HW5_Prob1

Haoyu_Zhao

May 7, 2017

Problem 1

a All the R code will shown in the report.

b

```
#Load the data
#Skip the step that convert the
#xlsx file into the csv file.
flavor <- read.csv("flavor.csv", sep=";")
#Delete the column that contains the variety
flavor <- flavor[,2:9]
#Compute the Covariance
Cov <- cov(flavor)
#do the PC analysis
flavor.pc <- princomp(flavor)
summary(flavor.pc)
```

```
## Importance of components:
##                            Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     8.3357813 4.6912363 3.1474171 1.55063523 1.29621790
## Proportion of Variance 0.6454225 0.2044213 0.0920153 0.02233422 0.01560657
## Cumulative Proportion  0.6454225 0.8498438 0.9418591 0.96419328 0.97979985
##                            Comp.6      Comp.7      Comp.8
## Standard deviation     1.1164306 0.717358144 0.643194062
## Proportion of Variance 0.0115775 0.004779952 0.003842691
## Cumulative Proportion  0.9913774 0.996157309 1.000000000
```

```
lds <- flavor.pc$loadings
sdevs <- flavor.pc$sdev
#the loading for the factor analysis
l1 <- lds[,1] * sdevs[1]
l2 <- lds[,2] * sdevs[2]
l3 <- lds[,3] * sdevs[3]

#For m=1, the loadings are
ld <- matrix(l1, 8, 1)
print(ld)
```

```
##             [,1]
## [1,] -5.04730017
## [2,] -4.32490311
## [3,] -3.38622039
## [4,] -0.66521459
## [5,] -0.90355038
## [6,] -0.04860147
## [7,] -1.13673758
## [8,] -3.35935823
```

```r
#The communality
print(ld * ld)
```

```
##              [,1]
## [1,] 25.475239039
## [2,] 18.704786937
## [3,] 11.466488504
## [4,]  0.442510448
## [5,]  0.816403290
## [6,]  0.002362103
## [7,]  1.292172315
## [8,] 11.285287703
```

```r
#The matrix Cov - LL'
Cov <- matrix(Cov, 8, 8)
temp <- Cov - ld %*% t(ld)
#The specific variance
sv <- temp * diag(8)
sv
```

```
##          [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]
## [1,] 3.644457 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
## [2,] 0.000000 8.441239 0.000000 0.000000 0.000000 0.000000 0.000000
## [3,] 0.000000 0.000000 5.300912 0.000000 0.000000 0.000000 0.000000
## [4,] 0.000000 0.000000 0.000000 8.550805 0.000000 0.000000 0.000000
## [5,] 0.000000 0.000000 0.000000 0.000000 2.130377 0.000000 0.000000
## [6,] 0.000000 0.000000 0.000000 0.000000 0.000000 2.865403 0.000000
## [7,] 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 1.791081
## [8,] 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
##         [,8]
## [1,] 0.00000
## [2,] 0.00000
## [3,] 0.00000
## [4,] 0.00000
## [5,] 0.00000
## [6,] 0.00000
## [7,] 0.00000
## [8,] 6.67242
```

```r
#the portion of variance
sdevs[1] / sum(sdevs)
```

```
##    Comp.1
## 0.3877419
```

```r
#For m=2, the loadings are
ld <- cbind(l1,l2)
print(ld)
```

```
##                                l1         l2
## OVERALL.LIKING          -5.04730017  1.3074429
## TEXTURE                 -4.32490311  2.0579311
## Sweetness               -3.38622039 -1.1337442
## Sourness                -0.66521459 -2.2846747
## Salty                   -0.90355038 -1.2070271
## Bitter                  -0.04860147 -1.2959698
```

```
## Umami                    -1.13673758 -0.8280021
## Overall.Flavor.Intensity -3.35935823 -2.3950067
```

```
#The communality
print(rowSums(ld *ld))
```

```
##           OVERALL.LIKING                  TEXTURE                Sweetness
##                27.184646                22.939867                12.751864
##                 Sourness                    Salty                   Bitter
##                 5.662249                 2.273318                 1.681900
##                    Umami Overall.Flavor.Intensity
##                 1.977760                17.021345
```

```
#The matrix Cov - LL'
Cov <- matrix(Cov, 8, 8)
temp <- Cov - ld %*% t(ld)
#The specific variance
sv <- temp * diag(8)
sv
```

```
##                          OVERALL.LIKING  TEXTURE Sweetness Sourness
## OVERALL.LIKING                 1.935051 0.000000  0.000000 0.000000
## TEXTURE                        0.000000 4.206158  0.000000 0.000000
## Sweetness                      0.000000 0.000000  4.015537 0.000000
## Sourness                       0.000000 0.000000  0.000000 3.331067
## Salty                          0.000000 0.000000  0.000000 0.000000
## Bitter                         0.000000 0.000000  0.000000 0.000000
## Umami                          0.000000 0.000000  0.000000 0.000000
## Overall.Flavor.Intensity       0.000000 0.000000  0.000000 0.000000
##                              Salty   Bitter    Umami
## OVERALL.LIKING           0.0000000 0.000000 0.000000
## TEXTURE                  0.0000000 0.000000 0.000000
