SVM Model

Code

Open Libraries

Hide

Hide

```
library(e1071)
library(rpart)
library("kernlab", lib.loc="~/Library/R/3.3/library")
```

```
Attaching package: ‘kernlab’

The following object is masked from ‘package:modeltools’:

    prior

The following object is masked from ‘package:ggplot2’:

    alpha
```

Hide

Hide

```
library("caret", lib.loc="~/Library/R/3.3/library")
```

Hide

Hide

```
NotNormData <- cbind(dfDataSetKNN[1],dfNumAndDummies)
str(NotNormData)
```

```
'data.frame':   6235 obs. of  139 variables:
 $ Enrolling                                                                 :
Factor w/ 2 levels "N","Y": 1 1 1 1 1 1 1 1 1 2 ...
 $ Sex.F                                                                     :
int  0 0 0 0 0 0 1 0 1 1 ...
 $ Sex.M                                                                     :
int  1 1 1 1 1 1 0 1 0 0 ...
 $ Expel.N                                                                   :
int  1 1 1 1 1 1 0 1 1 1 ...
 $ Expel.Y                                                                   :
int  0 0 0 0 0 0 1 0 0 0 ...
```

```
 $ First.Gen.N                                                                :
int  1 1 1 1 1 1 0 1 1 1 ...
 $ First.Gen.Y                                                                :
int  0 0 0 0 0 0 1 0 0 0 ...
 $ Challenge Tag.N                                                            :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Challenge Tag.Y                                                            :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Boettcher.Semi.N                                                           :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Boettcher.Semi.Y                                                           :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Boettcher.Final.N                                                          :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Boettcher.Final.Y                                                          :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Daniels.Semi.N                                                             :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Daniels.Semi.Y                                                             :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Daniels.Final.N                                                            :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Daniels.Final.Y                                                            :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Harvey.App.N                                                               :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Harvey.App.Y                                                               :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Harvey.Final.N                                                             :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Harvey.Final.Y                                                             :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ FC.App.N                                                                   :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ FC.App.Y                                                                   :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Thorson.App.N                                                              :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Thorson.App.Y                                                              :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Thorson.Admit.N                                                            :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Thorson.Admit.Y                                                            :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Summet.Participant.N                                                       :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Summet.Participant.Y                                                       :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Mines.Medal.N                                                              :
int  1 1 1 1 1 1 1 1 1 1 ...
```

```
 $ Mines.Medal.Y                                                    :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ SPS.N                                                            :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ SPS.Y                                                            :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Veteran.N                                                        :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Veteran.Y                                                        :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Legacy.N                                                         :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Legacy.Y                                                         :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Athlete.N                                                        :
int  1 1 1 1 1 1 1 1 1 1 ...
 $ Athlete.Y                                                        :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ State.CO                                                         :
int  1 0 0 1 0 1 0 0 0 0 ...
 $ State.Other                                                      :
int  0 1 1 0 1 0 1 1 1 1 ...
 $ Citizenship.Foreign National/International                       :
int  0 0 0 0 0 0 1 0 0 0 ...
 $ Citizenship.International                                        :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Citizenship.Missing                                              :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Citizenship.U.S. Citizen                                         :
int  1 1 1 1 1 1 0 1 1 1 ...
