Decision Tree with the Iris Data

Karen Mazidi

Using rpart

```
library(rpart)
tree_iris <- rpart(Species~., data=iris, method="class")
tree_iris
```
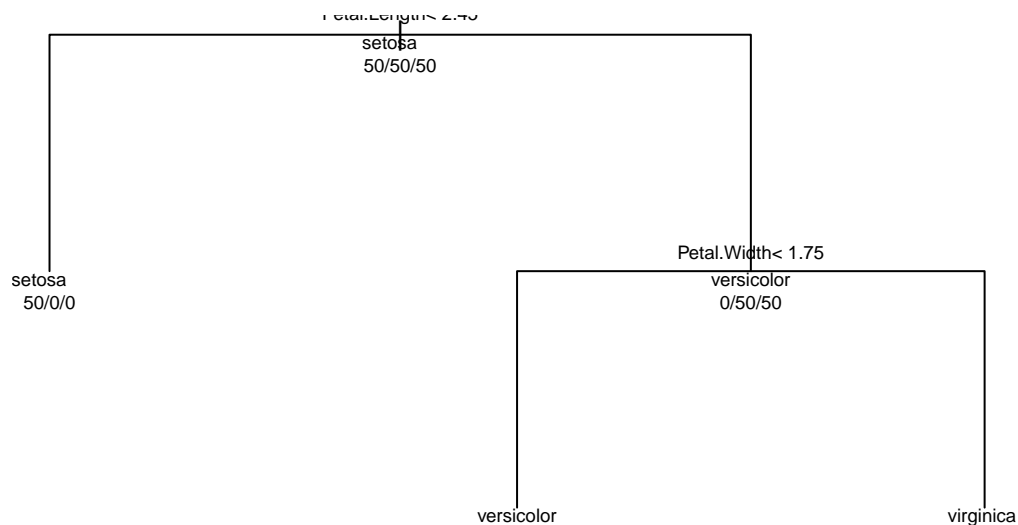
```
## n= 150
##
## node), split, n, loss, yval, (yprob)
##        * denotes terminal node
##
## 1) root 150 100 setosa (0.33333333 0.33333333 0.33333333)
##   2) Petal.Length< 2.45 50    0 setosa (1.00000000 0.00000000 0.00000000) *
##   3) Petal.Length>=2.45 100  50 versicolor (0.00000000 0.50000000 0.50000000)
##     6) Petal.Width< 1.75 54   5 versicolor (0.00000000 0.90740741 0.09259259) *
##     7) Petal.Width>=1.75 46   1 virginica (0.00000000 0.02173913 0.97826087) *
```

```
summary(tree_iris)
```

```
## Call:
## rpart(formula = Species ~ ., data = iris, method = "class")
##   n= 150
##
##      CP nsplit rel error xerror       xstd
## 1 0.50      0      1.00   1.14 0.05230679
## 2 0.44      1      0.50   0.67 0.06088788
## 3 0.01      2      0.06   0.10 0.03055050
##
## Variable importance
##  Petal.Width Petal.Length Sepal.Length  Sepal.Width
##           34           31           21           14
##
## Node number 1: 150 observations,    complexity param=0.5
##   predicted class=setosa      expected loss=0.6666667  P(node) =1
##     class counts:    50    50    50
##    probabilities: 0.333 0.333 0.333
##   left son=2 (50 obs) right son=3 (100 obs)
##   Primary splits:
##       Petal.Length < 2.45 to the left,  improve=50.00000, (0 missing)
##       Petal.Width  < 0.8  to the left,  improve=50.00000, (0 missing)
##       Sepal.Length < 5.45 to the left,  improve=34.16405, (0 missing)
##       Sepal.Width  < 3.35 to the right, improve=19.03851, (0 missing)
##   Surrogate splits:
##       Petal.Width  < 0.8  to the left,  agree=1.000, adj=1.00, (0 split)
##       Sepal.Length < 5.45 to the left,  agree=0.920, adj=0.76, (0 split)
##       Sepal.Width  < 3.35 to the right, agree=0.833, adj=0.50, (0 split)
##
## Node number 2: 50 observations
##   predicted class=setosa      expected loss=0  P(node) =0.3333333
```

```
##       class counts:     50      0      0
##    probabilities: 1.000 0.000 0.000
##
## Node number 3: 100 observations,    complexity param=0.44
##   predicted class=versicolor  expected loss=0.5  P(node) =0.6666667
##       class counts:      0     50     50
##    probabilities: 0.000 0.500 0.500
##   left son=6 (54 obs) right son=7 (46 obs)
##   Primary splits:
##       Petal.Width  < 1.75 to the left,  improve=38.969400, (0 missing)
##       Petal.Length < 4.75 to the left,  improve=37.353540, (0 missing)
##       Sepal.Length < 6.15 to the left,  improve=10.686870, (0 missing)
##       Sepal.Width  < 2.45 to the left,  improve= 3.555556, (0 missing)
##   Surrogate splits:
##       Petal.Length < 4.75 to the left,  agree=0.91, adj=0.804, (0 split)
##       Sepal.Length < 6.15 to the left,  agree=0.73, adj=0.413, (0 split)
##       Sepal.Width  < 2.95 to the left,  agree=0.67, adj=0.283, (0 split)
##
## Node number 6: 54 observations
##   predicted class=versicolor  expected loss=0.09259259  P(node) =0.36
##       class counts:      0     49      5
##    probabilities: 0.000 0.907 0.093
##
## Node number 7: 46 observations
##   predicted class=virginica   expected loss=0.02173913  P(node) =0.3066667
##       class counts:      0      1     45
##    probabilities: 0.000 0.022 0.978
```

```r
plot(tree_iris, uniform=TRUE)
text(tree_iris, use.n=TRUE, all=TRUE, cex=.6)
```


