SVM.R

patriciamaya

2020-11-03

```
#SVM for binary classification

#we will try to predict species of flowers (versicolor/ not versicolor)
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 ...
```

```
iris$Speciesclass[iris$Species == "versicolor"] <- "TRUE"
iris$Speciesclass[iris$Species != "versicolor"] <- "FALSE"
iris$Speciesclass <- factor(iris$Speciesclass)
iris$Species <- NULL
View(iris)

set.seed(123)
indx<- sample(2, nrow(iris), replace=T, prob=c(0.8,0.2))
train <- iris[indx==1, ]
test <- iris[indx==2, ]

library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.0.2
```

```
svmModel<- svm(Speciesclass ~ . , data=train, type="C-classification" , cost = 100, kernel = "linear", g
#C-classification - binary classification
#cost - cost of misclassification, if high-> not many misclassified points, margin can be small
       #if low -> make more mistakes, margn is larger
#to fin best cost and gamma values, use cross validation
svmModel
```

```
##
## Call:
## svm(formula = Speciesclass ~ ., data = train, type = "C-classification",
##     cost = 100, kernel = "linear", gamma = 1)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
```

```
##        cost:  100
##
## Number of Support Vectors:  73
```

```
svmModel$coefs #returns alpha_i * y_i (y_i - is the label--> class 1 and class -1)
```

```
##              [,1]
##  [1,]  100.000000
##  [2,]  100.000000
##  [3,]  100.000000
##  [4,]  100.000000
##  [5,]  100.000000
##  [6,]   63.950167
##  [7,]  100.000000
##  [8,]  100.000000
##  [9,]  100.000000
## [10,]  100.000000
## [11,]  100.000000
## [12,]  100.000000
## [13,]  100.000000
## [14,]  100.000000
## [15,]  100.000000
## [16,]  100.000000
## [17,]  100.000000
## [18,]    6.666156
## [19,]  100.000000
## [20,]  100.000000
## [21,]  100.000000
## [22,]  100.000000
## [23,]  100.000000
## [24,]  100.000000
## [25,]  100.000000
## [26,]  100.000000
## [27,]  100.000000
## [28,]  100.000000
## [29,]  100.000000
## [30,]  100.000000
## [31,]  100.000000
## [32,]  100.000000
## [33,]   42.399644
## [34,]  100.000000
## [35,]  100.000000
## [36,]  100.000000
## [37,]  100.000000
## [38,] -100.000000
## [39,] -100.000000
## [40,] -100.000000
## [41,] -100.000000
## [42,] -100.000000
## [43,] -100.000000
## [44,] -100.000000
## [45,] -100.000000
## [46,] -100.000000
## [47,] -100.000000
## [48,] -100.000000
```

```
## [49,] -100.000000
## [50,] -100.000000
## [51,] -100.000000
## [52,] -100.000000
## [53,]  -48.616666
## [54,] -100.000000
## [55,] -100.000000
## [56,] -100.000000
## [57,] -100.000000
## [58,] -100.000000
## [59,] -100.000000
## [60,] -100.000000
## [61,] -100.000000
## [62,] -100.000000
## [63,] -100.000000
## [64,] -100.000000
## [65,] -100.000000
## [66,] -100.000000
## [67,]  -64.399301
## [68,] -100.000000
## [69,] -100.000000
## [70,] -100.000000
## [71,] -100.000000
## [72,] -100.000000
## [73,] -100.000000
```

```
#alphas of non SV are zero
svmModel$SV  #since we have 4 input variables, our SV are in 4 dimensions
```

```
