Decision tree for the irish data

library(party)

## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

data(iris)  
head(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

str(iris)

## 'data.frame': 150 obs. of 5 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

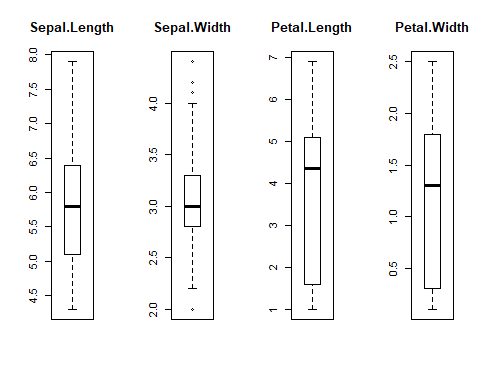
pairs(iris)



sum(is.na(iris))

## [1] 0

par(mfrow=c(1,4))  
for(i in 1:4){  
 boxplot(iris[i], main=colnames(iris)[i])  
}



## data partiotioning  
intraininglocal<-createDataPartition(iris$Species, p=0.75, list=F)  
train<-iris[intraininglocal,]  
dim(train)

## [1] 114 5

test<-iris[-intraininglocal,]  
dim(test)

## [1] 36 5

## model bulding  
model<-ctree(train$Species~., data=train)  
summary(model)

## Length Class Mode   
## 1 BinaryTree S4

pred<-predict(model, train)  
a=table(train$Species, pred)  
a

## pred  
## setosa versicolor virginica  
## setosa 38 0 0  
## versicolor 0 37 1  
## virginica 0 3 35

##Accuracy  
mean(pred==train$Species)\*100

## [1] 96.49123

## finding predicted values and confution matrix using test data  
pred\_val<-predict(model, newdata=test)  
confusionMatrix(test$Species, pred\_val)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction setosa versicolor virginica  
## setosa 12 0 0  
## versicolor 0 11 1  
## virginica 0 1 11  
##   
## Overall Statistics  
##   
## Accuracy : 0.9444   
## 95% CI : (0.8134, 0.9932)  
## No Information Rate : 0.3333   
## P-Value [Acc > NIR] : 1.728e-14   
##   
## Kappa : 0.9167   
##   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: setosa Class: versicolor Class: virginica  
## Sensitivity 1.0000 0.9167 0.9167  
## Specificity 1.0000 0.9583 0.9583  
## Pos Pred Value 1.0000 0.9167 0.9167  
## Neg Pred Value 1.0000 0.9583 0.9583  
## Prevalence 0.3333 0.3333 0.3333  
## Detection Rate 0.3333 0.3056 0.3056  
## Detection Prevalence 0.3333 0.3333 0.3333  
## Balanced Accuracy 1.0000 0.9375 0.9375

plot(model)

