task4_galaxy_recode

Code

data

Hide

```
samsung <- read.csv("galaxy_smallmatrix_labeled_9d.csv")
```

Hide

```
# create a new dataset that will be used for recoding sentiment
samsung_RC <- samsung
# recode sentiment to combine factor levels 0 & 1 and 4 & 5
samsung_RC$galaxysentiment <- recode(samsung_RC$galaxysentiment, '0' = 1, '1' = 1, '2' =
2, '3' = 3, '4' = 4, '5' = 4)
# inspect results
summary(samsung_RC)
```

```
     iphone       samsunggalaxy       sonyxperia       nokialumina          htcphone
 Min.   : 0.000   Min.   :0.00000   Min.   :0.00000   Min.   :0.000000   Min.   :  0.000
0
 1st Qu.: 1.000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:  0.000
0
 Median : 1.000   Median :0.00000   Median :0.00000   Median :0.000000   Median :  0.000
0
 Mean   : 2.209   Mean   :0.07133   Mean   :0.02409   Mean   :0.002324   Mean   :  0.137
5
 3rd Qu.: 1.000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:  0.000
0
 Max.   :46.000   Max.   :8.00000   Max.   :8.00000   Max.   :2.000000   Max.   :479.000
0
      ios          googleandroid      iphonecampos      samsungcampos        sonycampos
 Min.   :0.0000   Min.   :0.00000   Min.   :  0.0000   Min.   : 0.00000   Min.   :0.0000
00
 1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:  0.0000   1st Qu.: 0.00000   1st Qu.:0.0000
00
 Median :0.0000   Median :0.00000   Median :  0.0000   Median : 0.00000   Median :0.0000
00
 Mean   :0.1527   Mean   :0.03973   Mean   :  0.2891   Mean   : 0.05368   Mean   :0.0096
82
 3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:  0.0000   3rd Qu.: 0.00000   3rd Qu.:0.0000
00
 Max.   :6.0000   Max.   :6.00000   Max.   :156.0000   Max.   :65.00000   Max.   :8.0000
00
  nokiacampos          htccampos         iphonecamneg     samsungcamneg        sonycamne
g
 Min.   : 0.000000   Min.   :  0.0000   Min.   : 0.0000   Min.   : 0.00000   Min.   :0.0
00000
 1st Qu.: 0.000000   1st Qu.:  0.0000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.:0.0
00000
 Median : 0.000000   Median :  0.0000   Median : 0.0000   Median : 0.00000   Median :0.0
00000
 Mean   : 0.006893   Mean   :  0.1132   Mean   : 0.2345   Mean   : 0.05499   Mean   :0.0
02324
 3rd Qu.: 0.000000   3rd Qu.:  0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.00000   3rd Qu.:0.0
00000
 Max.   :17.000000   Max.   :156.0000   Max.   :31.0000   Max.   :17.00000   Max.   :5.0
00000
  nokiacamneg          htccamneg        iphonecamunc     samsungcamunc        sonycamunc
 Min.   : 0.000000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.00000   Min.   :0.00
0000
 1st Qu.: 0.000000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.:0.00
0000
 Median : 0.000000   Median : 0.0000   Median : 0.0000   Median : 0.00000   Median :0.00
0000
 Mean   : 0.006196   Mean   : 0.0931   Mean   : 0.2539   Mean   : 0.01867   Mean   :0.00
3098
 3rd Qu.: 0.000000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.00000   3rd Qu.:0.00
0000
 Max.   :12.000000   Max.   :21.0000   Max.   :45.0000   Max.   :11.00000   Max.   :4.00
0000
```

```
  nokiacamunc         htccamunc         iphonedispos     samsungdispos        sonydispos
 Min.   : 0.00000   Min.   : 0.00000   Min.   : 0.0000   Min.   : 0.00000   Min.   : 0.0
0000
 1st Qu.: 0.00000   1st Qu.: 0.00000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.: 0.0
0000
 Median : 0.00000   Median : 0.00000   Median : 0.0000   Median : 0.00000   Median : 0.0
0000
 Mean   : 0.00364   Mean   : 0.04709   Mean   : 0.7165   Mean   : 0.06305   Mean   : 0.0
1108
 3rd Qu.: 0.00000   3rd Qu.: 0.00000   3rd Qu.: 0.0000   3rd Qu.: 0.00000   3rd Qu.: 0.0
0000
 Max.   :10.00000   Max.   :20.00000   Max.   :66.0000   Max.   :51.00000   Max.   :32.0
0000
  nokiadispos          htcdispos         iphonedisneg      samsungdisneg        sonydisne
g
 Min.   : 0.000000   Min.   :  0.0000   Min.   :  0.0000   Min.   : 0.0000   Min.   : 0.
00000
 1st Qu.: 0.000000   1st Qu.:  0.0000   1st Qu.:  0.0000   1st Qu.: 0.0000   1st Qu.: 0.
00000
 Median : 0.000000   Median :  0.0000   Median :  0.0000   Median : 0.0000   Median : 0.
