Michigan Technological University

Michigan Technological University, Computer Science

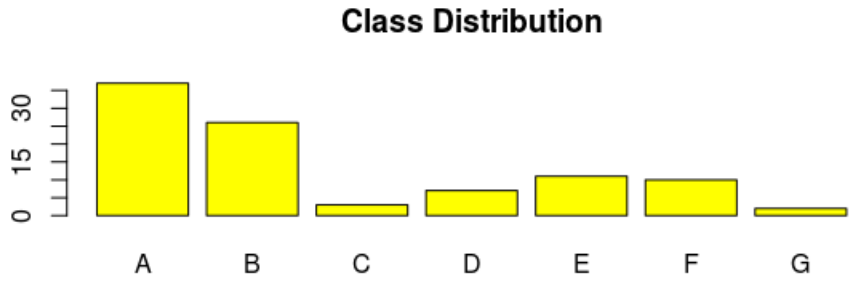
MA 5790 Combined Section - Predictive Modeling Assignment 6

Anil Silwal

December 14, 2018

Problem 13.2: Use the fatty acid data from the previous exercise set (Exercise 12.2).

a. Use the same data splitting approach (if any) and pre-processing steps that you did in the previous chapter. Using the same classification statistic as before, build models described in this chapter for these data. Which model has the best predictive ability? How does this optimal model's performance compare to the best linear model's performance? Would you infer that the data have nonlinear separation boundaries based on this comparison?
Solution:



Given this classification imbalance in oilType and data being very small, I didn't split the data instead I first trained the data and compared the trained result with the previous data.
However, the co-related data were removed before training the model.

I build the non-linear models described in given chapter for the fatty acids predictor and output generated by each model are below:

i. **Nonlinear Discriminant Analysis Model For Fatty Acids Predictor:**

```
> confusionMatrix(mdaFit)
Bootstrapped (25 reps) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    A    B    C    D    E    F    G
         A 34.3  0.1  0.0  0.0  0.0  0.0  0.9
         B  2.0 28.1  0.0  0.0  0.0  0.0  0.0
         C  0.0  0.2  2.5  0.1  0.0  0.0  0.0
         D  0.0  0.0  0.0  6.7  0.0  0.1  0.0
         E  0.8  0.0  0.0  0.0 11.9  0.2  0.0
         F  0.0  0.0  0.0  0.0  0.0 10.2  0.0
         G  0.0  0.1  0.0  0.0  0.0  0.0  1.7

 Accuracy (average) : 0.954
```

ii. **Nonlinear Discriminant Analysis Accuracy For Testing Data For Fatty Acids Predictor:**

```
> confusionMatrix(mdaPrediction,oilType)
Confusion Matrix and Statistics

          Reference
Prediction  A  B  C  D  E  F  G
         A 34  0  0  0  0  0  0
         B  2 26  0  0  0  0  0
         C  0  0  3  0  0  0  0
         D  0  0  0  7  0  0  0
         E  1  0  0  0 11  0  0
         F  0  0  0  0  0 10  0
         G  0  0  0  0  0  0  2

Overall Statistics

               Accuracy : 0.9688
                 95% CI : (0.9114, 0.9935)
    No Information Rate : 0.3854
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9585
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F Class: G
Sensitivity            0.9189   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Specificity            1.0000   0.9714  1.00000  1.00000   0.9882   1.0000  1.00000
Pos Pred Value         1.0000   0.9286  1.00000  1.00000   0.9167   1.0000  1.00000
Neg Pred Value         0.9516   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Prevalence             0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Rate         0.3542   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Prevalence   0.3542   0.2917  0.03125  0.07292   0.1250   0.1042  0.02083
Balanced Accuracy      0.9595   0.9857  1.00000  1.00000   0.9941   1.0000  1.00000
```

iii. **RDA Analysis Model For Fatty Acids Predictor:**

```
> confusionMatrix(rdaFit)
Bootstrapped (25 reps) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    A    B    C    D    E    F    G
         A 20.7  0.1  0.0  0.0  0.0  0.0  0.9
         B  1.4 16.3  0.0  0.0  0.0  0.0  0.0
         C  0.0  0.2  1.4  0.0  0.0  0.0  0.0
         D  0.0  0.0  0.0  3.7  0.0  0.1  0.0
         E  0.5  0.0  0.0  0.0  7.5  0.2  0.0
         F  0.0  0.0  0.0  0.0  0.0  5.9  0.0
         G 14.5 11.9  1.2  3.1  4.4  4.4  1.7

 Accuracy (average) : 0.5719
```

