HW_0818.R

User

Thu Sep 20 18:09:24 2018

```r
data(iris)

# Q1-(1)
round(prop.table(table(iris$Species)) * 100, digits=1)
```

```
##
##     setosa versicolor  virginica
##       33.3       33.3       33.3
```

```r
# Q1-(2)
normalize=function(x) {
  return((x-min(x)) / (max(x) - min(x)))
}
iris_n = as.data.frame(lapply(iris[1:4], normalize ))
head(iris_n)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1   0.22222222   0.6250000   0.06779661  0.04166667
## 2   0.16666667   0.4166667   0.06779661  0.04166667
## 3   0.11111111   0.5000000   0.05084746  0.04166667
## 4   0.08333333   0.4583333   0.08474576  0.04166667
## 5   0.19444444   0.6666667   0.06779661  0.04166667
## 6   0.30555556   0.7916667   0.11864407  0.12500000
```

```r
# Q1-(3)
set.seed(123)
c_train=sample(1:150, 100)
iris_train=iris_n[c_train, ] ; head(iris_train)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width
## 44    0.19444444   0.6250000   0.10169492  0.20833333
## 118   0.94444444   0.7500000   0.96610169  0.87500000
## 61    0.19444444   0.0000000   0.42372881  0.37500000
## 130   0.80555556   0.4166667   0.81355932  0.62500000
## 138   0.58333333   0.4583333   0.76271186  0.70833333
## 7     0.08333333   0.5833333   0.06779661  0.08333333
```

```r
iris_train_labels=iris[c_train, 5] ; head(iris_train_labels)
```

```
## [1] setosa     virginica  versicolor virginica  virginica  setosa
## Levels: setosa versicolor virginica
```

```r
iris_test=iris_n[-c_train, ] ; head(iris_test)
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width
## 2    0.16666667   0.4166667   0.06779661  0.04166667
## 4    0.08333333   0.4583333   0.08474576  0.04166667
## 10   0.16666667   0.4583333   0.08474576  0.00000000
## 13   0.13888889   0.4166667   0.06779661  0.00000000
## 19   0.38888889   0.7500000   0.11864407  0.08333333
## 21   0.30555556   0.5833333   0.11864407  0.04166667
```

```
iris_test_labels=iris[-c_train, 5] ; head(iris_test_labels)
```

```
## [1] setosa setosa setosa setosa setosa setosa
## Levels: setosa versicolor virginica
```

```
# Q2-(1)
library(class)
k_10 = knn(train=iris_train, test=iris_test, cl=iris_train_labels, k=10)
k_10
```

