MultiClass Classification

Karen Mazidi

In this notebook we look at multi-class classification with the iris data set, built into R. This is a well-known data set that takes measurements from 150 irises, equally divided into 3 species virginica, setosa and versicolor.

Data exploration

Explore the data with functions and graphics.

```
attach(iris)
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[1:4], pch = 21, bg = c("red", "yellow", "blue")[unclass(Species)])
```
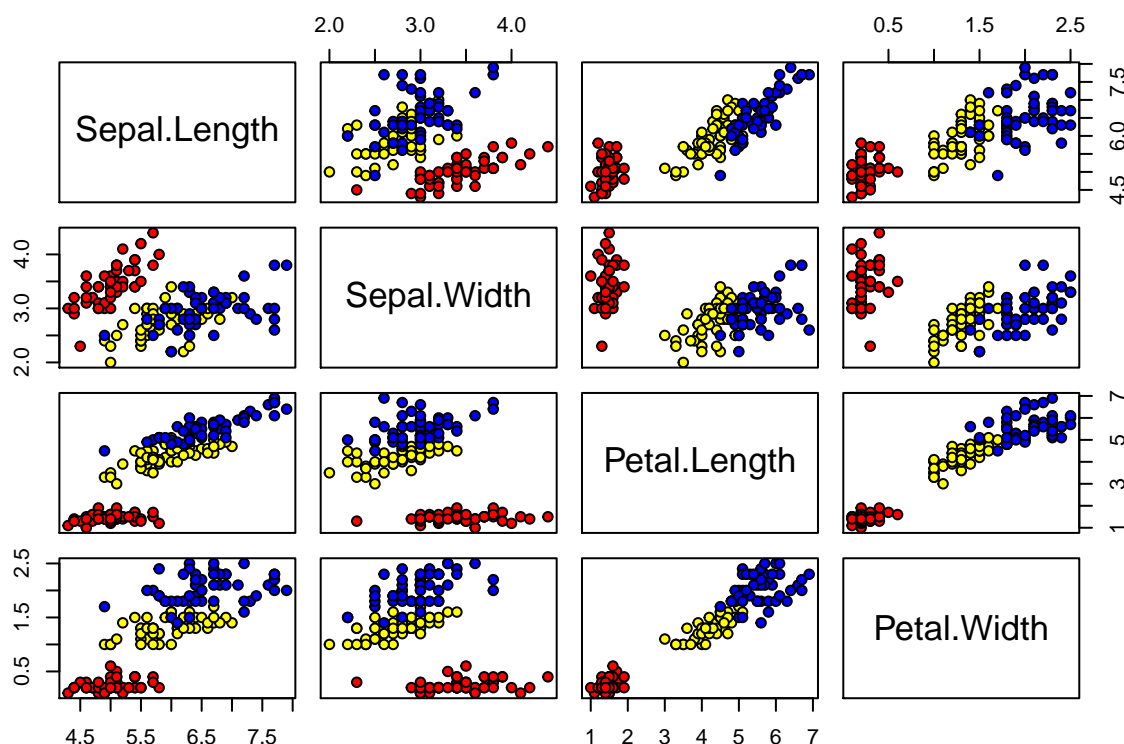

Let's see how well Petal.Length and Petal.Width separate the classes.

```
plot(Petal.Length, Petal.Width, pch=21, bg=c("red","yellow","blue")
     [unclass(Species)])
```
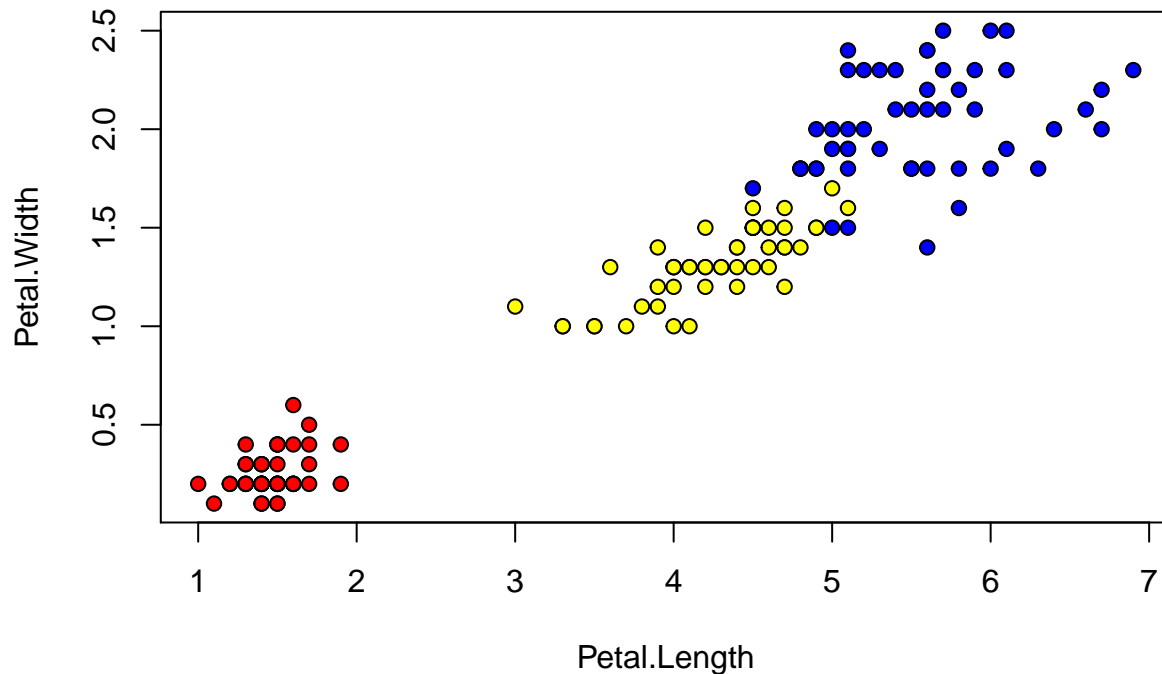
One versus all

In one versus all classification we will build 3 classifiers on 3 data sets:

- virginica versus not
- setosa versus not
- versicolor versus not

```r
# reclassify as virginica or not
iris_virginica <- iris
iris_virginica$Species <- as.factor(ifelse (iris_virginica$Species=="virginica",1,0))

# reclassify as setosa or not
iris_setosa <- iris
iris_setosa$Species <- as.factor(ifelse (iris_setosa$Species=="setosa",1,0))

# reclassify as versicolor or not
iris_versicolor <- iris
iris_versicolor$Species <- as.factor(ifelse (iris_versicolor$Species=="versicolor",1,0))
```

Function for logistic regression

We will write a function to handle repeated calls.

```r
fun <- function(df, i){
    train <- df[i,]
    test <- df[-i,]
    glm1 <- glm(Species~., data=train, family="binomial")
    probs <- predict(glm1, newdata=test)
    pred <- ifelse(probs>0.5, 1, 0)
    acc <- mean(pred==test$Species)
    print(paste("accuracy = ", acc))
```

```
    table(pred, test$Species)
}
```

Virginica

```
set.seed(1234)
i <- sample(1:150, 100, replace=FALSE)
fun(iris_virginica, i)
```

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## [1] "accuracy =  0.98"

##
## pred  0  1
##    0 32  0
##    1  1 17
```

Setosa

```
fun(iris_setosa, i)
```

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## [1] "accuracy =  1"

##
## pred  0  1
##    0 38  0
##    1  0 12
```

Versicolor

```
fun(iris_versicolor, i)
```

```
## [1] "accuracy =  0.62"

##
## pred  0  1
##    0 23 13
##    1  6  8
```