iv. **RDA Analysis Accuracy For Testing Data For Fatty Acids Predictor:**

```
> confusionMatrix(rdaPrediction,oilType)
Confusion Matrix and Statistics

          Reference
Prediction  A  B  C  D  E  F  G
         A 34  0  0  0  0  0  0
         B  2 26  0  0  0  0  0
         C  0  0  3  0  0  0  0
         D  0  0  0  7  0  0  0
         E  1  0  0  0 11  0  0
         F  0  0  0  0  0 10  0
         G  0  0  0  0  0  0  2

Overall Statistics

               Accuracy : 0.9688
                 95% CI : (0.9114, 0.9935)
    No Information Rate : 0.3854
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9585
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F Class: G
Sensitivity            0.9189   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Specificity            1.0000   0.9714  1.00000  1.00000   0.9882   1.0000  1.00000
Pos Pred Value         1.0000   0.9286  1.00000  1.00000   0.9167   1.0000  1.00000
Neg Pred Value         0.9516   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Prevalence             0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Rate         0.3542   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Prevalence   0.3542   0.2917  0.03125  0.07292   0.1250   0.1042  0.02083
Balanced Accuracy      0.9595   0.9857  1.00000  1.00000   0.9941   1.0000  1.00000
```

v. **Neural Network Model For Fatty Acids Predictor:**

```
> confusionMatrix(nnetFit)
Bootstrapped (25 reps) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    A    B    C    D    E    F    G
         A 34.2  0.7  1.0  0.3  0.0  0.0  0.0
         B  1.6 27.3  0.0  0.0  0.0  0.0  0.6
         C  0.0  0.6  1.5  0.0  0.0  0.0  0.0
         D  0.8  0.0  0.0  5.9  0.0  0.0  0.5
         E  0.5  0.0  0.0  0.0 11.9  0.0  0.1
         F  0.0  0.0  0.0  0.6  0.0 10.6  1.0
         G  0.0  0.0  0.0  0.0  0.0  0.0  0.5

 Accuracy (average) : 0.9171
```

vi. **Neural Network Accuracy For Testing Data For Fatty Acids Predictor:**

```
> confusionMatrix(nnetPrediction,oilType)
Confusion Matrix and Statistics

          Reference
Prediction  A  B  C  D  E  F  G
         A 36  0  0  0  0  0  0
         B  1 26  0  0  0  0  0
         C  0  0  3  0  0  0  0
         D  0  0  0  7  0  0  0
         E  0  0  0  0 11  0  0
         F  0  0  0  0  0 10  0
         G  0  0  0  0  0  0  2

Overall Statistics

               Accuracy : 0.9896
                 95% CI : (0.9433, 0.9997)
    No Information Rate : 0.3854
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9861
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F Class: G
Sensitivity            0.9730   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Specificity            1.0000   0.9857  1.00000  1.00000   1.0000   1.0000  1.00000
Pos Pred Value         1.0000   0.9630  1.00000  1.00000   1.0000   1.0000  1.00000
Neg Pred Value         0.9833   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Prevalence             0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Rate         0.3750   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Prevalence   0.3750   0.2812  0.03125  0.07292   0.1146   0.1042  0.02083
Balanced Accuracy      0.9865   0.9929  1.00000  1.00000   1.0000   1.0000  1.00000
```

vii. **Flexible Discriminant Analysis Model For Fatty Acids Predictor:**

```
> confusionMatrix(fdaTuned)
Bootstrapped (25 reps) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    A    B    C    D    E    F    G
         A 35.8  0.2  0.0  0.2  0.0  0.0  1.2
         B  0.9 27.8  0.0  0.0  0.0  0.0  0.1
         C  0.2  0.1  2.5  0.2  0.0  0.0  0.0
         D  0.0  0.0  0.0  6.1  0.0  0.3  0.1
         E  0.1  0.1  0.0  0.0 11.9  0.0  0.0
         F  0.0  0.1  0.0  0.1  0.0 10.2  0.2
         G  0.0  0.1  0.0  0.1  0.0  0.0  1.0

 Accuracy (average) : 0.954
```

viii. **Flexible Discriminant Analysis Model Accuracy For Testing Data For Fatty Acids Predictor:**

```
> confusionMatrix(fdaPrediction,oilType)
Confusion Matrix and Statistics

          Reference
Prediction  A  B  C  D  E  F  G
         A 37  0  0  0  0  0  0
         B  0 26  0  0  0  0  0
         C  0  0  3  0  0  0  0
         D  0  0  0  7  0  0  0
         E  0  0  0  0 11  0  0
         F  0  0  0  0  0 10  0
         G  0  0  0  0  0  0  2

Overall Statistics

               Accuracy : 1
                 95% CI : (0.9623, 1)
    No Information Rate : 0.3854
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F Class: G
Sensitivity            1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Specificity            1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Pos Pred Value         1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Neg Pred Value         1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Prevalence             0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Rate         0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Prevalence   0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Balanced Accuracy      1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
```

