

18BCE080_PRAC10

Ishan Tewari

29/04/2021

Loading iris dataset

```
library(datasets)
data(iris)
summary(iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300
## Median :5.800 Median :3.000 Median :4.350 Median :1.300
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
## Species
## setosa :50
## versicolor:50
## virginica :50
##
##
##
```

```
library(e1071)
library('caTools')
split = sample.split(iris$Species, SplitRatio = 0.7)
train_data = subset(iris, split==T)
test_data = subset(iris, split==F)
classifier = naiveBayes(Species~., data=train_data)
table(predict(classifier, test_data), test_data$Species, dnn=list('predicted', 'actual'))
```

```
##          actual
## predicted  setosa versicolor virginica
## setosa      15         0         0
## versicolor  0         15         2
## virginica   0         0        13
```

```
classifier$tables$Petal.Length
```

```
##          Petal.Length
## Y          [,1]      [,2]
## setosa    1.465714 0.1830186
## versicolor 4.220000 0.4714184
## virginica  5.577143 0.6073499
```

```
library(caret)
```

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

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

```
confusionMatrix(predict(classifier, test_data), test_data$Species)
```

```
## Confusion Matrix and Statistics
##
##          Reference
## Prediction  setosa versicolor virginica
## setosa      15         0         0
## versicolor  0         15         2
## virginica   0         0        13
##
## Overall Statistics
##
```

```

##          Accuracy : 0.9556
##          95% CI : (0.8485, 0.9946)
##    No Information Rate : 0.3333
##    P-Value [Acc > NIR] : < 2.2e-16
##
##          Kappa : 0.9333
##
##  McNemar's Test P-Value : NA
##
## Statistics by Class:
##
##          Class: setosa Class: versicolor Class: virginica
## Sensitivity          1.0000          1.0000          0.8667
## Specificity          1.0000          0.9333          1.0000
## Pos Pred Value       1.0000          0.8824          1.0000
## Neg Pred Value       1.0000          1.0000          0.9375
## Prevalence           0.3333          0.3333          0.3333
## Detection Rate       0.3333          0.3333          0.2889
## Detection Prevalence 0.3333          0.3778          0.2889
## Balanced Accuracy     1.0000          0.9667          0.9333

```