Carthage University
INSAT
Department of Mathematics & Computer Sciences

# Exercises Sheet: Principle Component Analysis (PCA)

**Exercise 1.** *Determine wether the following statements are true or false, justify your answer.*

1. *A positive correlation between two given statistical variables indicates that when one of the variables increases the other one increases too.*
2. *Changing the scale of measurement (e.g. from meters to kilometers) will change the covariance value.*
3. *Total variance or inertia of a given data matrix equals the number of records.*
4. *If $R_X$ denotes the correlation matrix associated to a given data matrix $X$, then $R_X$ is symmetric and non-negative definite.*
5. *The principle components are linear combination of the initial variables that form the data matrix $X$.*
6. *The principle components are mutually non-correlated.*

**Exercise 2.** *Consider the following data matrix:*

$$X = \begin{bmatrix} 1 & 1 & 3 \\ -1 & -1 & 3 \\ 1 & 1 & -3 \\ -1 & -1 & -3 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \end{bmatrix}$$

1. *Determine the standardized data matrix $Z$.*
2. *Deduce the correlation matrix $R_X$.*
3. *Determine the spectrum of $R_X$.*
4. *Deduce the principle components matrix $C_X$.*
5. *Decide how many principle components we should retain. Justify your decision.*
6. *Say whether we were able to predict the result of PCA earlier.*

**Exercise 3.** *Find below the record of the results of 6 cognitive tests conducted on 15 children of 10 years old. Each test has a score out of 5. Tests are as follows: CUB (cubes), PUZ (puzzles), CAL (mental calculus ), MEM (memory of digits), COM (Comprehension) and VOC (Vocabulary).*

| | *CUB* | *PUZ* | *CAL* | *MEM* | *COM* | *VOC* |
|---|---|---|---|---|---|---|
| *I1* | *5* | *5* | *4* | *0* | *1* | *1* |
| *I2* | *4* | *3* | *3* | *2* | *2* | *1* |
| *I3* | *2* | *1* | *2* | *3* | *2* | *2* |
| *I4* | *5* | *3* | *5* | *3* | *4* | *3* |
| *I5* | *4* | *4* | *3* | *2* | *3* | *2* |
| *I6* | *2* | *0* | *1* | *3* | *1* | *1* |
| *I7* | *3* | *3* | *4* | *2* | *4* | *4* |
| *I8* | *1* | *2* | *1* | *4* | *3* | *3* |
| *I9* | *0* | *1* | *0* | *3* | *1* | *0* |
| *I10* | *2* | *0* | *1* | *3* | *1* | *0* |
| *I11* | *1* | *2* | *1* | *1* | *0* | *1* |
| *I12* | *4* | *2* | *4* | *2* | *1* | *2* |
| *I13* | *3* | *2* | *3* | *3* | *2* | *3* |
| *I14* | *1* | *0* | *0* | *3* | *2* | *2* |
| *I15* | *2* | *1* | *1* | *2* | *3* | *2* |

*Use the statistical software "R" to answer the same questions of Exercise 2.*