task3_galaxy_PCA

Code

Hide

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

Hide

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

Train and Test Set:

Hide

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

```
[1] 0.7001781
```

Hide

```
# data = training and testing from iphoneDF (no feature selection)
# create object containing centered, scaled PCA components from training set
# excluded the dependent variable and set threshold to .95
preprocessParams <- preProcess(trainData_samsung[,-59], method=c("center", "scale", "pc
a"), thresh = 0.95)
print(preprocessParams)
```

```
Created from 9040 samples and 58 variables

Pre-processing:
  - centered (58)
  - ignored (0)
  - principal component signal extraction (58)
  - scaled (58)

PCA needed 25 components to capture 95 percent of the variance
```

Hide

```
# use predict to apply pca parameters, create training, exclude dependant
train.pca <- predict(preprocessParams, trainData_samsung[,-59])

# add the dependent to training
train.pca$galaxysentiment <- trainData_samsung$galaxysentiment

# use predict to apply pca parameters, create testing, exclude dependant
test.pca <- predict(preprocessParams, testData_samsung[,-59])

# add the dependent to training
test.pca$galaxysentiment <- testData_samsung$galaxysentiment

# inspect results
str(train.pca)
```

```
'data.frame':   9040 obs. of  26 variables:
 $ PC1            : num  -0.6816 -0.6027 -0.0928 0.1652 -0.6816 ...
 $ PC2            : num  0.1305 0.1013 -0.0533 -0.2391 0.1305 ...
 $ PC3            : num  -0.515 -0.174 -1.004 0.236 -0.515 ...
 $ PC4            : num  0.37 0.089 0.207 1.711 0.37 ...
 $ PC5            : num  0.063 0.1245 -0.0977 -0.3891 0.063 ...
 $ PC6            : num  0.044 -0.0141 -0.1374 0.7977 0.044 ...
 $ PC7            : num  -0.0997 -0.2746 -0.1397 0.0972 -0.0997 ...
 $ PC8            : num  -0.0438 -0.1754 -0.3818 -0.3953 -0.0438 ...
 $ PC9            : num  0.0241 0.0399 -0.1968 -0.074 0.0241 ...
 $ PC10           : num  -0.034 -0.0305 -0.0391 -0.0186 -0.034 ...
 $ PC11           : num  -0.0743 -0.1726 1.0532 -1.0311 -0.0743 ...
 $ PC12           : num  -0.222 -0.127 1.175 0.141 -0.222 ...
 $ PC13           : num  0.0141 0.1783 0.4664 -0.1372 0.0141 ...
 $ PC14           : num  -0.1372 -0.1234 1.7503 -0.0771 -0.1372 ...
 $ PC15           : num  -0.000969 -0.051562 -0.060297 -0.106255 -0.000969 ...
 $ PC16           : num  -0.0439 -0.1734 0.3044 0.1523 -0.0439 ...
 $ PC17           : num  0.125 0.481 -0.102 0.206 0.125 ...
 $ PC18           : num  -0.0329 -0.0716 -0.3775 -0.1383 -0.0329 ...
 $ PC19           : num  0.054 0.0859 -1.1555 0.0522 0.054 ...
 $ PC20           : num  0.0266 0.00666 -0.25682 0.27631 0.0266 ...
 $ PC21           : num  0.0258 0.0159 -1.7615 -0.0268 0.0258 ...
 $ PC22           : num  -0.0646 -0.0622 0.5349 0.1169 -0.0646 ...
 $ PC23           : num  0.01446 -0.00576 -0.09571 -0.4197 0.01446 ...
 $ PC24           : num  -0.0485 -0.0427 0.3178 0.2788 -0.0485 ...
 $ PC25           : num  0.0266 0.0207 0.3412 -0.4983 0.0266 ...
 $ galaxysentiment: Factor w/ 6 levels "0","1","2","3",..: 6 4 4 1 2 1 4 6 6 6 ...
```

