Decision tree

2022-11-04

```r
data(iris) #loading in built dataset Iris
summary(iris)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
## Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
## 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
## Median :5.800   Median :3.000   Median :4.350   Median :1.300
## Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
## 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
## Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##       Species
## setosa    :50
## versicolor:50
## virginica :50
##
##
##
```

```r
nrow <- nrow(iris); ncol <- ncol(iris)  #inline code
iris[1:4,]  # View first four rows of data set
```

```
##   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
```

#Load required package:

```r
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.1.3
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```r
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------ tidyverse 1.3.1 --
```

```
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## v purrr   0.3.4
```

```
## Warning: package 'tibble' was built under R version 4.1.3
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
```

```r
modFit <- train(Species ~., method = "rpart", data=iris) #Fit model
print(modFit$finalModel)    #Summarize model
```

```
## 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) *
```
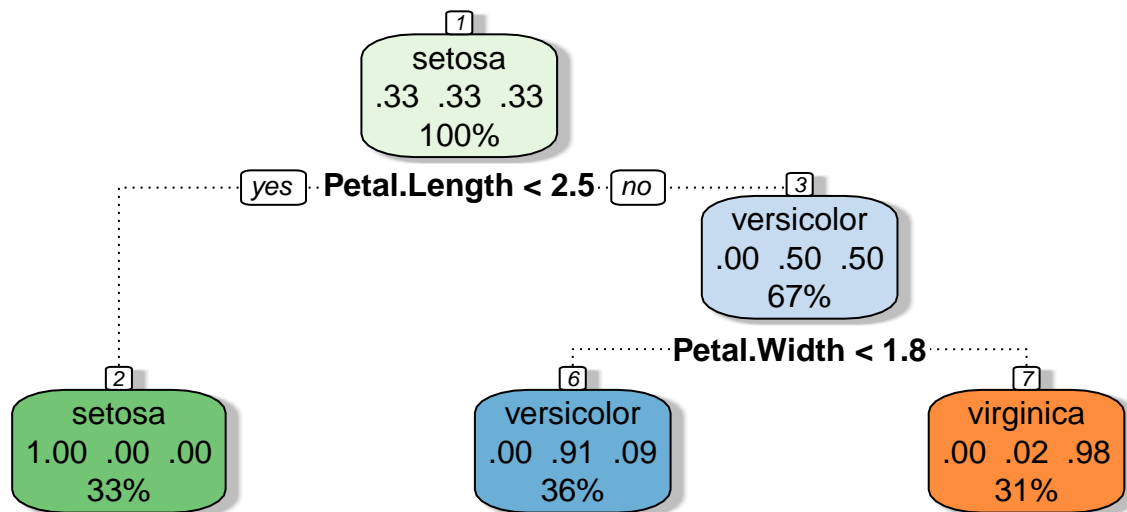
```r
library(rattle)
```

```
## Warning: package 'rattle' was built under R version 4.1.3
```

```
## Loading required package: bitops
```

```
## Rattle: A free graphical interface for data science with R.
## Version 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```r
fancyRpartPlot(modFit$finalModel) # Plot decision tree
```

```
                    ┌─┐
                    │1│
                  setosa
              .33  .33  .33
                   100%

    ┌yes┐  Petal.Length < 2.5  ┌no┐        ┌─┐
                                            │3│
                                        versicolor
                                      .00  .50  .50
                                           67%

                          Petal.Width < 1.8

┌─┐                    ┌─┐                        ┌─┐
│2│                    │6│                        │7│
setosa              versicolor                 virginica
1.00  .00  .00      .00  .91  .09           .00  .02  .98
    33%                  36%                     31%
```

Rattle 2022−Nov−04 19:18:52 Dell