KNN examples

Abhirup Sen

02/06/2021

```
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 ...
```

#step 1 : Split the data into train and test

```
library(caTools)
set.seed(1)
split <- sample.split(iris, SplitRatio = 0.7)
split
```

```
## [1]  TRUE  TRUE FALSE FALSE  TRUE
```

```
train <- subset(iris, split=="TRUE")
test <- subset(iris, split=="FALSE")
train
```

```
##     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
## 5            5.0         3.6          1.4         0.2     setosa
## 6            5.4         3.9          1.7         0.4     setosa
## 7            4.6         3.4          1.4         0.3     setosa
## 10           4.9         3.1          1.5         0.1     setosa
## 11           5.4         3.7          1.5         0.2     setosa
## 12           4.8         3.4          1.6         0.2     setosa
## 15           5.8         4.0          1.2         0.2     setosa
## 16           5.7         4.4          1.5         0.4     setosa
## 17           5.4         3.9          1.3         0.4     setosa
## 20           5.1         3.8          1.5         0.3     setosa
## 21           5.4         3.4          1.7         0.2     setosa
## 22           5.1         3.7          1.5         0.4     setosa
## 25           4.8         3.4          1.9         0.2     setosa
## 26           5.0         3.0          1.6         0.2     setosa
```

```
## 27          5.0         3.4          1.6         0.4     setosa
## 30          4.7         3.2          1.6         0.2     setosa
## 31          4.8         3.1          1.6         0.2     setosa
## 32          5.4         3.4          1.5         0.4     setosa
## 35          4.9         3.1          1.5         0.2     setosa
## 36          5.0         3.2          1.2         0.2     setosa
## 37          5.5         3.5          1.3         0.2     setosa
## 40          5.1         3.4          1.5         0.2     setosa
## 41          5.0         3.5          1.3         0.3     setosa
## 42          4.5         2.3          1.3         0.3     setosa
## 45          5.1         3.8          1.9         0.4     setosa
## 46          4.8         3.0          1.4         0.3     setosa
## 47          5.1         3.8          1.6         0.2     setosa
## 50          5.0         3.3          1.4         0.2     setosa
## 51          7.0         3.2          4.7         1.4 versicolor
## 52          6.4         3.2          4.5         1.5 versicolor
## 55          6.5         2.8          4.6         1.5 versicolor
## 56          5.7         2.8          4.5         1.3 versicolor
## 57          6.3         3.3          4.7         1.6 versicolor
## 60          5.2         2.7          3.9         1.4 versicolor
## 61          5.0         2.0          3.5         1.0 versicolor
## 62          5.9         3.0          4.2         1.5 versicolor
## 65          5.6         2.9          3.6         1.3 versicolor
## 66          6.7         3.1          4.4         1.4 versicolor
## 67          5.6         3.0          4.5         1.5 versicolor
## 70          5.6         2.5          3.9         1.1 versicolor
## 71          5.9         3.2          4.8         1.8 versicolor
## 72          6.1         2.8          4.0         1.3 versicolor
## 75          6.4         2.9          4.3         1.3 versicolor
## 76          6.6         3.0          4.4         1.4 versicolor
## 77          6.8         2.8          4.8         1.4 versicolor
## 80          5.7         2.6          3.5         1.0 versicolor
## 81          5.5         2.4          3.8         1.1 versicolor
## 82          5.5         2.4          3.7         1.0 versicolor
## 85          5.4         3.0          4.5         1.5 versicolor
## 86          6.0         3.4          4.5         1.6 versicolor
## 87          6.7         3.1          4.7         1.5 versicolor
## 90          5.5         2.5          4.0         1.3 versicolor
## 91          5.5         2.6          4.4         1.2 versicolor
## 92          6.1         3.0          4.6         1.4 versicolor
## 95          5.6         2.7          4.2         1.3 versicolor
## 96          5.7         3.0          4.2         1.2 versicolor
## 97          5.7         2.9          4.2         1.3 versicolor
## 100         5.7         2.8          4.1         1.3 versicolor
## 101         6.3         3.3          6.0         2.5  virginica
## 102         5.8         2.7          5.1         1.9  virginica
## 105         6.5         3.0          5.8         2.2  virginica
## 106         7.6         3.0          6.6         2.1  virginica
## 107         4.9         2.5          4.5         1.7  virginica
## 110         7.2         3.6          6.1         2.5  virginica
## 111         6.5         3.2          5.1         2.0  virginica
## 112         6.4         2.7          5.3         1.9  virginica
## 115         5.8         2.8          5.1         2.4  virginica
## 116         6.4         3.2          5.3         2.3  virginica
```