Hide

```
str(test.pca)
```

```
'data.frame':   3871 obs. of  26 variables:
 $ PC1            : num  -0.66 -0.68 -0.628 -0.618 -0.303 ...
 $ PC2            : num  0.1167 0.1294 0.1052 0.106 -0.0202 ...
 $ PC3            : num  -0.466 -0.513 -0.345 -0.345 -1.127 ...
 $ PC4            : num  0.33266 0.37607 0.23915 0.39107 -0.00571 ...
 $ PC5            : num  0.0765 0.059 0.0966 0.0325 -1.583 ...
 $ PC6            : num  -0.1327 0.0485 -0.19 0.0157 -0.3219 ...
 $ PC7            : num  0.08261 -0.09722 0.07429 0.00201 -0.29279 ...
 $ PC8            : num  -0.0261 -0.0563 -0.0661 -0.1718 -0.2244 ...
 $ PC9            : num  0.0171 0.0102 0.0178 -0.0404 -0.5823 ...
 $ PC10           : num  -0.0303 0.1052 -0.0292 -0.0612 0.1685 ...
 $ PC11           : num  -0.1544 -0.0781 -0.1838 0.3652 2.116 ...
 $ PC12           : num  -0.181 -0.223 -0.153 -0.459 4.572 ...
 $ PC13           : num  -0.00571 0.01452 0.029 -0.07072 -0.50579 ...
 $ PC14           : num  -0.13 -0.137 -0.125 -0.122 3.706 ...
 $ PC15           : num  -0.002886 -0.000101 -0.017586 -0.008876 -0.736172 ...
 $ PC16           : num  -0.0377 -0.0435 -0.0647 -0.0302 0.692 ...
 $ PC17           : num  0.1041 0.1224 0.182 0.173 0.0651 ...
 $ PC18           : num  0.0388 -0.0333 0.055 -0.1582 -0.1038 ...
 $ PC19           : num  0.0514 0.054 0.0587 0.0607 0.15 ...
 $ PC20           : num  -0.0972 0.0302 -0.1553 0.2822 0.2273 ...
 $ PC21           : num  0.0321 0.027 0.0325 0.0184 2.5417 ...
 $ PC22           : num  -0.1178 -0.0638 -0.1312 -0.0418 -0.0735 ...
 $ PC23           : num  0.0347 0.0115 0.0404 -0.0483 0.3563 ...
 $ PC24           : num  -0.0341 -0.0491 -0.0194 0.0719 0.1551 ...
 $ PC25           : num  0.0176 0.0306 -0.0221 -0.0553 0.1746 ...
 $ galaxysentiment: Factor w/ 6 levels "0","1","2","3",..: 6 6 2 5 1 4 2 6 5 3 ...
```

Hide

```
#c5
c5_samsung_PCA<- train(galaxysentiment ~., data = train.pca, method = "C5.0",trControl =
fitControl)
```

Compare Accuracy on Prediction Results:

Hide

```
#c5
prediction_c5_samsung_PCA<- predict(c5_iphone_smallMatrix_PCA, test.pca)
postResample(prediction_c5_iphone_smallMatrix_PCA, test.pca$galaxysentiment)
```

```
   Accuracy       Kappa
0.503229140 0.008097589
```

