

Fuzzy Logic and Principal Components Analysis

Shagul Faraj Karim

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Applied Mathematics and Computer Science

Eastern Mediterranean University
February 2016
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Cem Tanova
Acting Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Applied Mathematics and Computer Science

Prof. Dr. Nazım Mahmudov
Chair, Department of Mathematics

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Applied Mathematics and Computer Science.

Asst. Prof. Dr. Yucel Tandoğdu
Supervisor

Examining Committee

1. Prof. Dr. Rashad Aliyev
2. Prof. Dr. Agamriza Bashirov
3. Asst. Prof. Dr. Yucel Tandoğdu

ABSTRACT

Data analysis is the process of collecting and processing data with the aim of extracting significant and sound results to aid in decision making in almost every field where data collection is possible. However, when the number of variables involved in a process increase, processing of such data becomes more difficult. One way of alleviating such problems, is to reduce the number of variables to be processed in such a way that, the reduced version still represents great part of the variation in the data. This is achieved by the technique named Principal Component Analysis (PCA).

One other aspect considered in this study is the case when the interpretation of data is not very easy, as some data values may not definitely be assigned to a sub group of interest. Handling such situations is becoming possible through the theory of fuzzy logic. This enables the partial assignment of data to different sub groups, through the use of fuzzy membership functions. Using different fuzzy membership functions, it is possible to generate different membership data sets. Application of PCA to such data produced some interesting results that can be handy in selecting the type of the membership functions.

Keywords: Fuzzy logic, fuzzy set, fuzzy membership, covariance matrix, correlation matrix, principal component analysis.

ÖZ

Veri analizi, veri toplama, değerlendirme ve elde edilen sonuçların karar verme işlemlerinde kullanılması amacı ile veri elede edilebilecek her alanda kullanılan bir işlemdir. Ancak bir işlemde kullanılan değişken sayısı arttıkça, veri analizi daha zor hale gelir. Bu zorluğun üstesinden gelmenin bir yoluda, işlemi kontrol eden değişken sayısının, işlemdeki varyansın çok yüksek bir oranda temsil eileceği daha düşük bir boyuta indirgenmesidir. Bu amaca yönelik boyut indirgemesi Temel Bileşenler analizi yöntemi ile elde edilebilir.

Bu tezde üzerinde çalışılan diğer bir konu, bazı verilerin veri setini oluşturan alt kümelerden herhangi birine kesin tayininin mümkün olmadığı durumlardır. Kesin olmayan kümeler kuramı ile bu tür durumların çözümünde büyük ilerlemeler sağlanmıştır. Bu kuram çerçevesinde üyelik fonksiyonları kullanılarak verilerin farklı alt kümelere kısmi tayini yapılabilmektedir. Farklı üyelik fonksiyonları kullanılarak, farklı üyelik veri kümeleri üretmek mümkündür. Bu şekilde elde edilen veri kümelerinde temel bileşenler analizi yöntemleri uygulanmış ve tatmin edici sonuçlara ulaşılmıştır.

Anahtar kelimeler: Kesin olmayan mantık, kesin olmayan küme, kesin olmayan üyelik, kovaryans matrisi, korelasyon matrisi, temel bileşenler analizi.

DEDICATION

Dedicated to my dears:

-Father.

-Brothers and sister.

-Husband Ahmad, and

-Daughters Arina and Andea.

Shagul Faraj Karim

ACKNOWLEDGMENT

I am whole heartedly thankful and deepest gratitude to my supervisor Asst. Prof. Dr. *Yücel Tandođdu* for his excellent guidance, broad range of knowledge and unfailing patience, along with his encouragement, has been totally crucial component in the present study.

A special thanks to my Father *Faraj Karim*.

I would like to say thanks to my brothers *Mohamad* and *Aros Babani* for the love and support they have given to me.

Special thanks are also to my respectful friend Mr. *Yves Yannick Yameni*, for his time to discuss and help in my work for preparing this thesis. My sincere thanks are also expressed to all others who have helped me.

Finally my heartfelt thanks to my husband *Ahmad* for his patience and supports.

