RML\_050.R

darki

2020-04-02

# Rozdział 5 - Klasyfikacja K-najbliższych sąsiadów  
# Dane giełdowe indeksu S&P zmian dziennych  
  
rm(list=ls())  
library(class)  
library(ISLR)

## Warning: package 'ISLR' was built under R version 3.6.3

attach(Smarket)  
train=(Year<2005)  
Direction.2005=Direction[!train]  
train.X=cbind(Lag1,Lag2)[train,]  
test.X=cbind(Lag1,Lag2)[!train,]  
train.Direction=Direction[train]  
set.seed(1)  
knn.pred=knn(train.X,test.X,train.Direction,k=1)  
table(knn.pred,Direction.2005)

## Direction.2005  
## knn.pred Down Up  
## Down 43 58  
## Up 68 83

(83+43)/252

## [1] 0.5

knn.pred=knn(train.X,test.X,train.Direction,k=3)  
table(knn.pred,Direction.2005)

## Direction.2005  
## knn.pred Down Up  
## Down 48 54  
## Up 63 87

mean(knn.pred==Direction.2005)

## [1] 0.5357143

# Dane firmy ubezpieczeniowej  
  
dim(Caravan)

## [1] 5822 86

attach(Caravan)  
head(Caravan)

## MOSTYPE MAANTHUI MGEMOMV MGEMLEEF MOSHOOFD MGODRK MGODPR MGODOV MGODGE  
## 1 33 1 3 2 8 0 5 1 3  
## 2 37 1 2 2 8 1 4 1 4  
## 3 37 1 2 2 8 0 4 2 4  
## 4 9 1 3 3 3 2 3 2 4  
## 5 40 1 4 2 10 1 4 1 4  
## 6 23 1 2 1 5 0 5 0 5  
## MRELGE MRELSA MRELOV MFALLEEN MFGEKIND MFWEKIND MOPLHOOG MOPLMIDD  
## 1 7 0 2 1 2 6 1 2  
## 2 6 2 2 0 4 5 0 5  
## 3 3 2 4 4 4 2 0 5  
## 4 5 2 2 2 3 4 3 4  
## 5 7 1 2 2 4 4 5 4  
## 6 0 6 3 3 5 2 0 5  
## MOPLLAAG MBERHOOG MBERZELF MBERBOER MBERMIDD MBERARBG MBERARBO MSKA  
## 1 7 1 0 1 2 5 2 1  
## 2 4 0 0 0 5 0 4 0  
## 3 4 0 0 0 7 0 2 0  
## 4 2 4 0 0 3 1 2 3  
## 5 0 0 5 4 0 0 0 9  
## 6 4 2 0 0 4 2 2 2  
## MSKB1 MSKB2 MSKC MSKD MHHUUR MHKOOP MAUT1 MAUT2 MAUT0 MZFONDS MZPART  
## 1 1 2 6 1 1 8 8 0 1 8 1  
## 2 2 3 5 0 2 7 7 1 2 6 3  
## 3 5 0 4 0 7 2 7 0 2 9 0  
## 4 2 1 4 0 5 4 9 0 0 7 2  
## 5 0 0 0 0 4 5 6 2 1 5 4  
## 6 2 2 4 2 9 0 5 3 3 9 0  
## MINKM30 MINK3045 MINK4575 MINK7512 MINK123M MINKGEM MKOOPKLA PWAPART  
## 1 0 4 5 0 0 4 3 0  
## 2 2 0 5 2 0 5 4 2  
## 3 4 5 0 0 0 3 4 2  
## 4 1 5 3 0 0 4 4 0  
## 5 0 0 9 0 0 6 3 0  
## 6 5 2 3 0 0 3 3 0  
## PWABEDR PWALAND PPERSAUT PBESAUT PMOTSCO PVRAAUT PAANHANG PTRACTOR  
## 1 0 0 6 0 0 0 0 0  
## 2 0 0 0 0 0 0 0 0  
## 3 0 0 6 0 0 0 0 0  
## 4 0 0 6 0 0 0 0 0  
## 5 0 0 0 0 0 0 0 0  
## 6 0 0 6 0 0 0 0 0  
## PWERKT PBROM PLEVEN PPERSONG PGEZONG PWAOREG PBRAND PZEILPL PPLEZIER  
## 1 0 0 0 0 0 0 5 0 0  
## 2 0 0 0 0 0 0 2 0 0  
## 3 0 0 0 0 0 0 2 0 0  
## 4 0 0 0 0 0 0 2 0 0  
## 5 0 0 0 0 0 0 6 0 0  
## 6 0 0 0 0 0 0 0 0 0  
## PFIETS PINBOED PBYSTAND AWAPART AWABEDR AWALAND APERSAUT ABESAUT AMOTSCO  
## 1 0 0 0 0 0 0 1 0 0  
## 2 0 0 0 2 0 0 0 0 0  
## 3 0 0 0 1 0 0 1 0 0  
## 4 0 0 0 0 0 0 1 0 0  
## 5 0 0 0 0 0 0 0 0 0  
## 6 0 0 0 0 0 0 1 0 0  
## AVRAAUT AAANHANG ATRACTOR AWERKT ABROM ALEVEN APERSONG AGEZONG AWAOREG  
## 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  
## ABRAND AZEILPL APLEZIER AFIETS AINBOED ABYSTAND Purchase  
## 1 1 0 0 0 0 0 No  
## 2 1 0 0 0 0 0 No  
## 3 1 0 0 0 0 0 No  
## 4 1 0 0 0 0 0 No  
## 5 1 0 0 0 0 0 No  
## 6 0 0 0 0 0 0 No

summary(Purchase)

## No Yes   
## 5474 348

348/5822

## [1] 0.05977327

standardized.X=scale(Caravan[,-86])  
var(Caravan[,1])

## [1] 165.0378

var(Caravan[,2])

## [1] 0.1647078

var(standardized.X[,1])

## [1] 1

var(standardized.X[,2])

## [1] 1

test=1:1000  
train.X=standardized.X[-test,]  
test.X=standardized.X[test,]  
train.Y=Purchase[-test]  
test.Y=Purchase[test]  
set.seed(1)  
knn.pred=knn(train.X,test.X,train.Y,k=1)  
mean(test.Y!=knn.pred)

## [1] 0.118

mean(test.Y!="No")

## [1] 0.059

table(knn.pred,test.Y)

## test.Y  
## knn.pred No Yes  
## No 873 50  
## Yes 68 9

9/(68+9)

## [1] 0.1168831

knn.pred=knn(train.X,test.X,train.Y,k=3)  
table(knn.pred,test.Y)

## test.Y  
## knn.pred No Yes  
## No 920 54  
## Yes 21 5

5/26

## [1] 0.1923077

knn.pred=knn(train.X,test.X,train.Y,k=5)  
table(knn.pred,test.Y)

## test.Y  
## knn.pred No Yes  
## No 930 55  
## Yes 11 4

4/15

## [1] 0.2666667

glm.fit=glm(Purchase~.,data=Caravan,family=binomial,subset=-test)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

glm.probs=predict(glm.fit,Caravan[test,],type="response")  
glm.pred=rep("No",1000)  
glm.pred[glm.probs>.5]="Yes"  
table(glm.pred,test.Y)

## test.Y  
## glm.pred No Yes  
## No 934 59  
## Yes 7 0

glm.pred=rep("No",1000)  
glm.pred[glm.probs>.25]="Yes"  
table(glm.pred,test.Y)

## test.Y  
## glm.pred No Yes  
## No 919 48  
## Yes 22 11

11/(22+11)

## [1] 0.3333333