# Data 609 - HW 7

Avery Davidowitz

2023-04-26

### Import

```
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.2.3
```

### Ex 1

Use the svm() algorithm of the e1071 package to carry out the support vector machine for the PlantGrowth data set.

```
svm_model <- svm(group ~ weight, data = PlantGrowth)

summary(svm_model)
```

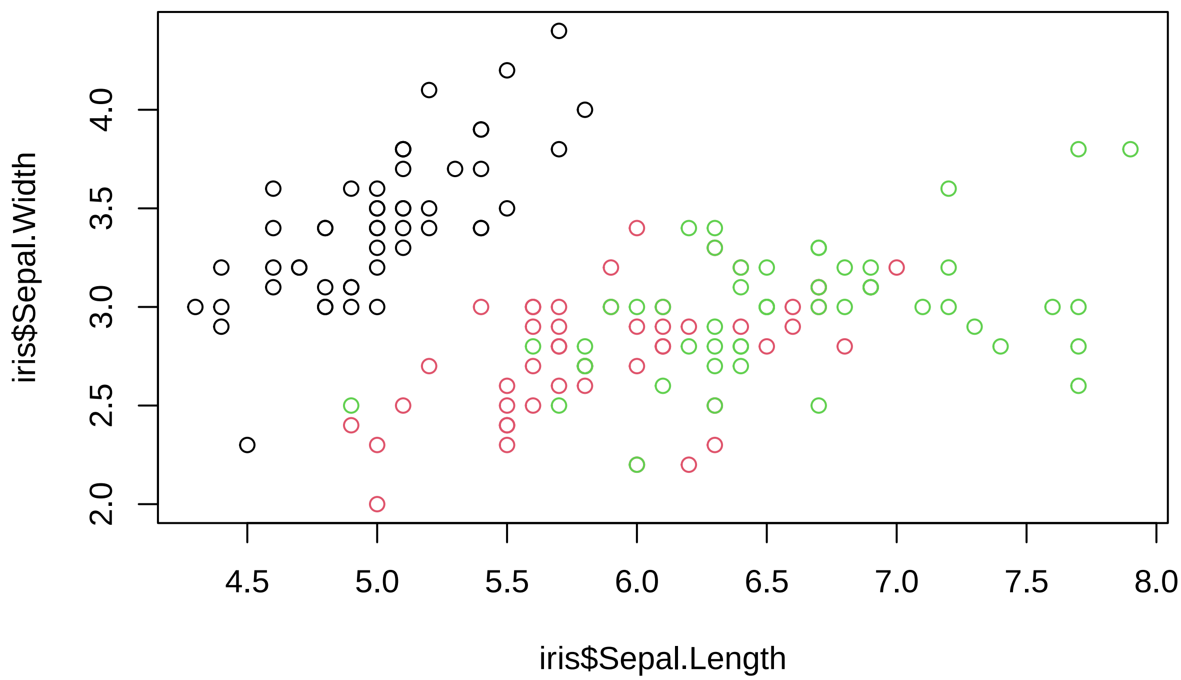
```
##
## Call:
## svm(formula = group ~ weight, data = PlantGrowth)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  radial
##        cost:  1
##
## Number of Support Vectors:  29
##
##  ( 10 9 10 )
##
##
## Number of Classes:  3
##
## Levels:
##  ctrl trt1 trt2
```

There are 29 support vectors used to classify the PlantGrowth data set out of the 30 data points.

