hw9.R

lukemcevoy

2021-12-04

```
#  First Name      : Luke
#  Last Name       : McEvoy
#  Id              : 10439121
#  purpose         : Applying SVM to Breast Cancer data

rm(list=ls())
# file<-file.choose()
file<-'/Users/lukemcevoy/Develop/stevens/f21/dataMining/week12/wisc_bc_ContinuousVar.csv'

bc_raw<-  read.csv(file, na.strings = "?",)
summary(bc_raw)
```

```
##        id              diagnosis          radius_mean      texture_mean
##  Min.   :     8670   Length:569         Min.   : 6.981   Min.   : 9.71
##  1st Qu.:   869218   Class :character   1st Qu.:11.700   1st Qu.:16.17
##  Median :   906024   Mode  :character   Median :13.370   Median :18.84
##  Mean   : 30371831                      Mean   :14.127   Mean   :19.29
##  3rd Qu.:  8813129                      3rd Qu.:15.780   3rd Qu.:21.80
##  Max.   :911320502                      Max.   :28.110   Max.   :39.28
##  perimeter_mean     area_mean      smoothness_mean   compactness_mean
##  Min.   : 43.79   Min.   : 143.5   Min.   :0.05263   Min.   :0.01938
##  1st Qu.: 75.17   1st Qu.: 420.3   1st Qu.:0.08637   1st Qu.:0.06492
##  Median : 86.24   Median : 551.1   Median :0.09587   Median :0.09263
##  Mean   : 91.97   Mean   : 654.9   Mean   :0.09636   Mean   :0.10434
##  3rd Qu.:104.10   3rd Qu.: 782.7   3rd Qu.:0.10530   3rd Qu.:0.13040
##  Max.   :188.50   Max.   :2501.0   Max.   :0.16340   Max.   :0.34540
##  concavity_mean    concave.points_mean symmetry_mean    fractal_dimension_mean
##  Min.   :0.00000   Min.   :0.00000     Min.   :0.1060   Min.   :0.04996
##  1st Qu.:0.02956   1st Qu.:0.02031     1st Qu.:0.1619   1st Qu.:0.05770
##  Median :0.06154   Median :0.03350     Median :0.1792   Median :0.06154
##  Mean   :0.08880   Mean   :0.04892     Mean   :0.1812   Mean   :0.06280
##  3rd Qu.:0.13070   3rd Qu.:0.07400     3rd Qu.:0.1957   3rd Qu.:0.06612
##  Max.   :0.42680   Max.   :0.20120     Max.   :0.3040   Max.   :0.09744
##    radius_se        texture_se      perimeter_se       area_se
##  Min.   :0.1115   Min.   :0.3602   Min.   : 0.757   Min.   :  6.802
##  1st Qu.:0.2324   1st Qu.:0.8339   1st Qu.: 1.606   1st Qu.: 17.850
##  Median :0.3242   Median :1.1080   Median : 2.287   Median : 24.530
##  Mean   :0.4052   Mean   :1.2169   Mean   : 2.866   Mean   : 40.337
##  3rd Qu.:0.4789   3rd Qu.:1.4740   3rd Qu.: 3.357   3rd Qu.: 45.190
##  Max.   :2.8730   Max.   :4.8850   Max.   :21.980   Max.   :542.200
##  smoothness_se      compactness_se      concavity_se     concave.points_se
##  Min.   :0.001713   Min.   :0.002252   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.005169   1st Qu.:0.013080   1st Qu.:0.01509   1st Qu.:0.007638
```

```
##  Median :0.006380   Median :0.020450   Median :0.02589   Median :0.010930
##  Mean   :0.007041   Mean   :0.025478   Mean   :0.03189   Mean   :0.011796
##  3rd Qu.:0.008146   3rd Qu.:0.032450   3rd Qu.:0.04205   3rd Qu.:0.014710
##  Max.   :0.031130   Max.   :0.135400   Max.   :0.39600   Max.   :0.052790
##   symmetry_se       fractal_dimension_se  radius_worst   texture_worst
##  Min.   :0.007882   Min.   :0.0008948    Min.   : 7.93   Min.   :12.02
##  1st Qu.:0.015160   1st Qu.:0.0022480    1st Qu.:13.01   1st Qu.:21.08
##  Median :0.018730   Median :0.0031870    Median :14.97   Median :25.41
##  Mean   :0.020542   Mean   :0.0037949    Mean   :16.27   Mean   :25.68
##  3rd Qu.:0.023480   3rd Qu.:0.0045580    3rd Qu.:18.79   3rd Qu.:29.72
##  Max.   :0.078950   Max.   :0.0298400    Max.   :36.04   Max.   :49.54
##  perimeter_worst    area_worst     smoothness_worst  compactness_worst
##  Min.   : 50.41   Min.   : 185.2   Min.   :0.07117   Min.   :0.02729
##  1st Qu.: 84.11   1st Qu.: 515.3   1st Qu.:0.11660   1st Qu.:0.14720
##  Median : 97.66   Median : 686.5   Median :0.13130   Median :0.21190
##  Mean   :107.26   Mean   : 880.6   Mean   :0.13237   Mean   :0.25427
##  3rd Qu.:125.40   3rd Qu.:1084.0   3rd Qu.:0.14600   3rd Qu.:0.33910
##  Max.   :251.20   Max.   :4254.0   Max.   :0.22260   Max.   :1.05800
##  concavity_worst  concave.points_worst symmetry_worst   fractal_dimension_worst
##  Min.   :0.0000   Min.   :0.00000      Min.   :0.1565   Min.   :0.05504
##  1st Qu.:0.1145   1st Qu.:0.06493      1st Qu.:0.2504   1st Qu.:0.07146
##  Median :0.2267   Median :0.09993      Median :0.2822   Median :0.08004
##  Mean   :0.2722   Mean   :0.11461      Mean   :0.2901   Mean   :0.08395
##  3rd Qu.:0.3829   3rd Qu.:0.16140      3rd Qu.:0.3179   3rd Qu.:0.09208
##  Max.   :1.2520   Max.   :0.29100      Max.   :0.6638   Max.   :0.20750
```

```
summary(bc_raw$diagnosis)
```

```
##    Length     Class      Mode
##       569 character character
```

```
table(bc_raw$diagnosis)
```

```
##
##   B   M
## 357 212
```

```
mmnorm2 <-function(x)
{z<-((x-min(x))/(max(x)-min(x)))
return(z)
}

bc<- data.frame(diagnosis=as.factor(bc_raw$diagnosis)
                ,       radius_mean =mmnorm2(   bc_raw$radius_mean)
                ,      texture_mean=mmnorm2(bc_raw$texture_mean)
                ,      perimeter_mean=mmnorm2(bc_raw$perimeter_mean)
                ,      area_mean=mmnorm2(bc_raw$area_mean)
                ,      smoothness_mean=mmnorm2(bc_raw$smoothness_mean)
                ,      compactness_mean=mmnorm2(bc_raw$compactness_mean)
                ,      concavity_mean=mmnorm2(bc_raw$concavity_mean)
                ,      concave.points_mean=mmnorm2(bc_raw$concave.points_mean)
                ,      symmetry_mean=mmnorm2(bc_raw$symmetry_mean)
                ,      fractal_dimension_mean=mmnorm2(bc_raw$fractal_dimension_mean)
                ,      radius_se=mmnorm2(bc_raw$radius_se)
                ,      perimeter_se=mmnorm2(bc_raw$perimeter_se)
                ,      texture_se=mmnorm2(bc_raw$texture_se)
```

```
                ,       area_se=mmnorm2(bc_raw$area_se)
                ,       smoothness_se=mmnorm2(bc_raw$smoothness_se)
                ,       compactness_se=mmnorm2(bc_raw$compactness_se)
                ,       concavity_se=mmnorm2(bc_raw$concavity_se)
                ,       concave.points_se =mmnorm2(bc_raw$concave.points_se)
                ,       symmetry_se=mmnorm2(bc_raw$symmetry_se)
                ,        fractal_dimension_se=mmnorm2(bc_raw$fractal_dimension_se)
                ,        radius_worst=mmnorm2(bc_raw$radius_worst)
                ,        texture_worst=mmnorm2(bc_raw$texture_worst)
                ,        perimeter_worst=mmnorm2(bc_raw$perimeter_worst)
                ,        area_worst=mmnorm2(bc_raw$area_worst)
