svm_lab.R

alexaubrey

Mon May 28 10:56:27 2018

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

```
## Warning: package 'caret' was built under R version 3.4.4
```

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

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

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unknown timezone 'zone/tz/2018c.
## 1.0/zoneinfo/America/Chicago'
```

```
data <- read.csv(file='/Users/alexaubrey/Documents/School/MSDS/PredAnalytics/inclass_ass
ignments/svm/vote2.csv',
                 head=TRUE, sep=",")

t1 = sample(1:435, 348)
t2 = setdiff(1:435, t1)
train = subset(data[t1,])
test = subset(data[t2,], select =-party)

mod <- svm(party ~ ., data=train)

cl <- data[t2,]$party

pred <- predict(mod, test)
pred
```

```
##         11         14         15         17         18         31 
## republican   democrat republican   democrat   democrat republican 
##         35         46         53         59         60         62 
##   democrat   democrat   democrat republican republican republican 
##         63         64         69         76         77         81 
##   democrat   democrat   democrat republican   democrat   democrat 
##         93         95         96        100        110        111 
##   democrat   democrat   democrat republican   democrat   democrat 
##        115        118        133        153        166        168 
##   democrat republican republican   democrat   democrat republican 
##        169        174        178        188        190        193 
## republican   democrat   democrat   democrat   democrat   democrat 
##        196        200        205        208        210        213 
## republican   democrat republican republican   democrat   democrat 
##        219        228        230        235        236        237 
##   democrat   democrat republican   democrat republican   democrat 
##        241        242        243        258        264        265 
## republican   democrat   democrat republican   democrat   democrat 
##        276        282        290        313        316        317 
##   democrat republican   democrat   democrat   democrat   democrat 
##        318        319        324        340        341        347 
##   democrat   democrat   democrat republican republican republican 
##        351        355        359        366        367        372 
##   democrat   democrat   democrat   democrat   democrat   democrat 
##        381        384        393        395        396        398 
##   democrat   democrat republican   democrat   democrat   democrat 
##        399        400        401        405        411        418 
##   democrat republican republican republican republican   democrat 
##        420        422        425 
##   democrat   democrat   democrat 
## Levels: democrat republican
```

```
confusionMatrix(pred, cl)
```

```
## Confusion Matrix and Statistics
## 
##             Reference
## Prediction   democrat republican
##   democrat         57          2
##   republican        2         26
##                                           
##                Accuracy : 0.954           
##                  95% CI : (0.8864, 0.9873)
##     No Information Rate : 0.6782          
##     P-Value [Acc > NIR] : 2.65e-10        
##                                           
##                   Kappa : 0.8947          
##  Mcnemar's Test P-Value : 1               
##                                           
##             Sensitivity : 0.9661          
##             Specificity : 0.9286          
##          Pos Pred Value : 0.9661          
##          Neg Pred Value : 0.9286          
##              Prevalence : 0.6782          
##          Detection Rate : 0.6552          
##    Detection Prevalence : 0.6782          
##       Balanced Accuracy : 0.9473          
##                                           
##        'Positive' Class : democrat        
## 
```

```
summary(mod)
```

```
## 
## Call:
## svm(formula = party ~ ., data = train)
## 
## 
## Parameters:
##    SVM-Type:  C-classification 
##  SVM-Kernel:  radial 
##        cost:  1 
##       gamma:  0.0625 
## 
## Number of Support Vectors:  100
## 
##  ( 54 46 )
## 
## 
## Number of Classes:  2 
## 
## Levels: 
##  democrat republican
```

```
library(caret)

y = train$party
x = subset(train, select=-party)

P_model <- train(x,y, method="svmPoly", tuneLength=5,
                 trControl=trainControl(method='repeatedcv', number=10, repeats=10))

L_model <- train(x,y, method="svmLinear", tuneLength=5,
                 trControl=trainControl(method='repeatedcv', number=10, repeats=10))

R_model <- train(x,y, method="svmRadial", tuneLength=5,
                 trControl=trainControl(method='repeatedcv', number=10, repeats=10))