 $ Citizenship.U.S. Permanent Resident/Green Card Holder            :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Citizenship.Undocumented/DACA                                    :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Ethnicity.American Indian or Alaska Native                       :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Ethnicity.Asian                                                  :
int  1 0 0 0 0 0 1 0 0 0 ...
 $ Ethnicity.Black or African American                              :
int  0 0 0 0 1 0 0 0 0 0 ...
 $ Ethnicity.Hispanic or Latino                                     :
int  0 0 0 0 0 0 0 1 0 0 ...
 $ Ethnicity.Missing                                                :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Ethnicity.Multiracial                                            :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Ethnicity.Native Hawaiian or Other Pacific Islander              :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Ethnicity.NotDeclared                                            :
int  0 0 0 0 0 0 0 0 0 0 ...
```

```
 $ Ethnicity.Unknown                                                                 :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Ethnicity.White                                                                   :
int  0 1 1 1 0 1 0 0 1 1 ...
 $ Major.App.Applied Mathematics & Statistics - Computational & Applied Mathematics:
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Applied Mathematics & Statistics - Statistics                           :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Chemical Engineering                                                    :
int  0 0 0 0 0 0 0 1 0 0 ...
 $ Major.App.Chemical Engineering - Biological Engineering Specialty                 :
int  0 0 0 0 0 0 0 0 1 0 ...
 $ Major.App.Chemistry                                                               :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Chemistry - Biochemistry Specialty                                      :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Chemistry - Environmental Chemistry Specialty                           :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Civil Engineering                                                       :
int  1 0 0 0 0 0 0 0 0 1 ...
 $ Major.App.Computer Science                                                        :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Economics                                                               :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Electrical Engineering                                                  :
int  0 0 0 0 0 0 1 0 0 0 ...
 $ Major.App.Engineering Physics                                                     :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Environmental Engineering                                               :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Geological Engineering                                                  :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Geophysical Engineering                                                 :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Mechanical Engineering                                                  :
int  0 1 1 1 1 0 0 0 0 0 ...
 $ Major.App.Metallurgical & Materials Engineering                                   :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Mining Engineering                                                      :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Missing                                                                 :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Petroleum Engineering                                                   :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ Major.App.Undecided                                                               :
int  0 0 0 0 0 1 0 0 0 0 ...
 $ First Contact.ACT                                                                 :
int  1 1 0 0 0 0 0 0 0 0 ...
 $ First Contact.Application                                                         :
int  0 0 0 0 0 0 0 0 0 0 ...
```

```
 $ First Contact.Athlete Form                                                :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.Campus Visit                                                :
int  0 0 0 1 0 1 0 0 1 1 ...
 $ First Contact.College Fair                                                :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.FUF                                                         :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.Girls Lead the Way                                          :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.GPA Form                                                    :
int  0 0 1 0 0 0 0 0 0 0 ...
 $ First Contact.Inquiry Form                                                :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.Mailing                                                     :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.Materials                                                   :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.MEP                                                         :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.Phone                                                       :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.Preview Mines                                               :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Contact.Royall Search                                               :
int  0 0 0 0 1 0 0 1 0 0 ...
 $ First Contact.SAT                                                         :
int  0 0 0 0 0 0 1 0 0 0 ...
 $ First Contact.TOEFL                                                       :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Visit.Campus Tour                                                   :
int  0 1 0 0 0 0 0 0 0 0 ...
 $ First Visit.Campus Visit                                                  :
int  0 0 0 1 0 1 0 0 1 1 ...
 $ First Visit.Class Shadow                                                  :
int  0 0 0 0 0 0 0 0 0 0 ...
 $ First Visit.Discover Mines                                                :
int  0 0 0 0 0 0 0 0 0 0 ...
  [list output truncated]
```

