

# smandumu\_A4

smandumu

4/23/2020

```
library(mlbench)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(caret)
```

```
## Loading required package: lattice
## Loading required package: ggplot2
```

```
library(gmodels)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
## v tibble  2.1.3      v purrr  0.3.3
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x purrr::lift()    masks caret::lift()
```

```
data(BreastCancer)
dim(BreastCancer)
```

```
## [1] 699  11
```

```
levels(BreastCancer$Class)
```

```
## [1] "benign"    "malignant"
```

```
str(BreastCancer)
```

```
## 'data.frame': 699 obs. of 11 variables:
## $ Id : chr "1000025" "1002945" "1015425" "1016277" ...
## $ Cl.thickness : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 5 5 3 6 4 8 1 2 2 4 ...
## $ Cell.size : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 1 4 1 8 1 10 1 1 1 2 ...
```

```
## $ Cell.shape      : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 1 4 1 8 1 10 1 2 1 1 ...
## $ Marg.adhesion   : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 1 5 1 1 3 8 1 1 1 1 ...
## $ Epith.c.size     : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 2 7 2 3 2 7 2 2 2 2 ...
## $ Bare.nuclei      : Factor w/ 10 levels "1","2","3","4",...: 1 10 2 4 1 10 10 1 1 1 ...
## $ Bl.cromatin       : Factor w/ 10 levels "1","2","3","4",...: 3 3 3 3 3 9 3 3 1 2 ...
## $ Normal.nucleoli: Factor w/ 10 levels "1","2","3","4",...: 1 2 1 7 1 7 1 1 1 1 ...
## $ Mitoses          : Factor w/ 9 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 5 1 ...
## $ Class            : Factor w/ 2 levels "benign","malignant": 1 1 1 1 1 2 1 1 1 1 ...
```

```
length(which(is.na(BreastCancer)))
```

```
## [1] 16
```

```
library(mice)
```

```
##
```

```
## Attaching package: 'mice'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      cbind, rbind
```

```
BreastCancer <- BreastCancer[,-1] %>% drop_na()
```

```
levels(BreastCancer$Class)
```

```
## [1] "benign"      "malignant"
```

```
set.seed(2020)
```

```
library(caTools)#Package has split function which is used to split our dataset into training and test d
```

```
split=sample.split(BreastCancer, SplitRatio = 0.7) # Splitting data into training and test dataset
```

```
trg_set=subset(BreastCancer,split==TRUE) # Training dataset
```

```
test_set=subset(BreastCancer,split==FALSE)# Test dataset
```

```
# SVM Model and Raw propability
```

```
library(e1071)
```

```
(svm_model <- svm(Class~., data=trg_set,method = 'svmlinear'))
```

```
##
```

```
## Call:
```

```
## svm(formula = Class ~ ., data = trg_set, method = "svmlinear")
```

```
##
```

```
##
```

```
## Parameters:
```

```
##      SVM-Type:  C-classification
```

```
##      SVM-Kernel: radial
```

```
##      cost: 1
```

```
##
```

```
## Number of Support Vectors: 136
```

```
(svm_pred <-predict(svm_model,test_set,type = "raw"))
```

```
##      1      3      10      11      13      20      21      23
##  benign  benign  benign  benign  benign  benign malignant  benign
##      31      32      34      42      43      45      52      53
##  benign  benign  benign malignant malignant malignant  benign malignant
##      55      62      63      65      72      73      75      82
## malignant  benign malignant  benign malignant  benign malignant  benign
##      83      85      92      93      95      102      103      105
##  benign malignant  benign  benign  benign  benign  benign malignant
```

```

##      112      113      115      122      123      125      132      133
## malignant malignant      benign      benign malignant malignant      benign malignant
##      135      143      144      147      154      155      157      166
##      benign malignant      benign malignant      benign      benign      benign      benign
##      167      169      176      177      179      186      187      189
## malignant      benign malignant      benign      benign      benign malignant malignant
##      196      197      199      206      207      209      216      217
##      benign malignant      benign malignant malignant      benign malignant      benign
##      219      226      227      229      237      238      240      247
## malignant      benign malignant      benign malignant malignant malignant malignant
##      248      251      258      259      261      268      269      271
## malignant      benign      benign      benign malignant malignant malignant malignant
##      279      280      282      289      290      292      302      303
##      benign malignant      benign      benign malignant      benign      benign malignant
##      305      312      313      315      324      325      327      334
## malignant      benign malignant      benign malignant      benign      benign malignant
##      335      337      344      345      347      354      355      357
## malignant malignant      benign malignant      benign malignant      benign      benign
##      364      365      367      374      375      377      384      385
##      benign      benign malignant      benign      benign      benign      benign      benign
##      387      394      395      397      404      405      407      415
## malignant      benign      benign      benign      benign      benign      benign malignant
##      416      418      425      426      428      435      436      438
##      benign      benign      benign malignant malignant malignant malignant      benign
##      445      446      448      455      456      458      465      466
##      benign      benign      benign      benign      benign malignant      benign malignant
##      468      475      476      478      485      486      488      495
## malignant      benign      benign      benign      benign      benign malignant      benign
##      496      498      505      506      508      515      516      518
##      benign      benign      benign      benign      benign malignant malignant      benign
##      525      526      528      535      536      538      545      546
##      benign      benign      benign      benign      benign      benign      benign      benign
##      548      555      556      558      565      566      568      575
##      benign      benign      benign      benign      benign malignant      benign malignant
##      576      578      585      586      588      595      596      598
##      benign      benign      benign      benign      benign malignant      benign      benign
##      605      606      608      615      616      619      626      627
## malignant malignant      benign      benign      benign      benign      benign malignant
##      629      636      637      639      646      647      649      656
##      benign      benign malignant      benign      benign      benign malignant      benign
##      657      659      666      667      669      676      677      679
##      benign malignant      benign      benign malignant      benign      benign      benign
##      686      687      689      696      697      699
##      benign      benign      benign      benign malignant malignant
## Levels: benign malignant

