Homework 3

Asif Hasan - 301376671

2023-11-17

1. Problem Set 13, Applications

```
library(FNN)
```

```
## Warning: package 'FNN' was built under R version 4.1.2
```

1(a).

```
# read the data
vehicle <- read.csv("vehicle.csv")
# print summary
summary(vehicle)
```

```
##   Compactness      Circularity    Distance.Circularity  Radius.Ratio
##  Min.   : 73.00   Min.   :33.00   Min.   : 40.00       Min.   :104.0
##  1st Qu.: 87.00   1st Qu.:40.00   1st Qu.: 70.00       1st Qu.:141.0
##  Median : 93.00   Median :44.00   Median : 80.00       Median :167.0
##  Mean   : 93.68   Mean   :44.86   Mean   : 82.09       Mean   :168.9
##  3rd Qu.:100.00   3rd Qu.:49.00   3rd Qu.: 98.00       3rd Qu.:195.0
##  Max.   :119.00   Max.   :59.00   Max.   :112.00       Max.   :333.0
##  Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio   Elongatedness
##  Min.   : 47.00       Min.   : 2.000          Min.   :112.0   Min.   :26.00
##  1st Qu.: 57.00       1st Qu.: 7.000          1st Qu.:146.2   1st Qu.:33.00
##  Median : 61.00       Median : 8.000          Median :157.0   Median :43.00
##  Mean   : 61.69       Mean   : 8.567          Mean   :168.8   Mean   :40.93
##  3rd Qu.: 65.00       3rd Qu.:10.000          3rd Qu.:198.0   3rd Qu.:46.00
##  Max.   :138.00       Max.   :55.000          Max.   :265.0   Max.   :61.00
##  Pr.Axis.Rectangularity Max.Length.Rectangularity
##  Min.   :17.00          Min.   :118
##  1st Qu.:19.00          1st Qu.:137
##  Median :20.00          Median :146
##  Mean   :20.58          Mean   :148
##  3rd Qu.:23.00          3rd Qu.:159
##  Max.   :29.00          Max.   :188
##  Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis
##  Min.   :130.0                    Min.   : 184.0
##  1st Qu.:167.0                    1st Qu.: 318.2
##  Median :178.5                    Median : 364.0
##  Mean   :188.6                    Mean   : 439.9
##  3rd Qu.:217.0                    3rd Qu.: 587.0
##  Max.   :320.0                    Max.   :1018.0
##  Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis
##  Min.   :109.0             Min.   : 59.00            Min.   : 0.000
##  1st Qu.:149.0             1st Qu.: 67.00            1st Qu.: 2.000
##  Median :173.0             Median : 71.50            Median : 6.000
##  Mean   :174.7             Mean   : 72.46            Mean   : 6.377
##  3rd Qu.:198.0             3rd Qu.: 75.00            3rd Qu.: 9.000
##  Max.   :268.0             Max.   :135.00            Max.   :22.000
##  Kurtosis.About.Minor.Axis Kurtosis.About.Major.Axis Hollows.Ratio
##  Min.   : 0.0              Min.   :176.0             Min.   :181.0
##  1st Qu.: 5.0              1st Qu.:184.0             1st Qu.:190.2
##  Median :11.0              Median :188.0             Median :197.0
##  Mean   :12.6              Mean   :188.9             Mean   :195.6
##  3rd Qu.:19.0              3rd Qu.:193.0             3rd Qu.:201.0
##  Max.   :41.0              Max.   :206.0             Max.   :211.0
##      class
##  Min.   :1.000
##  1st Qu.:1.250
##  Median :2.000
##  Mean   :2.478
##  3rd Qu.:3.000
##  Max.   :4.000
```

1(b).

```
# convert class to factor
vehicle$class <- factor(vehicle$class, labels=c("2D", "4D", "Bus", "Van"))
# print summary
summary(vehicle)
```

```
##   Compactness      Circularity    Distance.Circularity  Radius.Ratio
##  Min.   : 73.00   Min.   :33.00   Min.   : 40.00       Min.   :104.0
##  1st Qu.: 87.00   1st Qu.:40.00   1st Qu.: 70.00       1st Qu.:141.0
##  Median : 93.00   Median :44.00   Median : 80.00       Median :167.0
##  Mean   : 93.68   Mean   :44.86   Mean   : 82.09       Mean   :168.9
##  3rd Qu.:100.00   3rd Qu.:49.00   3rd Qu.: 98.00       3rd Qu.:195.0
##  Max.   :119.00   Max.   :59.00   Max.   :112.00       Max.   :333.0
##  Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio   Elongatedness
##  Min.   : 47.00       Min.   : 2.000          Min.   :112.0   Min.   :26.00
##  1st Qu.: 57.00       1st Qu.: 7.000          1st Qu.:146.2   1st Qu.:33.00
##  Median : 61.00       Median : 8.000          Median :157.0   Median :43.00
##  Mean   : 61.69       Mean   : 8.567          Mean   :168.8   Mean   :40.93
##  3rd Qu.: 65.00       3rd Qu.:10.000          3rd Qu.:198.0   3rd Qu.:46.00
##  Max.   :138.00       Max.   :55.000          Max.   :265.0   Max.   :61.00
##  Pr.Axis.Rectangularity Max.Length.Rectangularity
##  Min.   :17.00          Min.   :118
##  1st Qu.:19.00          1st Qu.:137
##  Median :20.00          Median :146
##  Mean   :20.58          Mean   :148
##  3rd Qu.:23.00          3rd Qu.:159
##  Max.   :29.00          Max.   :188
##  Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis
##  Min.   :130.0                    Min.   : 184.0
##  1st Qu.:167.0                    1st Qu.: 318.2
##  Median :178.5                    Median : 364.0
##  Mean   :188.6                    Mean   : 439.9
##  3rd Qu.:217.0                    3rd Qu.: 587.0
##  Max.   :320.0                    Max.   :1018.0
##  Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis
##  Min.   :109.0             Min.   : 59.00            Min.   : 0.000
##  1st Qu.:149.0             1st Qu.: 67.00            1st Qu.: 2.000
##  Median :173.0             Median : 71.50            Median : 6.000
##  Mean   :174.7             Mean   : 72.46            Mean   : 6.377
##  3rd Qu.:198.0             3rd Qu.: 75.00            3rd Qu.: 9.000
##  Max.   :268.0             Max.   :135.00            Max.   :22.000
##  Kurtosis.About.Minor.Axis Kurtosis.About.Major.Axis Hollows.Ratio   class
##  Min.   : 0.0              Min.   :176.0             Min.   :181.0   2D :212
##  1st Qu.: 5.0              1st Qu.:184.0             1st Qu.:190.2   4D :217
##  Median :11.0              Median :188.0             Median :197.0   Bus:218
##  Mean   :12.6              Mean   :188.9             Mean   :195.6   Van:199
##  3rd Qu.:19.0              3rd Qu.:193.0             3rd Qu.:201.0
##  Max.   :41.0              Max.   :206.0             Max.   :211.0
```