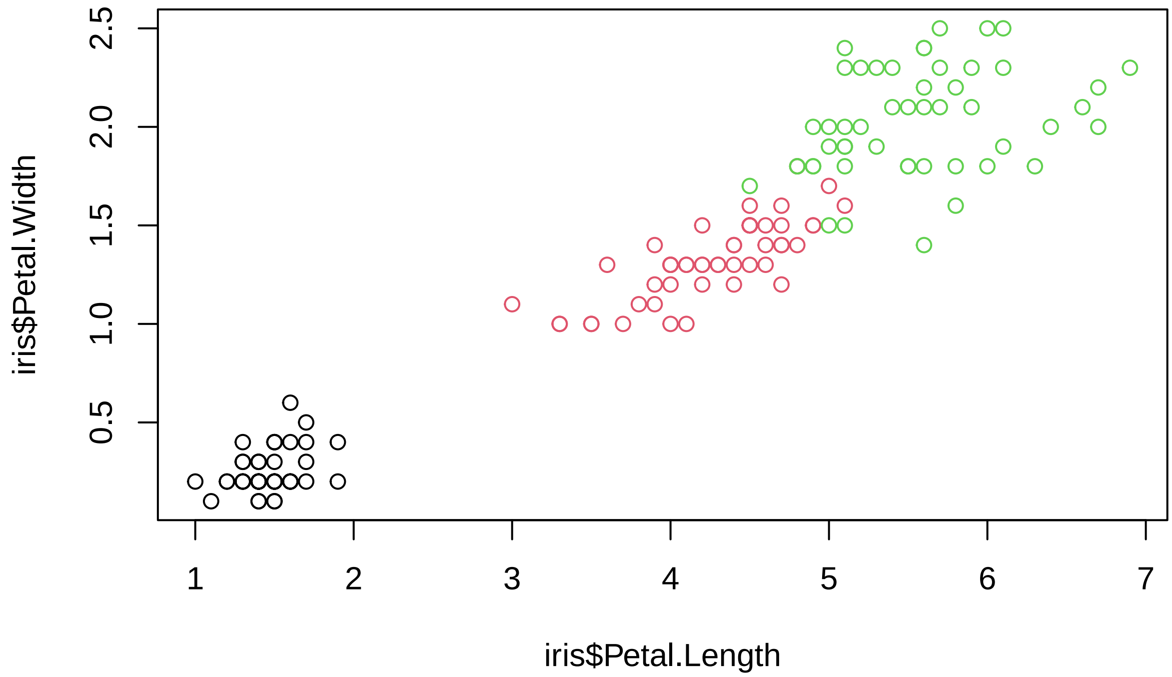
### Ex 2

Do a similar SVM analysis as that in the previous question using the iris data set. Discuss the number of support vectors/samples.

```
data(iris)

plot(iris$Sepal.Length, iris$Sepal.Width, col=iris$Species)
```




```
plot(iris$Petal.Length, iris$Petal.Width, col=iris$Species)
```




```
col<-c("Petal.Length", "Petal.Width", "Species")
irisubset <- iris[,col]
```

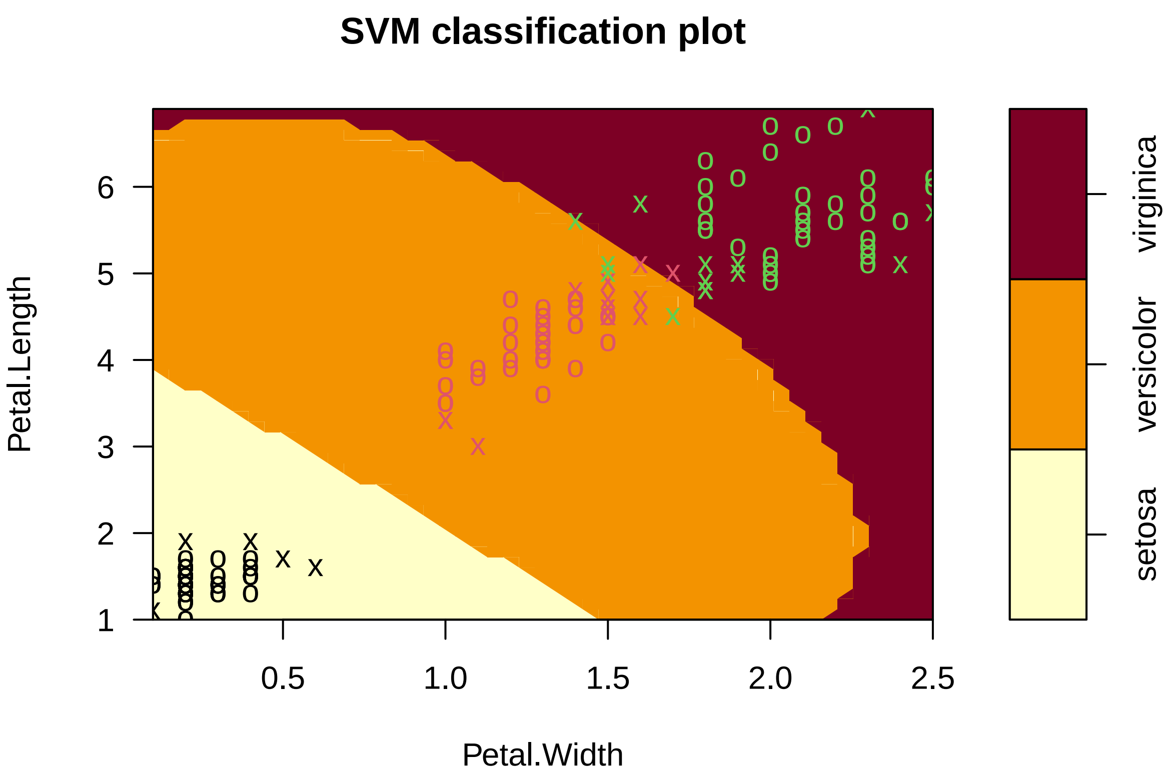
Looking at plots of the features it seems petal length and width will be much more useful to classify the data since they are already in visible groupings. Therefore I subset the data set.

```
svm2 <- svm(Species~., data = irisubset)
summary(svm2)
```

```
##
## Call:
## svm(formula = Species ~ ., data = irisubset)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  radial
##        cost:  1
##
## Number of Support Vectors:  37
##
##  ( 5 16 16 )
##
##
## Number of Classes:  3
```

```
##
## Levels:
##  setosa versicolor virginica
```

```
plot(svm2, irisubset)
```