00000
 Mean   : 0.009991   Mean   :  0.1579   Mean   :  0.6079   Mean   : 0.0663   Mean   : 0.
01247
 3rd Qu.: 0.000000   3rd Qu.:  0.0000   3rd Qu.:  0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.
00000
 Max.   :23.000000   Max.   :479.0000   Max.   :120.0000   Max.   :48.0000   Max.   :97.
00000
  nokiadisneg          htcdisneg        iphonedisunc     samsungdisunc        sonydisunc
 Min.   : 0.000000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.00000   Min.   : 0.0
00000
 1st Qu.: 0.000000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.: 0.0
00000
 Median : 0.000000   Median : 0.0000   Median : 0.0000   Median : 0.00000   Median : 0.0
00000
 Mean   : 0.007978   Mean   : 0.1163   Mean   : 0.5049   Mean   : 0.02719   Mean   : 0.0
04492
 3rd Qu.: 0.000000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.00000   3rd Qu.: 0.0
00000
 Max.   :18.000000   Max.   :48.0000   Max.   :77.0000   Max.   :42.00000   Max.   :11.0
00000
  nokiadisunc          htcdisunc         iphoneperpos    samsungperpos        sonyperpos
 Min.   : 0.000000   Min.   : 0.00000   Min.   : 0.000   Min.   : 0.00000   Min.   : 0.0
0000
 1st Qu.: 0.000000   1st Qu.: 0.00000   1st Qu.: 0.000   1st Qu.: 0.00000   1st Qu.: 0.0
0000
 Median : 0.000000   Median : 0.00000   Median : 0.000   Median : 0.00000   Median : 0.0
0000
 Mean   : 0.004957   Mean   : 0.06274   Mean   : 0.603   Mean   : 0.06204   Mean   : 0.0
0852
 3rd Qu.: 0.000000   3rd Qu.: 0.00000   3rd Qu.: 0.000   3rd Qu.: 0.00000   3rd Qu.: 0.0
0000
 Max.   :16.000000   Max.   :42.00000   Max.   :46.000   Max.   :57.00000   Max.   :14.0
0000
  nokiaperpos          htcperpos       iphoneperneg     samsungperneg        sonyperneg
```

```
 Min.   : 0.00000   Min.   : 0.000   Min.   : 0.0000   Min.   : 0.00000   Min.   : 0.000
000
 1st Qu.: 0.00000   1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.: 0.000
000
 Median : 0.00000   Median : 0.000   Median : 0.0000   Median : 0.00000   Median : 0.000
000
 Mean   : 0.01069   Mean   : 0.128   Mean   : 0.4855   Mean   : 0.06444   Mean   : 0.006
429
 3rd Qu.: 0.00000   3rd Qu.: 0.000   3rd Qu.: 0.0000   3rd Qu.: 0.00000   3rd Qu.: 0.000
000
 Max.   :21.00000   Max.   :65.000   Max.   :33.0000   Max.   :33.00000   Max.   :28.000
000
  nokiaperneg        htcperneg        iphoneperunc    samsungperunc        sonyperunc
 Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.000   Min.   : 0.00000   Min.   :0.00000
0
 1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.00000   1st Qu.:0.00000
0
 Median : 0.0000   Median : 0.0000   Median : 0.000   Median : 0.00000   Median :0.00000
0
 Mean   : 0.0103   Mean   : 0.1207   Mean   : 0.298   Mean   : 0.02207   Mean   :0.00216
9
 3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.: 0.00000   3rd Qu.:0.00000
0
 Max.   :18.0000   Max.   :33.0000   Max.   :31.000   Max.   :31.00000   Max.   :5.00000
0
  nokiaperunc          htcperunc          iosperpos         googleperpos        iosperne
g
 Min.   : 0.000000   Min.   : 0.00000   Min.   : 0.00000   Min.   : 0.00000   Min.   :
0.00000
 1st Qu.: 0.000000   1st Qu.: 0.00000   1st Qu.: 0.00000   1st Qu.: 0.00000   1st Qu.:
0.00000
 Median : 0.000000   Median : 0.00000   Median : 0.00000   Median : 0.00000   Median :
0.00000
 Mean   : 0.005577   Mean   : 0.06421   Mean   : 0.05948   Mean   : 0.03873   Mean   :
0.05786
 3rd Qu.: 0.000000   3rd Qu.: 0.00000   3rd Qu.: 0.00000   3rd Qu.: 0.00000   3rd Qu.:
0.00000
 Max.   :13.000000   Max.   :59.00000   Max.   :26.00000   Max.   :24.00000   Max.   :2
0.00000
  googleperneg        iosperunc         googleperunc      galaxysentiment
 Min.   : 0.00000   Min.   : 0.00000   Min.   : 0.00000   Min.   :1.000
 1st Qu.: 0.00000   1st Qu.: 0.00000   1st Qu.: 0.00000   1st Qu.:3.000
 Median : 0.00000   Median : 0.00000   Median : 0.00000   Median :4.000
 Mean   : 0.05848   Mean   : 0.03323   Mean   : 0.01619   Mean   :3.356
 3rd Qu.: 0.00000   3rd Qu.: 0.00000   3rd Qu.: 0.00000   3rd Qu.:4.000
 Max.   :25.00000   Max.   :10.00000   Max.   :24.00000   Max.   :4.000
```

