STA567 Homework7

Lina Lee

11/5/2019

Using 10-fold cross validation, fit and tune a support vector classifier for predicting the region of each oil sample.

```
tc <- trainControl("repeatedcv", number=10, repeats=10)
```

SVC linear kernal

```
Set.seed(123)

svc_linear_oliveoil <- train(region ~ .,
                            data=oliveoil,
                            method="svmLinear",
                            trControl=tc,
                            tuneGrid=data.frame(C=c(1,10,50,100)))

svc_linear_oliveoil$results[which.max(svc_linear_oliveoil$results$Accuracy),]

##   C  Accuracy     Kappa AccuracySD    KappaSD
## 1 1 0.9581801 0.9486454 0.02530692 0.03095497
```

Further investigate the tunring parameter

```
Set.seed(123)

svc_linear_oliveoil2 <- train(region ~ .,
                            data=oliveoil,
                            method="svmLinear",
                            trControl=tc,
                            tuneGrid=data.frame(C=c(0.5,1,3,5)))

svc_linear_oliveoil2$results[which.max(svc_linear_oliveoil2$results$Accuracy)
,]

##   C  Accuracy    Kappa AccuracySD    KappaSD
## 2 1 0.9573683 0.947635 0.02295777 0.02828307

# misclassification rate
1-max(svc_linear_oliveoil2$results$Accuracy)

## [1] 0.04263175
```

```
# fit the final model to make a confusion matrix later.
Set.seed(123)

svc_linear_final <- train(region ~ .,
                          data=oliveoil,
                          method="svmLinear",
                          tuneGrid=data.frame(C=1))
```

SVC polynomial kernal

```
Set.seed(123)

svc_poly_oliveoil <- train(region ~ .,
                           data=oliveoil,
                           method="svmPoly",
                           trControl=tc,
                           tuneGrid=expand.grid(C=c(.1,1,10,100),
                                                degree=2:4, scale=1))
# The best Tune and corresponding Accuracy
svc_poly_oliveoil$results[which.max(svc_poly_oliveoil$results$Accuracy),]

##     C degree scale  Accuracy     Kappa AccuracySD    KappaSD
## 1 0.1      2     1 0.9566513 0.9466193 0.02435038 0.02991958
```

Further investigate the tunring parameter

```
Set.seed(123)

svc_poly_oliveoil2 <- train(region ~ .,
                            data=oliveoil,
                            method="svmPoly",
                            trControl=tc,
                            tuneGrid=expand.grid(C=c(0.01,0.05,0.08,1),
                                                 degree=2:4, scale=1))

svc_poly_oliveoil2$results[which.max(svc_poly_oliveoil2$results$Accuracy),]

##      C degree scale  Accuracy     Kappa AccuracySD    KappaSD
## 4 0.05      2     1 0.9606675 0.9514781 0.02554258 0.03161607

# The misclassification rate
1-max(svc_poly_oliveoil$results$Accuracy)

## [1] 0.04334873

# fit the final model to make a confusion matrix later.
Set.seed(123)

svc_poly_final <- train(region ~ .,
                        data=oliveoil,
                        method="svmPoly",
                        trControl=tc,
```

```r
                          tuneGrid=expand.grid(C=0.05,
                                               degree=2, scale=1))
```

SVC radial kernal

```r
Set.seed(123)

svc_rad_oliveoil <- train(region ~ .,
                          data=oliveoil,
                          method="svmRadial",
                          trControl=tc,
                          tuneGrid=expand.grid(C=c(1,2,4,8,16,30),
                          sigma=c(0.01,0.1,0.16,0.64,0.99)))
# The best Tune and corresponding Accuracy
svc_rad_oliveoil$results[which.max(svc_rad_oliveoil$results$Accuracy),]
```

```
##   C sigma  Accuracy     Kappa AccuracySD   KappaSD
## 8 2  0.16 0.9642236 0.9559886 0.02194281 0.0269919
```

Further investigate the tunring parameter

```r
Set.seed(123)

svc_rad_oliveoil2 <- train(region ~ .,
                          data=oliveoil,
                          method="svmRadial",
                          trControl=tc,
                          tuneGrid=expand.grid(C=c(1,2,4,8,16),
                          sigma=c(0.1,0.05,0.08,0.16,0.2,0.4)))
# The best Tune and corresponding Accuracy
svc_rad_oliveoil2$results[which.max(svc_rad_oliveoil$results$Accuracy),]
```

```
##     C sigma  Accuracy     Kappa AccuracySD    KappaSD
## 21 8   0.1 0.9580324 0.9483582 0.02222101 0.02728349
```

```r
# The misclassification rate
1-max(svc_rad_oliveoil$results$Accuracy)
```

```
## [1] 0.03577638
```

```r
# fit the final model to make a confusion matrix later.

Set.seed(123)

svc_rad_final<- train(region ~ .,
                      data=oliveoil,
                      method="svmRadial",
                      trControl=tc,
                      tuneGrid=expand.grid(C=8,
                      sigma=0.08))
```

Report any tuning parameters chosen, as well as the overall misclassification rate.

