Lab10 - Support Vector Machines

Machine Learning usando o R - Análise Macro

Thalles Quinaglia Liduares

2022-08-23

Upload base de dados

```
setwd("C:\\Program Files\\R\\Dados\\ML")

data<-read.csv("letterdata.csv", stringsAsFactors = TRUE)

attach(data)
```

Upload pacotes

```
library(e1071)
library(caret)
```

```
## Carregando pacotes exigidos: ggplot2
```

```
## Carregando pacotes exigidos: lattice
```

Estrutura dos dados

```
str(data)
```

```
## 'data.frame':    20000 obs. of  17 variables:
##  $ letter: Factor w/ 26 levels "A","B","C","D",..: 20 9 4 14 7 19 2 1 10 13 ...
##  $ xbox  : int  2 5 4 7 2 4 4 1 2 11 ...
##  $ ybox  : int  8 12 11 11 1 11 2 1 2 15 ...
##  $ width : int  3 3 6 6 3 5 5 3 4 13 ...
##  $ height: int  5 7 8 6 1 8 4 2 4 9 ...
##  $ onpix : int  1 2 6 3 1 3 4 1 2 7 ...
##  $ xbar  : int  8 10 10 5 8 8 8 8 10 13 ...
##  $ ybar  : int  13 5 6 9 6 8 7 2 6 2 ...
##  $ x2bar : int  0 5 2 4 6 6 6 2 2 6 ...
##  $ y2bar : int  6 4 6 6 6 9 6 2 6 2 ...
##  $ xybar : int  6 13 10 4 6 5 7 8 12 12 ...
##  $ x2ybar: int  10 3 3 4 5 6 6 2 4 1 ...
##  $ xy2bar: int  8 9 7 10 9 6 6 8 8 9 ...
##  $ xedge : int  0 2 3 6 1 0 2 1 1 8 ...
##  $ xedgey: int  8 8 7 10 7 8 8 6 6 1 ...
##  $ yedge : int  0 4 3 2 5 9 7 2 1 1 ...
##  $ yedgex: int  8 10 9 8 10 7 10 7 7 8 ...
```

Normalizando os dados numericos

```
data[-1]<-scale(data[-1])

#summary(data)
```

Divisão da amostra entre treino e teste

```
set.seed(1608)

part_data<-floor(0.75*nrow(data))

treino_data <-sample(seq_len(nrow(data)), size = part_data)

treino<-data[treino_data, ]

teste<-data[-treino_data,]
```

Treinando o modelo

```
classf<-e1071::svm(letter~.,
                       data=treino,
                       type="C-classification",
                       kernel="linear",
                      cost=100)
```