Hide

```
summary(c5_iphone_smallMatrix_PCA)
```

```
Call:
(function (x, y, trials = 1, rules = FALSE, weights = NULL, control = C5.0Control(), cos
ts
 -0.18275823174015, -0.0789439504431598, -0.252977512491766, 5.42983713168266,
 0.195568925219517, -0.536719375413499, -0.199678919427778, -0.18275823174015, 0.00


C5.0 [Release 2.07 GPL Edition]     Mon Mar  9 17:50:40 2020
-------------------------------

Class specified by attribute `outcome'

Read 9083 cases (26 attributes) from undefined.data

Rules:

Rule 1: (621/4, lift 6.6)
    PC4 <= -0.5914555
    PC5 <= 0.3666646
    PC18 <= 0.8474609
    PC21 <= 1.959069
    ->  class 0  [0.992]

Rule 2: (46, lift 6.5)
    PC1 > 0.6845021
    PC17 > 0.07093082
    ->  class 0  [0.979]

Rule 3: (170/4, lift 6.4)
    PC2 <= -0.04554113
    PC3 > -1.951145
    ->  class 0  [0.971]

Rule 4: (97/3, lift 6.3)
    PC4 <= -0.9458116
    PC18 > -0.5910822
    PC18 <= 0.8474609
    PC21 > -0.3507631
    PC21 <= 1.959069
    ->  class 0  [0.960]

Rule 5: (250/12, lift 6.3)
    PC7 <= 1.749553
    PC9 > 0.09286059
    PC10 > -0.2345495
    PC14 <= 0.3078527
    PC17 > -1.320543
    PC17 <= 0.07093082
    PC23 > -0.3260998
    ->  class 0  [0.948]

Rule 6: (32/1, lift 6.2)
```

```
    PC6 <= 0.4673022
    PC11 > -0.02374925
    PC17 <= 0.07093082
    PC23 <= -0.3260998
    ->  class 0  [0.941]

Rule 7: (217/13, lift 6.2)
    PC4 <= -1.151087
    PC6 <= 8.392989
    PC21 <= 1.959069
    ->  class 0  [0.936]

Rule 8: (106/7, lift 6.1)
    PC5 <= 0.2339443
    PC9 > -0.02000268
    PC17 > 0.1170632
    PC18 > -0.4109737
    PC20 <= 0.01174213
    ->  class 0  [0.926]

Rule 9: (64/7, lift 5.8)
    PC2 > -0.04554113
    PC4 > -0.5914555
    PC16 > 0.7285371
    PC25 > -1.005402
    ->  class 0  [0.879]

Rule 10: (5, lift 5.7)
    PC2 > -0.04554113
    PC14 <= 0.1545502
    PC17 > 0.1170632
    PC20 > 0.019214
    PC23 <= -0.4745359
    ->  class 0  [0.857]

Rule 11: (10/1, lift 5.5)
    PC4 <= -0.5914555
    PC6 <= 0.2774508
    PC21 > 1.959069
    ->  class 0  [0.833]

Rule 12: (10/1, lift 5.5)
    PC12 <= -0.9229887
    PC21 <= -1.874327
    ->  class 0  [0.833]

Rule 13: (12/2, lift 5.2)
    PC9 <= -0.05301111
    PC17 > 0.1170632
    PC20 > 0.019214
    PC23 <= 0.441895
    ->  class 0  [0.786]

Rule 14: (36/1, lift 27.1)
```

```
    PC4 <= -0.5914555
    PC6 > 0.2774508
    PC21 > 1.959069
    PC23 <= 2.153342
    ->  class 2  [0.947]

Rule 15: (5/1, lift 20.4)
    PC2 > -0.04554113
    PC5 > 0.2339443
    PC17 > 0.1170632
    PC19 > 0.0681016
    PC19 <= 0.1761989
    PC20 <= 0.019214
    ->  class 2  [0.714]

Rule 16: (182, lift 10.9)
    PC1 > 0.6212639
    PC2 > -0.04554113
    PC10 <= 0.02375685
    PC17 > 0.1170632
    PC20 <= 0.019214
    ->  class 3  [0.995]

Rule 17: (130, lift 10.8)
    PC1 > 0.4197969
    PC18 > 0.8474609
    ->  class 3  [0.992]

Rule 18: (105/4, lift 10.4)
    PC4 <= -0.5914555
    PC21 > 1.959069
    PC23 > 2.153342
    ->  class 3  [0.953]

Rule 19: (70/3, lift 10.3)
    PC4 > -0.2593564
    PC5 <= 0.2339443
    PC9 <= -0.02000268
    PC16 <= 0.04571968
    PC17 > 0.1170632
    PC19 <= 0.4555009
    PC20 <= 0.01174213
    PC21 <= 0.09331081
    PC23 > -0.1284413
    ->  class 3  [0.944]

Rule 20: (31/1, lift 10.3)
    PC2 > -0.04554113
    PC4 > -0.2593564
    PC9 <= -0.02000268
    PC17 > 0.1170632
    PC20 <= 0.01174213
    PC23 > -0.1284413
    PC23 <= -0.08829965
```

```
    ->  class 3  [0.939]

Rule 21: (38/2, lift 10.1)
    PC4 > -0.2593564
    PC5 <= 0.2339443
    PC9 <= -0.02000268
    PC15 <= 0.0342439
    PC17 > 0.1170632
    PC20 <= 0.01174213
    PC22 > -0.05041727
