exercise_5

chenchunpeng

2020/3/21

```
# load data
data(iris)
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
colnames(iris)[1:5]=c("sepal_length","sepal_width","petal_length","petal_width","class")
iris$class=factor(iris$class)
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 ...
##  $ class       : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
# Randomly extract training and test sets
sample_iris=sample(150,110,replace = FALSE)
sample_iris
```

```
##   [1]  82  29  72   3 126  51 137  57 130  76  53  75  20  62  63 120  92  97
##  [19]  43  93 131  56   6  67  22  88 100  32  34  23 146  36  26  94 101 114
##  [37]  39  18  27 124 104 123  78  46  15 149 128 133  37 103   2 115  81  90
##  [55]   9  44   7  66 112 111  65  47  52  42  84  74  86 108  17 107  25  12
##  [73] 106  30  24  60   5  13 140  87  61 127  11 138  38 110  59  19  21 116
##  [91] 144   1 142 102  35  79  14  89  33 113  95 139 147 121  50  70  58 148
## [109]  71  48
```

```
iris_training=iris[sample_iris,]
iris_test=iris[-sample_iris,]
iris_training_labels=iris[sample_iris,]$class
iris_test_labels=iris[-sample_iris,]$class
table(iris_training$class)
```

```
##
##     setosa versicolor  virginica
##         40         36         34
```

```r
table(iris_test$class)
```

```
##
##     setosa versicolor  virginica
##         10         14         16
```

```r
#  Naive Bayes training and prediction
library(e1071)
iris_classifier=naiveBayes(iris_training,iris_training_labels)
iris_test_pred=predict(iris_classifier,iris_test)
iris_test_pred
```

```
##  [1] setosa     setosa     setosa     setosa     setosa     setosa
##  [7] setosa     setosa     setosa     setosa     versicolor versicolor
## [13] versicolor versicolor versicolor versicolor versicolor versicolor
## [19] versicolor versicolor versicolor versicolor versicolor versicolor
## [25] virginica  virginica  virginica  virginica  virginica  virginica
## [31] virginica  virginica  virginica  virginica  virginica  virginica
## [37] virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

```r
# Prediction result
library(gmodels)
CrossTable(iris_test_pred,iris_test_labels,prop.chisq = FALSE, prop.t = FALSE,
           prop.r = FALSE, dnn = c('predicted', 'actual'))
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Col Total |
## |-------------------------|
##
##
## Total Observations in Table:  40
##
##
##              | actual
##    predicted |     setosa | versicolor |  virginica |  Row Total |
## -------------|------------|------------|------------|------------|
##       setosa |         10 |          0 |          0 |         10 |
##              |      1.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|
##   versicolor |          0 |         14 |          0 |         14 |
##              |      0.000 |      1.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|
##    virginica |          0 |          0 |         16 |         16 |
##              |      0.000 |      0.000 |      1.000 |            |
## -------------|------------|------------|------------|------------|
## Column Total |         10 |         14 |         16 |         40 |
##              |      0.250 |      0.350 |      0.400 |            |
## -------------|------------|------------|------------|------------|
##
##
```

```r
# laplace = 1
iris_classifier2=naiveBayes(iris_training,iris_training_labels,laplace = 1)
iris_test_pred2=predict(iris_classifier2,iris_test)
iris_test_pred2
```

```
##  [1] setosa     setosa     setosa     setosa     setosa     setosa
##  [7] setosa     setosa     setosa     setosa     versicolor versicolor
## [13] versicolor versicolor versicolor versicolor versicolor versicolor
## [19] versicolor versicolor versicolor versicolor versicolor versicolor
## [25] virginica  virginica  virginica  virginica  virginica  virginica
## [31] virginica  virginica  virginica  virginica  virginica  virginica
## [37] virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

```r
CrossTable(iris_test_pred2,iris_test_labels,prop.chisq = FALSE, prop.t = FALSE,
           prop.r = FALSE, dnn = c('predicted', 'actual'))
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Col Total |
## |-------------------------|
##
##
## Total Observations in Table:  40
##
##
##              | actual
##    predicted |     setosa | versicolor |  virginica |  Row Total |
## -------------|------------|------------|------------|------------|
##       setosa |         10 |          0 |          0 |         10 |
##              |      1.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|
##   versicolor |          0 |         14 |          0 |         14 |
##              |      0.000 |      1.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|
##    virginica |          0 |          0 |         16 |         16 |
##              |      0.000 |      0.000 |      1.000 |            |
## -------------|------------|------------|------------|------------|
## Column Total |         10 |         14 |         16 |         40 |
##              |      0.250 |      0.350 |      0.400 |            |
## -------------|------------|------------|------------|------------|
##
##
```