SVM classifier	Model Parameters	Misclassification rate
Linear	C=1	0.04263
Polynomial	C=0.05,degree=2,scale=1	0.04334
Radial Kernels	C=8, sigma=0.1	0.03578

Among three classifier, Radial Kernels has the lowest misclassification rate. Therefore, the SVM with Radial Kernels perfromed best among the three classifiers.

From the confusion matrix, are there some regions where the classifier performs particularly well?

```r
x <- subset(oliveoil, select=-region)
y<- oliveoil$region
pred_linear <- predict(svc_linear_final,x)

table(pred_linear,y)
```

```
##                  y
## pred_linear        Apulia.north Calabria Apulia.south Sicily
##    Apulia.north              25        0            0      2
##    Calabria                   0       53            0      2
##    Apulia.south               0        1          202      3
##    Sicily                     0        2            4     29
##    Sardinia.inland            0        0            0      0
##    Sardinia.coast             0        0            0      0
##    Liguria.east               0        0            0      0
##    Liguria.west               0        0            0      0
##    Umbria                     0        0            0      0
##                  y
## pred_linear        Sardinia.inland Sardinia.coast Liguria.east Liguria.west
##    Apulia.north                  0              0            0            0
##    Calabria                      0              0            0            0
##    Apulia.south                  0              0            0            0
##    Sicily                        0              0            0            0
##    Sardinia.inland              65              0            0            0
##    Sardinia.coast                0             33            0            0
##    Liguria.east                  0              0           48            0
##    Liguria.west                  0              0            1           50
##    Umbria                        0              0            1            0
##                  y
## pred_linear        Umbria
##    Apulia.north          0
##    Calabria              0
##    Apulia.south          0
##    Sicily                0
```

```
##    Sardinia.inland         0
##    Sardinia.coast          0
##    Liguria.east            0
##    Liguria.west            0
##    Umbria                 51
```

In the regions of Apulia.north,Sicily Sardinia.inland, Sardinia.coast, Liguria.west, Umbria, all
the predicted region match with the actual region. the SVC Linear model perfroms well in
the regions.

```
pred_poly<- predict(svc_poly_final,x)
table(pred_poly,y)

##                     y
## pred_poly           Apulia.north Calabria Apulia.south Sicily
##    Apulia.north              24        0            0      0
##    Calabria                   1       54            0      4
##    Apulia.south               0        2          204      3
##    Sicily                     0        0            2     29
##    Sardinia.inland            0        0            0      0
##    Sardinia.coast             0        0            0      0
##    Liguria.east               0        0            0      0
##    Liguria.west               0        0            0      0
##    Umbria                     0        0            0      0
##                     y
## pred_poly           Sardinia.inland Sardinia.coast Liguria.east Liguria.west
##    Apulia.north                   0              0            0            0
##    Calabria                       0              0            0            0
##    Apulia.south                   0              0            0            0
##    Sicily                         0              0            0            0
##    Sardinia.inland               65              0            0            0
##    Sardinia.coast                 0             33            0            0
##    Liguria.east                   0              0           50            0
##    Liguria.west                   0              0            0           50
##    Umbria                         0              0            0            0
##                     y
## pred_poly           Umbria
##    Apulia.north          0
##    Calabria              0
##    Apulia.south          0
##    Sicily                0
##    Sardinia.inland       0
##    Sardinia.coast        0
##    Liguria.east          0
##    Liguria.west          0
##    Umbria               51
```

In the regions of Sicily, Sardinia.inland, Sardinia.coast, Liguria.east, Liguria.west, Umbria, all
the predicted region match with the actual regions. the SVC polynomial model perfroms
well in the regions.

```
pred_rad<- predict(svc_rad_final,x)
table(pred_rad,y)

##                    y
## pred_rad          Apulia.north Calabria Apulia.south Sicily
##    Apulia.north           25         0           0        0
##    Calabria                0        55           0        3
##    Apulia.south            0         1         204        3
##    Sicily                  0         0           2       30
##    Sardinia.inland         0         0           0        0
##    Sardinia.coast          0         0           0        0
##    Liguria.east            0         0           0        0
##    Liguria.west            0         0           0        0
##    Umbria                  0         0           0        0
##                    y
## pred_rad          Sardinia.inland Sardinia.coast Liguria.east Liguria.west
##    Apulia.north               0              0            0            0
##    Calabria                   0              0            0            0
##    Apulia.south               0              0            0            0
##    Sicily                     0              0            0            0
##    Sardinia.inland           65              0            0            0
##    Sardinia.coast             0             33            0            0
##    Liguria.east               0              0           50            0
##    Liguria.west               0              0            0           50
##    Umbria                     0              0            0            0
##                    y
## pred_rad          Umbria
##    Apulia.north        0
##    Calabria            0
##    Apulia.south        0
##    Sicily              0
##    Sardinia.inland     0
##    Sardinia.coast      0
##    Liguria.east        0
##    Liguria.west        0
##    Umbria             51
```

In the regions of Apulia.north, Sicily, Sardinia.inland, Sardinia.coast, Liguria.east, Liguria.west, Umbria, all the predicted region match with the actual regions.the SVC radial kernel model perfroms well in thoese regions.