Predição do modelo

```
prev.model<-predict(classf, newdata=teste[-1])
```

Acurácia do modelo

```
confusionMatrix(prev.model, teste$letter)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A   B   C   D   E   F   G   H   I   J   K   L   M   N   O   P   Q
##          A 165   2   0   0   0   0   0   0   0   1   0   0   0   0   2   0   0
##          B   0 148   0  12   3   1   0   4   0   0   1   0   1   1   0   1   1
##          C   0   0 171   0   6   0  13   4   0   0   1   1   1   0   4   0   0
##          D   0   1   0 170   0   0   5  10   2   0   2   1   0   1   2   1   1
##          E   0   1   5   0 148   3   1   1   0   0   3   6   0   0   0   0   5
##          F   0   1   0   0   2 185   2   4   2   3   0   0   0   0   0  19   0
##          G   0   4   5   0   9   3 146   3   0   0   6   2   0   0   1   7   1
##          H   0   5   4   0   0   1   1 131   0   0   0   1   6   1   8   1   0
##          I   0   0   0   0   0   2   0   0 159   5   0   0   0   0   0   0   0
##          J   4   1   0   2   0   5   1   1   7 155   0   0   0   0   0   0   1
##          K   4   1   4   2   2   0   1   7   0   0 150   1   0   2   0   0   0
##          L   1   0   0   0   3   0   2   0   3   0   2 163   0   0   0   0   1
##          M   0   0   0   1   0   0   1   1   0   0   1   0 182   1   7   0   0
##          N   1   0   0   2   0   1   0   2   0   2   0   1   2 201   2   0   0
##          O   0   0   1   5   0   0   4  11   0   2   1   0   0   4 150   2   5
##          P   0   0   0   1   0   5   0   4   0   0   0   0   0   0   1 188   0
##          Q   1   0   0   1   5   0   8   2   0   0   0   7   0   0   7   2 156
##          R   2   6   0   1   3   0   2  11   0   0   9   0   0   1   2   0   0
##          S   0   3   0   0   5   1   6   0  12   2   0   1   0   0   0   0  12
##          T   0   0   0   0   5   4   0   1   0   0   0   0   0   0   0   0   0
##          U   2   0   1   0   0   0   0   1   0   0   0   0   1   0   1   0   0
##          V   0   1   0   0   0   0   3   1   0   0   0   0   0   1   0   1   1
##          W   0   0   0   0   0   0   0   0   0   0   1   0   1   1  10   0   1
##          X   0   1   0   1   2   0   2   2   0   3   4   3   0   0   0   0   0
##          Y   2   0   0   0   0   1   0   3   1   0   0   0   0   0   0   1   0
##          Z   0   0   0   0   1   0   0   0   2   3   0   0   0   0   0   0   0
##           Reference
## Prediction   R   S   T   U   V   W   X   Y   Z
##          A   3   1   0   1   0   0   0   0   2
##          B   9   4   0   0   3   0   1   0   0
##          C   0   0   0   0   0   0   0   0   0
##          D   3   0   2   0   0   0   2   0   0
##          E   1  11   3   0   0   0   1   0   3
##          F   0   6   3   0   0   0   0   4   0
##          G   3   1   1   2   0   0   0   0   0
##          H   6   0   1   2   3   0   0   0   0
##          I   0   3   0   0   0   0   2   0   0
##          J   0   0   0   0   0   0   0   0   5
##          K   8   0   3   3   0   0   5   0   0
##          L   0   0   0   0   0   0   1   0   0
##          M   2   0   0   3   0   8   0   0   0
##          N   4   0   0   4   0   3   0   0   0
##          O   1   1   0   1   0   1   0   0   0
##          P   0   0   1   0   2   0   0   2   0
##          Q   0   5   0   0   0   0   0   2   1
##          R 158   0   0   0   0   0   1   0   0
##          S   0 131   2   0   0   0   1   1  15
##          T   1   1 164   1   0   0   4   1   2
##          U   0   0   1 184   0   1   0   0   0
##          V   0   0   0   0 154   0   0   2   0
##          W   0   0   0   1   3 194   0   0   0
```

```
##          X   0   1   1   0   0   0 181   0   0
##          Y   0   0   3   0   3   0   0 169   0
##          Z   0  12   1   0   0   0   1   0 153
##
## Overall Statistics
##
##                Accuracy : 0.8512
##                  95% CI : (0.841, 0.861)
##     No Information Rate : 0.0446
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8452
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E Class: F
## Sensitivity            0.9066   0.8457   0.8953   0.8586   0.7629   0.8726
## Specificity            0.9975   0.9913   0.9938   0.9931   0.9908   0.9904
## Pos Pred Value         0.9322   0.7789   0.8507   0.8374   0.7708   0.8009
## Neg Pred Value         0.9965   0.9944   0.9958   0.9942   0.9904   0.9943
## Prevalence             0.0364   0.0350   0.0382   0.0396   0.0388   0.0424
## Detection Rate         0.0330   0.0296   0.0342   0.0340   0.0296   0.0370
## Detection Prevalence   0.0354   0.0380   0.0402   0.0406   0.0384   0.0462
## Balanced Accuracy      0.9521   0.9185   0.9445   0.9259   0.8769   0.9315
##                      Class: G Class: H Class: I Class: J Class: K Class: L
## Sensitivity            0.7374   0.6422   0.8457   0.8807   0.8287   0.8717
## Specificity            0.9900   0.9917   0.9975   0.9944   0.9911   0.9973
## Pos Pred Value         0.7526   0.7661   0.9298   0.8516   0.7772   0.9261
## Neg Pred Value         0.9892   0.9849   0.9940   0.9956   0.9936   0.9950
## Prevalence             0.0396   0.0408   0.0376   0.0352   0.0362   0.0374
## Detection Rate         0.0292   0.0262   0.0318   0.0310   0.0300   0.0326
## Detection Prevalence   0.0388   0.0342   0.0342   0.0364   0.0386   0.0352
## Balanced Accuracy      0.8637   0.8169   0.9216   0.9375   0.9099   0.9345
##                      Class: M Class: N Class: O Class: P Class: Q Class: R
## Sensitivity            0.9381   0.9393   0.7614   0.8430   0.8432   0.7940
## Specificity            0.9948   0.9950   0.9919   0.9967   0.9915   0.9921
## Pos Pred Value         0.8792   0.8933   0.7937   0.9216   0.7919   0.8061
## Neg Pred Value         0.9975   0.9973   0.9902   0.9927   0.9940   0.9915
## Prevalence             0.0388   0.0428   0.0394   0.0446   0.0370   0.0398
## Detection Rate         0.0364   0.0402   0.0300   0.0376   0.0312   0.0316
## Detection Prevalence   0.0414   0.0450   0.0378   0.0408   0.0394   0.0392
## Balanced Accuracy      0.9665   0.9671   0.8767   0.9198   0.9174   0.8930
##                      Class: S Class: T Class: U Class: V Class: W Class: X
## Sensitivity            0.7401   0.8817   0.9109   0.9167   0.9372   0.9050
## Specificity            0.9874   0.9958   0.9983   0.9979   0.9962   0.9958
## Pos Pred Value         0.6823   0.8913   0.9583   0.9390   0.9151   0.9005
## Neg Pred Value         0.9904   0.9954   0.9963   0.9971   0.9973   0.9960
## Prevalence             0.0354   0.0372   0.0404   0.0336   0.0414   0.0400
## Detection Rate         0.0262   0.0328   0.0368   0.0308   0.0388   0.0362
## Detection Prevalence   0.0384   0.0368   0.0384   0.0328   0.0424   0.0402
## Balanced Accuracy      0.8637   0.9388   0.9546   0.9573   0.9667   0.9504
##                      Class: Y Class: Z
## Sensitivity            0.9337   0.8453
## Specificity            0.9971   0.9958
```

```
## Pos Pred Value         0.9235   0.8844
## Neg Pred Value         0.9975   0.9942
## Prevalence             0.0362   0.0362
## Detection Rate         0.0338   0.0306
## Detection Prevalence   0.0366   0.0346
## Balanced Accuracy      0.9654   0.9206
```

O modelo SVM proposto neste exercicio possui acurácia de 85.12%.