Hide

```
str(samsung_RC)
```

```
'data.frame':   12911 obs. of  59 variables:
 $ iphone          : int  1 1 1 0 1 2 1 1 4 1 ...
 $ samsunggalaxy   : int  0 0 1 0 0 0 0 0 0 0 ...
 $ sonyxperia      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokialumina     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htcphone        : int  0 0 0 1 0 0 0 0 0 0 ...
 $ ios             : int  0 0 0 0 0 0 0 0 0 0 ...
 $ googleandroid   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ iphonecampos    : int  0 0 1 0 0 1 0 0 0 0 ...
 $ samsungcampos   : int  0 0 1 0 0 0 0 0 0 0 ...
 $ sonycampos      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiacampos     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htccampos       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ iphonecamneg    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ samsungcamneg   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ sonycamneg      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiacamneg     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htccamneg       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ iphonecamunc    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ samsungcamunc   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ sonycamunc      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiacamunc     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htccamunc       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ iphonedispos    : int  0 1 0 0 0 0 2 0 0 0 ...
 $ samsungdispos   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ sonydispos      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiadispos     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htcdispos       : int  0 0 0 1 0 0 0 0 0 0 ...
 $ iphonedisneg    : int  0 1 0 0 0 0 0 0 0 0 ...
 $ samsungdisneg   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ sonydisneg      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiadisneg     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htcdisneg       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ iphonedisunc    : int  0 1 0 0 0 0 0 0 0 0 ...
 $ samsungdisunc   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ sonydisunc      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiadisunc     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htcdisunc       : int  0 0 0 1 0 0 0 0 0 0 ...
 $ iphoneperpos    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ samsungperpos   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ sonyperpos      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiaperpos     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htcperpos       : int  0 0 0 1 0 0 0 0 0 0 ...
 $ iphoneperneg    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ samsungperneg   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ sonyperneg      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiaperneg     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htcperneg       : int  0 0 0 1 0 0 0 0 0 0 ...
 $ iphoneperunc    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ samsungperunc   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ sonyperunc      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nokiaperunc     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ htcperunc       : int  0 0 0 1 0 0 0 0 0 0 ...
```

```
 $ iosperpos      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ googleperpos   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ iosperneg      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ googleperneg   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ iosperunc      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ googleperunc   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ galaxysentiment: num  4 3 3 1 1 1 3 4 4 4 ...
```