TABLE OF CONTENTS

ABSTRACT	iii
ÖZ.....	iv
DEDICATION	iii
ACKNOWLEDGMENT.....	vii
LIST OF TABLES.....	ix
LIST OF FIGURES	x
LIST OF SYMBOLS	xi
1 INTRODUCTION.....	1
2 LITERATURE REVIEW.....	3
3 CONCEPTS OF FUZZY SET THEORY.....	5
3.1 Classical Set Theory.....	5
3.2 Measurability of Sets.....	7
3.3 Algebra of Fuzzy Set Theory.....	9
3.3.1 Fuzzy interval and Arithmetic Intervals	12
4 CONCEPTS OF PRINCIPAL COMPONENT ANALYSIS	16
4.1 Basic Idea of Principal Components	16
4.2 Algebraic Review.....	17
4.2.1 Spectral decomposition.....	20
4.2.2 Singular value decomposition.....	20
4.2.3 Rotation.....	21
4.3 Statistical Review.....	21
4.3.1 Covariance Matrix.....	21
4.3.2 Correlation Matrix	26

4.4 Computation of Principal Components.....	27
4.4.1 Theoretical Background	27
5 FUZZY LOGIC AND PRINCIPAL COMPONENTS ANALYSIS	35
5.1 Fuzzy Principal Components Analysis.....	35
5.1.1 Fuzzification Procedure and its Algorithms	35
5.1.2 Application of PCA Concepts to Fuzzy Data	39
6 CONCLUSION.....	49
REFERENCES.....	511

LIST OF TABLES

Table 4.1: Student grades.....	30
Table 4.2: Eigenvalues obtained from the covariance matrix, and percentage of variation they represent.....	33
Table 4.3: Eigenvalues obtained from the correlation matrix, and percentage of variation they represent.....	33
Table 4.4: Correlation coefficient values between the PCs and random variables, obtained from the covariance and correlation matrices.	34
Table 5.1: Fuzzy membership data obtained from BMI values.....	40
Table 5.2: Fuzzy membership values from different membership functions random variables.....	45
Table 5.3: Correlation coefficient values between the PCs and random variables, obtained from the covariance and correlation matrices.....	48

LIST OF FIGURES

Figure 1: Linear fuzzy membership function for Body Mass Index.....	10
Figure 2: Fuzzy membership graph for the sweetness in the {80-120} gram fuzzy intervals.....	12
Figure 3: Confidence interval in two-dimensional space.....	13
Figure 4: Eigenvalues of example 1 forming an elbow.....	29
Figure 5.1: Standard normal quintile – quintile plot for X_1	41
Figure 5.2: Fuzzy membership values obtained from 5 different memberships function.....	45
Figure 5.3: Eigenvalues of example 2 forming an elbow.....	47

LIST OF SYMBOLS

λ	Eigenvalue
\mathbf{e}	Eigenvector
\mathbf{x}'	Transpose of a vector
\mathbf{x}	Vector
μ	Population mean
Σ	Population covariance matrix
\mathbf{S}	Sample covariance matrix
ρ	Population correlation coefficient matrix
\mathbf{R}	Sample correlation coefficient matrix
σ	Standard deviation of a random variable
Λ	Diagonal matrix of eigenvalues
ρ_{Y_i, X_j}	Correlation between the i^{th} PC Y_i and its j^{th} variable X_j
Y_{iS}	i^{th} Principal component computed using a sample covariance matrix
Y_{iR}	i^{th} Principal component computed using a sample correlation matrix
PC	Principal components
PCA	Principal components analysis

Chapter 1

INTRODUCTION

Various statistical techniques are used in data analysis, starting with the univariate case progressing towards the multivariate data analysis. Principal component analysis (PCA) is a statistical technique that provides the facility of dimension reduction, when available data is multivariate [1] and [2]. This is done without losing much from the inherent characteristics of the data. One other important issue in data processing arises when the data is inhomogeneous and has to be divided into sub groups. Until recently it was assumed that an element in one sub group can not belong to another group, meaning the sub sets are disjoint. However, there are situations when an element can be considered as a partial member of two sub-sets. In such cases a fuzzy interval is defined between neighboring subsets, where elements in this interval are attributed membership values according to some membership function. That is each element of the fuzzy interval will have $a\%$ membership to one sub-set, and $(1-a)\%$ to the neighboring sub-set. Elements outside the fuzzy interval are considered as “crisp”, belonging to only one sub-set with membership value 1 and not belonging to any other sub-set, meaning membership value 0 [3].

In this thesis an attempt is made to establish a relationship between the fuzzy membership values generated under different conditions assumed to be represented by different variables, and processing them through the PCA to achieve the reduction of variables for easy manipulation.

Chapter 3 summarizes the necessary concepts of classical set theory, algebra of fuzzy set theory, with examples. A brief review of fuzzy interval and arithmetic intervals is presented.

In chapter 4 fundamentals of the principal component analysis are explained in fair detail, and examples used to further clarify certain important concepts. Since PCA uses certain matrix algebra, a review of this is given. A brief summary of multivariate statistics is also introduced, as this provides the essential input to PCA. Covariance (**S**) and correlation (**R**) matrices computed from raw data are especially important in PCA. The work of researchers who contributed in the establishment of PCA, starting with K. Pearson and H. Hotelling are considered. Theory behind the determination of PCs based on the eigenvalues and eigenvectors of **S** or **R** matrices is briefly explained. Properties and interpretation of PCs are given.

