MAST90138 Week 2 Lab

Goals: Get familiar with `R`: matrix operations, basic descriptive statistics. Please first install `R` and `RStudio` if you haven't already done so.

Task 1: get familiar with matrix operations

Using the function `matrix`, Create the matrices

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 0 & 2 & 2 \end{pmatrix}$$

and

$$B = \begin{pmatrix} -1 & 5 \\ 1 & 1 \\ 3 & 2 \end{pmatrix}$$

1. Compute A^T, $diag(A)$, A^TA and A^TB.

```
> A=matrix(c(1,1,2,1,0,1,0,2,2),nrow=3,byrow=T)
> B=matrix(c(-1,5,1,1,3,2),nrow=3,byrow=T)
> At=t(A)
> diagA=diag(A)
> AtA=At %*% A
> AtB=At %*% B
```

2. Compute the matrix D whose elements are the elements of A to the power 3. Compare this matrix with the matrix $E = A^3$

```
> D=A^3
> D=A*A*A
> E=A%*%A%*%A
```

3. Compute the dimension, the trace, the determinant, the eigenvalues and the eigenvectors of A. Create three vectors, v1, v2 and v3 which are respectively the first, second and third eigenvectors of A. Create three scalars lambda1, lambda2 and lambda3 which are respectively the first, second and third eigenvalues of A. Also compute the rank of A. For the rank, you need to use the package `Matrix`

```
> dim(A)
> trA=sum(diagA)
> det(A)
> EE=eigen(A)
> Evect=EE$vectors
> v1=Evect[,1]
> v2=Evect[,2]
> v3=Evect[,3]
> Eval=EE$values
> lambda1=Eval[1]
```

```
> lambda2=Eval[2]
> lambda3=Eval[3]
> library(Matrix)
> rkA=rankMatrix(A)
```

Task 2: elementary descriptive statistics

The file `google_review_ratings.csv` taken from `https://archive.ics.uci.edu/ml/datasets/Tarvel+Review+Ratings#`, contains data populated by capturing user ratings from Google reviews (Dennis: I have corrected one seemingly erroneous entry in the dataset). Reviews on attractions from 24 categories across Europe are considered. Google user rating ranges from 1 to 5 and average user rating per category is calculated. This data set contains the reviews over the $p = 24$ categories of $n = 5456$ individuals.

1. Open the file to see its structure and check how the following information is formatted in the file:
   Attribute 1 : Unique user id Attribute 2 : Average ratings on churches Attribute 3 : Average ratings on resorts Attribute 4 : Average ratings on beaches Attribute 5 : Average ratings on parks Attribute 6 : Average ratings on theatres Attribute 7 : Average ratings on museums Attribute 8 : Average ratings on malls Attribute 9 : Average ratings on zoo Attribute 10 : Average ratings on restaurants Attribute 11 : Average ratings on pubs/bars Attribute 12 : Average ratings on local services Attribute 13 : Average ratings on burger/pizza shops Attribute 14 : Average ratings on hotels/other lodgings Attribute 15 : Average ratings on juice bars Attribute 16 : Average ratings on art galleries Attribute 17 : Average ratings on dance clubs Attribute 18 : Average ratings on swimming pools Attribute 19 : Average ratings on gyms Attribute 20 : Average ratings on bakeries Attribute 21 : Average ratings on beauty & spas Attribute 22 : Average ratings on cafes Attribute 23 : Average ratings on view points Attribute 24 : Average ratings on monuments Attribute 25 : Average ratings on gardens

2. Set the R working directory to the directory of the subject you create for yourself.

   ```
   > setwd("~/Dropbox/MAST 90138 S2 2022/Labs/Week 2")
   ```

   (This should just be your own directory, whatever it is)

3. Use R commands to create, using instructions that read the above file, a 5456×24 data matrix `X` whose ith row, for $i = 1, \ldots, 5456$, is the vector of 24 reviews for the ith individual.

   (Use the "File –> Import data set –> from text readr" in RStudio, but it may confuse students since the resulting object is a `tibble`.)

   ```
   > X=as.matrix(google_review_ratings[,2:25])
   ```

   or

   ```
   > Data= read.table(file="google_review_ratings.csv",sep=",", header=T)
   > X=as.matrix(Data[,2:25])
   ```

   or

```
> Data = read.csv(file="google_review_ratings.csv")
> X=as.matrix(Data[,2:25])
```

(Use the `head()` and `class()` function to view the imported dataset.)

4. Compute the mean vector, the covariance matrix and the correlation matrix of the rating data.

```
> barX=colMeans(X)
> covX=cov(X)
> corX=cor(X)
```

(The 12th and 24th numbers may be "NA" due to missing values, but that's fine, just tell the students about that)

5. Draw pairwise scatterplots for the first 10 categories.

```
> pairs(X[,1:10], pch = ".")
```

We see that it is time consuming to create scatterplots even of a subgroup of the data. We need better tools to visualize multivariate data.