

Churn_nearest_neighbour_classification.R

S Varatharajan

Wed Oct 10 14:28:36 2018

```
setwd("C:/Users/tsraj/Desktop/Acadgild project")
library(readr)
churn <- read_csv("churn.csv")

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   `Day Mins` = col_double(),
##   `Eve Mins` = col_double(),
##   `Night Mins` = col_double(),
##   `Intl Mins` = col_double(),
##   `Day Charge` = col_double(),
##   `Eve Charge` = col_double(),
##   `Night Charge` = col_double(),
##   `Intl Charge` = col_double(),
##   State = col_character(),
##   Phone = col_character()
## )

## See spec(...) for full column specifications.

View(churn)
library(doParallel)

## Loading required package: foreach

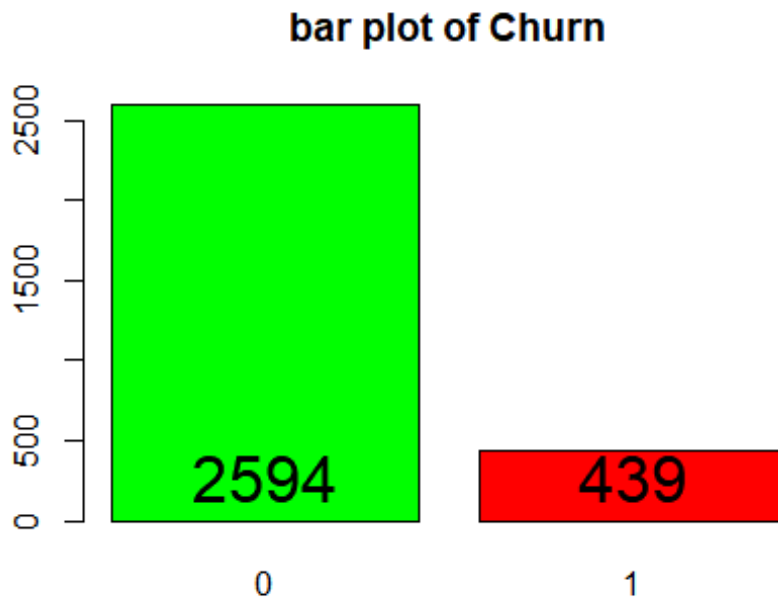
## Loading required package: iterators

## Loading required package: parallel

registerDoParallel()
set.seed(12345)
train<-churn[1:2500,]
test<-churn[2501:3033,]
mydata1 <- rbind(test,train)
mydata2<-mydata1[, -19]
mydata3<-mydata2[, -19]
mydata<-mydata3[, -19]
#Nearest Neighbour classification

barplot(table(mydata$Churn), col= c("green", "red"), main='bar plot of
Churn')
```

```
text(barplot(table(mydata$Churn), col =c('green' , 'red'), main='bar plot of Churn'), 0,table(mydata$Churn), cex =2 , pos =3)
```