Chapter 5 summarizes the fuzzification and defuzzification processes, with subsequent applications. In the application part, topics presented in chapters 3 and 4 are combined to show that the fuzzy membership data represented by different variables can be used in PC analysis. This is achieved by using numeric examples. First example with the same type fuzzy membership function, with different fuzzy intervals is tested. Obtained results were not satisfactory, as the generated variables turned out to have extremely high correlation. Second example where different fuzzy membership functions were applied for the same fuzzy interval produced better results, with variables still highly correlated, but not as in the first example. This enabled the application of PCA for dimension reduction and interpretation of obtained results.

Chapter 2

LITERATURE REVIEW

Fuzzy principal component analysis is a topic that is gaining momentum with increasing research devoted to this field recently. While PCA is a well-established methodology mainly applied in dimension reduction of high dimensional data, fuzzy logic has come into focus only in mid 1970s and rapidly gained ground in application fields where the establishment of a deterministic model is very difficult or impossible. As a result new terminology like fuzzy sets, fuzzy logic, fuzzy systems, and fuzzy control theories are introduced and found application in various fields. With the advent, of the fuzzy logic introduced in 1965 by Zadeh Lotfi. A., it developed to a new tool of processing data which are not defined explicitly but belong to a certain interval. Some interesting and useful sources that are referred to during this study on this topic are [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16] and [17].

Since this study attempts to establish a link between the fuzzy logic and the PCA by creating multi variate data via the fuzzy membership functions, and processing this data by PCA, brief review of literature is also undertaken. It is widely accepted that the work of KARL Pearson in 1901 on the geometric representation of points aiming to determine the direction with the highest number of points in a given scatter plot, formed the base for the idea of principal components analysis. Harold Hotelling developed the foundation of PCA in 1933 as it is commonly accepted today [18]. In

1939 Girshick Meyer produced good results about multivariate hypothesis testing based on the PCA [27].

About two decades after Meyer, this in 1963, Anderson T.W contributed in the development of the principal component analysis, by establishing the computation of principal components using the covariance matrix as well as the asymptotic properties of the roots of the characteristic polynomial [18].

In 1967 Jeffers studied the properties of eigenvalues and corresponding eigenvectors in his article titled “Two case studies in the application of principal component analysis “[28].

In 1974 Baxter proved that the graphic plot of principal component scores makes it easy to be understood.

Chapter 3

CONCEPTS OF FUZZY SET THEORY

Fuzzy set theory utilizes concepts of set theory in mathematics. Given a set and any entity, we can say that this entity is a member of or is not a member of the set. From probabilistic point of view, the membership of an entity to a set cannot be measured, since the membership does not depend on chance. Partial membership cannot be handled by the classical set theory concepts. However, the concepts of set theory are needed in fuzzy set theory. A brief review of set theory will be given.

3.1 Classical Set Theory

This section is a review of the ordinary set theory. It forms the foundation to the concepts of the fuzzy sets which will be used in fuzzy logic.

Let's consider a nonempty set S , called the universal set and it is made of elements defined within a particular context. Each element from the universal set is usually called an element or a member of the set. A subset of S , is made of the union of several elements of S . The following notations have the corresponding meaning defined as follow [20].

Definition 1:

- a) $x \in S$: The element x belongs to the set S .
- b) $x \notin S$: The element x doesn't belong to the set S .
- c) $A \subset S$: Shows that A is a subset of S .

In general the notation $A \subset S$ implies that A is strictly included in S , which means that there exists at least an element of S which is not in A . Besides that, if A can also be equal to S the following notation is used $A \subseteq S$.

\mathbb{R}^n which is called n -dimensional Euclidean space is the most useful universe for

this study. In this space, a vector \mathbf{x} is represented by $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

Definition 2: Considering a subset $A \subseteq \mathbb{R}^n$, A is said to be convex if for given

two vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ in A and for any $\alpha \in [0,1]$, $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in A$.

For given two sets A_1 and A_2 , the difference between them is defined by $A_1 - A_2 = \{x \in A_1 \mid x \notin A_2\}$. Considering the particular case where the set A_1 is in the universal set S , the difference $S - A_2$ is called the complement of A_2 with respect to S . This complement set is usually denoted by \bar{A}_2 or A_2' .

Remark: If U is a universe and if A is included in U then the following equalities hold,

$$\bar{\bar{A}} = A, \bar{U} = \emptyset \text{ and } \bar{\emptyset} = U.$$

Considered a set S and subsets A_1 and A_2 , the following operation are defined

- a) $kA_1 = \{kx \mid x \in A_1\}$: Multiplication of a set.
- b) $A_1 \cup A_2 = \{x \mid x \in A_1 \text{ or } x \in A_2\}$: Union of two sets.
- c) $A_1 \cap A_2 = \{x \mid x \in A_1 \text{ and } x \in A_2\}$: Intersection of two sets.

