q1.R

elry

2021-06-26

```
library(e1071);
library(ggplot2);

RNGversion("3.5.2");
```

```
## Warning in RNGkind("Mersenne-Twister", "Inversion", "Rounding"): non-uniform
## 'Rounding' sampler used
```

```
set.seed(1987);

wdbc <- read.csv(
  "https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data",
  col.names=c(
    "patientid",
    "outcome",
    "radius_mean",
    "texture_mean",
    "perimeter_mean",
    "area_mean",
    "smoothness_mean",
    "compactness_mean",
    "concavity_mean",
    "concavepoints_mean",
    "symmetry_mean",
    "fractaldimension_mean",
    "radius_error",
    "texture_error",
    "perimeter_error",
    "area_error",
    "smoothness_error",
    "compactness_error",
    "concavity_error",
    "concavepoints_error",
    "symmetry_error",
    "fractaldimension_error",
    "radius_worst",
    "texture_worst",
    "perimeter_worst",
    "area_worst",
    "smoothness_worst",
    "compactness_worst",
    "concavity_worst",
    "concavepoints_worst",
```

```
    "symmetry_worst",
    "fractaldimension_worst")
);

wdbc <- as.data.frame(unclass(wdbc),stringsAsFactors=T)

head(wdbc);
```

```
##   patientid outcome radius_mean texture_mean perimeter_mean area_mean
## 1    842517       M       20.57        17.77         132.90    1326.0
## 2  84300903       M       19.69        21.25         130.00    1203.0
## 3  84348301       M       11.42        20.38          77.58     386.1
## 4  84358402       M       20.29        14.34         135.10    1297.0
## 5    843786       M       12.45        15.70          82.57     477.1
## 6    844359       M       18.25        19.98         119.60    1040.0
##   smoothness_mean compactness_mean concavity_mean concavepoints_mean
## 1         0.08474          0.07864         0.0869            0.07017
## 2         0.10960          0.15990         0.1974            0.12790
## 3         0.14250          0.28390         0.2414            0.10520
## 4         0.10030          0.13280         0.1980            0.10430
## 5         0.12780          0.17000         0.1578            0.08089
## 6         0.09463          0.10900         0.1127            0.07400
##   symmetry_mean fractaldimension_mean radius_error texture_error
## 1        0.1812               0.05667       0.5435        0.7339
## 2        0.2069               0.05999       0.7456        0.7869
## 3        0.2597               0.09744       0.4956        1.1560
## 4        0.1809               0.05883       0.7572        0.7813
## 5        0.2087               0.07613       0.3345        0.8902
## 6        0.1794               0.05742       0.4467        0.7732
##   perimeter_error area_error smoothness_error compactness_error concavity_error
## 1           3.398      74.08         0.005225           0.01308         0.01860
## 2           4.585      94.03         0.006150           0.04006         0.03832
## 3           3.445      27.23         0.009110           0.07458         0.05661
## 4           5.438      94.44         0.011490           0.02461         0.05688
## 5           2.217      27.19         0.007510           0.03345         0.03672
## 6           3.180      53.91         0.004314           0.01382         0.02254
##   concavepoints_error symmetry_error fractaldimension_error radius_worst
## 1             0.01340        0.01389               0.003532        24.99
## 2             0.02058        0.02250               0.004571        23.57
## 3             0.01867        0.05963               0.009208        14.91
## 4             0.01885        0.01756               0.005115        22.54
## 5             0.01137        0.02165               0.005082        15.47
## 6             0.01039        0.01369               0.002179        22.88
##   texture_worst perimeter_worst area_worst smoothness_worst compactness_worst
## 1         23.41          158.80     