PCA

Hide

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

Hide

```
iphone_smallMatrix_PCA <- iphoneDFBig
```

Train and Test Set:

Hide

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

```
[1] 0.7001465
```

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_iphoneDFBig[,-59], method=c("center", "scale",
"pca"), thresh = 0.95)
print(preprocessParams)
```

Hide

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

# add the dependent to training
train.pca$iphonesentiment <- trainData_iphoneDFBig$iphonesentiment

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

# add the dependent to training
test.pca$iphonesentiment <- testData_iphoneDFBig$iphonesentiment

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

```
'data.frame':    9083 obs. of  26 variables:
 $ PC1            : num  0.708 0.536 0.642 -0.933 0.426 ...
 $ PC2            : num  -0.04457 -0.01077 -0.0344 0.3211 0.00265 ...
 $ PC3            : num  0.0825 -0.0445 0.0324 -1.2297 -0.1255 ...
 $ PC4            : num  -0.587 -0.127 -0.407 4.03 0.134 ...
 $ PC5            : num  0.127 0.228 0.168 -3.289 0.299 ...
 $ PC6            : num  -0.183 -0.359 -0.247 9.104 -0.432 ...
 $ PC7            : num  0.0989 0.217 0.1281 -1.6555 0.1992 ...
 $ PC8            : num  0.109 0.011 0.0405 5.2398 -0.1929 ...
 $ PC9            : num  -0.0489 -0.0437 -0.0429 -0.2676 -0.0215 ...
 $ PC10           : num  0.02621 0.00438 0.0166 -0.03889 -0.00241 ...
 $ PC11           : num  -0.1036 -0.0639 -0.0904 -0.2311 -0.0719 ...
 $ PC12           : num  -0.1289 -0.1403 -0.1375 0.0693 -0.159 ...
 $ PC13           : num  0.141 0.239 0.175 -1.265 0.225 ...
 $ PC14           : num  0.12734 -0.00727 0.07946 0.81924 0.00402 ...
 $ PC15           : num  -0.00633 -0.0321 -0.02504 -0.07638 -0.07509 ...
 $ PC16           : num  0.0494 0.0538 0.0654 0.1774 0.1143 ...
 $ PC17           : num  0.0622 -0.2513 -0.0876 -0.0363 -0.4897 ...
 $ PC18           : num  -0.0797 0.4076 0.1324 0.4854 0.6994 ...
 $ PC19           : num  -0.0172 -0.2356 -0.1127 -0.1942 -0.3827 ...
 $ PC20           : num  0.01405 -0.01589 0.01716 -0.00747 0.03424 ...
 $ PC21           : num  -0.02375 0.06051 -0.00759 0.10619 -0.04603 ...
 $ PC22           : num  0.01724 -0.00425 -0.02552 -0.12752 -0.12361 ...
 $ PC23           : num  -0.072 0.1022 -0.026 0.0933 0.0299 ...
 $ PC24           : num  0.0135 -0.073 0.0776 0.0522 0.2416 ...
 $ PC25           : num  0.0324 0.1693 -0.1004 -0.1783 -0.4315 ...
 $ iphonesentiment: Factor w/ 6 levels "0","1","2","3",..: 1 1 1 5 1 1 4 1 1 1 ...
```

Hide

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

```
'data.frame':    3890 obs. of  26 variables:
 $ PC1           : num   0.642 0.708 -2.038 0.708 -20.891 ...
 $ PC2           : num   -0.0344 -0.0446 0.4915 -0.0446 4.8429 ...
 $ PC3           : num   0.0324 0.0825 -1.9511 0.0825 -1.8112 ...
 $ PC4           : num   -0.407 -0.587 7.097 -0.587 -4.136 ...
 $ PC5           : num   0.168 0.127 0.779 0.127 -0.784 ...
 $ PC6           : num   -0.247 -0.183 -0.596 -0.183 5.43 ...
 $ PC7           : num   0.1281 0.0989 3.2864 0.0989 -2.1174 ...
 $ PC8           : num   0.0405 0.109 -1.3145 0.109 -7.9419 ...
 $ PC9           : num   -0.0429 -0.0489 0.1719 -0.0489 0.7767 ...
 $ PC10          : num   0.0166 0.0262 -0.1893 0.0262 0.8399 ...
 $ PC11          : num   -0.0904 -0.1036 1.0452 -0.1036 -1.9983 ...
 $ PC12          : num   -0.138 -0.129 -0.343 -0.129 -0.141 ...
 $ PC13          : num   0.175 0.141 -0.457 0.141 -0.744 ...
 $ PC14          : num   0.0795 0.1273 0.1677 0.1273 1.3906 ...
 $ PC15          : num   -0.02504 -0.00633 0.20063 -0.00633 0.30032 ...
 $ PC16          : num   0.0654 0.0494 -0.0541 0.0494 -0.0313 ...
 $ PC17          : num   -0.0876 0.0622 0.1242 0.0622 -0.4449 ...
 $ PC18          : num   0.1324 -0.0797 -0.3896 -0.0797 -0.082 ...
 $ PC19          : num   -0.1127 -0.0172 0.2318 -0.0172 0.4166 ...
 $ PC20          : num   0.0172 0.0141 -0.0894 0.0141 0.0358 ...
 $ PC21          : num   -0.00759 -0.02375 0.15021 -0.02375 -0.38734 ...
 $ PC22          : num   -0.0255 0.0172 0.1192 0.0172 0.1915 ...
 $ PC23          : num   -0.026 -0.072 0.0807 -0.072 -0.0616 ...
 $ PC24          : num   0.0776 0.0135 -0.3547 0.0135 0.7918 ...
 $ PC25          : num   -0.1004 0.0324 0.7742 0.0324 0.1675 ...
 $ iphonesentiment: Factor w/ 6 levels "0","1","2","3",..: 1 1 5 1 1 1 1 1 1 1 ...
```

Hide

```
#c5
c5_iphone_smallMatrix_PCA <- train(iphonesentiment ~., data = train.pca, method = "C5.0"
,trControl = fitControl)
```

Compare Accuracy on Prediction Results:

Hide

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

```
Accuracy    Kappa
0.766838 0.548541
```

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, 0.0069106
5


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
```