ix. **Support Vector Machines Model for Fatty Acids Predictor:**

```
> confusionMatrix(svmRModel)
Bootstrapped (25 reps) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    A    B    C    D    E    F    G
         A 36.2  0.1  0.0  0.1  0.0  0.0  0.1
         B  0.3 27.0  0.0  0.0  0.0  0.0  0.0
         C  0.1  1.4  2.5  0.0  0.0  0.0  0.0
         D  0.0  0.0  0.0  6.6  0.0  0.3  0.0
         E  0.3  0.0  0.0  0.0 11.9  0.0  0.8
         F  0.0  0.0  0.0  0.1  0.0 10.2  0.0
         G  0.0  0.0  0.0  0.0  0.0  0.0  1.7

 Accuracy (average) : 0.962
```

x. **Support Vector Machines Model Accuracy For Testing Data For Fatty Acids Predictor:**

```
> confusionMatrix(svmPrediction,oilType)
Confusion Matrix and Statistics

          Reference
Prediction  A  B  C  D  E  F  G
         A 37  0  0  0  0  0  0
         B  0 26  0  0  0  0  0
         C  0  0  3  0  0  0  0
         D  0  0  0  7  0  0  0
         E  0  0  0  0 11  0  0
         F  0  0  0  0  0 10  0
         G  0  0  0  0  0  0  2

Overall Statistics

               Accuracy : 1
                 95% CI : (0.9623, 1)
    No Information Rate : 0.3854
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F Class: G
Sensitivity            1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Specificity            1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Pos Pred Value         1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Neg Pred Value         1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Prevalence             0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Rate         0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Prevalence   0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Balanced Accuracy      1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
```

xi. **K-Nearest Neighbors Model for Fatty Acids Predictor:**

```
> confusionMatrix(knnFit)
Bootstrapped (25 reps) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    A    B    C    D    E    F    G
         A 36.2  0.9  0.0  0.2  0.0  0.0  0.2
         B  0.5 26.5  0.0  0.0  0.0  0.0  0.3
         C  0.1  1.2  2.5  0.0  0.0  0.0  0.0
         D  0.0  0.0  0.0  6.3  0.0  0.0  0.0
         E  0.2  0.0  0.0  0.0 11.9  0.2  0.3
         F  0.0  0.0  0.0  0.2  0.0 10.4  0.0
         G  0.0  0.0  0.0  0.0  0.0  0.0  1.7

 Accuracy (average) : 0.9551
```

xii. **K-Nearest Neighbors Model Accuracy For Testing Data For Fatty Acids Predictor:**

```
> confusionMatrix(knnPrediction,oilType)
Confusion Matrix and Statistics

          Reference
Prediction  A  B  C  D  E  F  G
         A 37  0  0  0  0  0  0
         B  0 26  0  0  0  0  0
         C  0  0  3  0  0  0  0
         D  0  0  0  7  0  0  0
         E  0  0  0  0 11  0  0
         F  0  0  0  0  0 10  0
         G  0  0  0  0  0  0  2

Overall Statistics

               Accuracy : 1
                 95% CI : (0.9623, 1)
    No Information Rate : 0.3854
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F Class: G
Sensitivity            1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Specificity            1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Pos Pred Value         1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Neg Pred Value         1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Prevalence             0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Rate         0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Prevalence   0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Balanced Accuracy      1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
```

xiii. **Naive Bayes Model for Fatty Acids Predictor:**

```
> confusionMatrix(nbFit)
Bootstrapped (25 reps) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    A    B    C    D    E    F    G
         A 35.5  0.5  0.0  0.0  0.3  0.0  0.5
         B  0.5 24.7  0.0  0.0  0.0  0.5  0.0
         C  0.0  1.1  2.7  0.0  0.0  1.3  0.0
         D  0.3  0.0  0.0  4.5  0.0  0.0  0.0
         E  0.5  1.1  0.0  0.0 12.7  0.0  0.0
         F  0.0  0.0  0.0  0.0  0.0  7.4  0.0
         G  1.6  0.8  0.0  1.1  0.5  0.8  1.1

 Accuracy (average) : 0.8859
```

