Machine Learning

Ravi Kumar Tiwari

14 June 2016

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.2.5
```

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

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

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 3.2.5
```

```
## Loading required package: rpart
```

```
library(rattle)
```

```
## Warning: package 'rattle' was built under R version 3.2.5
```

```
## Rattle: A free graphical interface for data mining with R.
## Version 4.1.0 Copyright (c) 2006-2015 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```
nrow(iris)
```

```
## [1] 150
```

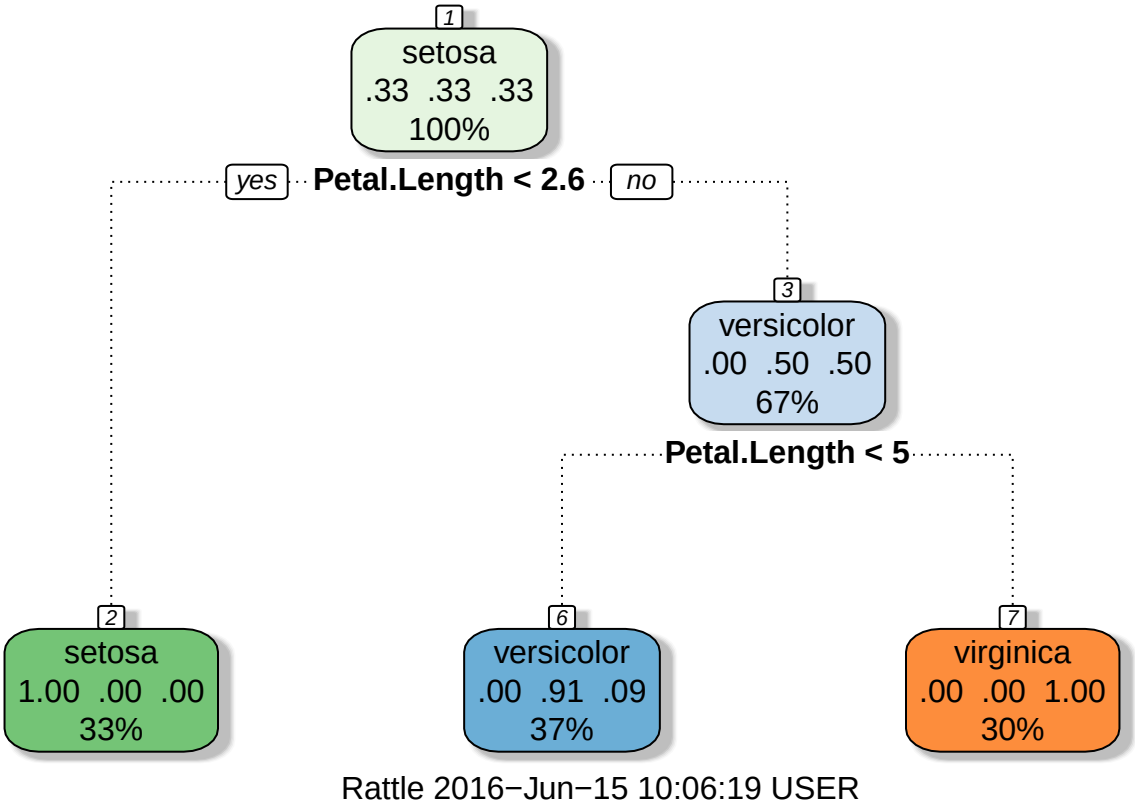
```
inTrain <- createDataPartition(iris$Species, p = 0.6, list = FALSE)
trainData <- iris[inTrain,]
testData <- iris[-inTrain,]
```

```
treeModel <- train(Species ~ ., data = trainData, method = "rpart")
preClass <- predict(treeModel, newdata = testData)
confusionMatrix(preClass, testData$Species)
```

