ica11_shuangyu_zhao

shuangyu_zhao

2023-02-21

```
library(ISLR2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
oj_scale <- scale(select_if(OJ[, 10:14], is.numeric))
```

```
split_pct <- 0.75
n <- split_pct * nrow(oj_scale)
row_samp <- sample(1:nrow(oj_scale), n, replace = FALSE)
train <- oj_scale[row_samp,]
test <- oj_scale[-row_samp,]
train.Y <- OJ[row_samp,]$Purchase
test.Y <- OJ[-row_samp,]$Purchase
```

```
library(class)
library(caret)
```

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

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
knn_mod <- knn(train, test, cl = train.Y , k = 5, prob = TRUE)
confusionMatrix(knn_mod , reference = as.factor(test.Y))
```

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

```
##
##           Reference
## Prediction  CH  MM
##         CH 143  24
##         MM  23  78
##
##                Accuracy : 0.8246
##                  95% CI : (0.7737, 0.8682)
##     No Information Rate : 0.6194
##     P-Value [Acc > NIR] : 2.164e-13
##
##                   Kappa : 0.6273
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.8614
##             Specificity : 0.7647
##          Pos Pred Value : 0.8563
##          Neg Pred Value : 0.7723
##              Prevalence : 0.6194
##          Detection Rate : 0.5336
##    Detection Prevalence : 0.6231
##       Balanced Accuracy : 0.8131
##
##        'Positive' Class : CH
##
```

```
knn_mod2 <- knn(train, test, cl = train.Y , k = 7, prob = TRUE)
confusionMatrix(knn_mod2 , reference = as.factor(test.Y))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  CH  MM
##         CH 142  20
##         MM  24  82
##
##                Accuracy : 0.8358
##                  95% CI : (0.7859, 0.8781)
##     No Information Rate : 0.6194
##     P-Value [Acc > NIR] : 7.877e-15
##
##                   Kappa : 0.6544
##
##  Mcnemar's Test P-Value : 0.6511
##
##             Sensitivity : 0.8554
##             Specificity : 0.8039
##          Pos Pred Value : 0.8765
##          Neg Pred Value : 0.7736
##              Prevalence : 0.6194
##          Detection Rate : 0.5299
##    Detection Prevalence : 0.6045
##       Balanced Accuracy : 0.8297
```

```
##
##        'Positive' Class : CH
##
```