1(c).

```
corMatrix <- cor(vehicle[, -19])
corMatrix
```

```
##                                Compactness Circularity Distance.Circularity
## Compactness                     1.00000000  0.69286923            0.7924440
## Circularity                     0.69286923  1.00000000            0.7984920
## Distance.Circularity            0.79244402  0.79849200            1.0000000
## Radius.Ratio                    0.69165856  0.62277837            0.7716439
## Pr.Axis.Aspect.Ratio            0.09322213  0.14969187            0.1615292
## Max.Length.Aspect.Ratio         0.14824919  0.24746673            0.2643086
## Scatter.Ratio                   0.81300326  0.86036714            0.9072801
## Elongatedness                  -0.78864702 -0.82875480           -0.9123072
## Pr.Axis.Rectangularity          0.81343702  0.85792532            0.8953261
## Max.Length.Rectangularity       0.67614317  0.96577578            0.7745238
## Scaled.Variance.Along.Major.Axis  0.76441546  0.80849631            0.8644323
## Scaled.Variance.Along.Minor.Axis  0.81863161  0.85267941            0.8896611
## Scaled.Radius.of.Gyration       0.58534709  0.93608041            0.7058181
## Skewness.About.Major.Axis      -0.25298400  0.05866929           -0.2316598
## Skewness.About.Minor.Axis       0.23369282  0.14843285            0.1164777
## Kurtosis.About.Minor.Axis       0.15677928 -0.01548177            0.2645222
## Kurtosis.About.Major.Axis       0.29778044 -0.11304723            0.1479800
## Hollows.Ratio                   0.36555185  0.03867702            0.3354525
##                                Radius.Ratio Pr.Axis.Aspect.Ratio
## Compactness                     0.69165856           0.09322213
## Circularity                     0.62277837           0.14969187
## Distance.Circularity            0.77164394           0.16152916
## Radius.Ratio                    1.00000000           0.66540656
## Pr.Axis.Aspect.Ratio            0.66540656           1.00000000
## Max.Length.Aspect.Ratio         0.44804838           0.64809643
## Scatter.Ratio                   0.73846048           0.10606364
## Elongatedness                  -0.79255741          -0.18508583
## Pr.Axis.Rectangularity          0.71149706           0.07998168
## Max.Length.Rectangularity       0.57015440           0.12916829
## Scaled.Variance.Along.Major.Axis  0.79758793           0.27470468
## Scaled.Variance.Along.Minor.Axis  0.72578142           0.09218201
## Scaled.Radius.of.Gyration       0.53906903           0.12440783
## Skewness.About.Major.Axis      -0.18241222           0.15230355
## Skewness.About.Minor.Axis       0.05053425          -0.05680698
## Kurtosis.About.Minor.Axis       0.17429937          -0.03417976
## Kurtosis.About.Major.Axis       0.38075719           0.23819525
## Hollows.Ratio                   0.46927820           0.26738123
##                                Max.Length.Aspect.Ratio Scatter.Ratio
## Compactness                                 0.14824919   0.813003257
## Circularity                                 0.24746673   0.860367138
## Distance.Circularity                        0.26430861   0.907280057
## Radius.Ratio                                0.44804838   0.738460476
## Pr.Axis.Aspect.Ratio                        0.64809643   0.106063643
## Max.Length.Aspect.Ratio                     1.00000000   0.166769266
## Scatter.Ratio                               0.16676927   1.000000000
## Elongatedness                              -0.18004430  -0.973385344
## Pr.Axis.Rectangularity                      0.16186089   0.992088347
## Max.Length.Rectangularity                   0.30594269   0.810647613
## Scaled.Variance.Along.Major.Axis            0.31942961   0.951862122
## Scaled.Variance.Along.Minor.Axis            0.14460641   0.996318013
## Scaled.Radius.of.Gyration                   0.18979054   0.800931523
## Skewness.About.Major.Axis                   0.29454772  -0.028618420
```

```
## Skewness.About.Minor.Axis                    0.01521789   0.074347173
## Kurtosis.About.Minor.Axis                    0.04337924   0.210706455
## Kurtosis.About.Major.Axis                   -0.02731663   0.004053166
## Hollows.Ratio                                0.14391873   0.119949827
##                                 Elongatedness Pr.Axis.Rectangularity
## Compactness                       -0.78864702             0.81343702
## Circularity                       -0.82875480             0.85792532
## Distance.Circularity              -0.91230719             0.89532606
## Radius.Ratio                      -0.79255741             0.71149706
## Pr.Axis.Aspect.Ratio              -0.18508583             0.07998168
## Max.Length.Aspect.Ratio           -0.18004430             0.16186089
## Scatter.Ratio                     -0.97338534             0.99208835
## Elongatedness                      1.00000000            -0.95051244
## Pr.Axis.Rectangularity            -0.95051244             1.00000000
## Max.Length.Rectangularity         -0.77553091             0.81330473
## Scaled.Variance.Along.Major.Axis  -0.93839190             0.93826639
## Scaled.Variance.Along.Minor.Axis  -0.95652176             