

Caravan_KNN

Nabil Momin

2024-06-10

```
library(corrgram)
library(corrplot)

## corrplot 0.92 loaded

library(caTools)
library(Amelia)

## Loading required package: Rcpp

## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.8.2, built: 2024-04-10)
## ## Copyright (C) 2005-2024 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##

library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

#### Getting the data

library(ISLR)

head(Caravan)

## MOSTYPE MAANTHUI MGEMOMV MGEMLEEF MOSHOOFD MGODRK MGODPR MGODOV MGODGE
MRELGE
## 1 33 1 3 2 8 0 5 1 3
7
## 2 37 1 2 2 8 1 4 1 4
6
## 3 37 1 2 2 8 0 4 2 4
```

3										
## 4	9	1	3	3	3	2	3	2	4	
5										
## 5	40	1	4	2	10	1	4	1	4	
7										
## 6	23	1	2	1	5	0	5	0	5	
0										
##	MRELSA	MRELOV	MFALLEEN	MFGEKIND	MFWEKIND	MOPLHOOG	MOPLMIDD	MOPLLAAG		
MBERHOOG										
## 1	0	2	1	2	6	1	2	7		
1										
## 2	2	2	0	4	5	0	5	4		
0										
## 3	2	4	4	4	2	0	5	4		
0										
## 4	2	2	2	3	4	3	4	2		
4										
## 5	1	2	2	4	4	5	4	0		
0										
## 6	6	3	3	5	2	0	5	4		
2										
##	MBERZELF	MBERBOER	MBERMIDD	MBERARBG	MBERARBO	MSKA	MSKB1	MSKB2	MSKC	MSKD
## 1	0	1	2	5	2	1	1	2	6	1
## 2	0	0	5	0	4	0	2	3	5	0
## 3	0	0	7	0	2	0	5	0	4	0
## 4	0	0	3	1	2	3	2	1	4	0
## 5	5	4	0	0	0	9	0	0	0	0
## 6	0	0	4	2	2	2	2	2	4	2
##	MHHUUR	MHKOOP	MAUT1	MAUT2	MAUT0	MZFONDS	MZPART	MINKM30	MINK3045	MINK4575
## 1	1	8	8	0	1	8	1	0	4	5
## 2	2	7	7	1	2	6	3	2	0	5
## 3	7	2	7	0	2	9	0	4	5	0
## 4	5	4	9	0	0	7	2	1	5	3
## 5	4	5	6	2	1	5	4	0	0	9
## 6	9	0	5	3	3	9	0	5	2	3
##	MINK7512	MINK123M	MINKGEM	MKOOKLA	PWAPART	PWABEDR	PWALAND	PPERSAUT		
PBESAUT										
## 1	0	0	4	3	0	0	0	6		
0										
## 2	2	0	5	4	2	0	0	0		
0										
## 3	0	0	3	4	2	0	0	6		
0										
## 4	0	0	4	4	0	0	0	6		
0										
## 5	0	0	6	3	0	0	0	0		
0										
## 6	0	0	3	3	0	0	0	6		
0										
##	PMOTSCO	PVRAAUT	PAANHANG	PTRACTOR	PWERKT	PBROM	PLEVEN	PPERSONG	PGEZONG	

```
## 1      0      0      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0      0      0
## 3      0      0      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0      0      0
## PWAOREG PBRAND PZEILPL PPLEZIER PFIETS PINBOED PBYSTAND AWAPART AWABEDR
## 1      0      5      0      0      0      0      0      0      0
## 2      0      2      0      0      0      0      0      2      0
## 3      0      2      0      0      0      0      0      1      0
## 4      0      2      0      0      0      0      0      0      0
## 5      0      6      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0      0      0
## AWALAND APERSAUT ABESAUT AMOTSCO AVRAAUT AAANHANG ATRACTOR AWERKT ABROM
## 1      0      1      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0      0      0
## 3      0      1      0      0      0      0      0      0      0
## 4      0      1      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0      0      0
## 6      0      1      0      0      0      0      0      0      0
## ALEVEN APERSONG AGEZONG AWAOREG ABRAND AZEILPL APLEZIER AFIETS AINBOED
## 1      0      0      0      0      1      0      0      0      0
## 2      0      0      0      0      1      0      0      0      0
## 3      0      0      0      0      1      0      0      0      0
## 4      0      0      0      0      1      0      0      0      0
## 5      0      0      0      0      1      0      0      0      0
## 6      0      0      0      0      0      0      0      0      0
## ABYSTAND Purchase
## 1      0      No
## 2      0      No
## 3      0      No
## 4      0      No
## 5      0      No
## 6      0      No
```