Remark:

- $A_1 \cap S = A_1, \quad A_1 \cup S = S, \quad A_1 \cap \emptyset = \emptyset, \quad A_1 \cap A_1' = \emptyset.$
- If $A_1 \cap A_2 = \emptyset$ then the sets A_1 and A_2 are said to be disjoint.

Definition 3:

The characteristic function of a set S , indicates whether an element belongs to that set or not. It is defined and denoted by,

$$\mathbf{X}_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases} \quad (3.1.1)$$

Considered a set S and subsets A_1 and A_2 , the following properties hold

- $\mathbf{X}_{A_1 \cup A_2}(x) = \max\{\mathbf{X}_{A_1}(x), \mathbf{X}_{A_2}(x)\}$
- $\mathbf{X}_{A_1 \cap A_2}(x) = \min\{\mathbf{X}_{A_1}(x), \mathbf{X}_{A_2}(x)\}$
- $\mathbf{X}_{A_1'}(x) = 1 - \mathbf{X}_{A_1}(x)$

3.2 Measurability of Sets

This section defines the measurability of classical sets. The measure theory as defined here is useful in fuzzy set theory as well [10], [14], [16] and [17].

Given the universal set \mathbf{S} and a family of subsets of \mathbf{S} being \mathbf{A} (\mathbf{A} is nonempty), and non-negative real valued function defined on \mathbf{A} , $\mu: \mathbf{A} \rightarrow [0, \infty]$, the set $B \in \mathbf{A}$ is a null set with respect to μ if $\mu(B) = 0$. That is to say $\mu(B) = \{\mu(b) | b \in B\}$.

Properties of the measure μ

1. Additivity property. For any collection of sets $\{A_1, A_2, \dots, A_n\}$ in \mathbf{A} satisfying

$\bigcup_{i=1}^n A_i \in \mathbf{A}$ and $A_i \cap A_j = \emptyset, i \neq j, i, j = 1, \dots, n$, when $\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$, μ is said to be additive. If $n = \infty$ μ said to be countably additive.

2. $A \in \mathbf{A}, B \in \mathbf{A}, A \subseteq B, B - A \in \mathbf{A}$, and $\mu(B) < \infty$ implies $\mu(B - A) = \mu(B) - \mu(A)$, meaning that μ is *subtractive*.

When properties 1 and 2 are satisfied, and there exist C a nonempty set, $C \in \mathbf{A}$ such that $\mu(C) < \infty$, then μ is said to be a measure on \mathbf{A} .

When $\mu(A) = 0$ then μ called *trivial measure*. If $\mu(A) =$ number of elements in \mathbf{A} then μ is a *natural measure*. A set $A \in \mathbf{A}$ is said to have a *finite measure* if $\mu(A) < \infty$.

Given $\{A_i\}_{i=1}^{\infty} \in \mathbf{A}$ such that $A \subseteq \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < \infty$ for all $i = 1, \dots$.

μ is σ -finite on \mathbf{A} if every set of \mathbf{A} has σ -finite measure.

When $B \in \mathbf{A}, A \subseteq B$, and $\mu(B) = 0$ they imply $\mu(A) = 0$. Then μ is *complete measure*.

Let $[A \in \mathbf{A}, B \in \mathbf{A}, \text{ and } A \subseteq B] \rightarrow \mu(A) \leq \mu(B)$, then μ is *monotone*.

For any $A, A_1, \dots, A_n \in \mathbf{A}$ where $A = \bigcup_{i=1}^n A_i$, if $\mu(A) \leq \sum_{i=1}^n \mu(A_i)$ then μ is *countably sub additive*.

Let μ be a measure on a universe \mathbf{A} ; let a set $B \in \mathbf{A}$ and a sequence of set $\{B_j\} \subset \mathbf{A}$.

- The measure μ is said to be **continuous from below at B** , if $B_1 \subseteq B_2 \subseteq \dots$

and $\lim_{j \rightarrow \infty} B_j = B$ imply all together that $\lim_{j \rightarrow \infty} \mu(B_j) = \mu(B)$.

- The measure μ is said to be **continuous from above at B** , if $\{B_j\} \subset \mathbf{A}$,

$B_1 \supseteq B_2 \supseteq \dots$, $\mu(B_1) < \infty$, $\lim_{j \rightarrow \infty} B_j = B$, imply all together that

$$\lim_{j \rightarrow \infty} \mu(B_j) = \mu(B) [9,10].$$

- If μ is continuous from below and from above at B , or on \mathbf{A} , then μ is said to be continuous.