0.99234619
## Scaled.Radius.of.Gyration         -0.76614632             0.79828199
## Skewness.About.Major.Axis          0.10487465            -0.01633331
## Skewness.About.Minor.Axis         -0.05334635             0.08234563
## Kurtosis.About.Minor.Axis         -0.18518873             0.21308767
## Kurtosis.About.Major.Axis         -0.11204551            -0.02205601
## Hollows.Ratio                     -0.21672508             0.09851910
##                                 Max.Length.Rectangularity
## Compactness                                   0.676143173
## Circularity                                   0.965775776
## Distance.Circularity                          0.774523840
## Radius.Ratio                                  0.570154405
## Pr.Axis.Aspect.Ratio                          0.129168289
## Max.Length.Aspect.Ratio                       0.305942689
## Scatter.Ratio                                 0.810647613
## Elongatedness                                -0.775530913
## Pr.Axis.Rectangularity                        0.813304735
## Max.Length.Rectangularity                     1.000000000
## Scaled.Variance.Along.Major.Axis              0.747151993
## Scaled.Variance.Along.Minor.Axis              0.798071495
## Scaled.Radius.of.Gyration                     0.866478540
## Skewness.About.Major.Axis                     0.040339043
## Skewness.About.Minor.Axis                     0.137473338
## Kurtosis.About.Minor.Axis                     0.001183035
## Kurtosis.About.Major.Axis                    -0.107718191
## Hollows.Ratio                                 0.076769617
##                                 Scaled.Variance.Along.Major.Axis
## Compactness                                          0.76441546
## Circularity                                          0.80849631
## Distance.Circularity                                 0.86443228
## Radius.Ratio                                         0.79758793
## Pr.Axis.Aspect.Ratio                                 0.27470468
## Max.Length.Aspect.Ratio                              0.31942961
## Scatter.Ratio                                        0.95186212
## Elongatedness                                       -0.93839190
## Pr.Axis.Rectangularity                               0.93826639
## Max.Length.Rectangularity                            0.74715199
## Scaled.Variance.Along.Major.Axis                     1.00000000
```

```
## Scaled.Variance.Along.Minor.Axis                      0.94988787
## Scaled.Radius.of.Gyration                             0.78096248
## Skewness.About.Major.Axis                             0.11129593
## Skewness.About.Minor.Axis                             0.03913734
## Kurtosis.About.Minor.Axis                             0.19194816
## Kurtosis.About.Major.Axis                             0.01118902
## Hollows.Ratio                                         0.08553952
##                                  Scaled.Variance.Along.Minor.Axis
## Compactness                                           0.818631612
## Circularity                                           0.852679411
## Distance.Circularity                                  0.889661065
## Radius.Ratio                                          0.725781417
## Pr.Axis.Aspect.Ratio                                  0.092182007
## Max.Length.Aspect.Ratio                               0.144606405
## Scatter.Ratio                                         0.996318013
## Elongatedness                                        -0.956521755
## Pr.Axis.Rectangularity                                0.992346188
## Max.Length.Rectangularity                             0.798071495
## Scaled.Variance.Along.Major.Axis                      0.949887874
## Scaled.Variance.Along.Minor.Axis                      1.000000000
## Scaled.Radius.of.Gyration                             0.797539612
## Skewness.About.Major.Axis                            -0.019872473
## Skewness.About.Minor.Axis                             0.076435424
## Kurtosis.About.Minor.Axis                             0.202893719
## Kurtosis.About.Major.Axis                             0.005379651
## Hollows.Ratio                                         0.105432191
##                                  Scaled.Radius.of.Gyration
## Compactness                                       0.58534709
## Circularity                                       0.93608041
## Distance.Circularity                              0.70581805
## Radius.Ratio                                      0.53906903
## Pr.Axis.Aspect.Ratio                              0.12440783
## Max.Length.Aspect.Ratio                           0.18979054
## Scatter.Ratio                                     0.80093152
## Elongatedness                                    -0.76614632
## Pr.Axis.Rectangularity                            0.79828199
## Max.Length.Rectangularity                         0.86647854
## Scaled.Variance.Along.Major.Axis                  0.78096248
## Scaled.Variance.Along.Minor.Axis                  0.79753961
## Scaled.Radius.of.Gyration                         1.00000000
## Skewness.About.Major.Axis                         0.19120943
