

Support Vector Machine (SVM)

Eduardo Abdala

5/7/2020

SVM is used when you have binary linearly separable data. The target vector can be either 1 or -1. It uses linear regression concepts to draw frontiers so the machine can classify data.

The frontiers take the form $f(x) = \beta_0 + x^T \beta = 0$.

If you know the sign of the above function, you can predict the response variable.

This gives you $G(x) = \text{sign}(\beta_0 + x^T \beta)$. The goal is to maximize the distance from the separating line to the margin boundary, which is the shortest distance from the boundary line to a point. Alternatively, you can minimize the norm of the coefficients: $\min \|B\|$, subject to $y_i(\beta_0 + x_i^T \beta) \geq 1, i = 1, \dots, N$.

```
# iris <- read.csv("~/DATA 4319/iris_data.csv")

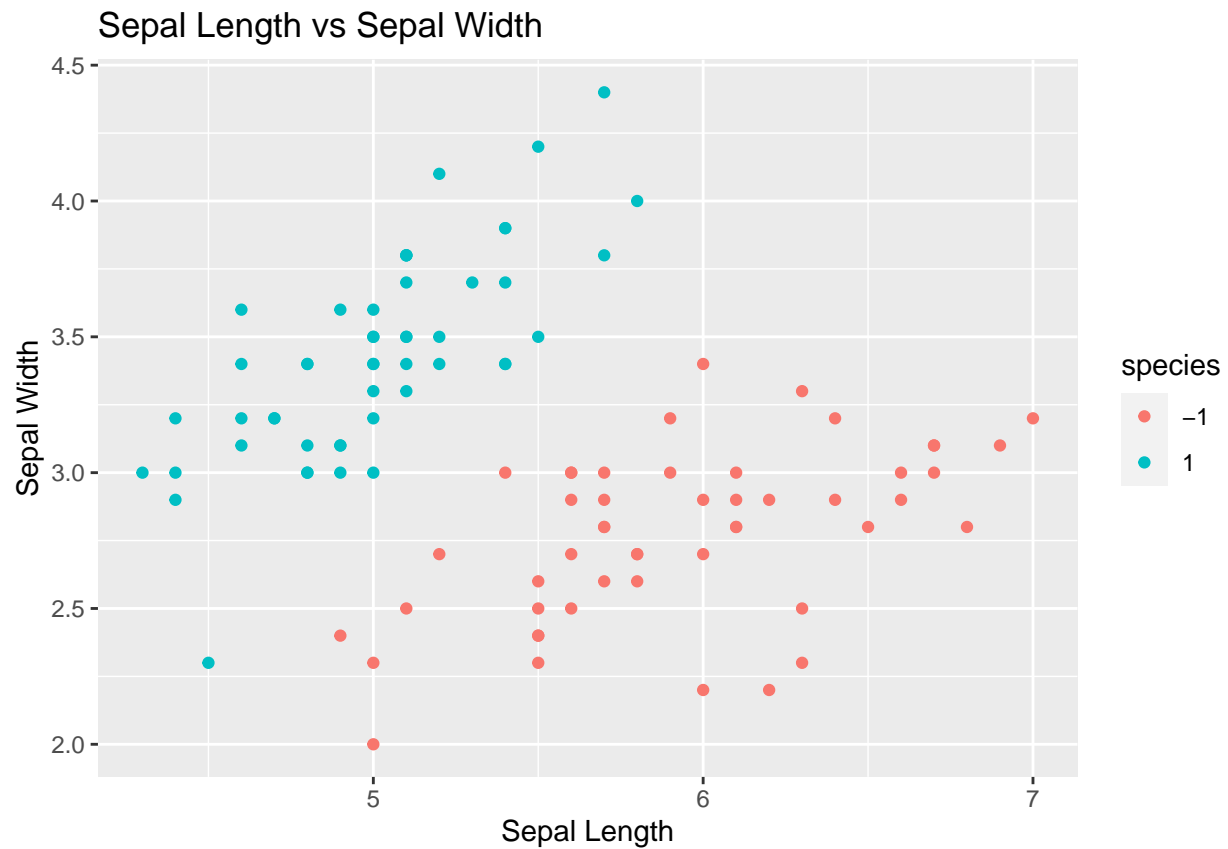
# # removing third species
iris <- iris[1:100,]
# # removing petal length and width
drops <- c("Petal.Length", "Petal.Width")
iris <- iris[ , !(names(iris) %in% drops)]

# Renaming columns and creating vectors

Species <- as.factor(ifelse(Species=="setosa", 1, -1))
species <- Species[1:100]
slength <- Sepal.Length[1:100]
swidth <- Sepal.Width[1:100]
# plength <- Petal.Length
# pwidth <- Petal.Width
new <- data.frame(species, slength, swidth)

svm1 <- ksvm(species~slength+swidth, data = new, kernel="vanilladot", cross = 5)

## Setting default kernel parameters
```



```
preds1<-predict(svm1,newdata=new) # Fitted values
summary(preds1) # Standard output
```

```
## -1  1
## 51 49
```