checking the variance

```
var(Caravan[,1])
```

```
## [1] 165.0378
```

```
var(Caravan[,2])
```

```
## [1] 0.1647078
```

standardizing the scale for the columns

```
View(Caravan)
```

```
standard.caravan <- scale(Caravan[1:85])
```

```

var(standard.caravan[,1])
## [1] 1
var(standard.caravan[,2])
## [1] 1
final.caravan <- cbind(standard.caravan,Caravan[86])
View(final.caravan)

#### Making the train and test

sample <- sample.split(final.caravan,SplitRatio = 0.7)
train <- subset(final.caravan,sample == TRUE)
test <- subset(final.caravan,sample == FALSE)

#### Making the model

library(class)

model.purchase <- knn(train[1:85],test[1:85],train$Purchase,k=1)
error.rate <- mean(test$Purchase != model.purchase)

#### Elbow method

error.rate <- NULL
model.purchase <- NULL

for (i in 1:40){
  model.purchase <- knn(train[1:85],test[1:85],train$Purchase,k=i)
  error.rate[i] <- mean(test$Purchase != model.purchase)
}

k.values <- 1:40

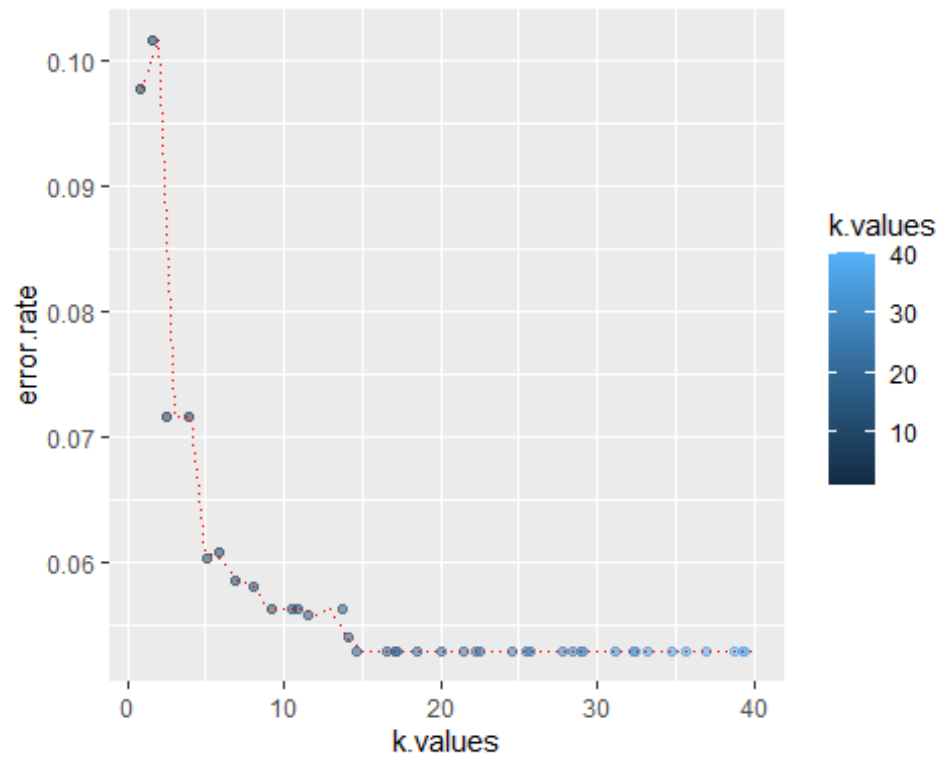
error.df <- data.frame(error.rate,k.values)
View(error.df)

#### ggplot using the error.rate

ggplot(error.df,aes(k.values,error.rate)) +

```

```
geom_point(position=position_jitter(w=1, h=0),aes(color=k.values),alpha=0.5)  
+ geom_line(lty='dotted',color='red')
```



This is a much better data set to implement the KNN and we can see the error rate is high but it eventually it goes down and remains the same