## Sweetness                0.0000000 0.000000 0.000000
## Sourness                 0.0000000 0.000000 0.000000
## Salty                    0.6734624 0.000000 0.000000
## Bitter                   0.0000000 1.185866 0.000000
## Umami                    0.0000000 0.000000 1.105493
## Overall.Flavor.Intensity 0.0000000 0.000000 0.000000
##                          Overall.Flavor.Intensity
## OVERALL.LIKING                          0.0000000
## TEXTURE                                 0.0000000
## Sweetness                               0.0000000
## Sourness                                0.0000000
## Salty                                   0.0000000
## Bitter                                  0.0000000
## Umami                                   0.0000000
## Overall.Flavor.Intensity                0.9363624
```

```
#the portion of variance
sdevs[1] / sum(sdevs)
```

```
##    Comp.1
## 0.3877419
```

```
sdevs[2] / sum(sdevs)
```

```
##    Comp.2
```

```
## 0.2182146
```

```
#For m=3
ld <- cbind(l1,l2,l3)
t(ld) %*% ld
```

```
##              l1            l2            l3
## l1 6.948525e+01  0.000000e+00  3.747003e-16
## l2 0.000000e+00  2.200770e+01 -1.050549e-14
## l3 3.747003e-16 -1.050549e-14  9.906234e+00
```

```
print(ld)
```

```
##                                  l1         l2          l3
## OVERALL.LIKING           -5.04730017  1.3074429  0.73606614
## TEXTURE                  -4.32490311  2.0579311 -1.86881816
## Sweetness                -3.38622039 -1.1337442  1.69603162
## Sourness                 -0.66521459 -2.2846747 -1.57587743
## Salty                    -0.90355038 -1.2070271 -0.39811211
## Bitter                   -0.04860147 -1.2959698 -0.59291472
## Umami                    -1.13673758 -0.8280021 -0.03365409
## Overall.Flavor.Intensity -3.35935823 -2.3950067  0.02954650
```

```
#The communality
print(rowSums(ld *ld))
```

```
##           OVERALL.LIKING                  TEXTURE                Sweetness
##                27.726439                26.432349                15.628388
##                 Sourness                    Salty                   Bitter
##                 8.145639                 2.431811                 2.033448
##                    Umami Overall.Flavor.Intensity
##                 1.978892                17.022218
```

```
#The matrix Cov - LL'
Cov <- matrix(Cov, 8, 8)
temp <- Cov - ld %*% t(ld)
#The specific variance
sv <- temp * diag(8)
sv
```

```
##                           OVERALL.LIKING   TEXTURE Sweetness  Sourness
## OVERALL.LIKING                 1.393257 0.0000000  0.000000 0.0000000
## TEXTURE                        0.000000 0.7136771  0.000000 0.0000000
## Sweetness                      0.000000 0.0000000  1.139013 0.0000000
## Sourness                       0.000000 0.0000000  0.000000 0.8476773
## Salty                          0.000000 0.0000000  0.000000 0.0000000
## Bitter                         0.000000 0.0000000  0.000000 0.0000000
## Umami                          0.000000 0.0000000  0.000000 0.0000000
## Overall.Flavor.Intensity       0.000000 0.0000000  0.000000 0.0000000
##                                  Salty    Bitter   Umami
## OVERALL.LIKING           0.0000000 0.0000000 0.00000
## TEXTURE                  0.0000000 0.0000000 0.00000
## Sweetness                0.0000000 0.0000000 0.00000
## Sourness                 0.0000000 0.0000000 0.00000
## Salty                    0.5149692 0.0000000 0.00000
## Bitter                   0.0000000 0.8343177 0.00000
## Umami                    0.0000000 0.0000000 1.10436
```

```
## Overall.Flavor.Intensity 0.0000000 0.0000000 0.00000
##                          Overall.Flavor.Intensity
## OVERALL.LIKING                          0.0000000
## TEXTURE                                 0.0000000
## Sweetness                               0.0000000
## Sourness                                0.0000000
## Salty                                   0.0000000
## Bitter                                  0.0000000
## Umami                                   0.0000000
## Overall.Flavor.Intensity                0.9354894
```

```
#the portion of variance
sdevs[1] / sum(sdevs)
```

```
##    Comp.1
## 0.3877419
```

```
sdevs[2] / sum(sdevs)
```

```
##    Comp.2
## 0.2182146
```

```
sdevs[3] / sum(sdevs)
```

```
##    Comp.3
## 0.1464033
```

```
#Then we do the factor analysis using the MLE.
#We provide the result of m = 1,2,3.
flavor.fa1 <- factanal(flavor, factors = 1)
flavor.fa2 <- factanal(flavor, factors = 2)
flavor.fa3 <- factanal(flavor, factors = 3)
#print the result
flavor.fa1
```

```