3.3 Algebra of Fuzzy Set Theory

In the section 3.1, the characteristic function of a set S was defined as in (3.1.1). This formula actually indicates the membership or the non-membership of an element to the set S . The aim of this section is to generalize the definition of the characteristic function on the case which an element has only a partial membership on a set [3], [7], [8], [9] and [11]. This generalized characteristic function is illustrated via the following examples.

Example 3.1: Let all adult human constitute all members of the universal set. The body mass index (BMI) is used to divide adults in to 4 groups as below

$$BMI = \begin{cases} < 20 & \text{thin} \\ 20 - 25 & \text{normal} \\ 25 - 30 & \text{overweight} \\ > 30 & \text{obese} \end{cases}$$

Although the boundary values defining the category of the person depending on the weight, the boundary values cannot be so sharp. There must be a transition interval around each boundary value. This interval indicates that a person cannot belong 100% to one category or the other separated by the boundary. Therefore, the category of a person in this interval is not crisp, but it is fuzzy.

If we assume the fuzzy relation between categories is linear, and decide the size or width of the fuzzy interval, then the membership of a person to adjacent categories can be determined.

In example 1 let us assume the following Fuzzy intervals are valid

$$BMI = \begin{cases} < 18 & \text{thin} \\ 18 - 22 & \text{fuzzy} \\ 22 - 23 & \text{normal} \\ 23 - 27 & \text{fuzzy} \\ 27 - 28 & \text{overweight} \\ 28 - 32 & \text{fuzzy} \\ > 32 & \text{obese} \end{cases}$$

Figure 1 shows the fuzzy relation between categories as given above.

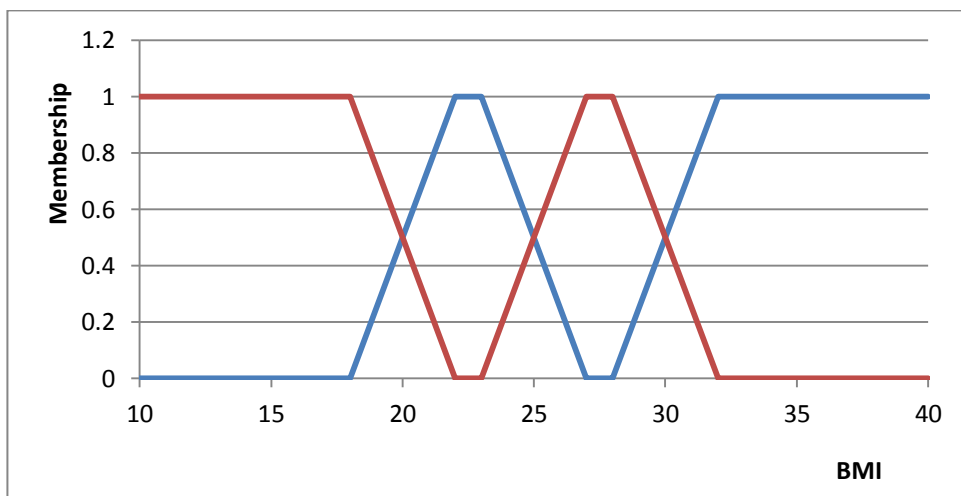


Figure 1: Linear fuzzy membership function for Body Mass Index

Based on the assumed fuzzy intervals, the linear membership function is computed as $\mu_1 = 0.25x - 4.5$ when the slope is positive, and $\mu_2 = -0.25x + 5.5$ when the slope is negative in the fuzzy interval (18, 22). For example if the BMI of a person is $x=19$, then this person is 25% in the normal category and 75% in the thin category.

Example 3.2: Main ingredients of baking 1 kg cake are 300 grams flour, 300 grams sugar, 300 grams butter, and 5 eggs. Assume the sweetness of a cake can be classified according to sugar content as

$$\text{Sweetness} = \begin{cases} < 100 & \textit{not sweet} \\ 100 - 200 & \textit{medium} \\ 200 - 300 & \textit{sweet} \\ > 300 & \textit{very sweet} \end{cases}$$

As explained in example 3.1, there will be a fuzzy interval in changing from one category to another. Let us assume the following fuzzy intervals

$$\text{Sweetness} = \begin{cases} < 80 & \textit{not sweet} \\ 80 - 120 & \textit{fuzzy} \\ 120 - 180 & \textit{medium} \\ 180 - 220 & \textit{fuzzy} \\ 220 - 280 & \textit{sweet} \\ 280 - 320 & \textit{fuzzy} \\ > 320 & \textit{very sweet} \end{cases}$$

If it is assumed that the fuzzy relationship is represented by the following function

$$\mu = \left\{ (x, f_\mu(x)) \mid x \in [a, b], f_\mu(x) = \frac{1}{1 + |x - y|} \right\}, \text{ then the fuzzy membership graph is}$$

as given in figure 2. Here a, b are the lower and upper boundaries for the fuzzy

intervals, and y is the center point for $f_\mu(x)$, or intersection of two curves

