2018R1 High-Dimensional Data Analysis (STAT5103) Assignment 1

Yiu Chung WONG 1155017920

1. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}$$

```r
A <- matrix(c(1,2,3,2,1,2,3,2,1), 3, 3)
A
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    2    1    2
## [3,]    3    2    1
```

(a) Find the eigenvalues (λ_1, λ_2, λ_3) and their corresponding normalized eigenvectors (x_1, x_2, x_3)

```r
eigen(A)
```

```
## eigen() decomposition
## $values
## [1]  5.7015621 -0.7015621 -2.0000000
##
## $vectors
##            [,1]       [,2]          [,3]
## [1,] -0.6059128  0.3645129  7.071068e-01
## [2,] -0.5154991 -0.8568901 -5.551115e-16
## [3,] -0.6059128  0.3645129 -7.071068e-01
```

(b) Let P = (x_1, x_2, x_3), a matrix formed by using the eigenvectors as columns. Find PDP' where D is a diagonal matrix with the eigenvalues (λ_1, λ_2, λ_3) on its diagonal.

```r
P <- eigen(A)$vectors
D <- solve(P) %*% A %*% P
P %*% D %*% solve(P)
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    2    1    2
## [3,]    3    2    1
```

(c) Find the singular value decomposition of A.

```r
svd(A)
```

```
## $d
## [1] 5.7015621 2.0000000 0.7015621
##
```

```
## $u
##            [,1]          [,2]       [,3]
## [1,] -0.6059128  7.071068e-01  0.3645129
## [2,] -0.5154991  3.330669e-16 -0.8568901
## [3,] -0.6059128 -7.071068e-01  0.3645129
##
## $v
##            [,1]          [,2]       [,3]
## [1,] -0.6059128 -7.071068e-01 -0.3645129
## [2,] -0.5154991  3.330669e-16  0.8568901
## [3,] -0.6059128  7.071068e-01 -0.3645129
```

2. Let the eigenvalue-eigenvector pairs of a 2x2 symmetric matrix A be

$$\left(3, \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}\right) \, and \left(5, \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\right)$$

(a) Find A.

```r
x = 1/sqrt(2)
P = matrix(c(x, -x, x, x), 2, 2)
D = diag(c(3,5))
A = P %*% D %*% solve(P)
A
```

```
##      [,1] [,2]
## [1,]    4    1
## [2,]    1    4
```

(b) Find the determinant of A.

```r
det(A)
```

```
## [1] 15
```

(c) Find the trace of A.

```r
sum(diag(A))
```

```
## [1] 8
```

3. Using x_3 - x_9 of the US Crime Data, compute

(a) the sample mean ($\bar{x}$),

```r
crime <- readxl::read_excel("uscrime.xlsx", sheet = 1, skip = 1)
sapply(crime[, 3:9], mean)
```

```
##     Murder       Rape    Robbery    Assault   Burglary    Larcery Autothieft
##      6.858     15.616    101.510    135.420    930.800   1943.640    367.860
```

(b) the sample covariance matrix (S), and

$cov_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N-1}$

```
cov(crime[, 3:9]) %>% round(., 2)
```

```
##            Murder    Rape  Robbery  Assault  Burglary   Larcery Autothieft
## Murder      14.81   14.70   119.68   213.15    384.46    176.95      84.36
## Rape        14.70   54.00   369.52   348.61   1804.50   3132.74     646.41
## Robbery    119.68  369.52  8316.23  3501.22  20485.88  28234.76   11232.25
## Assault    213.15  348.61  3501.22  4647.11  12816.31  15324.75    4495.59
## Burglary   384.46 1804.50 20485.88 12816.31 130356.94 205309.15   50455.50
## Larcery    176.95 3132.74 28234.76 15324.75 205309.15 503857.62   78605.91
## Autothieft  84.36  646.41 11232.25  4495.59  50455.50  78605.91   39843.96
```

(c) the sample correlation matrix (R).

$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$

```
cor(crime[, 3:9]) %>% round(., 2)
```

```
##            Murder Rape Robbery Assault Burglary Larcery Autothieft
## Murder       1.00 0.52    0.34    0.81     0.28    0.06       0.11
## Rape         0.52 1.00    0.55    0.70     0.68    0.60       0.44
## Robbery      0.34 0.55    1.00    0.56     0.62    0.44       0.62
## Assault      0.81 0.70    0.56    1.00     0.52    0.32       0.33
## Burglary     0.28 0.68    0.62    0.52     1.00    0.80       0.70
## Larcery      0.06 0.60    0.44    0.32     0.80    1.00       0.55
## Autothieft   0.11 0.44    0.62    0.33     0.70    0.55       1.00
```