xiv. **Naive Bayes Model Accuracy For Testing Data For Fatty Acids Predictor:**

```
> confusionMatrix(nbPrediction,oilType)
Confusion Matrix and Statistics

          Reference
Prediction  A  B  C  D  E  F  G
         A 37  0  0  0  0  0  0
         B  0 26  0  0  0  0  0
         C  0  0  3  0  0  0  0
         D  0  0  0  7  0  0  0
         E  0  0  0  0 11  0  0
         F  0  0  0  0  0 10  0
         G  0  0  0  0  0  0  2

Overall Statistics

               Accuracy : 1
                 95% CI : (0.9623, 1)
    No Information Rate : 0.3854
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E Class: F Class: G
Sensitivity            1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Specificity            1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Pos Pred Value         1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Neg Pred Value         1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
Prevalence             0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Rate         0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Detection Prevalence   0.3854   0.2708  0.03125  0.07292   0.1146   0.1042  0.02083
Balanced Accuracy      1.0000   1.0000  1.00000  1.00000   1.0000   1.0000  1.00000
```

Comparison table for different Linear Model's Performance for predicting oilType(from previous assignment)

#	Models	Accuracy	Kappa
1	Logistic Regression (LR)	0.8824	0.835
2	Linear Discriminant Analysis(LDA)	0.8824	0.8404
3	Partial Least Square Discriminant Analysis(PLS-DA)	**0.9412**	0.9183
4	Penalized LR	0.8824	0.8404
5	Penalized LDA	0.1176	-0.02
6	Nearest Shrinkage Centroids(NSC)	**0.9412**	0.9183

Comparison table for different Non-Linear Model's Performance for predicting oilType

#	Models	Accuracy
1a.	Non-Linear Discriminant Analysis Model (LDA)	0.9688
1b.	RDA	0.9688
2	Neural Network Model(NN)	0.9896
3	Flexible Discriminant Analysis(FDA)	**1**
4	Support Vector Machine Model (SVM)	**1**
5	K-Nearest Neighbors Model (KNN)	**1**
6	Naive Bayes Model	**1**

Comparing the accuracy results between the linear and non-linear models, I found that both models does optimal performance for this data. The best of optimal model for linear is 0.9412 whereas non-linear is 1 which is almost same.

Since both model type performs best for this data so, I would infer that data may not have non-linear separation boundaries based on this comparison.

b. Which oil type does the optimal model most accurately predict? Least accurately predict?

Solution

In this exercise, we have 7-classes in response thus, I would choose accuracy metric to find the optimal model in this exercise.

Comparison table for different Non-Linear Model's Performance for predicting oilType

#	Models	Accuracy
1a.	Non-Linear Discriminant Analysis Model (LDA)	0.9688
1b.	RDA	0.9688
2	Neural Network Model(NN)	0.9896
3	Flexible Discriminant Analysis(FDA)	**1**
4	Support Vector Machine Model (SVM)	**1**
5	K-Nearest Neighbors Model (KNN)	**1**
6	Naive Bayes Model	**1**

The comparison shows that Flexible Discriminant Analysis(FDA), Support Vector Machine Model (SVM), K-Nearest Neighbors Model (KNN),Naive Bayes Model model shows best accuracy value among all other models. So, I would go with the either of these models that gives accuracy 1 as optimal model.

Again, the comparison shows that LDA, RDA model shows least accuracy value among all other models which is 0.9688

c. Of the models presented in this chapter, which performs best on these data? Which oil type does the model most accurately predict? Least accurately predict?

Solution 12.2(b)
The comparison shows that both Partial Least Square Discriminant Analysis(PLS-DA) and Nearest Shrinkage Centroids(NSC) shows best accuracy or kappa value among all other models
So, one of those models performs best on these data.