This SVM generates 37 support vectors. This means that of the 120 points 37 are used in calculating the margins. The plot still shows a very nice job of classifying. Can we do better though?

### Ex 3

Use the iris data set to select 80% of the sameples for training svm(), the use the rest of the 20% for validation. Discuss your results.

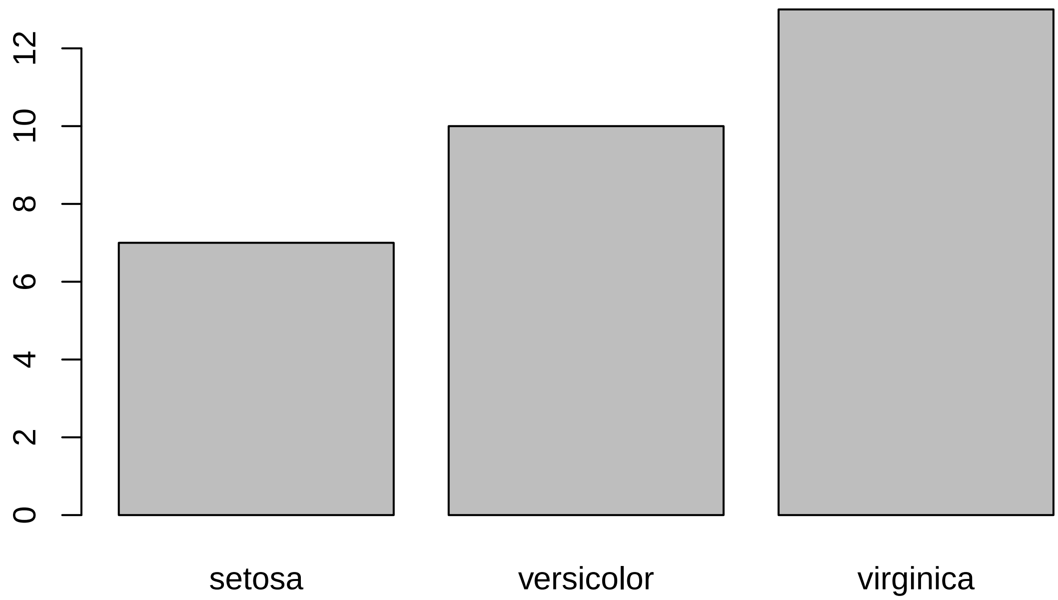
```
s<-sample(150, 150 * .8)
iris_train<-irisubset[s,]
iris_test<-irisubset[-s,]

svm3 <- svm(Species ~., data = iris_train, kernel = "linear", scale = FALSE)
print(svm3)
```

```
##
## Call:
## svm(formula = Species ~ ., data = iris_train, kernel = "linear",
##     scale = FALSE)
```

```
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
##        cost:  1
##
## Number of Support Vectors:  22
```

```
plot(svm3, iris_train)
```


```
iris_predict <- predict(svm3, iris_test[,col], type="class")
plot(iris_predict)
```




```
table(iris_predict, iris_test[,3])
```

```
##
## iris_predict setosa versicolor virginica
##   setosa          7          0         0
##   versicolor      0         10         0
##   virginica       0          2        11
```

```
mean(iris_predict== iris_test[,3])
```

```
## [1] 0.9333333
```

Splitting up into training and test shows that this data set is well suited for generalizing the SVM classifications with 93% accuracy.