Parrtitions created with 75% of data for training and 25% of data for testing.

Hide

Hide

```
set.seed(9850)
inTrainingK <- createDataPartition(NotNormData$Enrolling, p = 0.75, list = FALSE)
trainingK <- NotNormData[inTraining, ]
testingK <- NotNormData[-inTraining, ]
```

Hide

Hide

```
trainingK
```

	Enrolling <fctr>	Se… <int>	Se… <int>	Expel.N <int>	Expel.Y <int>	First.Gen.N <int>	First.Gen.Y <int>	Challenge Tag.N <int>
1	N	0	1	1	0	1	0	1
2	N	0	1	1	0	1	0	1
4	N	0	1	1	0	1	0	1
5	N	0	1	1	0	1	0	1
8	N	0	1	1	0	1	0	1
9	N	1	0	1	0	1	0	1
12	Y	0	1	1	0	1	0	1
13	N	1	0	1	0	1	0	1
16	N	0	1	1	0	1	0	1
18	N	0	1	1	0	1	0	1

1-10 of 4,677 rows | 1-10 of 139 columns Previous **1** 2 3 4 5 6 … 100 Next

Hide

Hide

```
testingK
```

	Enrolling <fctr>	Se… <int>	Se… <int>	Expel.N <int>	Expel.Y <int>	First.Gen.N <int>	First.Gen.Y <int>	Challenge Tag.N <int>
3	N	0	1	1	0	1	0	1
6	N	0	1	1	0	1	0	1
7	N	1	0	0	1	0	1	1
10	Y	1	0	1	0	1	0	1
11	N	0	1	1	0	1	0	1

14	N	0	1	1	0	1	0	1	
15	N	0	1	1	0	1	0	1	
17	Y	0	1	1	0	0	1	1	
31	N	0	1	1	0	1	0	1	
34	N	0	1	1	0	1	0	1	

1-10 of 1,558 rows | 1-10 of 139 columns Previous **1** 2 3 4 5 6 … 100 Next

SVM with out scaling before model

Hide

Hide

```
model_ksvm <- ksvm(Enrolling ~., trainingK, kernel = "vanilladot")
```

```
 Setting default kernel parameters
```

```
Variable(s) `' constant. Cannot scale data.
```

Hide

Hide

```
model_ksvm_classifier <- ksvm(Enrolling ~., data=trainingK, kernel = "vanilladot",sca
led=FALSE)
```

```
 Setting default kernel parameters
```

Hide

Hide

```
model_ksvm_predictor <- predict(model_ksvm_classifier,testingK)
model_ksvm_classifier
```

```
Support Vector Machine object of class "ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 1

Linear (vanilla) kernel function.

Number of Support Vectors : 697

Objective Function Value : -3032.78
Training error : 0.088732
```

Hide

Hide

```
table(model_ksvm_predictor, testingK$Enrolling)
```

```
model_ksvm_predictor    N    Y
                   N 1159   91
                   Y   70  238
```

Hide

Hide

```
agreement <- model_ksvm_predictor == testingK$Enrolling
table(agreement)
```

```
agreement
FALSE  TRUE
  161  1397
```

Hide

Hide

```
prop.table(table(agreement))
```

```
agreement
    FALSE      TRUE
0.1033376 0.8966624
```

SVM with Scaling before Model

Hide

Hide

```
model_ksvm <- ksvm(Enrolling ~., training, kernel = "vanilladot")
```

```
 Setting default kernel parameters
```

```
Variable(s) `' constant. Cannot scale data.
```

Hide

Hide

```
model_ksvm_classifier <- ksvm(Enrolling ~., data=training, kernel = "vanilladot", sca
le=TRUE)
```

```
 Setting default kernel parameters
```

```
Variable(s) `' constant. Cannot scale data.
```

Hide

Hide

```
model_ksvm_predictor <- predict(model_ksvm_classifier,testing)
model_ksvm_classifier
```

```
Support Vector Machine object of class "ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 1

Linear (vanilla) kernel function.

Number of Support Vectors : 1057

Objective Function Value : -972.2043
Training error : 0.085097
```

Hide

Hide

```
head(model_ksvm_predictor)
```

```
[1] N N N Y N N
Levels: N Y
```

Hide

Hide

```
table(model_ksvm_predictor, testing$Enrolling)
```

```
model_ksvm_predictor    N    Y
                   N 1179  104
                   Y   50  225
```

Hide

Hide

```
agreement <- model_ksvm_predictor == testing$Enrolling
table(agreement)
```

```
agreement
FALSE  TRUE
  154  1404
```

Hide

Hide

```
prop.table(table(agreement))
```

```
agreement
     FALSE       TRUE
0.09884467 0.90115533
```

Confusion Matrix

Hide

Hide

```
svm_table <- table(testing$Enrolling, model_ksvm_predictor)
svm_table
```

```
   model_ksvm_predictor
       N    Y
  N 1179   50
  Y  104  225
```