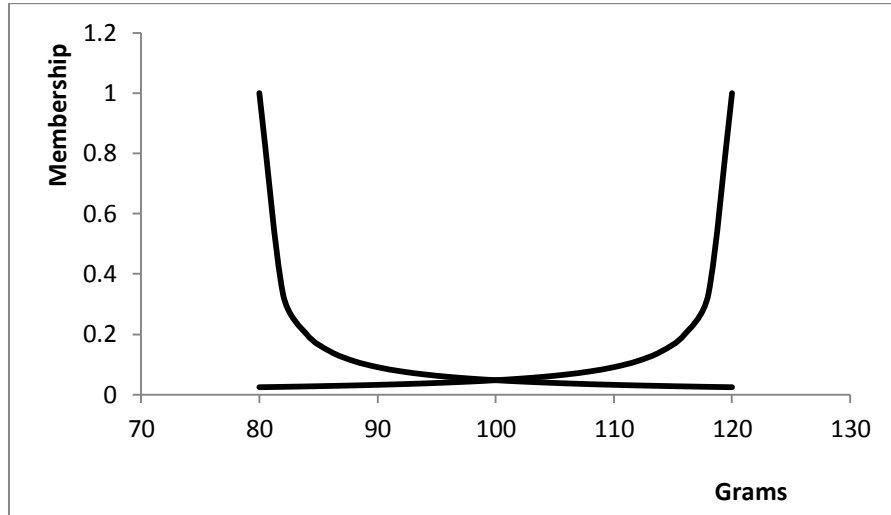


Figure 2: Fuzzy membership graph for the sweetness in the 80-120 gram fuzzy intervals.

3.3.1 Fuzzy interval and Arithmetic Intervals

In this section, the study assumes that a member set value v is uncertain. Nevertheless, it is assumed belonging to a certain interval, called range. It follows that $\underline{v} \leq v \leq \bar{v}$, with $[\underline{v}, \bar{v}] \subset \mathbb{R}$ called the confidence interval for the value of v . Moreover, if $\underline{v} = \bar{v}$ then the interval $[\underline{v}, \bar{v}] = [v, v] = v$. This topic is mainly concerned by close and bounded interval. They are graphically represent as given in figure 3. [3], [8], [9] and [11].

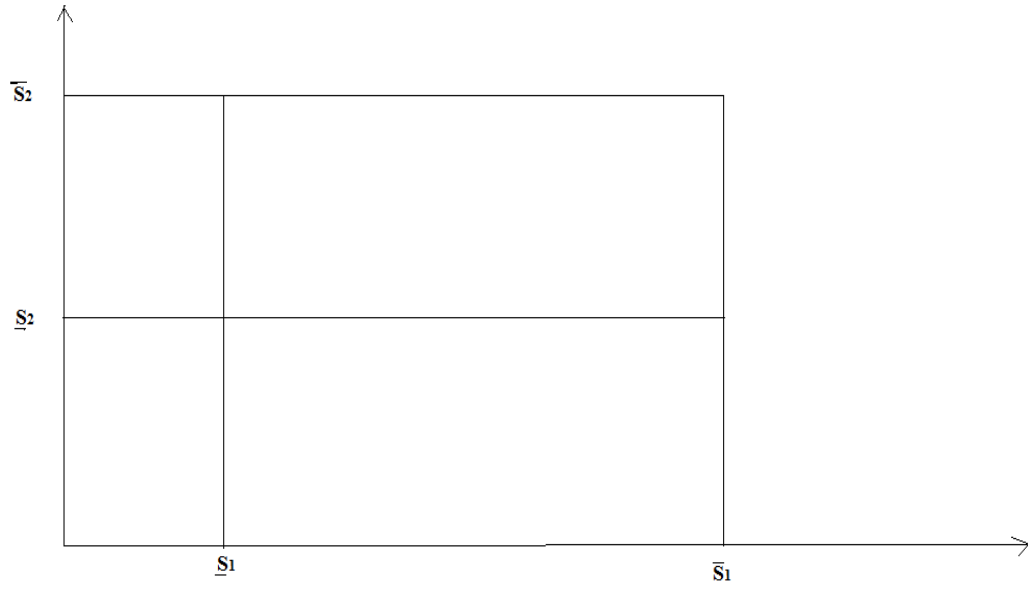


Figure 3: Confidence interval in two-dimensional space.

The operation and rules which are important for the interval of confidence are given in what follows [7], [8] and [9].

Definition: Consider two confidence intervals $[\underline{v}_1, \bar{v}_1]$ and $[\underline{v}_2, \bar{v}_2]$. The following relations and properties hold [9].