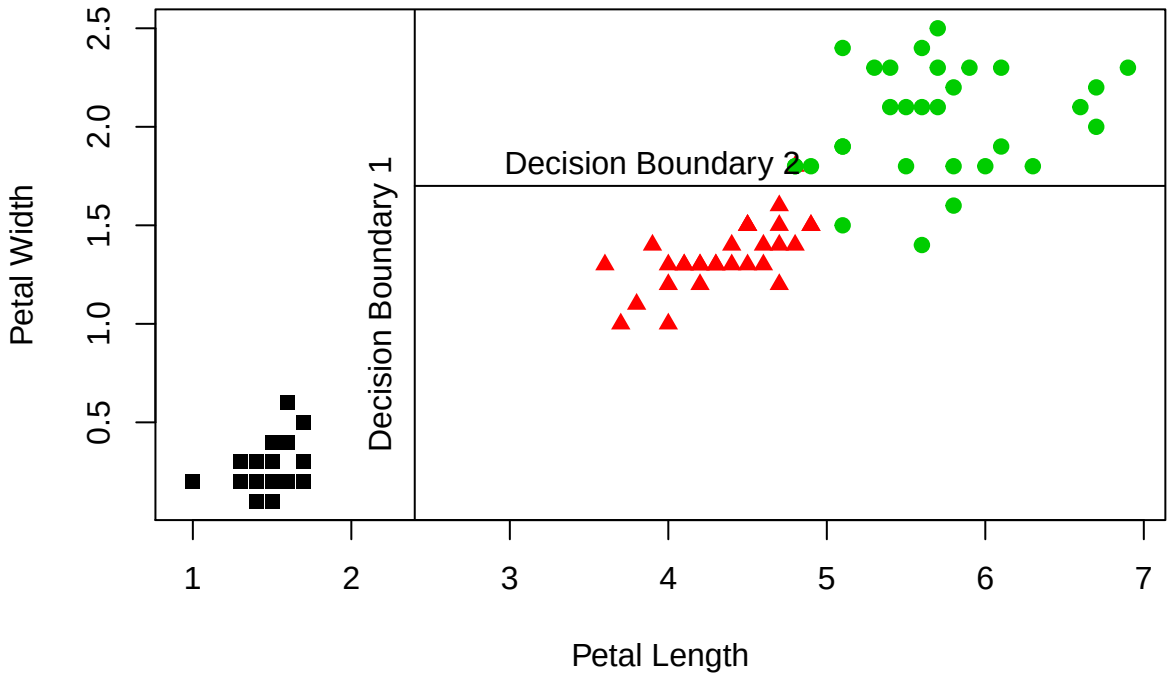
```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   setosa versicolor virginica
##   setosa         20          0         0
##   versicolor      0         18         3
```

```
##   virginica       0          2        17
##
## Overall Statistics
##
##                Accuracy : 0.9167
##                  95% CI : (0.8161, 0.9724)
##     No Information Rate : 0.3333
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.875
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: setosa Class: versicolor Class: virginica
## Sensitivity                 1.0000            0.9000           0.8500
## Specificity                 1.0000            0.9250           0.9500
## Pos Pred Value              1.0000            0.8571           0.8947
## Neg Pred Value              1.0000            0.9487           0.9268
## Prevalence                  0.3333            0.3333           0.3333
## Detection Rate              0.3333            0.3000           0.2833
## Detection Prevalence        0.3333            0.3500           0.3167
## Balanced Accuracy           1.0000            0.9125           0.9000
```

```
fancyRpartPlot(treeModel$finalModel)
```



```
preClassType <- ifelse(trainData$Species == "setosa", 15,
                   ifelse(trainData$Species== "versicolor", 17, 19))
plot(trainData$Petal.Length, trainData$Petal.Width, col = trainData$Species,
     pch = preClassType, xlab = "Petal Length", ylab = "Petal Width")
abline(v = 2.4)
segments(2.4, 1.7, 7.2, 1.7 )
text(x = 2.2, y = 1.1, labels = "Decision Boundary 1", srt = 90)
text(x = 3.9, y = 1.8, labels = "Decision Boundary 2")
```