##     Sepal.Length Sepal.Width Petal.Length  Petal.Width
## 2    -1.14301072 -0.05031691  -1.35775785 -1.325369188
## 9    -1.75988984 -0.28448405  -1.35775785 -1.325369188
## 10   -1.14301072  0.18385024  -1.30021365 -1.457687399
## 13   -1.26638655 -0.05031691  -1.35775785 -1.457687399
## 14   -1.88326566 -0.05031691  -1.53039046 -1.457687399
## 25   -1.26638655  0.88635167  -1.07003683 -1.325369188
## 26   -1.01963490 -0.05031691  -1.24266944 -1.325369188
## 30   -1.38976237  0.41801738  -1.24266944 -1.325369188
## 35   -1.14301072  0.18385024  -1.30021365 -1.325369188
## 39   -1.75988984 -0.05031691  -1.41530205 -1.325369188
## 42   -1.63651402 -1.68948691  -1.41530205 -1.193050977
## 102  -0.03262832 -0.75281834   0.77137766  0.924040399
## 103   1.57125738 -0.05031691   1.23173129  1.188676821
## 105   0.83100244 -0.05031691   1.17418709  1.320995032
## 108   1.81800903 -0.28448405   1.46190810  0.791722188
## 109   1.07775409 -1.22115262   1.17418709  0.791722188
## 112   0.70762662 -0.75281834   0.88646607  0.924040399
## 113   1.20112991 -0.05031691   1.00155448  1.188676821
## 117   0.83100244 -0.05031691   1.00155448  0.791722188
## 119   2.31151232 -0.98698548   1.80717332  1.453313243
## 120   0.21412333 -1.92365405   0.71383346  0.394767555
## 122  -0.27937996 -0.51865119   0.65628926  1.056358610
## 123   2.31151232 -0.51865119   1.69208491  1.056358610
## 124   0.58425080 -0.75281834   0.65628926  0.791722188
## 127   0.46087498 -0.51865119   0.59874505  0.791722188
```

```
## 128   0.33749915 -0.05031691   0.65628926  0.791722188
## 129   0.70762662 -0.51865119   1.05909868  1.188676821
## 130   1.69463320 -0.05031691   1.17418709  0.527085766
## 131   1.94138485 -0.51865119   1.34681969  0.924040399
## 133   0.70762662 -0.51865119   1.05909868  1.320995032
## 134   0.58425080 -0.51865119   0.77137766  0.394767555
## 135   0.33749915 -0.98698548   1.05909868  0.262449344
## 136   2.31151232 -0.05031691   1.34681969  1.453313243
## 138   0.70762662  0.18385024   1.00155448  0.791722188
## 143  -0.03262832 -0.75281834   0.77137766  0.924040399
## 147   0.58425080 -1.22115262   0.71383346  0.924040399
## 150   0.09074751 -0.05031691   0.77137766  0.791722188
## 51    1.44788156  0.41801738   0.54120085  0.262449344
## 52    0.70762662  0.41801738   0.42611245  0.394767555
## 53    1.32450574  0.18385024   0.65628926  0.394767555
## 54   -0.40275579 -1.68948691   0.13839143  0.130131133
## 55    0.83100244 -0.51865119   0.48365665  0.394767555
## 56   -0.15600414 -0.51865119   0.42611245  0.130131133
## 57    0.58425080  0.65218452   0.54120085  0.527085766
## 58   -1.14301072 -1.45531977  -0.26441799 -0.266823500
## 60   -0.77288325 -0.75281834   0.08084723  0.262449344
## 62    0.09074751 -0.05031691   0.25347984  0.394767555
## 64    0.33749915 -0.28448405   0.54120085  0.262449344
## 66    1.07775409  0.18385024   0.36856824  0.262449344
## 70   -0.27937996 -1.22115262   0.08084723 -0.134505289
## 71    0.09074751  0.41801738   0.59874505  0.791722188
## 72    0.33749915 -0.51865119   0.13839143  0.130131133
## 74    0.33749915 -0.51865119   0.54120085 -0.002187078
## 75    0.70762662 -0.28448405   0.31102404  0.130131133
## 76    0.95437827 -0.05031691   0.36856824  0.262449344
## 77    1.20112991 -0.51865119   0.59874505  0.262449344
## 78    1.07775409 -0.05031691   0.71383346  0.659403977
## 79    0.21412333 -0.28448405   0.42611245  0.394767555
## 80   -0.15600414 -0.98698548  -0.14932958 -0.266823500
## 83   -0.03262832 -0.75281834   0.08084723 -0.002187078
## 84    0.21412333 -0.75281834   0.77137766  0.527085766
## 85   -0.52613161 -0.05031691   0.42611245  0.394767555
## 86    0.21412333  0.88635167   0.42611245  0.527085766
## 90   -0.40275579 -1.22115262   0.13839143  0.130131133
## 92    0.33749915 -0.05031691   0.48365665  0.262449344
## 93   -0.03262832 -0.98698548   0.13839143 -0.002187078
## 94   -1.01963490 -1.68948691  -0.26441799 -0.266823500
## 95   -0.27937996 -0.75281834   0.25347984  0.130131133
## 96   -0.15600414 -0.05031691   0.25347984 -0.002187078
## 97   -0.15600414 -0.28448405   0.25347984  0.130131133
## 98    0.46087498 -0.28448405   0.31102404  0.130131133
## 99   -0.89625908 -1.22115262  -0.43705060 -0.134505289
## 100  -0.15600414 -0.51865119   0.19593563  0.130131133
```