```
## 117          6.5         3.0          5.5         1.8  virginica
## 120          6.0         2.2          5.0         1.5  virginica
## 121          6.9         3.2          5.7         2.3  virginica
## 122          5.6         2.8          4.9         2.0  virginica
## 125          6.7         3.3          5.7         2.1  virginica
## 126          7.2         3.2          6.0         1.8  virginica
## 127          6.2         2.8          4.8         1.8  virginica
## 130          7.2         3.0          5.8         1.6  virginica
## 131          7.4         2.8          6.1         1.9  virginica
## 132          7.9         3.8          6.4         2.0  virginica
## 135          6.1         2.6          5.6         1.4  virginica
## 136          7.7         3.0          6.1         2.3  virginica
## 137          6.3         3.4          5.6         2.4  virginica
## 140          6.9         3.1          5.4         2.1  virginica
## 141          6.7         3.1          5.6         2.4  virginica
## 142          6.9         3.1          5.1         2.3  virginica
## 145          6.7         3.3          5.7         2.5  virginica
## 146          6.7         3.0          5.2         2.3  virginica
## 147          6.3         2.5          5.0         1.9  virginica
## 150          5.9         3.0          5.1         1.8  virginica
```

```
test
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 3            4.7         3.2          1.3         0.2     setosa
## 4            4.6         3.1          1.5         0.2     setosa
## 8            5.0         3.4          1.5         0.2     setosa
## 9            4.4         2.9          1.4         0.2     setosa
## 13           4.8         3.0          1.4         0.1     setosa
## 14           4.3         3.0          1.1         0.1     setosa
## 18           5.1         3.5          1.4         0.3     setosa
## 19           5.7         3.8          1.7         0.3     setosa
## 23           4.6         3.6          1.0         0.2     setosa
## 24           5.1         3.3          1.7         0.5     setosa
## 28           5.2         3.5          1.5         0.2     setosa
## 29           5.2         3.4          1.4         0.2     setosa
## 33           5.2         4.1          1.5         0.1     setosa
## 34           5.5         4.2          1.4         0.2     setosa
## 38           4.9         3.6          1.4         0.1     setosa
## 39           4.4         3.0          1.3         0.2     setosa
## 43           4.4         3.2          1.3         0.2     setosa
## 44           5.0         3.5          1.6         0.6     setosa
## 48           4.6         3.2          1.4         0.2     setosa
## 49           5.3         3.7          1.5         0.2     setosa
## 53           6.9         3.1          4.9         1.5 versicolor
## 54           5.5         2.3          4.0         1.3 versicolor
## 58           4.9         2.4          3.3         1.0 versicolor
## 59           6.6         2.9          4.6         1.3 versicolor
## 63           6.0         2.2          4.0         1.0 versicolor
## 64           6.1         2.9          4.7         1.4 versicolor
## 68           5.8         2.7          4.1         1.0 versicolor
## 69           6.2         2.2          4.5         1.5 versicolor
## 73           6.3         2.5          4.9         1.5 versicolor
## 74           6.1         2.8          4.7         1.2 versicolor
```

```
## 78          6.7         3.0          5.0         1.7 versicolor
## 79          6.0         2.9          4.5         1.5 versicolor
## 83          5.8         2.7          3.9         1.2 versicolor
## 84          6.0         2.7          5.1         1.6 versicolor
## 88          6.3         2.3          4.4         1.3 versicolor
## 89          5.6         3.0          4.1         1.3 versicolor
## 93          5.8         2.6          4.0         1.2 versicolor
## 94          5.0         2.3          3.3         1.0 versicolor
## 98          6.2         2.9          4.3         1.3 versicolor
## 99          5.1         2.5          3.0         1.1 versicolor
## 103         7.1         