Hide

Hide

```
confusionMatrix(svm_table)
```

```
Confusion Matrix and Statistics

   model_ksvm_predictor
       N    Y
  N 1179   50
  Y  104  225

               Accuracy : 0.9012
                 95% CI : (0.8852, 0.9155)
    No Information Rate : 0.8235
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6843
 Mcnemar's Test P-Value : 1.947e-05

            Sensitivity : 0.9189
            Specificity : 0.8182
         Pos Pred Value : 0.9593
         Neg Pred Value : 0.6839
             Prevalence : 0.8235
         Detection Rate : 0.7567
   Detection Prevalence : 0.7888
      Balanced Accuracy : 0.8686

       'Positive' Class : N
```

SVM RBF or Bassian Kernel Model

Hide

Hide

```
model_classifier_rbf <- ksvm(Enrolling ~., data=training, kernel = "rbfdot")
```

```
Variable(s) `' constant. Cannot scale data.
```

Hide

Hide

```
model_predictions_rbf <- predict(model_classifier_rbf, testing)
agreement_rbf <- model_predictions_rbf == testing$Enrolling
table(agreement_rbf)
```

```
agreement_rbf
FALSE  TRUE
  151  1407
```

Hide

Hide

```
prop.table(table(agreement_rbf))
```

```
agreement_rbf
     FALSE       TRUE
0.09691913 0.90308087
```

Confusion Matrix

Hide

Hide

```
rbf_table <- table(testing$Enrolling, model_predictions_rbf)
rbf_table
```

```
   model_predictions_rbf
      N    Y
  N 1182   47
  Y  104  225
```

Hide

Hide

```
confusionMatrix(rbf_table)
```

```
Confusion Matrix and Statistics

   model_predictions_rbf
       N    Y
  N 1182   47
  Y  104  225

               Accuracy : 0.9031
                 95% CI : (0.8873, 0.9173)
    No Information Rate : 0.8254
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6894
 Mcnemar's Test P-Value : 5.184e-06

            Sensitivity : 0.9191
            Specificity : 0.8272
         Pos Pred Value : 0.9618
         Neg Pred Value : 0.6839
             Prevalence : 0.8254
         Detection Rate : 0.7587
   Detection Prevalence : 0.7888
      Balanced Accuracy : 0.8732

       'Positive' Class : N
```

SVM Polynomial Kernel Model

Hide

Hide

```
model_classifier_poly=ksvm(Enrolling~.,data=training,kernel="polydot",gama=1,cost=1)
```

```
 Setting default kernel parameters
```

```
Variable(s) `' constant. Cannot scale data.
```

Hide

Hide

```
model_predictions_poly <- predict(model_classifier_poly, testing)
agreement_poly <- model_predictions_poly == testing$Enrolling
table(agreement_poly)
```

```
agreement_poly
FALSE  TRUE
  154  1404
```

Hide

Hide

```
prop.table(table(agreement_poly))
```

```
agreement_poly
     FALSE       TRUE
0.09884467 0.90115533
```

Confusion Matrix

Hide

Hide

```
poly_table <- table(testing$Enrolling, model_predictions_poly)
poly_table
```

```
  model_predictions_poly
     N    Y
  N 1179   50
  Y  104  225
```

Hide

Hide

```
confusionMatrix(poly_table)
```

```
Confusion Matrix and Statistics

   model_predictions_poly
       N    Y
  N 1179   50
  Y  104  225

               Accuracy : 0.9012
                 95% CI : (0.8852, 0.9155)
    No Information Rate : 0.8235
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6843
 Mcnemar's Test P-Value : 1.947e-05

            Sensitivity : 0.9189
            Specificity : 0.8182
         Pos Pred Value : 0.9593
         Neg Pred Value : 0.6839
             Prevalence : 0.8235
         Detection Rate : 0.7567
   Detection Prevalence : 0.7888
      Balanced Accuracy : 0.8686

       'Positive' Class : N
```