Hide

```
# make galaxysentiment a factor
samsung_RC$galaxysentiment <- as.factor(samsung_RC$galaxysentiment)
```

samsung_RC

Hide

```
# convert variable types, categorical
#samsung_RC$galaxysentiment <- as.factor(samsung_RC$galaxysentiment)
```

Hide

```
# create 10-fold cross validation fitcontrol
fitControl <- trainControl(method = "cv", number = 10)
```

Train and Test Set:

Hide

```
# Create Train and Test Set for samsung_RC
# create 75% sample of row indices
in_training <-createDataPartition(samsung_RC$galaxysentiment, p = .7, list = FALSE)
# create 75% sample of data and save it to trainData
trainData_samsung_RC <- samsung_RC[in_training, ]
 # create 25% sample of data and save it to test_data
testData_samsung_RC <- samsung_RC[-in_training, ]
# verify split percentages
nrow(trainData_samsung_RC) / nrow(samsung_RC)
```

```
[1] 0.7001007
```

Hide

```
#c5
c5_samsung_RC <- train(galaxysentiment ~., data = trainData_samsung_RC, method = "C5.0",
trControl = fitControl)
```

Compare Accuracy on Prediction Results:

Hide

```
#c5
prediction_c5_samsung_RC <- predict(c5_samsung_RC, testData_samsung_RC)
postResample(prediction_c5_samsung_RC, testData_samsung_RC$galaxysentiment)
```

```
 Accuracy     Kappa
0.8416839 0.5901211
```

