# R Codes kNN

#Predict the new data point

install.packages("ipred")

library(ipred)

library(mlbench)

library(klaR)

#Read the data

mower <- read.csv("C:/MA 299/R/mower.csv")

#Standardize column 2 and column 3

mower[,2:3]=scale(mower[,2:3])#standardized column 2 and column 3

#Define the data point to be predicted (data point 25)

new <- mower[25,]

#Define train set (the first 24 data points)

train <- mower[1:24,]

#Perform kNN and prediction

KNN=ipredknn(ownership~income+lotsize,data=train,k=3)

#Show result

result=predict(KNN,new,"class")

result

#Cross validation (train and validation sets)

install.packages("DMwR")

#For the newer version of R, you may need to install DMwR2 instead.

library(DMwR)

#Read the data

mower24 <- read.csv("C:/MA 299/R/mower24.csv")

#Standardized column 2 and column 3

mower24[,2:3]=scale(mower24[,2:3])

#Split in train (75% of original data) + validation sets

idxs = sample(1:nrow(mower24), as.integer(0.75\*nrow(mower24)))

trainmower24 = mower24[idxs,]

testmower24 = mower24[-idxs,]

#kNN with k = 3 without normalizing the data because it is done earlier

nn3 = kNN(ownership~income+lotsize,trainmower24,testmower24,norm = FALSE, k = 3)

table(testmower24[, "ownership"], nn3)

#Cross validation (train and validation sets) 10 runs

library(DMwR)

mower24 <- read.csv("C:/MA 299/R/mower24.csv")

mower24[,2:3]=scale(mower24[,2:3])

r<-rnorm(10)

for (i in 1:10) {

idxs = sample(1:nrow(mower24), as.integer(0.75\*nrow(mower24)))

trainmower24 = mower24[idxs,]

testmower24 = mower24[-idxs,]

nn3 = kNN(ownership~income+lotsize,trainmower24,testmower24,norm = FALSE, k = 1)

Table = table(testmower24[, "ownership"], nn3)

a = Table[1,1]

b = Table[1,2]

c = Table[2,1]

d = Table[2,2]

rate = (b+c)/(a+b+c+d)

r[i]=rate}

r

mean(r)

#k-fold

#Note: You need to install Java in your computer as well.

#Go to http://www.java.com/en/download/.

install.packages("rJava")

install.packages("RWekajars")

install.packages("RWeka")

library(rJava)

library(RWeka)

#Read the data

mower24 <- read.csv("C:/MA 299/R/mower24.csv")

#standardized column 2 and column 3

mower24[,2:3]=scale(mower24[,2:3])

#k-fold cross validation for kNN

classifier = IBk(ownership~income+lotsize,data=mower24,control

= Weka\_control(K = 3, X = TRUE))

evaluate\_Weka\_classifier(classifier, numFolds = 24)