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

```
accuracy_10 = sum(k_10==iris_test_labels)
accuracy_10
```

```
## [1] 47
```

```
# Q2-(2)
library(gmodels)
CrossTable(x=iris_test_labels, y=k_10, prop.chisq=FALSE)
```

```
##
##
##    Cell Contents
## |-----------------------|
## |                     N |
## |          N / Row Total |
## |          N / Col Total |
## |        N / Table Total |
## |-----------------------|
##
##
## Total Observations in Table:   50
##
##
##                  | k_10
## iris_test_labels |    setosa | versicolor |  virginica |  Row Total |
## -----------------|-----------|-----------|-----------|-----------|
##           setosa |        14 |         0 |         0 |        14 |
##                  |     1.000 |     0.000 |     0.000 |     0.280 |
##                  |     1.000 |     0.000 |     0.000 |           |
##                  |     0.280 |     0.000 |     0.000 |           |
## -----------------|-----------|-----------|-----------|-----------|
##       versicolor |         0 |        17 |         2 |        19 |
##                  |     0.000 |     0.895 |     0.105 |     0.380 |
##                  |     0.000 |     0.944 |     0.111 |           |
##                  |     0.000 |     0.340 |     0.040 |           |
## -----------------|-----------|-----------|-----------|-----------|
##        virginica |         0 |         1 |        16 |        17 |
##                  |     0.000 |     0.059 |     0.941 |     0.340 |
##                  |     0.000 |     0.056 |     0.889 |           |
##                  |     0.000 |     0.020 |     0.320 |           |
## -----------------|-----------|-----------|-----------|-----------|
##     Column Total |        14 |        18 |        18 |        50 |
##                  |     0.280 |     0.360 |     0.360 |           |
## -----------------|-----------|-----------|-----------|-----------|
##
##
```

```
# Q3-(1)
accuracy_k=rep(0,100)
    # accuracy_k=NULL
for(kk in c(1:nrow(iris_train))) {
  knn_k = knn(train=iris_train, test=iris_test, cl=iris_train_labels, k=kk)
  accuracy_k[kk]=sum(knn_k==iris_test_labels)/length(iris_test_labels)
}
accuracy_k
```

```
##    [1] 0.92 0.92 0.94 0.92 0.94 0.94 0.94 0.94 0.92 0.96 0.94 0.94 0.94 0.92
##   [15] 0.94 0.92 0.92 0.90 0.92 0.90 0.92 0.90 0.90 0.92 0.90 0.90 0.90 0.90
##   [29] 0.90 0.92 0.92 0.92 0.92 0.92 0.92 0.92 0.92 0.92 0.92 0.90 0.90 0.90
##   [43] 0.90 0.90 0.90 0.90 0.90 0.90 0.90 0.90 0.90 0.90 0.90 0.90 0.90 0.88
##   [57] 0.90 0.88 0.84 0.86 0.86 0.84 0.82 0.84 0.82 0.80 0.80 0.80 0.82 0.78
##   [71] 0.80 0.80 0.80 0.82 0.78 0.78 0.78 0.74 0.76 0.72 0.72 0.72 0.72 0.72
##   [85] 0.72 0.72 0.72 0.72 0.68 0.64 0.68 0.66 0.64 0.60 0.62 0.62 0.44 0.28
##   [99] 0.28 0.28
```

```
# Q3-(2)
test_k<-data.frame(k=c(1:nrow(iris_train)),accuracy=accuracy_k)
test_k
```

```
##       k accuracy
## 1    1    0.92
## 2    2    0.92
## 3    3    0.94
## 4    4    0.92
## 5    5    0.94
## 6    6    0.94
## 7    7    0.94
## 8    8    0.94
## 9    9    0.92
## 10  10    0.96
## 11  11    0.94
## 12  12    0.94
## 13  13    0.94
## 14  14    0.92
## 15  15    0.94
## 16  16    0.92
## 17  17    0.92
## 18  18    0.90
## 19  19    0.92
## 20  20    0.90
## 21  21    0.92
## 22  22    0.90
## 23  23    0.90
## 24  24    0.92
## 25  25    0.90
## 26  26    0.90
## 27  27    0.90
## 28  28    0.90
## 29  29    0.90
## 30  30    0.92
## 31  31    0.92
## 32  32    0.92
## 33  33    0.92
## 34  34    0.92
## 35  35    0.92
## 36  36    0.92
## 37  37    0.92
## 38  38    0.92
## 39  39    0.92
## 40  40    0.90
## 41  41    0.90
## 42  42    0.90
## 43  43    0.90
## 44  44    0.90
## 45  45    0.90
## 46  46    0.90
## 47  47    0.90
## 48  48    0.90
## 49  49    0.90
## 50  50    0.90
## 51  51    0.90
## 52  52    0.90
## 53  53    0.90
## 54  54    0.90
## 55  55    0.90
## 56  56    0.88
```

```
## 57    57     0.90
## 58    58     0.88
## 59    59     0.84
## 60    60     0.86
## 61    61     0.86
## 62    62     0.84
## 63    63     0.82
## 64    64     0.84
## 65    65     0.82
## 66    66     0.80
## 67    67     0.80
## 68    68     0.80
## 69    69     0.82
## 70    70     0.78
## 71    71     0.80
## 72    72     0.80
## 73    73     0.80
## 74    74     0.82
## 75    75     0.78
## 76    76     0.78
## 77    77     0.78
## 78    78     0.74
## 79    79     0.76
## 80    80     0.72
## 81    81     0.72
## 82    82     0.72
## 83    83     0.72
## 84    84     0.72
## 85    85     0.72
## 86    86     0.72
## 87    87     0.72
## 88    88     0.72
## 89    89     0.68
## 90    90     0.64
## 91    91     0.68
## 92    92     0.66
## 93    93     0.64
## 94    94     0.60
## 95    95     0.62
## 96    96     0.62
## 97    97     0.44
## 98    98     0.28
## 99    99     0.28
## 100  100     0.28
```

```
# Q3-(3)
test_k[test_k$accuracy %in% max(accuracy_k), "k"]
```

```
## [1] 10
```