train, test, cl,k = 1)

knn_model.5 = knn(train, test, cl,k = 5)

knn_model.20 = knn(train, test, cl,k = 20)

table(knn_model.20, test$Credit.Risks)
table(knn_model.5, test$Credit.Risks)
```

```
######70-30

n.points <- 1000 # number of rows in the dataset
sampling.rate <- 0.7

num.test.set.labels <- n.points * (1 - sampling.rate)
training <- sample(1:n.points, sampling.rate * n.points,
                   replace=FALSE)

train <- subset(dt[training, ], select = c("X.","Age","Job","Credit.amount'

testing <- setdiff(1:n.points, training)
test <- subset(dt[testing, ], select = c("X.","Age","Job","Credit.amount","'


cl = train$Credit.Risks


knn_model.1 = knn(train, test, cl,k = 1)

knn_model.5 = knn(train, test, cl,k = 5)

knn_model.20 = knn(train, test, cl,k = 20)

table(knn_model.20, test$Credit.Risks)
table(knn_model.5, test$Credit.Risks)
```

```
knn_model.20   1   2
           1 147   8
           2   0  45
```

```
knn_model.5   1   2
          1 147   9
          2   0  44
```

```
knn_model.20   1   2
           1 199  31
           2   0  70
```

```
knn_model.5   1   2
          1 198  11
```

```
2   1  90
```