Analyzing the result,
The model accurately predicts: Class D, Class E, and Class F.
And, the model least accurately predicts: Class A and Class B
And, the model does not predict at all to: Class C and Class G

```
###################################
# question 13.2 for predicting Oil Type
###################################
library(caret)
library(AppliedPredictiveModeling)

data(oil)
# use ?hepatic to see more details


library(MASS)
set.seed(975)

barplot(table(oilType),col=c("yellow"), main="Class Distribution")

#this gives 0 predictor with zero-variance
nearZeroVar(fattyAcids,saveMetrics =TRUE)

#remove the correlation between the predictors
highCorM¡-findCorrelation(cor(fattyAcids),cutoff = .75)
filteredCorFatty ¡- fattyAcids[,-highCorM]

####### Nonlinear Discriminant Analysis ##########

ctrl ¡- trainControl(summaryFunction = defaultSummary)
set.seed(476)
mdaFit ¡- train(x = filteredCorFatty,
                y = oilType,
                method = "mda",
                metric = "Accuracy",
                tuneGrid = expand.grid(.subclasses = 1:3),
                trControl = ctrl)

mdaPrediction¡-predict(mdaFit,filteredCorFatty)
confusionMatrix(mdaPrediction,oilType)


rdaParameter ¡- expand.grid(.gamma = c(0,1),
                            .lambda = c(0, .1, 1, 2))
set.seed(476)
rdaFit ¡- train(x = filteredCorFatty,
                y = oilType,
                method = "rda",
                metric = "Accuracy",
                tuneGrid = rdaParameter,
                trControl = ctrl)

rdaPrediction¡-predict(rdaFit,filteredCorFatty)
confusionMatrix(rdaPrediction,oilType)

############### Neural Networks #############

library(nnet)
set.seed(476)
nnetGrid ¡- expand.grid(.size = 1:10, .decay = c(0, .1, 1, 2))

maxSize ¡- max(nnetGrid$.size)
numWts ¡- 1*(maxSize * (6 + 1) + maxSize + 1) ## 6 is the number of predictors
nnetFit ¡- train(x = filteredCorFatty,
                 y = oilType,
                 method = "nnet",
                 metric = "Accuracy",
                 preProc = c("center", "scale", "spatialSign"),
                 tuneGrid = nnetGrid,
```

```
                  trace = FALSE,
                  maxit = 2000,
                  MaxNWts = numWts,
                  trControl = ctrl)


nnetPrediction¡-predict(nnetFit,filteredCorFatty)
confusionMatrix(nnetPrediction,oilType)

########## Flexible Discriminant Analysis ############

library(MASS)
set.seed(476)
marsGrid ¡- expand.grid(.degree = 1:2, .nprune = 2:38)
fdaTuned ¡- train(x = filteredCorFatty,
                  y = oilType,
                  method = "fda",
                  metric = "Accuracy",
                  # Explicitly declare the candidate models to test
                  tuneGrid = marsGrid,
                  trControl = ctrl)

fdaPrediction¡-predict(fdaTuned,filteredCorFatty)
confusionMatrix(fdaPrediction,oilType)


############## Support Vector Machines ##########

library(MASS)
set.seed(476)
library(kernlab)
library(caret)

sigmaRangeReduced ¡- sigest(as.matrix(filteredCorFatty))

## Given a range of values for the "sigma" inverse width parameter
## in the Gaussian Radial Basis kernel for use with SVM.
## The estimation is based on the data to be used.
svmRGridReduced ¡- expand.grid(.sigma = sigmaRangeReduced[1],
                               .C = 2 (seq(-4, 6)))
set.seed(476)
svmRModel ¡- train(x = filteredCorFatty,
                   y = oilType,
                   method = "svmRadial",
                   metric = "Accuracy",
                   preProc = c("center", "scale"),
                   tuneGrid = svmRGridReduced,
                   fit = FALSE,
                   trControl = ctrl)

svmPrediction¡-predict(svmRModel,filteredCorFatty)
confusionMatrix(svmPrediction,oilType)

############ K-Nearest Neighbors #############
library(caret)
set.seed(476)
knnFit ¡- train(x = filteredCorFatty,
                y = oilType,
                method = "knn",
                metric = "Accuracy",
                preProc = c("center", "scale"),
                ##tuneGrid = data.frame(.k = c(4*(0:5)+1, 20*(1:5)+1, 50*(2:9)+1)), ## 21 is the best
                tuneGrid = data.frame(.k = 1:50),
                trControl = ctrl)
```

```
knnPrediction¡-predict(knnFit,filteredCorFatty)
confusionMatrix(knnPrediction,oilType)



########## Naive Bayes ##########
library(klaR)
set.seed(476)
nbFit ¡- train( x = filteredCorFatty,
                y = oilType,
                method = "nb",
                metric = "Accuracy",
                ## preProc = c("center", "scale"),
                # tuneGrid = data.frame(.k = c(4*(0:5)+1, 20*(1:5)+1, 50*(2:9)+1)), ## 21 is the best
                tuneGrid = data.frame(.fL = 2,.usekernel = TRUE,.adjust = TRUE),
                trControl = ctrl)


nbPrediction¡-predict(nbFit,filteredCorFatty)
confusionMatrix(nbPrediction,oilType)



###################################
###################################
```

References:

1. Applied Predictive Modeling : @authors Max Kuhn. Kjell Johnson
2. https://archive.ics.uci.edu/ml/index.php