#proportion

```
round(prop.table(table(mydata$Churn))*100,digits = 2)
```

```
##
```

```
##      0      1
```

```
## 85.53 14.47
```

```
names(mydata)
```

```
## [1] "Account Length" "VMail Message" "Day Mins" "Eve Mins"
## [5] "Night Mins" "Intl Mins" "CustServ Calls" "Churn"
## [9] "Int'l Plan" "VMail Plan" "Day Calls" "Day Charge"
## [13] "Eve Calls" "Eve Charge" "Night Calls" "Night Charge"
## [17] "Intl Calls" "Intl Charge"
```

```
normalize<-function(x){return((x-min(x))/(max(x)-min(x)))}
```

```
mydata_n<-as.data.frame(lapply(mydata[2:18],normalize))
```

```
str(mydata)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 3033 obs. of 18 variables:
## $ Account Length: int 121 101 115 168 90 70 138 43 117 108 ...
## $ VMail Message : int 41 0 0 0 0 0 0 20 0 ...
## $ Day Mins : num 216 125 179 183 168 ...
## $ Eve Mins : num 242 257 271 179 139 ...
## $ Night Mins : num 147 193 246 293 138 ...
```

```
## $ Intl Mins      : num  9.6 13.4 16.4 9.9 13 12.5 13.2 12.6 8.7 10.9 ...
## $ CustServ Calls: int   1 0 2 2 1 3 1 0 2 2 ...
## $ Churn          : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Int'l Plan     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ VMail Plan     : int   1 0 0 0 0 0 0 0 1 0 ...
## $ Day Calls      : int  95 66 114 131 96 105 107 125 98 105 ...
## $ Day Charge     : num  36.6 21.2 30.4 31.1 28.5 ...
## $ Eve Calls      : int  92 85 96 73 104 135 120 88 107 114 ...
## $ Eve Charge     : num  20.6 21.9 23 15.2 11.8 ...
## $ Night Calls    : int  108 115 94 100 87 65 119 106 147 98 ...
## $ Night Charge   : num   6.61 8.69 11.07 13.18 6.23 ...
## $ Intl Calls     : int   3 4 5 5 1 9 4 3 3 3 ...
## $ Intl Charge    : num   2.59 3.62 4.43 2.67 3.51 3.38 3.56 3.4 2.35 2.94
...
```

```
str(mydata_n)
```

```
## 'data.frame':    3033 obs. of  17 variables:
## $ VMail.Message : num  0.804 0 0 0 0 ...
## $ Day.Mins      : num  0.614 0.356 0.509 0.522 0.477 ...
## $ Eve.Mins      : num  0.665 0.707 0.745 0.493 0.382 ...
## $ Night.Mins    : num  0.333 0.457 0.599 0.725 0.31 ...
## $ Intl.Mins     : num  0.48 0.67 0.82 0.495 0.65 0.625 0.66 0.63 0.435
0.545 ...
## $ CustServ.Calls: num  0.111 0 0.222 0.222 0.111 ...
## $ Churn         : num  0 0 0 0 0 0 0 0 0 0 ...
## $ Int.l.Plan    : num  0 0 0 0 0 0 0 0 0 0 ...
## $ VMail.Plan    : num  1 0 0 0 0 0 0 0 1 0 ...
## $ Day.Calls     : num  0.576 0.4 0.691 0.794 0.582 ...
## $ Day.Charge    : num  0.614 0.356 0.509 0.522 0.478 ...
## $ Eve.Calls     : num  0.548 0.506 0.571 0.435 0.619 ...
## $ Eve.Charge    : num  0.665 0.707 0.745 0.493 0.382 ...
## $ Night.Calls   : num  0.528 0.577 0.43 0.472 0.38 ...
## $ Night.Charge  : num  0.333 0.457 0.6 0.726 0.31 ...
## $ Intl.Calls    : num  0.1579 0.2105 0.2632 0.2632 0.0526 ...
## $ Intl.Charge   : num  0.48 0.67 0.82 0.494 0.65 ...
```

```
mydata_train<-mydata_n[1:2800,]
mydata_test<-mydata_n[2801:3033,]
mydata_train_labels<-mydata_n[1:2800,7]
mydata_test_labels<-mydata_n[2801:3033,7]