1- Equality of interval: $[\underline{v}_1, \bar{v}_1] = [\underline{v}_2, \bar{v}_2]$ if and only if $\underline{v}_1 = \underline{v}_2$ and $\bar{v}_1 = \bar{v}_2$.

2- Intersection: $[\underline{v}_1, \bar{v}_1] \cap [\underline{v}_2, \bar{v}_2] = [\max\{\underline{v}_1, \underline{v}_2\}; \min\{\bar{v}_1, \bar{v}_2\}]$.

3- Union: $[\underline{v}_1, \bar{v}_1] \cup [\underline{v}_2, \bar{v}_2] = [\min\{\underline{v}_1, \underline{v}_2\}; \max\{\bar{v}_1, \bar{v}_2\}]$. The union is defined

provided that $[\underline{v}_1, \bar{v}_1] \cap [\underline{v}_2, \bar{v}_2] \neq \emptyset$, otherwise the union result is not an interval, and

therefore it is undefined.

4- Inequality:

a) The interval $[\underline{v}_1, \bar{v}_1]$ is less than the interval $[\underline{v}_2, \bar{v}_2]$ i.e. $[\underline{v}_1, \bar{v}_1] < [\underline{v}_2, \bar{v}_2]$, if

and only if $\bar{v}_1 < \underline{v}_2$.

b) The interval $[\underline{v}_1, \bar{v}_1]$ is greater than the interval $[\underline{v}_2, \bar{v}_2]$ i.e. $[\underline{v}_1, \bar{v}_1] > [\underline{v}_2, \bar{v}_2]$, if and only if $\bar{v}_1 > \underline{v}_2$.

5- Inclusion: $[\underline{v}_1, \bar{v}_1] \subseteq [\underline{v}_2, \bar{v}_2]$ if and only if $\underline{v}_2 \leq \underline{v}_1$ and $\bar{v}_1 \leq \bar{v}_2$.

6- The width for a given interval $[\underline{v}_1, \bar{v}_1]$, is defined and denoted by

$w\{[\underline{v}_1, \bar{v}_1]\} = \bar{v}_1 - \underline{v}_1$. An important note is that, a singleton v has a width

$w\{v\} = w\{[v, v]\} = v - v = 0$; for all $v \in \mathbb{R}$.

7- Absolute value: $||[\underline{v}_1, \bar{v}_1]|| = \max\{|\underline{v}_1|, |\bar{v}_1|\}$. It is important to note that a singleton v

has an absolute value defined by $||[v, v]|| = \max\{|v|, |v|\} = |v|$, for all $v \in \mathbb{R}$.

8- Midpoint or mean: For a given interval $[\underline{v}_1, \bar{v}_1]$, it is defined and denoted by

$$m\{[\underline{v}_1, \bar{v}_1]\} = \frac{1}{2}(\underline{v}_1 + \bar{v}_1).$$

9- Symmetric: A given interval $[\underline{v}_1, \bar{v}_1]$ is symmetric if and only if $\underline{v}_1 = -\bar{v}_1$ or

$$m\{[\underline{v}_1, \bar{v}_1]\} = 0.$$

Behind the definitions mentioned above, there exists arithmetic for the interval of confidence defined as follow.

Definition: Consider $[\underline{v}_1, \bar{v}_1]$, $[\underline{v}_2, \bar{v}_2]$ and $[\underline{v}_3, \bar{v}_3]$ to be 3 intervals of confidence. The following operations are defined [8], [14] and [15].

1- Addition: $[\underline{v}_1, \bar{v}_1] + [\underline{v}_2, \bar{v}_2] = [\underline{v}_1 + \underline{v}_2, \bar{v}_1 + \bar{v}_2]$.

2- Subtraction: $[\underline{v}_1, \bar{v}_1] - [\underline{v}_2, \bar{v}_2] = [\underline{v}_1 - \underline{v}_2, \bar{v}_1 - \bar{v}_2]$.

3- Reciprocal:

a) If $0 \notin [\underline{v}_1, \bar{v}_1]$ then $[\underline{v}_1, \bar{v}_1]^{-1} = \left[\frac{1}{\underline{v}_1}, \frac{1}{\bar{v}_1} \right]$.

b) If $0 \in [\underline{v}_1, \bar{v}_1]$ then $[\underline{v}_1, \bar{v}_1]^{-1}$ is undefined.

4- Multiplication: $[\underline{v}_1, \bar{v}_1] \cdot [\underline{v}_2, \bar{v}_2] = [\underline{u}, \bar{u}]$. Where $\underline{u} = \min \{ \underline{v}_1 \underline{v}_2, \underline{v}_1 \bar{v}_2, \bar{v}_1 \underline{v}_2, \bar{v}_1 \bar{v}_2 \}$

and $\bar{u} = \max \{ \underline{v}_1 \underline{v}_2, \underline{v}_1 \bar{v}_2, \bar{v}_1 \underline{v}_2, \bar{v}_1 \bar{v}_2 \}$.