                ,        smoothness_worst=mmnorm2(bc_raw$smoothness_worst)
                ,        compactness_worst=mmnorm2(bc_raw$compactness_worst)
                ,        concavity_worst=mmnorm2(bc_raw$concavity_worst)
                ,        concave.points_worst=mmnorm2(bc_raw$concave.points_worst)
                ,        symmetry_worst=mmnorm2(bc_raw$symmetry_worst)
                ,        fractal_dimension_worst=mmnorm2(bc_raw$fractal_dimension_worst)
)



# create test and training dataset
index <- seq (1,nrow(bc),by=3)
test<-bc[index,]
training<-bc[-index,]
ncol(test)
```

```
## [1] 31
```

```
ncol(training)
```

```
## [1] 31
```

```
library(e1071)
?svm()
## svm
svm.model <- svm( diagnosis~., data = training  )
svm.pred <- predict(svm.model,  test)

table(actual=test[,5],svm.pred )
```

```
##                     svm.pred
## actual               B M
##   0.0330010604453871 1 0
##   0.0370731707317073 1 0
##   0.0428419936373277 1 0
##   0.0446235418875928 1 0
##   0.0497985153764581 1 0
##   0.0534040296924708 1 0
##   0.0553128313891834 1 0
##   0.0577306468716861 1 0
##   0.0613361611876988 1 0
##   0.0620996818663839 1 0
##   0.0625238600212089 1 0
```

```
##   0.0714316012725345 1 0
##   0.0719406150583245 1 0
##   0.0754612937433722 1 0
##   0.0755461293743372 2 0
##   0.0769459172852598 1 0
##   0.0806362672322375 1 0
##   0.0809331919406151 1 0
##   0.087423117709438  1 0
##   0.0898409331919406 1 0
##   0.0913679745493107 1 0
##   0.0925132555673383 1 0
##   0.0942099681866384 1 0
##   0.0945917285259809 1 0
##   0.0947189819724284 1 0
##   0.0960763520678685 1 0
##   0.0972216330858961 1 0
##   0.0974337221633086 1 0
##   0.0977306468716861 1 0
##   0.101124072110286  1 0
##   0.102269353128314  1 0
##   0.102863202545069  1 0
##   0.102905620360551  0 1
##   0.103753976670201  1 0
##   0.104559915164369  1 0
##   0.106299045599152  1 0
##   0.106977730646872  1 0
##   0.107359490986214  1 0
##   0.109650053022269  1 0
##   0.109946977730647  1 0
##   0.110031813361612  1 0
##   0.110201484623542  1 0
##   0.111474019088017  1 0
##   0.111516436903499  1 0
##   0.112916224814422  1 0
##   0.115206786850477  1 0
##   0.115673382820785  1 0
##   0.11983032873807   1 0
##   0.121993637327678  1 0
##   0.122332979851538  1 0
##   0.12237539766702   1 0
##   0.122460233297985  1 0
##   0.127168610816543  1 0
##   0.128313891834571  1 0
##   0.129713679745493  1 0
##   0.129968186638388  1 0
##   0.130477200424178  1 0
##   0.132258748674443  1 0
##   0.133361611876988  1 0
##   0.133700954400848  1 0
##   0.135821845174973  1 0
##   0.140996818663839  0 1
##   0.142905620360551  1 0
##   0.143414634146341  1 0
##   0.143541887592789  1 0
```

```
##   0.146553552492047  1 0
##   0.147783669141039  1 0
##   0.147868504772004  1 0
##   0.148589607635207  1 0
##   0.149692470837752  1 0
##   0.151770943796394  1 0
##   0.152279957582185  1 0
##   0.152958642629905  1 0
##   0.15389183457052   1 0
##   0.15457051961824   1 0
##   0.15728525980912   1 0
##   0.158176033934252  1 0
##   0.159448568398727  1 0
##   0.159703075291622  1 0
##   0.161908801696713  1 0
##   0.164029692470838  0 1
##   0.164199363732768  1 0
##   0.167041357370095  1 0
##   0.167507953340403  1 0