## Skewness.About.Minor.Axis                         0.16800281
## Kurtosis.About.Minor.Axis                        -0.05621812
## Kurtosis.About.Major.Axis                        -0.22736551
## Hollows.Ratio                                    -0.11780842
##                                  Skewness.About.Major.Axis
## Compactness                                      -0.25298400
## Circularity                                       0.05866929
## Distance.Circularity                             -0.23165977
## Radius.Ratio                                     -0.18241222
## Pr.Axis.Aspect.Ratio                              0.15230355
## Max.Length.Aspect.Ratio                           0.29454772
## Scatter.Ratio                                    -0.02861842
## Elongatedness                                     0.10487465
```

```
## Pr.Axis.Rectangularity                        -0.01633331
## Max.Length.Rectangularity                      0.04033904
## Scaled.Variance.Along.Major.Axis               0.11129593
## Scaled.Variance.Along.Minor.Axis              -0.01987247
## Scaled.Radius.of.Gyration                      0.19120943
## Skewness.About.Major.Axis                      1.00000000
## Skewness.About.Minor.Axis                     -0.09253891
## Kurtosis.About.Minor.Axis                     -0.12615735
## Kurtosis.About.Major.Axis                     -0.75414191
## Hollows.Ratio                                 -0.80539854
##                                 Skewness.About.Minor.Axis
## Compactness                                    0.23369282
## Circularity                                    0.14843285
## Distance.Circularity                           0.11647771
## Radius.Ratio                                   0.05053425
## Pr.Axis.Aspect.Ratio                          -0.05680698
## Max.Length.Aspect.Ratio                        0.01521789
## Scatter.Ratio                                  0.07434717
## Elongatedness                                 -0.05334635
## Pr.Axis.Rectangularity                         0.08234563
## Max.Length.Rectangularity                      0.13747334
## Scaled.Variance.Along.Major.Axis               0.03913734
## Scaled.Variance.Along.Minor.Axis               0.07643542
## Scaled.Radius.of.Gyration                      0.16800281
## Skewness.About.Major.Axis                     -0.09253891
## Skewness.About.Minor.Axis                      1.00000000
## Kurtosis.About.Minor.Axis                     -0.03936531
## Kurtosis.About.Major.Axis                      0.11693185
## Hollows.Ratio                                  0.09914923
##                                 Kurtosis.About.Minor.Axis
## Compactness                                   0.156779278
## Circularity                                  -0.015481769
## Distance.Circularity                          0.264522153
## Radius.Ratio                                  0.174299370
## Pr.Axis.Aspect.Ratio                         -0.034179762
## Max.Length.Aspect.Ratio                       0.043379243
## Scatter.Ratio                                 0.210706455
## Elongatedness                                -0.185188734
## Pr.Axis.Rectangularity                        0.213087671
## Max.Length.Rectangularity                     0.001183035
## Scaled.Variance.Along.Major.Axis              0.191948164
## Scaled.Variance.Along.Minor.Axis              0.202893719
## Scaled.Radius.of.Gyration                    -0.056218120
## Skewness.About.Major.Axis                    -0.126157345
## Skewness.About.Minor.Axis                    -0.039365312
## Kurtosis.About.Minor.Axis                     1.000000000
## Kurtosis.About.Major.Axis                     0.079089910
## Hollows.Ratio                                 0.204923603
##                                 Kurtosis.About.Major.Axis Hollows.Ratio
## Compactness                                   0.297780441    0.36555185
## Circularity                                  -0.113047232    0.03867702
## Distance.Circularity                          0.147979965    0.33545253
## Radius.Ratio                                  0.380757189    0.46927820
## Pr.Axis.Aspect.Ratio                          0.238195253    0.26738123
```

```
## Max.Length.Aspect.Ratio                  -0.027316627    0.14391873
## Scatter.Ratio                             0.004053166    0.11994983
## Elongatedness                            -0.112045506   -0.21672508
## Pr.Axis.Rectangularity                   -0.022056010    0.09851910
## Max.Length.Rectangularity                -0.107718191    0.07676962
## Scaled.Variance.Along.Major.Axis          0.011189021    0.08553952
## Scaled.Variance.Along.Minor.Axis          0.005379651    0.10543219
## Scaled.Radius.of.Gyration                -0.227365512   -0.11780842
## Skewness.About.Major.Axis                -0.754141913   -0.80539854
## Skewness.About.Minor.Axis                 0.116931854    0.09914923
## Kurtosis.About.Minor.Axis                 0.079089910    0.20492360
## Kurtosis.About.Major.Axis                 1.000000000    0.89409781
## Hollows.Ratio                             0.894097812    1.00000000
```