str(mydata_train)
```

```
## 'data.frame':    2800 obs. of  17 variables:
## $ VMail.Message : num  0.804 0 0 0 0 ...
## $ Day.Mins      : num  0.614 0.356 0.509 0.522 0.477 ...
## $ Eve.Mins      : num  0.665 0.707 0.745 0.493 0.382 ...
## $ Night.Mins    : num  0.333 0.457 0.599 0.725 0.31 ...
## $ Intl.Mins     : num  0.48 0.67 0.82 0.495 0.65 0.625 0.66 0.63 0.435
0.545 ...
## $ CustServ.Calls: num  0.111 0 0.222 0.222 0.111 ...
```

```

## $ Churn      : num  0 0 0 0 0 0 0 0 0 0 ...
## $ Int.l.Plan : num  0 0 0 0 0 0 0 0 0 0 ...
## $ VMail.Plan : num  1 0 0 0 0 0 0 0 1 0 ...
## $ Day.Calls  : num  0.576 0.4 0.691 0.794 0.582 ...
## $ Day.Charge : num  0.614 0.356 0.509 0.522 0.478 ...
## $ Eve.Calls  : num  0.548 0.506 0.571 0.435 0.619 ...
## $ Eve.Charge : num  0.665 0.707 0.745 0.493 0.382 ...
## $ Night.Calls : num  0.528 0.577 0.43 0.472 0.38 ...
## $ Night.Charge : num  0.333 0.457 0.6 0.726 0.31 ...
## $ Intl.Calls  : num  0.1579 0.2105 0.2632 0.2632 0.0526 ...
## $ Intl.Charge : num  0.48 0.67 0.82 0.494 0.65 ...

str(mydata_train_labels)

## num [1:2800] 0 0 0 0 0 0 0 0 0 0 ...

str(mydata_test)

## 'data.frame': 233 obs. of 17 variables:
## $ VMail.Message : num  0 0.314 0.725 0 0 ...
## $ Day.Mins      : num  0.893 0.415 0.216 0.558 0.376 ...
## $ Eve.Mins      : num  0.595 0.525 0.477 0.629 0.636 ...
## $ Night.Mins    : num  0.525 0.517 0.375 0.754 0.781 ...
## $ Intl.Mins     : num  0.64 0.565 0.41 0.61 0.435 0.585 0.38 0.495 0.385
0.295 ...
## $ CustServ.Calls: num  0.222 0 0 0.333 0.111 ...
## $ Churn         : num  1 0 0 0 0 0 0 0 0 0 ...
## $ Int.l.Plan    : num  0 0 0 0 0 0 0 0 1 0 ...
## $ VMail.Plan    : num  0 1 1 0 0 0 0 1 0 0 ...
## $ Day.Calls     : num  0.624 0.533 0.618 0.521 0.576 ...
## $ Day.Charge    : num  0.893 0.415 0.216 0.558 0.376 ...
## $ Eve.Calls     : num  0.899 0.768 0.875 0.488 0.44 ...
## $ Eve.Charge    : num  0.595 0.525 0.478 0.629 0.636 ...
## $ Night.Calls   : num  0.514 0.345 0.444 0.43 0.528 ...
## $ Night.Charge  : num  0.525 0.518 0.375 0.754 0.781 ...
## $ Intl.Calls    : num  0.211 0.368 0.684 0.211 0.526 ...
## $ Intl.Charge   : num  0.641 0.565 0.409 0.609 0.435 ...

str(mydata_test_labels)

## num [1:233] 1 0 0 0 0 0 0 0 0 0 ...

library(class)
#Apply knn
mydata_test_pred<-knn(train = mydata_train,test = mydata_test,
cl=mydata_train_labels,k=53)
summary(mydata_test_pred)

## 0 1
## 194 39

```

```
#Evalualte model
```

```
library(gmodels)
```

```
CrossTable(x=mydata_test_labels, y=mydata_test_pred,prop.chisq = FALSE)
```

```
##
```

```
##
```

```
##      Cell Contents
```

	N
N / Row Total	
N / Col Total	
N / Table Total	

```
##
```

```
##
```

```
## Total Observations in Table:  233
```

```
##
```

```
##
```

mydata_test_labels	mydata_test_pred		Row Total
	0	1	
0	194	0	194
	1.000	0.000	0.833
	1.000	0.000	
	0.833	0.000	
1	0	39	39
	0.000	1.000	0.167
	0.000	1.000	
	0.000	0.167	
Column Total	194	39	233
	0.833	0.167	

```
##
```

```
##
```