3.0          5.9         2.1  virginica
## 104         6.3         2.9          5.6         1.8  virginica
## 108         7.3         2.9          6.3         1.8  virginica
## 109         6.7         2.5          5.8         1.8  virginica
## 113         6.8         3.0          5.5         2.1  virginica
## 114         5.7         2.5          5.0         2.0  virginica
## 118         7.7         3.8          6.7         2.2  virginica
## 119         7.7         2.6          6.9         2.3  virginica
## 123         7.7         2.8          6.7         2.0  virginica
## 124         6.3         2.7          4.9         1.8  virginica
## 128         6.1         3.0          4.9         1.8  virginica
## 129         6.4         2.8          5.6         2.1  virginica
## 133         6.4         2.8          5.6         2.2  virginica
## 134         6.3         2.8          5.1         1.5  virginica
## 138         6.4         3.1          5.5         1.8  virginica
## 139         6.0         3.0          4.8         1.8  virginica
## 143         5.8         2.7          5.1         1.9  virginica
## 144         6.8         3.2          5.9         2.3  virginica
## 148         6.5         3.0          5.2         2.0  virginica
## 149         6.2         3.4          5.4         2.3  virginica
```

```
train$Species
```

```
##  [1] setosa     setosa     setosa     setosa     setosa     setosa
##  [7] setosa     setosa     setosa     setosa     setosa     setosa
## [13] setosa     setosa     setosa     setosa     setosa     setosa
## [19] setosa     setosa     setosa     setosa     setosa     setosa
## [25] setosa     setosa     setosa     setosa     setosa     setosa
## [31] versicolor versicolor versicolor versicolor versicolor versicolor
## [37] versicolor versicolor versicolor versicolor versicolor versicolor
## [43] versicolor versicolor versicolor versicolor versicolor versicolor
## [49] versicolor versicolor versicolor versicolor versicolor versicolor
## [55] versicolor versicolor versicolor versicolor versicolor versicolor
## [61] virginica  virginica  virginica  virginica  virginica  virginica
## [67] virginica  virginica  virginica  virginica  virginica  virginica
## [73] virginica  virginica  virginica  virginica  virginica  virginica
## [79] virginica  virginica  virginica  virginica  virginica  virginica
## [85] virginica  virginica  virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

#step 2: Feature Scaling

```
train_scale <- scale(train[,1:4])
test_scale <- scale(test[,1:4])
```

#step 3 : Fitting _KNN to the training dataset and predicting the test set

```
library(class)
?knn
```

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

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=1)
table(test$Species,test$Species_predicted)
```

```
##
##              setosa versicolor virginica
##   setosa         20          0         0
##   versicolor      0         17         3
##   virginica       0          2        18
```

#step 4 : Model evaluation - choose the rigth k

#calculate the error

```
misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.916666666666667"
```

#K = 3

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=3)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.883333333333333"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=5)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.883333333333333"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=7)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.9"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=10)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.933333333333333"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=13)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.916666666666667"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=15)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.933333333333333"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=17)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.933333333333333"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=19)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.916666666666667"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=21)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.9"
```

```
test$Species_predicted <- knn(train = train_scale,
                              test = test_scale,
                              cl = train$Species,
                              k=20)

misClassError <- mean(test$Species_predicted != test$Species)
print(paste('Accuracy=',1-misClassError))
```

```
## [1] "Accuracy= 0.9"
```