##   0.168865323435843  1 0
##   0.168950159066808  1 0
##   0.169416755037116  1 0
##   0.174591728525981  1 0
##   0.175864262990456  1 0
##   0.176330858960764  0 1
##   0.177094379639449  1 0
##   0.178960763520679  0 1
##   0.181378579003181  1 0
##   0.181675503711559  1 0
##   0.182014846235419  0 1
##   0.183160127253446  2 0
##   0.185408271474019  1 0
##   0.188292682926829  1 0
##   0.189395546129374  1 0
##   0.190498409331919  1 0
##   0.190965005302227  1 0
##   0.191177094379639  1 0
##   0.191898197242842  0 1
##   0.19219512195122   1 0
##   0.19338282078473   1 0
##   0.193425238600212  0 1
##   0.194655355249205  1 0
##   0.20746553552492   1 0
##   0.20763520678685   0 1
##   0.210901378579003  1 0
##   0.211113467656416  1 0
##   0.211664899257688  0 1
##   0.212386002120891  1 0
##   0.213191940615058  1 0
##   0.217773064687169  0 1
##   0.217857900318134  1 0
##   0.218579003181336  1 1
##   0.220233297985154  1 0
##   0.221633085896076  1 0
```

```
##   0.225365853658537  1 0
##   0.226765641569459  0 1
##   0.227953340402969  1 0
##   0.228632025450689  0 1
##   0.230668080593849  1 0
##   0.231601272534464  1 0
##   0.231898197242842  1 0
##   0.241484623541888  0 1
##   0.244326617179215  0 1
##   0.260911983032874  0 1
##   0.267232237539767  0 1
##   0.273594909862142  1 0
##   0.27558854718982   0 1
##   0.279787910922587  0 1
##   0.282629904559915  0 1
##   0.283987274655355  0 1
##   0.285980911983033  0 1
##   0.292428419936373  0 1
##   0.294634146341463  1 0
##   0.302905620360551  0 1
##   0.32271474019088   0 1
##   0.322841993637328  0 1
##   0.332895015906681  0 1
##   0.333107104984093  0 1
##   0.333361611876988  0 1
##   0.334931071049841  0 1
##   0.341251325556734  0 1
##   0.342778366914104  0 1
##   0.347910922587487  0 1
##   0.355037115588547  0 1
##   0.35677624602333   0 1
##   0.359066808059385  0 1
##   0.359787910922587  0 1
##   0.36373276776246   0 1
##   0.36627783669141   0 1
##   0.380275715800636  0 1
##   0.380699893955461  0 1
##   0.385365853658537  0 1
##   0.392152704135737  0 1
##   0.395546129374337  0 2
##   0.400636267232238  0 1
##   0.403181336161188  0 1
##   0.407423117709438  0 1
##   0.415482502651113  0 1
##   0.424814422057264  0 1
##   0.426086956521739  0 1
##   0.432025450689289  0 1
##   0.44475079533404   0 1
##   0.450689289501591  0 1
##   0.454082714740191  0 1
##   0.455355249204666  0 1
##   0.457900318133616  0 1
##   0.462566277836691  0 1
##   0.467232237539767  0 1
```

```
##   0.469353128313892  0 1
##   0.473594909862142  0 1
##   0.475715800636267  0 1
##   0.479533404029692  0 1
##   0.486320254506893  0 1
##   0.493107104984093  0 1
##   0.527041357370095  0 1
##   0.535949098621421  0 1
##   0.566489925768823  0 1
##   0.567762460233298  0 1
##   0.579215270413574  0 1
##   0.68016967126193   0 1
##   0.893531283138918  0 1
```

```
SVM_wrong<- (test$diagnosis!=svm.pred)
SVM_wrong
```

```
##   [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
##  [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [85] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [97] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
rate<-sum(SVM_wrong)/length(SVM_wrong)
rate
```

```
## [1] 0.01578947
```