```
svmModel$index  #what instances are SV
```

```
##  [1]   2   6   7   9  10  17  18  22  25  29  32  84  85  86  87  88  90  91  94
## [20]  95  96  98  99 100 102 103 104 105 106 107 108 109 110 111 115 118 121  40
## [39]  41  42  43  44  45  46  47  48  50  52  53  55  56  57  59  60  61  62  63
## [58]  64  65  68  69  70  71  72  74  75  76  77  78  79  80  81  82
```

```r
#classifier - w x + b
# w = sum ( alpha_i * y_i * x_i) in SV
w<- t(svmModel$coefs) %*% as.matrix(train[svmModel$index, 1:4])
w
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width
## [1,]    0.2097539   0.3829103    -4.209019    1.568842
```

```r
b<- - svmModel$rho #negative of intercept
b
```

```
## [1] 0.692505
```

```r
#w x + b, if >0 - assign to positive class, o.w - neg class

svmModel$decision.values  #predicted classes, assign given the sign
```

```
##      FALSE/TRUE
## 1    2.00254574
## 2    0.88886013
## 3    1.38431103
## 6    3.06941829
## 7    1.90761496
## 9    0.51925472
## 10   0.68477673
## 12   1.41804663
## 13   0.58225762
## 14   0.84074241
## 15   3.55462028
## 17   3.62691726
## 18   2.27722036
## 19   2.68056137
## 22   2.83245218
## 23   2.61037124
## 25   0.99992241
## 26   0.64203853
## 27   2.03125164
## 28   1.89509889
## 29   1.82450766
## 30   0.96618681
## 33   2.88022008
## 34   3.60001907
## 35   0.95945135
## 36   1.61946943
## 37   2.26963203
## 38   1.87398132
## 39   0.86859543
## 40   1.65320503
## 41   2.38466721
## 42  -0.29456384
## 43   1.28852737
## 44   2.79056685
## 45   2.48491918
## 46   1.13160686
## 47   2.35369416
## 48   1.21300840
```

```
## 49   2.34695871
## 51   0.67600669
## 52   1.03786347
## 53   0.43003797
## 54  -0.99165677
## 55   0.09055274
## 56  -0.57484486
## 57   1.21182669
## 58  -0.82165880
## 60   0.16647280
## 61  -1.90834426
## 62   0.87641632
## 63  -1.86600716
## 64  -0.24124220
## 66   0.78838128
## 69  -1.12565199
## 70  -0.94977144
## 71   1.28412367
## 72   0.24974040
## 73  -1.02132516
## 74  -1.00055741
## 75   0.13736581
## 76   0.54648743
## 77  -0.36708770
## 78   0.56619073
## 79   0.28025401
## 80  -0.42505324
## 81  -1.05229056
## 82  -1.18759044
## 83  -0.19130911
## 84  -0.70125175
## 85   0.29865266
## 86   1.60475847
## 90  -0.57172483
## 91  -1.19393245
## 92   0.10809851
## 93  -0.54064982
## 94  -0.99969688
## 95  -0.39861449
## 96  -0.01146332
## 97   0.05324533
## 98   0.07351003
## 99   0.14496179
## 100 -0.01734589
## 101  1.87202663
## 102  0.05891634
## 103  0.53822809
## 105  0.76071013
## 108 -0.98940492
## 109 -1.32396242
## 110  2.64990078
## 112 -0.02826582
## 113  0.99994339
## 115  1.64225540
```

```
## 116  2.12026250
## 117  0.08013587
## 119 -0.95446662
## 120 -1.88638147
## 121  1.72240297
## 122  0.75845063
## 123 -1.07980907
## 124  0.22263064
## 125  1.31916393
## 127  0.54004346
## 128  0.78867277
## 129  0.31292517
## 130 -0.66384238
## 131 -0.61401890
## 133  0.58759979
## 134 -0.67017674
## 135 -2.12551277
## 136  1.00039517
## 138  0.25817395
## 140  1.38121199
## 141  1.86263059
## 142  2.34868545
## 143  0.05891634
## 144  1.41172560
## 146  1.93548897
## 147 -0.06200142
## 148  1.04760933
## 149  2.33696392
## 150  0.44606751
```