Practice and Assignment 8

Decision Trees, Naïve Bayes, k-NN classifier

1. Recheck and revise from last Lab

(a) Create a table to compare Accuracy, Sensitivity and Specificity for the Training set and Test set using various training partitions of 40%, 50%, 60%, 70% and 80% for the file “ClassificationSimplecases.csv” using C5.0 (Please don’t enter the values in the table manually/directly).

2. Repeat 1(a) using Naïve Bayes classifier and compare the performance with those of decision tree.

Hints:

library(caret)

library(e1071)

#Read the file

#delete the RID column

nbmod <- naiveBayes(buys ~ ., data = bh[t.idx,])

…..

3. The following code for k-NN Classifier may not work. Either correct it or write your own code for k-NN classifier.

library(class)

library(e1071)

library(caTools)

#function to find the Euclidean distance

eucludian<-function(p1,p2){

dist1 = sqrt(sum((p1-p2)\*\*2))

return(dist1)

}

#knn function

knn1<-function(td,tp,k){

for(i in 1:nrow(td)){

td$distances[i]=eucludian(td[i,1:4],tp)

}

td=td[order(td$distance),]

df=td[1:k,5]

return(as.character(names(which.max(table(df)))))

}

knn2<-function(td,tp,k){

for(i in 1:nrow(tp)){

tp$pred[i]=knn1(td,tp[i,c(1:4)],k)

}

print(tp)

misClassError <- mean(tp$pred != tp$Species)

print(paste('Accuracy =', 1-misClassError))

cm <- table(test\_cl$Species, classifier\_knn)

cm

}

d=iris

split <- sample.split(d$Species, SplitRatio = 0.7)

train\_cl <- subset(d, split == "TRUE")

test\_cl <- subset(d, split == "FALSE")

train\_cl

test\_cl

print(nrow(train\_cl))

print(nrow(test\_cl))

knn2(train\_cl,test\_cl,3)

4. Change and Run the code for the file “RidingMowers larger dataset.xls” (attached) for k=3 and generate the Confusion matrix.