Hide

```
summary(c5_samsung_RC)
```

```
Call:
(function (x, y, trials = 1, rules = FALSE, weights = NULL, control = C5.0Control(), cos
ts
 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1, 0,
0, 0,
 0, 0, 0,


C5.0 [Release 2.07 GPL Edition]     Mon Mar  9 20:19:57 2020
-------------------------------

Class specified by attribute `outcome'

Read 9039 cases (59 attributes) from undefined.data

34 attributes winnowed
Estimated importance of remaining attributes:

     33%  iphone
     10%  iphonedispos
      9%  iphoneperpos
      8%  iphonedisunc
      8%  googleandroid
      7%  sonyxperia
      7%  iphonedisneg
      7%  samsunggalaxy
      5%  samsungcamneg
      2%  htccampos
      2%  iphonecampos
      1%  samsungperpos
      1%  iphoneperneg
     <1%  iphonecamunc
     <1%  htcphone
     <1%  ios
     <1%  samsungcampos
     <1%  iphonecamneg
     <1%  samsungdispos
     <1%  samsungdisneg
     <1%  samsungperneg
     <1%  iphoneperunc
     <1%  iosperneg
     <1%  iosperunc

Decision tree:

iphone <= 0:
:...googleandroid > 0: 2 (48/7)
:   googleandroid <= 0:
:   :...samsungperneg <= 0: 1 (614/45)
:       samsungperneg > 0: 4 (18/8)
iphone > 0:
:...samsungcamneg > 4: 1 (85/6)
```

```
samsungcamneg <= 4:
:...sonyxperia > 0:
    :...iphonedisneg > 0: 4 (8/1)
    :   iphonedisneg <= 0:
    :   :...samsungperpos <= 0: 1 (144/17)
    :       samsungperpos > 0: 4 (24/6)
    sonyxperia <= 0:
    :...googleandroid > 0:
        :...iphonedispos <= 0: 3 (99/11)
        :   iphonedispos > 0: 4 (12/5)
        googleandroid <= 0:
        :...samsunggalaxy > 0:
            :...iphoneperpos > 0: 4 (13)
            :   iphoneperpos <= 0:
            :   :...iphonecamunc <= 0: 3 (145/20)
            :       iphonecamunc > 0: 1 (8/3)
            samsunggalaxy <= 0:
            :...htccampos > 0:
                :...iphonedisunc <= 1:
                :   :...iphone <= 2: 1 (34/5)
                :   :   iphone > 2: 4 (6)
                :   iphonedisunc > 1:
                :   :...htcphone <= 1: 3 (18/4)
                :       htcphone > 1: 4 (8/3)
                htccampos <= 0:
                :...iphonedispos <= 0:
                    :...iphonedisunc <= 0: 4 (5903/985)
                    :   iphonedisunc > 0:
                    :   :...iphoneperneg > 0: 4 (45/5)
                    :       iphoneperneg <= 0:
                    :       :...iphoneperunc > 0: 4 (11)
                    :           iphoneperunc <= 0:
                    :           :...iphonedisneg > 0: 4 (23/8)
                    :               iphonedisneg <= 0:
                    :               :...iphoneperpos > 0: 4 (10/4)
                    :                   iphoneperpos <= 0:
                    :                   :...iphone <= 3: 3 (63/12)
                    :                       iphone > 3: 4 (6/1)
                    iphonedispos > 0:
                    :...iphoneperpos > 0:
                        :...iphonecamneg <= 1: 4 (885/136)
                        :   iphonecamneg > 1:
                        :   :...iphonedisunc > 0: 4 (97/14)
                        :       iphonedisunc <= 0:
                        :       :...iphoneperpos <= 1: 4 (7/1)
                        :           iphoneperpos > 1: 1 (9/1)
                        iphoneperpos <= 0:
                        :...iphonedisneg <= 0:
                            :...iphone <= 1:
                            :   :...iphonecampos <= 0: 3 (201/29)
                            :   :   iphonecampos > 0:
                            :   :   :...iphonedisunc <= 0: 4 (23/3)
                            :   :       iphonedisunc > 0: 3 (10/3)
                            :   iphone > 1:
```

```
                                :   :...iphonedisunc <= 0: 4 (43/8)
                                :       iphonedisunc > 0:
                                :       :...iphonedispos <= 2: 3 (22/6)
                                :           iphonedispos > 2: 4 (11/1)
                                iphonedisneg > 0:
                                :...iphonedispos > 1: 4 (107/10)
                                    iphonedispos <= 1:
                                    :...iphonedisunc <= 0: 4 (56/8)
                                        iphonedisunc > 0:
                                        :...iphone > 1: 4 (185/5)
                                            iphone <= 1:
                                            :...iphoneperneg <= 0: 3 (34/9)
                                                iphoneperneg > 0: 4 (4/1)


Evaluation on training data (9039 cases):

        Decision Tree
      ----------------
      Size      Errors

        38 1391(15.4%)   <<


       (a)   (b)   (c)   (d)    <-classified as
      ----  ----  ----  ----
       817     2     8   628    (a): class 1
         2    41     2   270    (b): class 2
         9     1   498   315    (c): class 3
        66     4    84  6292    (d): class 4


    Attribute usage:

    100.00% iphone
     97.11% googleandroid
     92.48% samsungcamneg
     91.54% sonyxperia
     88.36% samsunggalaxy
     87.02% iphonedispos
     86.53% htccampos
     73.33% iphonedisunc
     21.45% iphoneperpos
     11.04% iphonecamneg
     10.78% iphonedisneg
      6.99% samsungperneg
      2.59% iphonecampos
      2.17% iphoneperneg
      1.86% samsungperpos
      1.69% iphonecamunc
      1.25% iphoneperunc
      0.29% htcphone
```

```
Time: 0.2 secs
```