##
## Call:
## factanal(x = flavor, factors = 1)
##
## Uniquenesses:
##           OVERALL.LIKING                  TEXTURE                Sweetness
##                    0.635                    0.804                    0.352
##                 Sourness                    Salty                   Bitter
##                    0.668                    0.352                    0.814
##                    Umami Overall.Flavor.Intensity
##                    0.397                    0.005
##
## Loadings:
##                          Factor1
## OVERALL.LIKING           0.604
## TEXTURE                  0.443
## Sweetness                0.805
## Sourness                 0.576
## Salty                    0.805
## Bitter                   0.431
## Umami                    0.776
## Overall.Flavor.Intensity 0.997
##
```

```
##                Factor1
## SS loadings      3.972
## Proportion Var   0.497
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 317.85 on 20 degrees of freedom.
## The p-value is 1.8e-55
```

```
flavor.fa2
```

```
##
## Call:
## factanal(x = flavor, factors = 2)
##
## Uniquenesses:
##           OVERALL.LIKING                  TEXTURE                Sweetness
##                    0.005                    0.346                    0.331
##                 Sourness                    Salty                   Bitter
##                    0.286                    0.129                    0.273
##                    Umami Overall.Flavor.Intensity
##                    0.316                    0.141
##
## Loadings:
##                         Factor1 Factor2
## OVERALL.LIKING                    0.997
## TEXTURE                           0.807
## Sweetness                0.323    0.752
## Sourness                 0.845
## Salty                    0.867    0.346
## Bitter                   0.835   -0.170
## Umami                    0.624    0.543
## Overall.Flavor.Intensity 0.681    0.629
##
##                Factor1 Factor2
## SS loadings      3.125   3.048
## Proportion Var   0.391   0.381
## Cumulative Var   0.391   0.772
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 86.93 on 13 degrees of freedom.
## The p-value is 5.38e-13
```

```
flavor.fa3
```

```
##
## Call:
## factanal(x = flavor, factors = 3)
##
## Uniquenesses:
##           OVERALL.LIKING                  TEXTURE                Sweetness
##                    0.005                    0.281                    0.022
##                 Sourness                    Salty                   Bitter
##                    0.232                    0.131                    0.265
##                    Umami Overall.Flavor.Intensity
##                    0.327                    0.093
##
```

```
## Loadings:
##                          Factor1 Factor2 Factor3
## OVERALL.LIKING                     0.880   0.468
## TEXTURE                            0.831   0.167
## Sweetness                 0.154    0.364   0.906
## Sourness                  0.873
## Salty                     0.839    0.219   0.341
## Bitter                    0.834   -0.198
## Umami                     0.583    0.397   0.418
## Overall.Flavor.Intensity  0.602    0.379   0.633
##
##                 Factor1 Factor2 Factor3
## SS loadings       2.891   1.987   1.766
## Proportion Var    0.361   0.248   0.221
## Cumulative Var    0.361   0.610   0.830
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 34.24 on 7 degrees of freedom.
## The p-value is 1.56e-05
```

c The factorization of the PCA and MLE approach are much differnet. First, the PCA uses the covariance matrix, but in the 'factanal' in R laguage, it uses the correlation matrix to approximate the model.

When we look at the portion of variance explained by each factor, we can see that the PCA approach is much better then the MLE approach. There are 2 reasons, the first is that PCA aims to find the direction or loadings that can bestly explain the variance so it is absolutely the best if we just look at the portion of variance. The second reason is that in the 'factanal' function if R language, it uses the corralation matrix to compute, but in the PCA, we use the covariance matrix to compute. This will cause some bias on the certain factor and thus increase the portion of variance that the first few factors can analysis.

d In the PCA approach, the loadings of the first factor is a combination of all the features, and 1 of the features in the takes small weights. The loadings of the second factor contrast the first 2 features with the other features, the savor. The third factor contrast the first and the third features with other features.