```

```
CrossTable(svm_pred,test_set$Class)
```

```

##
##
##      Cell Contents
## |-----|
## |                      N |
## | Chi-square contribution |
## |                      N / Row Total |

```

```
## |          N / Col Total |
## |          N / Table Total |
## |-----|
##
##
## Total Observations in Table:  206
##
##
##          | test_set$Class
##      svm_pred |      benign | malignant | Row Total |
## -----|-----|-----|-----|
##      benign |         126 |         7 |         133 |
##              |        22.749 |        37.332 |         0.646 |
##              |         0.947 |         0.053 |         0.646 |
##              |         0.984 |         0.090 |         0.646 |
##              |         0.612 |         0.034 |         0.646 |
## -----|-----|-----|-----|
##      malignant |          2 |         71 |          73 |
##              |        41.447 |        68.016 |         0.354 |
##              |         0.027 |         0.973 |         0.354 |
##              |         0.016 |         0.910 |         0.354 |
##              |         0.010 |         0.345 |         0.354 |
## -----|-----|-----|-----|
## Column Total |         128 |          78 |          206 |
##              |         0.621 |         0.379 |         0.621 |
## -----|-----|-----|-----|
##
##
```

```
confusionMatrix(svm_pred,test_set$Class)
```

```
## Confusion Matrix and Statistics
##
##          Reference
## Prediction  benign malignant
##      benign      126         7
##      malignant     2        71
##
##          Accuracy : 0.9563
##          95% CI : (0.9187, 0.9798)
##      No Information Rate : 0.6214
##      P-Value [Acc > NIR] : <2e-16
##
##          Kappa : 0.906
##
##      Mcnemar's Test P-Value : 0.1824
##
##          Sensitivity : 0.9844
##          Specificity : 0.9103
##      Pos Pred Value : 0.9474
##      Neg Pred Value : 0.9726
##          Prevalence : 0.6214
##      Detection Rate : 0.6117
##      Detection Prevalence : 0.6456
##      Balanced Accuracy : 0.9473
```

```
##
##      'Positive' Class : benign
##
# Building model with different "C" parameter(1,2.5, 6.25)
library(kernlab)

##
## Attaching package: 'kernlab'
## The following object is masked from 'package:purrr':
##
##      cross
## The following object is masked from 'package:ggplot2':
##
##      alpha
Grid_Search <- expand.grid(.C = c(1,2.5,6.25))
(svm_model2 <- train(Class~. , data = trg_set, method = "svmLinear",tuneGrid = Grid_Search))

## Warning in .local(x, ...): Variable(s) `` constant. Cannot scale data.
## Warning in .local(x, ...): Variable(s) `` constant. Cannot scale data.
## Warning in .local(x, ...): Variable(s) `` constant. Cannot scale data.
## Support Vector Machines with Linear Kernel
##
## 477 samples
## 9 predictor
## 2 classes: 'benign', 'malignant'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 477, 477, 477, 477, 477, 477, ...
## Resampling results across tuning parameters:
##
##  C      Accuracy  Kappa
##  1.00  0.9369405  0.8552979
##  2.50  0.9371889  0.8558567
##  6.25  0.9374374  0.8564458
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was C = 6.25.
svm_pred2 <- predict(svm_model2,test_set, type = "raw")
CrossTable(svm_pred2,test_set$Class)

##
##
##      Cell Contents
## |-----|
## |                      N |
## | Chi-square contribution |
## |          N / Row Total |
## |          N / Col Total |
## |          N / Table Total |
```

```
## |-----|
##
##
## Total Observations in Table: 206
##
##
##      | test_set$Class
## svm_pred2 |    benign | malignant | Row Total |
## -----|-----|-----|-----|
##      benign |      124 |       10 |      134 |
##              |    19.932 |    32.709 |           |
##              |     0.925 |     0.075 |     0.650 |
##              |     0.969 |     0.128 |           |
##              |     0.602 |     0.049 |           |
## -----|-----|-----|-----|
##      malignant |         4 |        68 |       72 |
##                |    37.096 |    60.875 |           |
##                |     0.056 |     0.944 |     0.350 |
##                |     0.031 |     0.872 |           |
##                |     0.019 |     0.330 |           |
## -----|-----|-----|-----|
## Column Total |      128 |       78 |      206 |
##              |     0.621 |     0.379 |           |
## -----|-----|-----|-----|
##
##
```

```
confusionMatrix(svm_pred2,test_set$Class)
```

```
## Confusion Matrix and Statistics
##
##      Reference
## Prediction  benign malignant
##    benign      124         10
##    malignant     4          68
##
##      Accuracy : 0.932
##      95% CI : (0.8886, 0.9623)
##    No Information Rate : 0.6214
##    P-Value [Acc > NIR] : <2e-16
##
##      Kappa : 0.8534
##
##    Mcnemar's Test P-Value : 0.1814
##
##      Sensitivity : 0.9688
##      Specificity : 0.8718
##    Pos Pred Value : 0.9254
##    Neg Pred Value : 0.9444
##      Prevalence : 0.6214
##    Detection Rate : 0.6019
##    Detection Prevalence : 0.6505
##      Balanced Accuracy : 0.9203
##
##    'Positive' Class : benign
```

##