Variables with correlations between ± 0.7 and ± 0.9:

- Compactness and Circularity is correlated with Distance.Circularity, Scatter.Ratio, Elongatedness, Pr.Axis.Rectangularity, Scaled.Variance.Along.Major.Axis, and Scaled.Variance.Along.Minor.Axis.
- Distance.Circularity is correlated with Radius.Ratio, Pr.Axis.Rectangularity, Max.Length.Rectangularity, Scaled.Variance.Along.Major.Axis, Scaled.Variance.Along.Minor.Axis, and Scaled.Radius.of.Gyration.
- Radius.Ratio is correlated with Scatter.Ratio, Elongatedness, Pr.Axis.Rectangularity, Scaled.Variance.Along.Major.Axis, and Scaled.Variance.Along.Minor.Axis.
- Scatter.Ratio, Elongatedness, and Pr.Axis.Rectangularity is correlated with Max.Length.Rectangularity and Scaled.Radius.of.Gyration.
- Max.Length.Rectangularity, Scaled.Variance.Along.Major.Axis, Scaled.Variance.Along.Minor.Axis, and Scaled.Radius.of.Gyration are correlated with each other.
- Skewness.About.Major.Axis, Kurtosis.About.Major.Axis and Hollows.Ratio are correlated with each other.

Variables with correlations greater than ± 0.9:

- Circularity is highly correlated with Max.Length.Rectangularity, and Scaled.Radius.of.Gyration.
- Distance.Circularity is highly correlated with Scatter.Ratio and Elongatedness.
- Scatter.Ratio, Elongatedness, Pr.Axis.Rectangularity, Scaled.Variance.Along.Major.Axis, and Scaled.Variance.Along.Minor.Axis are highly correlated with each other.

2.

```
# sample the data into train and test sets
set.seed(46685326, kind="Mersenne-Twister")
perm <- sample(x=nrow(vehicle))
# print set 1
set1 <- vehicle[which(perm <= 3*nrow(vehicle)/4), ]
head(set1, n = 6)
```

```
##   Compactness Circularity Distance.Circularity Radius.Ratio
## 1          95          48                  83          178
## 2          91          41                  84          141
## 3         104          50                 106          209
## 4          93          41                  82          159
## 5          85          44                  70          205
## 7          97          43                  73          173
##   Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio Elongatedness
## 1                   72                      10           162            42
## 2                   57                       9           149            45
## 3                   66                      10           207            32
## 4                   63                       9           144            46
## 5                  103                      52           149            45
## 7                   65                       6           153            42
##   Pr.Axis.Rectangularity Max.Length.Rectangularity
## 1                     20                      159
## 2                     19                      143
## 3                     23                      158
## 4                     19                      143
## 5                     19                      144
## 7                     19                      143
##   Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis
## 1                              176                              379
## 2                              170                              330
## 3                              223                              635
## 4                              160                              309
## 5                              241                              325
## 7                              176                              361
##   Scaled.Radius.of.Gyration Skewness.About.Major.Axis Skewness.About.Minor.Axis
## 1                       184                        70                         6
## 2                       158                        72                         9
## 3                       220                        73                        14
## 4                       127                        63                         6
## 5                       188                       127                         9
## 7                       172                        66                        13
##   Kurtosis.About.Minor.Axis Kurtosis.About.Major.Axis Hollows.Ratio class
## 1                        16                       187           197   Van
## 2                        14                       189           199   Van
## 3                         9                       188           196    4D
## 4                        10                       199           207   Van
## 5                        11                       180           183   Bus
## 7                         1                       200           204   Bus
```

```
# print set 2
set2 <- vehicle[which(perm > 3*nrow(vehicle)/4), ]
head(set2, n = 6)
```

```
##    Compactness Circularity Distance.Circularity Radius.Ratio
## 6          107          57                  106          172
## 11          86          36                   70          143
## 12          90          34                   66          136
## 14          89          42                   85          144
## 15          94          49                   79          203
## 16          96          55                  103          201
##    Pr.Axis.Aspect.Ratio Max.Length.Aspect.Ratio Scatter.Ratio Elongatedness
## 6                    50                       6           255            26
## 11                   61                       9           133            50
## 12                   55                       6           123            54
## 14                   58                      10           152            44
## 15                   71                       5           174            37
## 16                   65                       9           204            32
##    Pr.Axis.Rectangularity Max.Length.Rectangularity
## 6                      28                      169
## 11                     18                      130
## 12                     17                      118
## 14                     19                      144
## 15                     21                      154
## 16                     23                      166
##    Scaled.Variance.Along.Major.Axis Scaled.Variance.Along.Minor.Axis
## 6                               280                              957
## 11                              153                              266
## 12                              148                              224
## 14                              173                              345
## 15                              196                              465
## 16                              227                              624
##    Scaled.Radius.of.Gyration Skewness.About.Major.Axis
## 6                        264                        85
## 11                       127                        66
## 12                       118                        65
## 14                       161                        72
## 15                       206                        71
## 16                       246                        74
##    Skewness.About.Minor.Axis Kurtosis.About.Minor.Axis
## 6                          5                         9
## 11                         2                        10
## 12                         5                        26
## 14                         8                        13
## 15                         6                         2
## 16                         6                         2
##    Kurtosis.About.Major.Axis Hollows.Ratio class
## 6                        181           183   Bus
## 11                       194           202   Van
## 12                       196           202    4D
## 14                       187           197   Van
## 15                       197           199   Bus
## 16                       186           194    2D
```