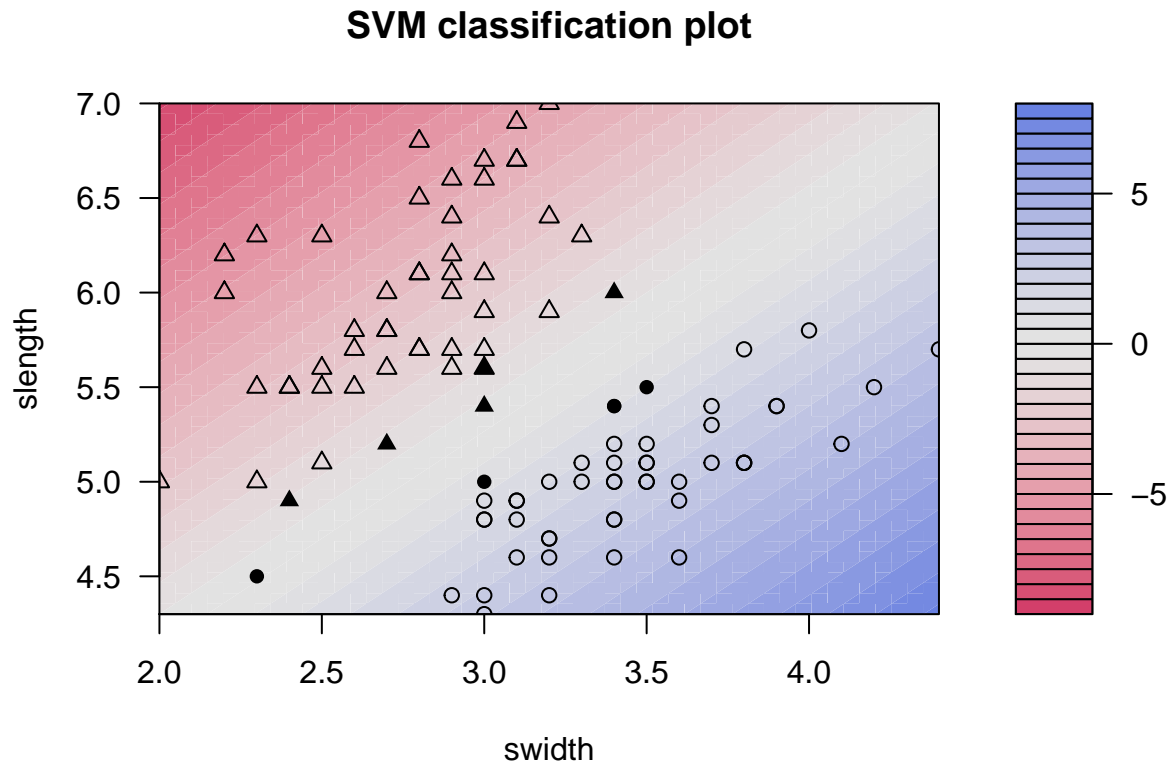
```
table(new$species,preds1) # Confusion table
```

```
##      preds1
##      -1  1
## -1 50  0
##  1  1 49
```

```
prop.table(table(new$species,preds1),1) # Percentage
```

```
##      preds1
##      -1    1
## -1 1.00 0.00
##  1 0.02 0.98
```

```
plot(svm1,data=new) # Plot separating hyperplane
```



Here, can see both the confusion matrix (correct vs type I and II error) for the model. We misclassified one flower, probably the circle on the lower left getting mistaken for a triangle (this is confirmed with the confusion matrix). The black shapes represent the values that make the margin. The line that is made when maximizing the distance to the middle is called the separating hyperplane.