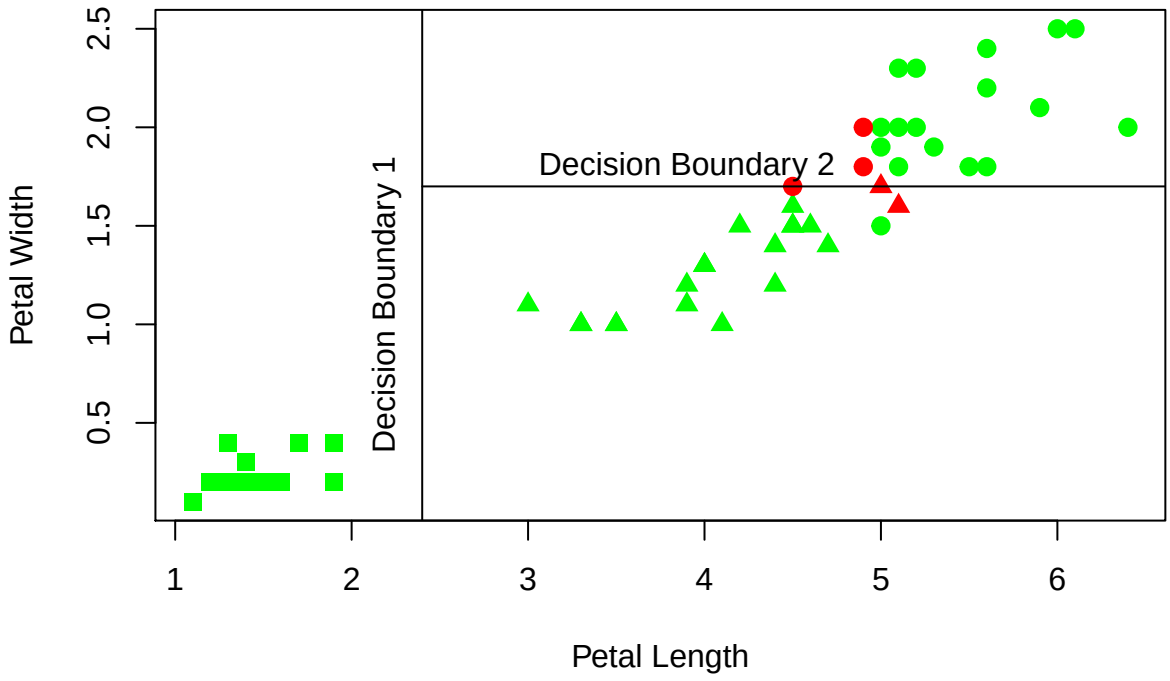

```
varImp(treeModel)
```

```
## rpart variable importance
##
##              Overall
## Petal.Length  100.00
## Petal.Width    95.22
## Sepal.Length   41.91
## Sepal.Width     0.00
```

```
predResult <- ifelse(preClass == testData$Species, "green", "red")

preClass <- ifelse(testData$Species == "setosa", 15,
                   ifelse(testData$Species== "versicolor", 17, 19))
```

```
plot(testData$Petal.Length, testData$Petal.Width, col = predResult,
     pch = preClass, cex = 1.2, xlab = "Petal Length", ylab = "Petal Width")
abline(v = 2.4)
segments(2.4, 1.7, 7.2, 1.7 )
text(x = 2.2, y = 1.1, labels = "Decision Boundary 1", srt = 90)
text(x = 3.9, y = 1.8, labels = "Decision Boundary 2")
```



Cross-validation

```
drawPoly <- function(x,y, col){
  polygon(c(x,x+1,x+1,x, x), c(y,y,y+0.5,y+0.5,y), col = col)
}


plot(1, type="n", axes=F, xlab="", ylab="", xlim = c(0,6), ylim= c(0,8),
     main = "5 fold cross validation illustration")
for (i in 1:5){
  for (j in 1:5) {
    col <- ifelse(i==j, "red", "green")
    drawPoly(i,j, col)
  }
}
```

```r
drawPoly(1,6, "red")
drawPoly(1,7, "green")

for(i in 1:5){
  text(i+0.5, 0.5, label = paste("k = ", i, sep = ""))
}

for(i in 1:5){
  text(0.5, i+0.25, label = i)
}

text(3, 7.3, label = "Training Data")
text(3, 6.3, label = "Test Data")
```