3(a).

```
# scale the X variables using mean and sd from training set
scale.1 <- function(x1, x2) {
  for (col in 1:ncol(x1)) {
    a <- mean(x2[, col])
    b <- sd(x2[, col])
    x1[, col] <- (x1[, col] - a) / b
  }
  x1
}
x.set1.unscaled <- as.matrix(set1[, -19])
x.set2.unscaled <- as.matrix(set2[, -19])
x.set1 <- scale.1(x.set1.unscaled, x.set1.unscaled)
x.set2 <- scale.1(x.set2.unscaled, x.set1.unscaled)
# fit a knn using m/k= 1
model.knn <- knn(train = x.set1, test = x.set2, cl=set1[, 19], k=1)
# create confusion matrix
table(model.knn, set2[, 19], dnn = c("Predicted", "Observed"))
```

```
##          Observed
## Predicted 2D 4D Bus Van
##       2D  25 18   1   0
##       4D  26 29   1   2
##       Bus  0  1  46   6
##       Van  4  6   1  46
```

There is a high misclassification rate between classes “2D” and “4D” which indicates that these two classes are harder to separate as these two classes overlap each other. Class “Bus” has the lowest misclassifications which indicates that this class is easier to separate from other classes; however, there is a small overlap with class “Van”. We also see a small number misclassificatios in class “Van” which indicates that this class is also easier to separate from other classes; however, there is a small overlap with classes “2D” and “4D”.

3(b).

```
# calculate misclassification rate
(misclass.model.knn <-
    mean(ifelse(model.knn == set2[, 19], yes = 0, no = 1)))
```

```
## [1] 0.3113208
```

```
# calculate aprox. se of misclassification
(se <- sqrt(misclass.model.knn * (1 - misclass.model.knn) / nrow(set2)))
```

```
## [1] 0.03180128
```

2. Problem Set 14, Applications

```
library(nnet)
```

```
## Warning: package 'nnet' was built under R version 4.1.2
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.1.2
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.1.2
```

```
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.1.2
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.1.2
```

```
## Loaded glmnet 4.1-7
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.1.2
```

1(a).

```
# scale the values in training set between 0 and 1
rescale <- function(x1, x2) {
  for (col in 1:ncol(x1)) {
    a <- min(x2[, col])
    b <- max(x2[, col])
    x1[, col] <- (x1[, col] - a) / (b - a)
  }
  x1
}
set1.rescale <- data.frame(cbind(rescale(set1[, -19], set1[, -19]),
                                 class = set1$class))
set2.rescale <- data.frame(cbind(rescale(set2[, -19], set1[, -19]),
                                class = set2$class))
#summary(set1.rescale)
#summary(set2.rescale)
```

```
summary(set1.rescale)[, 1:3]
```

```
##   Compactness      Circularity     Distance.Circularity
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.3261   1st Qu.:0.2692   1st Qu.:0.4286
##  Median :0.4348   Median :0.4231   Median :0.5714
##  Mean   :0.4483   Mean   :0.4512   Mean   :0.5973
##  3rd Qu.:0.5652   3rd Qu.:0.6154   3rd Qu.:0.8000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
```

```
summary(set2.rescale)[, 1:3]
```

```
##   Compactness       Circularity     Distance.Circularity
##  Min.   :0.06522   Min.   :0.0000   Min.   :0.1429
##  1st Qu.:0.30435   1st Qu.:0.2692   1st Qu.:0.4286
##  Median :0.39130   Median :0.4615   Median :0.5643
##  Mean   :0.45324   Mean   :0.4713   Mean   :0.6133
##  3rd Qu.:0.60870   3rd Qu.:0.6923   3rd Qu.:0.8286
##  Max.   :0.93478   Max.   :0.9615   Max.   :1.0286
```

1(b)

i.

```
# fit logistic regression
mod.logit <- multinom(
  data = set1.rescale, formula = class ~ .,
  trace = TRUE
)
```

```
## # weights:  80 (57 variable)
## initial  value 878.910625
## iter  10 value 540.126072
## iter  20 value 264.002381
## iter  30 value 222.740935
## iter  40 value 211.568230
## iter  50 value 203.944730
## iter  60 value 200.943742
## iter  70 value 199.287545
## iter  80 value 197.947131
## iter  90 value 196.809842
## iter 100 value 196.152373
## final  value 196.152373
## stopped after 100 iterations
```

```
#summary(mod.logit)
# run anova on the model
Anova(mod.logit)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: class
##                                 LR Chisq Df Pr(>Chisq)
## Compactness                       55.741  3  4.770e-12 ***
## Circularity                       18.992  3  0.0002744 ***
## Distance.Circularity              19.143  3  0.0002554 ***
## Radius.Ratio                     131.680  3  < 2.2e-16 ***
## Pr.Axis.Aspect.Ratio             141.223  3  < 2.2e-16 ***
## Max.Length.Aspect.Ratio           15.429  3  0.0014845 **
## Scatter.Ratio                      2.759  3  0.4302978
## Elongatedness                      4.330  3  0.2279767
## Pr.Axis.Rectangularity             8.232  3  0.0414498 *
## Max.Length.Rectangularity         28.668  3  2.630e-06 ***
## Scaled.Variance.Along.Major.Axis  31.655  3  6.186e-07 ***
## Scaled.Variance.Along.Minor.Axis   1.058  3  0.7871357
## Scaled.Radius.of.Gyration         26.215  3  8.599e-06 ***
## Skewness.About.Major.Axis         28.790  3  2.479e-06 ***
## Skewness.About.Minor.Axis         13.571  3  0.0035514 **
## Kurtosis.About.Minor.Axis          8.339  3  0.0395014 *
## Kurtosis.About.Major.Axis         99.194  3  < 2.2e-16 ***
## Hollows.Ratio                     61.632  3  2.634e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Considering a significance level of $\alpha \leq 0.05$ as being significantly important, the predictors Compactness, Circularity, Distance.Circularity, Radius.Ratio, Pr.Axis.Aspect.Ratio, Max.Length.Aspect.Ratio, Pr.Axis.Rectangularity, Max.Length.Rectangularity, Scaled.Variance.Along.Major.Axis, Scaled.Radius.of.Gyration, Skewness.About.Major.Axis, Skewness.About.Minor.Axis, Kurtosis.About.Minor.Axis, Kurtosis.About.Major.Axis, and Hollows.Ratio appear to be important. On the other hand, Scatter.Ratio, Elongatedness, and Scaled.Variance.Along.Minor.Axis seem to be deemed unimportant.

ii.