Using tree

```r
library(tree)
tree_iris2 <- tree(Species~., data=iris)
tree_iris2
```

```
## node), split, n, deviance, yval, (yprob)
##       * denotes terminal node
##
##  1) root 150 329.600 setosa ( 0.33333 0.33333 0.33333 )
##    2) Petal.Length < 2.45 50   0.000 setosa ( 1.00000 0.00000 0.00000 ) *
##    3) Petal.Length > 2.45 100 138.600 versicolor ( 0.00000 0.50000 0.50000 )
##      6) Petal.Width < 1.75 54  33.320 versicolor ( 0.00000 0.90741 0.09259 )
##       12) Petal.Length < 4.95 48   9.721 versicolor ( 0.00000 0.97917 0.02083 )
##         24) Sepal.Length < 5.15 5   5.004 versicolor ( 0.00000 0.80000 0.20000 ) *
##         25) Sepal.Length > 5.15 43   0.000 versicolor ( 0.00000 1.00000 0.00000 ) *
##       13) Petal.Length > 4.95 6   7.638 virginica ( 0.00000 0.33333 0.66667 ) *
##      7) Petal.Width > 1.75 46   9.635 virginica ( 0.00000 0.02174 0.97826 )
##       14) Petal.Length < 4.95 6   5.407 virginica ( 0.00000 0.16667 0.83333 ) *
##       15) Petal.Length > 4.95 40   0.000 virginica ( 0.00000 0.00000 1.00000 ) *
```

```r
summary(tree_iris2)
```

```
##
## Classification tree:
## tree(formula = Species ~ ., data = iris)
## Variables actually used in tree construction:
## [1] "Petal.Length" "Petal.Width"  "Sepal.Length"
## Number of terminal nodes:  6
## Residual mean deviance:  0.1253 = 18.05 / 144
## Misclassification error rate: 0.02667 = 4 / 150
```

```r
plot(tree_iris2)
text(tree_iris2, cex=0.5, pretty=0)
```
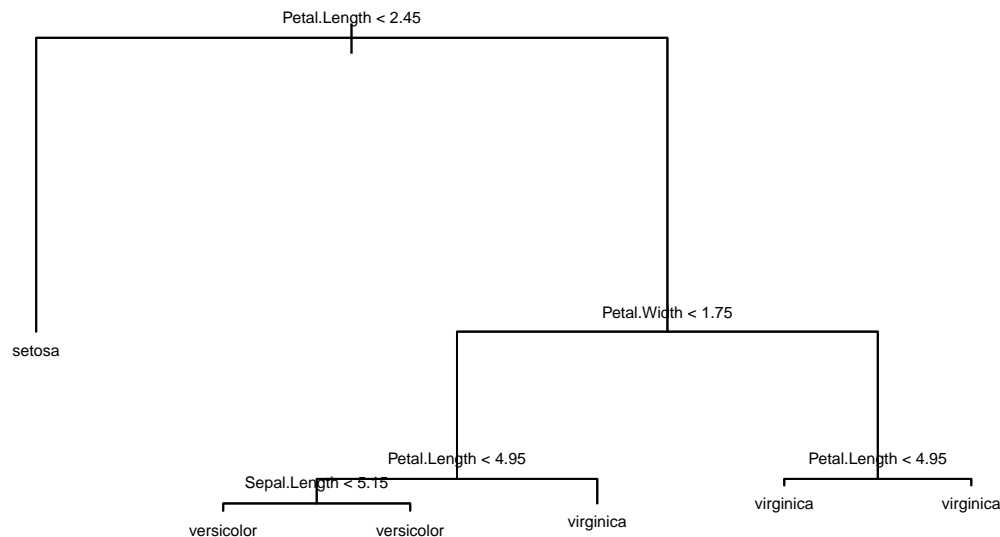

train and test

```r
set.seed(1958)
i <- sample(150, 100, replace=FALSE)
train <- iris[i,]
test <- iris[-i,]
tree_iris3 <- tree(Species~., data=train)
```

```
pred <- predict(tree_iris3, newdata=test, type="class")
table(pred, test$Species)
```

```
##
## pred         setosa versicolor virginica
##    setosa        16          0         0
##    versicolor     0         15         3
##    virginica      0          0        16
```

```
mean(pred==test$Species)
```

```
## [1] 0.94
```