

## EXPERIMENT 13

**Problem Statement:** Implement decision trees using 'iris' dataset using package party and 'rpart'.

**Iris dataset:** The Iris Dataset contains four features (length and width of sepals and petals) of 50 samples of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). ... The dataset is often used in data mining, classification and clustering examples and to test algorithms.

**Party package:** A Laboratory for Recursive Partytioning

A computational toolbox for recursive partitioning. The core of the package is ctree(), an implementation of conditional inference trees which embed tree-structured regression models into a well defined theory of conditional inference procedures.

**Rpart package:** Rpart is a powerful machine learning library in R that is used for building classification and regression trees. This library implements recursive partitioning and is very easy to use.

Setting a seed in R means to initialize a pseudorandom number generator. Most of the simulation

Methods in satatistics require the possibility to generate pseudorandam numbers that mimic the

Properties of independent generations of a uniform distribution in the interval (0,1) (0,1) (0,1).

**Plot(model):** To use R's regression diagnostic plots, we set up the regression model as an object and create a plotting environment of two rows and two columns. Then we use the plot() command, treating the model as an argument.

**predict(model, newdata, "class"):** This function enables us to predict the labels of the data values on the basis of the trained model.

**Code:**

```
str(iris)
set.seed(1234)
ind <- sample(2, nrow(iris), replace=TRUE, prob=c(0.7, 0.3))
print(ind)
trainData <- iris[ind==1,]
```

```

testData <- iris[ind==2,]

print(testData)

library(party)
myFormula <- Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width
iris_ctree <- ctree(myFormula, data=trainData)
table(predict(iris_ctree, newdata=testData), testData$Species)

print(iris_ctree)

plot(iris_ctree)

library(rpart)

model <- rpart(Species ~., data = iris)
par(xpd = NA) # otherwise on some devices the text is clipped
plot(model)
text(model, digits = 3)
print(model, digits = 2)
newdata <- data.frame(
  Sepal.Length = 6.5, Sepal.Width = 3.0,
  Petal.Length = 5.2, Petal.Width = 2.0
)

predict(model, newdata, "class")

```

### Output:

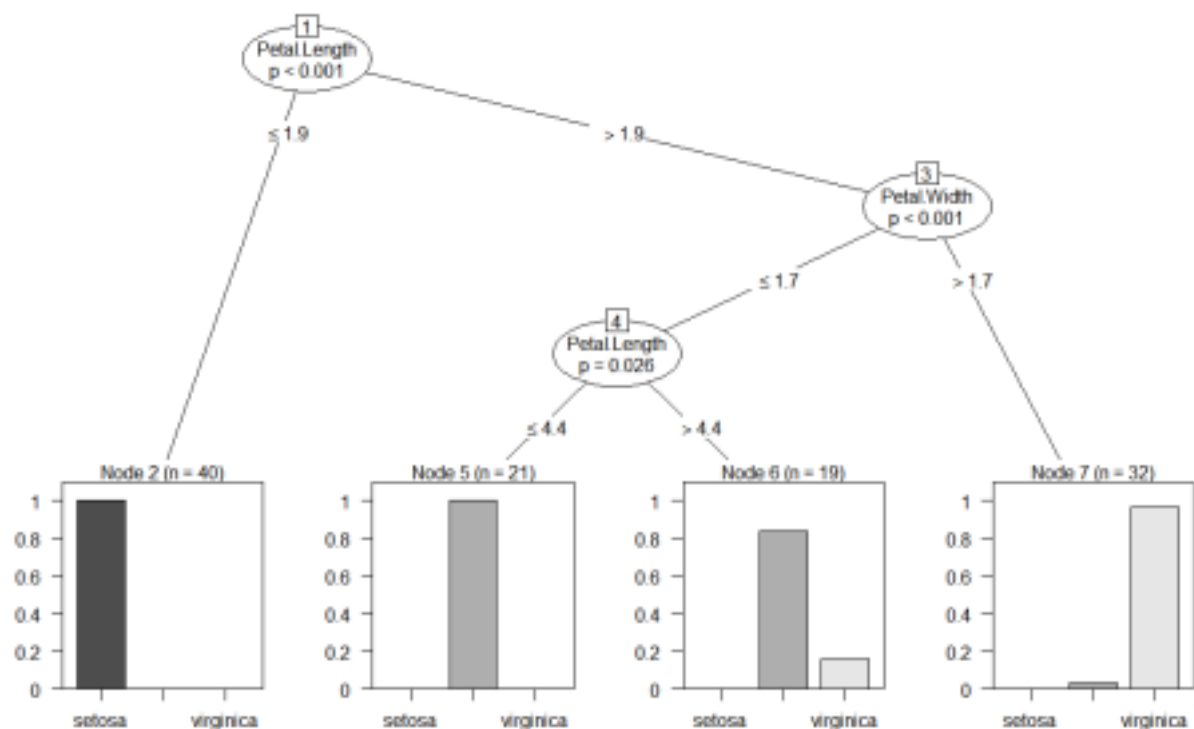
```

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 ...

```

```
> print(ind)
[1] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 2 1 1 2
[40] 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1
[79] 1 1 2 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 2
[118] 1 1 2 2 2 2 2 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 2 1 2 1 1 1 1 2 1 2 1

> print(testData)
  sepal.Length sepal.width Petal.Length Petal.width species
5           5.0         3.6         1.4         0.2   setosa
14          4.3         3.0         1.1         0.1   setosa
16          5.7         4.4         1.5         0.4   setosa
26          5.0         3.0         1.6         0.2   setosa
28          5.2         3.5         1.5         0.2   setosa
29          5.2         3.4         1.4         0.2   setosa
36          5.0         3.2         1.2         0.2   setosa
39          4.4         3.0         1.3         0.2   setosa
40          5.1         3.4         1.5         0.2   setosa
50          5.0         3.3         1.4         0.2   setosa
53          6.9         3.1         4.9         1.5 versicolor
```



```
1) root 150 100 setosa (0.333 0.333 0.333)
2) Petal.Length < 2.5 50 0 setosa (1.000 0.000 0.000) * 3)
Petal.Length >= 2.5 100 50 versicolor (0.000 0.500 0.500) 6) Petal.Width <
1.8 54 5 versicolor (0.000 0.907 0.093) * 7) Petal.Width >= 1.8 46 1
virginica (0.000 0.022 0.978) * 1
virginica
Levels: setosa versicolor virginica
```