```
# compute train and test error
pred.class.1 <- predict(mod.logit,
                        newdata = set1.rescale,
                        type = "class")
(mul.misclass.train <- mean(ifelse(pred.class.1 == set1$class,
                                   yes = 0, no = 1))
)
```

```
## [1] 0.159306
```

```
pred.class.2 <- predict(mod.logit,
                        newdata = set2.rescale,
                        type = "class")
(mul.misclass.test <- mean(ifelse(pred.class.2 == set2$class,
                                  yes = 0, no = 1))
)
```

```
## [1] 0.2122642
```

```
# compute se for test set
(se.logit <- sqrt(mul.misclass.test * (1 - mul.misclass.test) / nrow(set2)))
```

```
## [1] 0.02808411
```

The test error for logistic regression using 'multinom()' is lower than that of the optimal KNN, suggesting that logistic regression performs better in making predictions for this case.

iii.

```
# create confusion matrix
table(set2$class, pred.class.2, dnn = c("Obs", "Pred"))
```

```
##      Pred
## Obs   2D 4D Bus Van
##   2D  34 19   1   1
##   4D  18 33   1   2
##   Bus  1  1  47   0
##   Van  1  0   0  53
```

The confusion matrix shows that there is a high misclassification rate in classes '2D' and '4D' and it seems hard to seperate these two classes from each other as we see that 19 cases (out of total 55) of class '2D' are misclassified as class '4D' and 18 cases (out of total 54) of class '4D' are misclassified as class '2D'. On the other hand, classes 'Bus' and 'Van' have a very low miscalssification rate where we see only two misclassifed cases (out of total 49) for class 'Bus' and only one miscalssified case (out of total 54) for "Van'.

2(a).

```
# fit lasso
model.logit.lasso <- cv.glmnet(
  x = as.matrix(set1.rescale[, 1:18]),
  y = set1.rescale[, 19], family = "multinomial"
)
#model.logit.lasso
#plot(model.logit.lasso)
# find nonzero lasso coefficients
c <- coef(model.logit.lasso, s = model.logit.lasso$lambda.min)
cmat <- cbind(
  as.matrix(c[[1]]), as.matrix(c[[2]]),
  as.matrix(c[[3]]), as.matrix(c[[4]])
)
cmat != 0
```

```
##                                    1     1     1     1
## (Intercept)                       TRUE  TRUE  TRUE  TRUE
## Compactness                       TRUE  TRUE  TRUE  TRUE
## Circularity                       TRUE  TRUE FALSE FALSE
## Distance.Circularity              TRUE  TRUE  TRUE  TRUE
## Radius.Ratio                      TRUE  TRUE  TRUE  TRUE
## Pr.Axis.Aspect.Ratio              TRUE  TRUE  TRUE  TRUE
## Max.Length.Aspect.Ratio           TRUE  TRUE  TRUE  TRUE
## Scatter.Ratio                    FALSE FALSE FALSE  TRUE
## Elongatedness                     TRUE  TRUE  TRUE  TRUE
## Pr.Axis.Rectangularity           FALSE  TRUE FALSE  TRUE
## Max.Length.Rectangularity         TRUE  TRUE  TRUE  TRUE
## Scaled.Variance.Along.Major.Axis  TRUE FALSE  TRUE FALSE
## Scaled.Variance.Along.Minor.Axis  TRUE  TRUE FALSE FALSE
## Scaled.Radius.of.Gyration         TRUE  TRUE  TRUE  TRUE
## Skewness.About.Major.Axis         TRUE  TRUE  TRUE  TRUE
## Skewness.About.Minor.Axis         TRUE  TRUE  TRUE  TRUE
## Kurtosis.About.Minor.Axis         TRUE  TRUE  TRUE  TRUE
## Kurtosis.About.Major.Axis         TRUE  TRUE  TRUE  TRUE
## Hollows.Ratio                     TRUE  TRUE  TRUE  TRUE
```

The pattern is somewhat similar to ANOVA, because the variables that consistently appear in all four logistic regression models with LASSO are the same variables identified as important in ANOVA; and the only exception is the variable 'Elongatedness,' which, while deemed unimportant in ANOVA, appears to influence all four logistic regression models in the LASSO.

2(b).