P_model
```

```
## Support Vector Machines with Polynomial Kernel
##
## 348 samples
##  16 predictor
##   2 classes: 'democrat', 'republican'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 10 times)
## Summary of sample sizes: 313, 313, 313, 313, 314, 313, ...
## Resampling results across tuning parameters:
##
##   degree  scale  C     Accuracy   Kappa
##   1       1e-03  0.25  0.5976471  0.0000000
##   1       1e-03  0.50  0.9001933  0.7911016
##   1       1e-03  1.00  0.9062185  0.8083037
##   1       1e-03  2.00  0.9197059  0.8352513
##   1       1e-03  4.00  0.9372353  0.8709602
##   1       1e-02  0.25  0.9254622  0.8467928
##   1       1e-02  0.50  0.9389580  0.8744853
##   1       1e-02  1.00  0.9493697  0.8956958
##   1       1e-02  2.00  0.9539496  0.9056536
##   1       1e-02  4.00  0.9539496  0.9056536
##   1       1e-01  0.25  0.9539496  0.9056536
##   1       1e-01  0.50  0.9539496  0.9056536
##   1       1e-01  1.00  0.9510924  0.8995266
##   1       1e-01  2.00  0.9476303  0.8918289
##   1       1e-01  4.00  0.9519580  0.9001492
##   1       1e+00  0.25  0.9502017  0.8968885
##   1       1e+00  0.50  0.9508151  0.8978749
##   1       1e+00  1.00  0.9447647  0.8851421
##   1       1e+00  2.00  0.9447479  0.8852161
##   1       1e+00  4.00  0.9450504  0.8858992
##   1       1e+01  0.25  0.9441765  0.8839706
##   1       1e+01  0.50  0.9450756  0.8859183
##   1       1e+01  1.00  0.9450504  0.8861611
##   1       1e+01  2.00  0.9464958  0.8889975
##   1       1e+01  4.00  0.9450588  0.8861976
##   2       1e-03  0.25  0.9004790  0.7916753
##   2       1e-03  0.50  0.9065042  0.8088639
##   2       1e-03  1.00  0.9197059  0.8352513
##   2       1e-03  2.00  0.9372353  0.8709602
##   2       1e-03  4.00  0.9433025  0.8831123
##   2       1e-02  0.25  0.9392437  0.8750453
##   2       1e-02  0.50  0.9496555  0.8962981
##   2       1e-02  1.00  0.9530840  0.9038061
##   2       1e-02  2.00  0.9536639  0.9050365
##   2       1e-02  4.00  0.9510840  0.8995163
##   2       1e-01  0.25  0.9507815  0.8984090
##   2       1e-01  0.50  0.9508151  0.8983798
##   2       1e-01  1.00  0.9539664  0.9046134
##   2       1e-01  2.00  0.9502437  0.8969935
##   2       1e-01  4.00  0.9402353  0.8760584
##   2       1e+00  0.25  0.9114286  0.8160624
##   2       1e+00  0.50  0.9114286  0.8160624
##   2       1e+00  1.00  0.9114286  0.8160624
##   2       1e+00  2.00  0.9114286  0.8160624
##   2       1e+00  4.00  0.9114286  0.8160624
##   2       1e+01  0.25  0.8746050  0.7389639
##   2       1e+01  0.50  0.8746050  0.7389639
##   2       1e+01  1.00  0.8746050  0.7389639
##   2       1e+01  2.00  0.8746050  0.7389639
##   2       1e+01  4.00  0.8746050  0.7389639
##   3       1e-03  0.25  0.9027815  0.8015838
##   3       1e-03  0.50  0.9125210  0.8201674
##   3       1e-03  1.00  0.9312017  0.8583209
##   3       1e-03  2.00  0.9412605  0.8790588
##   3       1e-03  4.00  0.9513697  0.9000570
##   3       1e-02  0.25  0.9444370  0.8854573
##   3       1e-02  0.50  0.9522353  0.9019535
##   3       1e-02  1.00  0.9533782  0.9044194
##   3       1e-02  2.00  0.9516555  0.9007358
##   3       1e-02  4.00  0.9562605  0.9098823
##   3       1e-01  0.25  0.9571597  0.9113940
##   3       1e-01  0.50  0.9522605  0.9008440
##   3       1e-01  1.00  0.9413445  0.8777338
##   3       1e-01  2.00  0.9393277  0.8736585
##   3       1e-01  4.00  0.9393277  0.8736585
##   3       1e+00  0.25  0.9390756  0.8724795
##   3       1e+00  0.50  0.9390756  0.8724795
##   3       1e+00  1.00  0.9390756  0.8724795
##   3       1e+00  2.00  0.9390756  0.8724795
##   3       1e+00  4.00  0.9390756  0.8724795
##   3       1e+01  0.25  0.9387899  0.8723004
##   3       1e+01  0.50  0.9387899  0.8723004
##   3       1e+01  1.00  0.9387899  0.8723004
##   3       1e+01  2.00  0.9387899  0.8723004
##   3       1e+01  4.00  0.9387899  0.8723004
## 
## Accuracy was used to select the optimal model using the largest value.
## The final values used for the model were degree = 3, scale = 0.1 and C
##  = 0.25.
```

```
L_model
```

```
## Support Vector Machines with Linear Kernel
##
## 348 samples
##  16 predictor
##   2 classes: 'democrat', 'republican'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 10 times)
## Summary of sample sizes: 314, 313, 313, 313, 313, 314, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.9463109  0.8885893
##
## Tuning parameter 'C' was held constant at a value of 1
```

```
R_model
```

```
## Support Vector Machines with Radial Basis Function Kernel
##
## 348 samples
##  16 predictor
##   2 classes: 'democrat', 'republican'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 10 times)
## Summary of sample sizes: 313, 313, 314, 313, 313, 314, ...
## Resampling results across tuning parameters:
##
##   C     Accuracy   Kappa
##   0.25  0.9474286  0.8913286
##   0.50  0.9554370  0.9079967
##   1.00  0.9522773  0.9017700
##   2.00  0.9505546  0.8981330
##   4.00  0.9525630  0.9019353
##
## Tuning parameter 'sigma' was held constant at a value of 0.05068646
## Accuracy was used to select the optimal model using the largest value.
## The final values used for the model were sigma = 0.05068646 and C = 0.5.
```

```
pred <- predict(R_model, test)
confusionMatrix(pred, cl)
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   democrat republican
##   democrat         57          1
##   republican        2         27
##
##                Accuracy : 0.9655
##                  95% CI : (0.9025, 0.9928)
##     No Information Rate : 0.6782
##     P-Value [Acc > NIR] : 2.587e-11
##
##                   Kappa : 0.9217
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9661
##             Specificity : 0.9643
##          Pos Pred Value : 0.9828
##          Neg Pred Value : 0.9310
##              Prevalence : 0.6782
##          Detection Rate : 0.6552
##    Detection Prevalence : 0.6667
##       Balanced Accuracy : 0.9652
##
##        'Positive' Class : democrat
##
```