1956.0           0.1238            0.1866
## 2         25.53          152.50     1709.0           0.1444            0.4245
## 3         26.50           98.87      567.7           0.2098            0.8663
## 4         16.67          152.20     1575.0           0.1374            0.2050
## 5         23.75          103.40      741.6           0.1791            0.5249
## 6         27.66          153.20     1606.0           0.1442            0.2576
##   concavity_worst concavepoints_worst symmetry_worst fractaldimension_worst
## 1          0.2416              0.1860         0.2750                0.08902
## 2          0.4504              0.2430         0.3613                0.08758
## 3          0.6869              0.2575         0.6638                0.17300
```

```
## 4          0.4000              0.1625         0.2364                0.07678
## 5          0.5355              0.1741         0.3985                0.12440
## 6          0.3784              0.1932         0.3063                0.08368
```

```
summary(wdbc);
```

```
##    patientid         outcome  radius_mean      texture_mean   perimeter_mean
##  Min.   :     8670   B:357   Min.   : 6.981   Min.   : 9.71   Min.   : 43.79
##  1st Qu.:   869222   M:211   1st Qu.:11.697   1st Qu.:16.18   1st Qu.: 75.14
##  Median :   906157           Median :13.355   Median :18.86   Median : 86.21
##  Mean   : 30423820           Mean   :14.120   Mean   :19.31   Mean   : 91.91
##  3rd Qu.:  8825022           3rd Qu.:15.780   3rd Qu.:21.80   3rd Qu.:103.88
##  Max.   :911320502           Max.   :28.110   Max.   :39.28   Max.   :188.50
##    area_mean      smoothness_mean   compactness_mean  concavity_mean
##  Min.   : 143.5   Min.   :0.05263   Min.   :0.01938   Min.   :0.00000
##  1st Qu.: 420.2   1st Qu.:0.08629   1st Qu.:0.06481   1st Qu.:0.02954
##  Median : 548.8   Median :0.09587   Median :0.09252   Median :0.06140
##  Mean   : 654.3   Mean   :0.09632   Mean   :0.10404   Mean   :0.08843
##  3rd Qu.: 782.6   3rd Qu.:0.10530   3rd Qu.:0.13040   3rd Qu.:0.12965
##  Max.   :2501.0   Max.   :0.16340   Max.   :0.34540   Max.   :0.42680
##  concavepoints_mean symmetry_mean    fractaldimension_mean  radius_error
##  Min.   :0.00000    Min.   :0.1060   Min.   :0.04996       Min.   :0.1115
##  1st Qu.:0.02031    1st Qu.:0.1619   1st Qu.:0.05770       1st Qu.:0.2324
##  Median :0.03345    Median :0.1792   Median :0.06152       Median :0.3240
##  Mean   :0.04875    Mean   :0.1811   Mean   :0.06277       Mean   :0.4040
##  3rd Qu.:0.07373    3rd Qu.:0.1956   3rd Qu.:0.06612       3rd Qu.:0.4773
##  Max.   :0.20120    Max.   :0.3040   Max.   :0.09744       Max.   :2.8730
##  texture_error    perimeter_error    area_error      smoothness_error
##  Min.   :0.3602   Min.   : 0.757   Min.   :  6.802   Min.   :0.001713
##  1st Qu.:0.8331   1st Qu.: 1.605   1st Qu.: 17.850   1st Qu.:0.005166
##  Median :1.1095   Median : 2.285   Median : 24.485   Median :0.006374
##  Mean   :1.2174   Mean   : 2.856   Mean   : 40.138   Mean   :0.007042
##  3rd Qu.:1.4743   3rd Qu.: 3.337   3rd Qu.: 45.017   3rd Qu.:0.008151
##  Max.   :4.8850   Max.   :21.980   Max.   :542.200   Max.   :0.031130
##  compactness_error  concavity_error   concavepoints_error symmetry_error
##  Min.   :0.002252   Min.   :0.00000   Min.   :0.000000    Min.   :0.007882
##  1st Qu.:0.013048   1st Qu.:0.01506   1st Qu.:0.007634    1st Qu.:0.015128
##  Median :0.020435   Median :0.02587   Median :0.010920    Median :0.018725
##  Mean   :0.025437   Mean   :0.03186   Mean   :0.011789    Mean   :0.020526
##  3rd Qu.:0.032218   3rd Qu.:0.04176   3rd Qu.:0.014710    3rd Qu.:0.023398
##  Max.   :0.135400   Max.   :0.39600   Max.   :0.052790    Max.   :0.078950
##  fractaldimension_error  radius_worst   texture_worst   perimeter_worst
##  Min.   :0.0008948      Min.   : 7.93   Min.   :12.02   Min.   : 50.41
##  1st Qu.:0.0022445      1st Qu.:13.01   1st Qu.:21.09   1st Qu.: 84.10
##  Median :0.0031615      Median :14.96   Median :25.43   Median : 97.66
##  Mean   :0.0037907      Mean   :16.25   Mean   :25.69   Mean   :107.13
##  3rd Qu.:0.0045258      3rd Qu.:18.77   3rd Qu.:29.76   3rd Qu.:125.17
##  Max.   :0.0298400      Max.   :36.04   Max.   :49.54   Max.   :251.20
##    area_worst     smoothness_worst  compactness_worst concavity_worst
##  Min.   : 185.2   Min.   :0.07117   Min.   :0.02729   Min.   :0.0000
##  1st Qu.: 515.0   1st Qu.:0.11660   1st Qu.:0.14690   1st Qu.:0.1145
##  Median : 685.5   Median :0.13130   Median :0.21185   Median :0.2266
##  Mean   : 878.6   Mean   :0.13232   Mean   :0.25354   Mean   :0.2714
##  3rd Qu.:1073.5   3rd Qu.:0.14600   3rd Qu.:0.33760   3rd Qu.:0.3814
##  Max.   :4254.0   Max.   :0.22260   Max.   :1.05800   Max.   :1.2520
```

```
##  concavepoints_worst symmetry_worst   fractaldimension_worst
##  Min.   :0.00000     Min.   :0.1565   Min.   :0.05504
##  1st Qu.:0.06473     1st Qu.:0.2504   1st Qu.:0.07141
##  Median :0.09984     Median :0.2821   Median :0.08002
##  Mean   :0.11434     Mean   :0.2898   Mean   :0.08388
##  3rd Qu.:0.16132     3rd Qu.:0.3177   3rd Qu.:0.09206
##  Max.   :0.29100     Max.   :0.6638   Max.   :0.20750
```

```
any(is.na(wdbc));
```

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

```
mybreast <- wdbc[,!(names(wdbc) %in% c("patientid"))];

# Q3 Quais as quantidades de registros Benignos e Malignos?
table(mybreast$outcome);
```

```
##
##   B   M
## 357 211
```

```
# Anomaly detection
# outcome to numeric
mybreast$outcome = as.factor(as.numeric(mybreast$outcome) - 1);

# Separe os dados normais, Benignos dos Malignos
mybreast_B <- mybreast[mybreast$outcome == 0,];
mybreast_M <- mybreast[mybreast$outcome == 1,];

# Check que o nr de linhas bate com os valores B / M
nrow(mybreast_B);
```

```
## [1] 357
```

```
nrow(mybreast_M);
```

```
## [1] 211
```

```
svm <- svm(
  outcome~.,
  data = mybreast_B,
  scale = TRUE,
  kernel = "radial",
  type = "one-classification"
);
```

```
## Warning in Ops.factor(yorig, ret$fitted): '-' not meaningful for factors
```

```
print(svm);
```

```
##
## Call:
## svm(formula = outcome ~ ., data = mybreast_B, kernel = "radial",
##     type = "one-classification", scale = TRUE)
##
##
## Parameters:
##    SVM-Type:  one-classification
```

```
##  SVM-Kernel:  radial
##        gamma:  0.03333333
##           nu:  0.5
##
## Number of Support Vectors:  182
```

```
summary(svm);
```

```
##
## Call:
## svm(formula = outcome ~ ., data = mybreast_B, kernel = "radial",
##     type = "one-classification", scale = TRUE)
##
##
## Parameters:
##    SVM-Type:  one-classification
##  SVM-Kernel:  radial
##        gamma:  0.03333333
##           nu:  0.5
##
## Number of Support Vectors:  182
##
##
##
##
## Number of Classes: 1
```

```
predict_test <- predict(svm, mybreast_M);
table(predict_test);
```

```
## predict_test
## FALSE  TRUE
##   209     2
```

```
cat('Anomaly Detected (FALSE):', table(predict_test)[1]/sum(table(predict_test))*100, ' %');
```

```
## Anomaly Detected (FALSE): 99.05213  %
```

```
svm2 <- svm(
  outcome~.,
  data = mybreast,
  scale = TRUE,
  kernel = "radial"
);

summary(svm2);
```

```
##
## Call:
## svm(formula = outcome ~ ., data = mybreast, kernel = "radial", scale = TRUE)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  radial
##        cost:  1
##
```

```
## Number of Support Vectors:  119
##
##  ( 60 59 )
##
##
## Number of Classes:  2
##
## Levels:
##  0 1
```

```
predict_test2 <- predict(svm2, mybreast_M);

c_matrix = table(predict_test2, mybreast_M$outcome)
print(c_matrix);
```

```
##
## predict_test2   0   1
##             0   0   7
##             1   0 204
```

```
acc <- sum(diag(c_matrix))/sum(c_matrix)*100;
cat('Accuracy: ', acc, ' %', "\n")
```

```
## Accuracy:  96.68246  %
```