```
# compute train and test error
lasso.pred.train <- predict(
  object = model.logit.lasso, type = "class",
  s = model.logit.lasso$lambda.min,
  newx = as.matrix(set1.rescale[, 1:18])
)
(lassomisclass.train <-
    mean(ifelse(lasso.pred.train == set1$class, yes = 0, no = 1)))
```

```
## [1] 0.1671924
```

```
lasso.pred.test <- predict(model.logit.lasso,
                           type = "class",
                           s = model.logit.lasso$lambda.min,
                           newx = as.matrix(set2.rescale[, 1:18])
)
(lassomisclass.test <-
    mean(ifelse(lasso.pred.test == set2$class, yes = 0, no = 1)))
```

```
## [1] 0.2169811
```

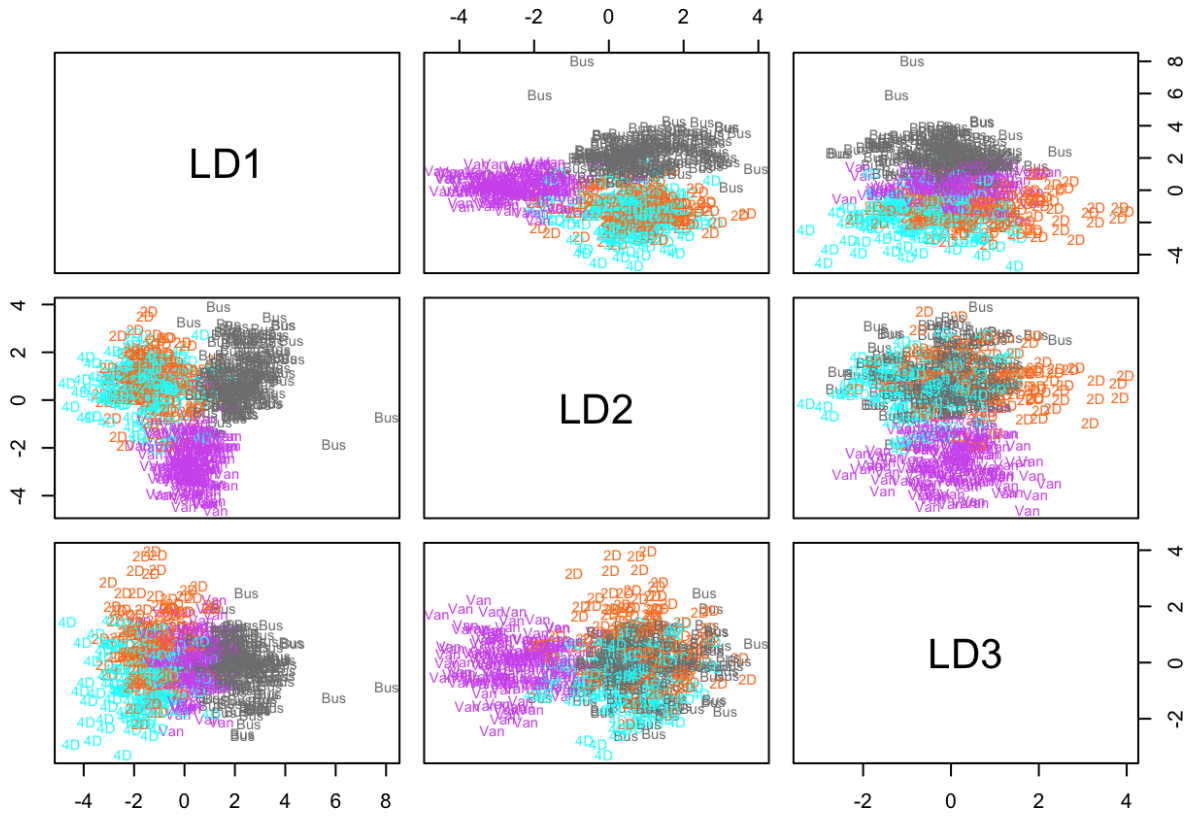
```
# compute standard error for test set
(se.lasso <- sqrt(lassomisclass.test * (1 - lassomisclass.test) / nrow(set2)))
```

```
## [1] 0.0283093
```

The test error for the LASSO version of logistic regression closely mirrors the test error observed in logistic regression using 'multinom()'. Therefore, the LASSO version of logistic regression outperforms KNN in terms of predictive accuracy, although its performance is comparable to logistic regression using 'multinom()'.

3(a).

```
# fit lda
model.lda <- lda(x = set1[, -19], grouping = set1$class)
# plot results
class.col <- ifelse(set1$class=="2D", y = 53, n =
                      ifelse(set1$class=="4D", y = 68, n =
                               ifelse(set1$class=="Bus", y = 203, n = 464)))
plot(model.lda, col = colors()[class.col])
```



LD1 seems to mostly separate classes 'Bus' and 'Van'.

LD2 seems to mostly seperate classes 'Bus' and '4D'.

LD3 seems to mostly separate classes 'Van' and '2D'.

3(b).

```
# compute train and test error
lda.pred.train <- predict(model.lda, newdata = set1[, -19])$class
(ldamisclass.train <- mean(ifelse(lda.pred.train == set1$class, yes = 0, no = 1)))
```

```
## [1] 0.1971609
```

```
lda.pred.test <- predict(model.lda, newdata = set2[, -19])$class
(ldamisclass.test <- mean(ifelse(lda.pred.test == set2$class, yes = 0, no = 1)))
```

```
## [1] 0.1981132
```

```
# compute se for test set
(se.lda <- sqrt(ldamisclass.test * (1 - ldamisclass.test) / nrow(set2)))
```

```
## [1] 0.02737444
```

The test error for Linear Discriminant Analysis (LDA) is the lowest among all the models, suggesting that LDA outperforms KNN, logistic regression using ‘multinom()’, and the LASSO version of logistic regression in terms of predictive accuracy.