

SVM examples

Abhirup Sen

24/05/2021

```
library("e1071")
```

```
data(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 ...
```

SPlit the data into test and train

```
library(caTools)
set.seed(125)

split <- sample.split(iris, SplitRatio = 0.7)
train <- subset(iris, split == "TRUE" )
test <- subset(iris, split == "FALSE")
```

```
?svm
```

```
## starting httpd help server ... done
```

```
svm_model <- svm(Species ~ ., data = train, method = "class")
summary(svm_model)
```

```
##
## Call:
## svm(formula = Species ~ ., data = train, method = "class")
##
##
## Parameters:
##   SVM-Type:  C-classification
##   SVM-Kernel: radial
##   cost: 1
##
```

```
## Number of Support Vectors: 42
##
## ( 6 18 18 )
##
##
## Number of Classes: 3
##
## Levels:
## setosa versicolor virginica
```

```
test$Species_predicted <- predict(svm_model, newdata = test, type = "class")
```

```
View(test)
```

```
#lets check the prediction errors
```

```
cm<- table(test$Species, test$Species_predicted)
library(caret)
```

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

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

```
confusionMatrix(cm)
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##
```

```
##           setosa versicolor virginica
## setosa         20          0          0
## versicolor      0          19         1
## virginica       0           1        19
```

```
##
```

```
## Overall Statistics
```

```
##
```

```
##           Accuracy : 0.9667
##           95% CI : (0.8847, 0.9959)
## No Information Rate : 0.3333
## P-Value [Acc > NIR] : < 2.2e-16
```

```
##
```

```
##           Kappa : 0.95
```

```
##
```

```
## McNemar's Test P-Value : NA
```

```
##
```

```
## Statistics by Class:
```

```
##
```

```
##           Class: setosa Class: versicolor Class: virginica
## Sensitivity           1.0000           0.9500           0.9500
## Specificity           1.0000           0.9750           0.9750
## Pos Pred Value         1.0000           0.9500           0.9500
## Neg Pred Value         1.0000           0.9750           0.9750
```

## Prevalence	0.3333	0.3333	0.3333
## Detection Rate	0.3333	0.3167	0.3167
## Detection Prevalence	0.3333	0.3333	0.3333
## Balanced Accuracy	1.0000	0.9625	0.9625

Let's see how tuning works

```
attach(iris)
```

```
x <- subset(iris, select = -Species)
y <- Species
```

```
svm_model_1 <- svm(Species ~., data = iris)
summary(svm_model_1)
```

```
##
## Call:
## svm(formula = Species ~ ., data = iris)
##
##
## Parameters:
##   SVM-Type:  C-classification
##   SVM-Kernel: radial
##         cost:  1
##
## Number of Support Vectors:  51
##
## ( 8 22 21 )
##
##
## Number of Classes:  3
##
## Levels:
##   setosa versicolor virginica
```

```
pred <- predict(svm_model_1, x)
table(pred, y)
```

```
##
##           y
## pred      setosa versicolor virginica
##   setosa      50          0          0
##   versicolor  0          48          2
##   virginica   0          2         48
```

```
confusionMatrix(table(pred, y))
```

```
## Confusion Matrix and Statistics
##
##           y
## pred      setosa versicolor virginica
##   setosa      50          0          0
```

```
##   versicolor      0      48      2
##   virginica       0       2     48
##
## Overall Statistics
##
##           Accuracy : 0.9733
##           95% CI : (0.9331, 0.9927)
##       No Information Rate : 0.3333
##       P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.96
##
## Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##           Class: setosa Class: versicolor Class: virginica
## Sensitivity           1.0000           0.9600           0.9600
## Specificity           1.0000           0.9800           0.9800
## Pos Pred Value        1.0000           0.9600           0.9600
## Neg Pred Value        1.0000           0.9800           0.9800
## Prevalence            0.3333           0.3333           0.3333
## Detection Rate        0.3333           0.3200           0.3200
## Detection Prevalence  0.3333           0.3333           0.3333
## Balanced Accuracy     1.0000           0.9700           0.9700
```

Finding the right cost and gamma function

```
svm_model_tune <- tune(svm, train.x = x, train.y=y,
                      kernel = "radial", ranges = list(cost = 10^(-1:2), gamma = c(.5,1,2)))
print(svm_model_tune)
```

```
##
## Parameter tuning of 'svm':
##
## - sampling method: 10-fold cross validation
##
## - best parameters:
##   cost gamma
##     1   0.5
##
## - best performance: 0.03333333
```

```
svm_model_post_tunning <- svm(Species ~ ., data = iris, kernel="radial", cost=1 , gamma = 0.5 )
summary(svm_model_post_tunning)
```

```
##
## Call:
## svm(formula = Species ~ ., data = iris, kernel = "radial", cost = 1,
##     gamma = 0.5)
##
##
```

```
## Parameters:
##   SVM-Type:  C-classification
##   SVM-Kernel: radial
##       cost:  1
##
## Number of Support Vectors:  59
##
## ( 11 23 25 )
##
## Number of Classes:  3
##
## Levels:
##   setosa versicolor virginica
```

```
pred <- predict(svm_model_post_tunning, x)
table (pred, y)
```

```
##           y
## pred      setosa versicolor virginica
##   setosa      50          0          0
##   versicolor   0          48          2
##   virginica    0          2          48
```

```
confusionMatrix(table (pred, y))
```

```
## Confusion Matrix and Statistics
```

```
##
##           y
## pred      setosa versicolor virginica
##   setosa      50          0          0
##   versicolor   0          48          2
##   virginica    0          2          48
##
```

```
## Overall Statistics
```

```
##
##           Accuracy : 0.9733
##           95% CI : (0.9331, 0.9927)
##   No Information Rate : 0.3333
##   P-Value [Acc > NIR] : < 2.2e-16
##
```

```
##           Kappa : 0.96
##
```

```
## McNemar's Test P-Value : NA
##
```

```
## Statistics by Class:
```

```
##
##           Class: setosa Class: versicolor Class: virginica
## Sensitivity           1.0000           0.9600           0.9600
## Specificity           1.0000           0.9800           0.9800
## Pos Pred Value        1.0000           0.9600           0.9600
## Neg Pred Value        1.0000           0.9800           0.9800
## Prevalence            0.3333           0.3333           0.3333
```

## Detection Rate	0.3333	0.3200	0.3200
## Detection Prevalence	0.3333	0.3333	0.3333
## Balanced Accuracy	1.0000	0.9700	0.9700