    PC23 > -0.1718248
    ->  class 3  [0.925]

Rule 22: (27/3, lift 9.4)
    PC17 > 0.1170632
    PC20 > 0.01772627
    PC20 <= 0.019214
    PC22 > -0.05041727
    PC23 > -0.1718248
    ->  class 3  [0.862]

Rule 23: (11/2, lift 8.4)
    PC5 <= 0.2339443
    PC9 > -0.02000268
    PC17 > 0.1170632
    PC18 <= -0.6602482
    PC19 <= 0.4555009
    PC20 <= 0.019214
    PC22 > -0.05041727
    PC23 > -0.1718248
    ->  class 3  [0.769]

Rule 24: (9/3, lift 6.9)
    PC4 > -0.9458116
    PC4 <= -0.5914555
    PC5 > 0.3666646
    PC19 > 1.15318
    ->  class 3  [0.636]

Rule 25: (11/4, lift 6.7)
    PC3 <= -1.951145
    PC4 > -0.5914555
    PC9 <= 0.1191311
    PC23 > -2.621573
    ->  class 3  [0.615]

Rule 26: (176, lift 9.0)
    PC6 > 8.392989
    ->  class 4  [0.994]

Rule 27: (92, lift 8.9)
    PC3 > -1.972841
    PC3 <= -1.951145
    ->  class 4  [0.989]
```

```
Rule 28: (200/9, lift 8.6)
    PC2 > 0.1977609
    PC3 > -1.951145
    PC4 > -0.5914555
    PC9 <= 0.09286059
    PC13 <= 1.944546
    PC16 <= 0.7285371
    PC17 <= 0.1170632
    PC18 <= 1.576866
    PC19 <= 0.257059
    PC23 > -0.3260998
    ->  class 4  [0.950]

Rule 29: (10, lift 8.3)
    PC3 <= -1.951145
    PC5 <= 0.5094531
    PC7 > 1.126188
    PC21 > -1.874327
    ->  class 4  [0.917]

Rule 30: (8, lift 8.1)
    PC4 > -0.5914555
    PC16 <= 0.7285371
    PC17 > 0.1170632
    PC17 <= 0.4868065
    PC20 <= 0.019214
    PC21 > -0.5795799
    PC22 > -0.05041727
    PC23 <= -0.1718248
    ->  class 4  [0.900]

Rule 31: (6, lift 7.9)
    PC4 > -0.5914555
    PC14 <= 0.1545502
    PC16 <= 0.7285371
    PC17 > 0.1170632
    PC20 > 0.019214
    PC20 <= 0.05844372
    PC23 > -0.4745359
    ->  class 4  [0.875]

Rule 32: (5, lift 7.7)
    PC3 > -1.951145
    PC7 > 1.749553
    PC17 <= 0.1170632
    PC21 <= -0.1933176
    PC23 > -0.3260998
    ->  class 4  [0.857]

Rule 33: (3, lift 7.2)
    PC3 <= -1.951145
    PC5 > 0.5094531
    PC7 <= -8.162539
```

```
    ->  class 4  [0.800]

Rule 34: (7/1, lift 7.0)
    PC3 <= -1.951145
    PC4 > -0.5914555
    PC9 <= 0.1191311
    PC23 <= -2.621573
    ->  class 4  [0.778]

Rule 35: (14/5, lift 5.6)
    PC4 > -0.5914555
    PC6 > 0.4673022
    PC11 > -0.02374925
    PC17 <= 0.07093082
    PC23 <= -0.3260998
    ->  class 4  [0.625]

Rule 36: (22/10, lift 4.9)
    PC2 > -0.04554113
    PC4 > -0.5914555
    PC11 <= 0.4507242
    PC16 <= -0.07373721
    PC17 > 0.1170632
    ->  class 4  [0.542]

Rule 37: (8061/7064, lift 1.1)
    PC4 > -0.5914555
    ->  class 4  [0.124]

Rule 38: (8907/3629, lift 1.0)
    PC6 <= 8.392989
    ->  class 5  [0.593]

Default class: 5


Evaluation on training data (9083 cases):

           Rules
     ----------------
       No      Errors

       38 2028(22.3%)   <<


      (a)   (b)   (c)   (d)   (e)   (f)    <-classified as
     ----  ----  ----  ----  ----  ----
      915                 2     8   449    (a): class 0
        2                       1   270    (b): class 1
        2          39     2     2   273    (c): class 2
        6               526           300    (d): class 3
       11           2     5   338   652    (e): class 4
       23                 8    10  5237    (f): class 5
```

```
    Attribute usage:

    100.00% PC6
     98.30% PC4
     11.55% PC18
     10.94% PC17
     10.51% PC21
      9.01% PC23
      8.61% PC5
      7.57% PC9
      7.46% PC2
      5.46% PC3
      4.78% PC20
      4.76% PC10
      4.00% PC16
      3.94% PC1
      3.25% PC19
      2.95% PC7
      2.87% PC14
      2.20% PC13
      0.92% PC22
      0.75% PC11
      0.70% PC25
      0.42% PC15
      0.11% PC12


Time: 0.5 secs
```