5- Division: $[\underline{v}_1, \bar{v}_1] / [\underline{v}_2, \bar{v}_2] = [\underline{v}_1, \bar{v}_1] \cdot [\underline{v}_2, \bar{v}_2]^{-1}$, provided that $[\underline{v}_2, \bar{v}_2] \notin 0$.

Chapter 4

CONCEPTS OF PRINCIPAL COMPONENT ANALYSIS

4.1 Basic Idea of Principal Components

Principal components analysis is a multivariate technique of dimension reduction, for a given multivariate or high dimension data. A multivariate data of p variables and n observations can be represented as follows [1], [2] and [19].

$$\begin{array}{cccc}
 & X_1 & X_2 & \cdots & X_p \\
 \text{observation 1} & x_{11} & x_{12} & \cdots & x_{1p} \\
 \text{observation 2} & x_{21} & x_{22} & \cdots & x_{2p} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 \text{observation n} & x_{n1} & x_{n2} & \cdots & x_{np}
 \end{array} \tag{4.1.1}$$

Principal components (PC) are some linear combinations of the p variables. It is represented as follows

$$\begin{aligned}
 Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
 Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
 &\vdots \\
 Y_p &= a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p
 \end{aligned} \tag{4.1.2}$$

The equation (4.1.2) can also be represented in a matrix form as follows

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \rightarrow \mathbf{y} = \mathbf{A} \cdot \mathbf{X} \tag{4.1.3}$$

The transition from a $(n \times p)$ raw data given by the equation (4.1.1) to p linear combinations given by equations (4.1.2) and (4.1.3) is done through some algebraic and statistical considerations (operations, properties, transformations, assumptions). In order to understand what a PC is, there is a need to define the necessary algebraic and statistical concepts.

4.2 Algebraic Review

Understanding principle component analysis (PCA), require basic knowledge of matrix algebra. Starting with the basic definition of the vector, any vector is denoted by a bold lower case letter, and may have n elements [2], [20]. For example

$\mathbf{a} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}$ is a column vector with n rows. The transpose of this vector is a row

vector denoted by $\mathbf{a}' = [a_{11} \ a_{12} \ \cdots \ a_{1n}]$

A matrix is made up of more than one vector. If there are n rows and p column than the matrix is denoted as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}. \quad (4.2.1)$$

Some of the vector and matrix operation that will widely be used in this thesis are briefly explained below.

Trace of a square matrix: It's the sum of the diagonal elements. If matrix \mathbf{A} is $n \times n$

, its trace $tr(\mathbf{A}) = \sum_{i=1}^n a_{ii}$.

Inverse of a square matrix: Given square matrix \mathbf{A} , its inverse is denoted by \mathbf{A}^{-1} and has the property $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Here \mathbf{I} is identity matrix. Some useful properties of the inverse of a matrix are [1], [2], [20] and [21]

- Inverse of a symmetric matrix is also symmetric
- Inverse of the transpose of \mathbf{A} is the transpose of \mathbf{A}^{-1} , $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$
- The inverse of the product of matrices \mathbf{A} , \mathbf{B} , \mathbf{C} can be written as the product of the inverse of these matrices in reverse order. $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$
- If $k \neq 0$ is a scalar then $(k\mathbf{A})^{-1} = \frac{1}{k}\mathbf{A}^{-1}$
- If \mathbf{A} diagonal matrix, then its inverse will be the reciprocals of the diagonal elements.

If matrix \mathbf{A} is not square then the left inverse \mathbf{A}_l^{-1} and right inverse \mathbf{A}_r^{-1} can be computed. If \mathbf{A} is $n \times p$ matrix and $n < p$ then $\mathbf{A}_r^{-1} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ and $\mathbf{A}\mathbf{A}_r^{-1} = \mathbf{I}_p$. Here \mathbf{I}_p is $p \times p$ identity matrix.

If $n > p$ $\mathbf{A}_l^{-1} = (\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}'$ and $\mathbf{A}\mathbf{A}_l^{-1} = \mathbf{I}_n$. Here \mathbf{I}_n is $n \times n$ identity matrix.

Positive definite matrix: Consider a symmetric matrix \mathbf{A} and any vector \mathbf{x} , such that the product $\mathbf{x}'\mathbf{A}\mathbf{x}$ is defined, then \mathbf{A} is said to be positive definite matrix if and only if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$. In case $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ the symmetric \mathbf{A} is said to be positive semi-definite.

Theorem (Perron–Frobenius): If a positive definite matrix has elements which are all positive, then the elements of its first eigenvector are positive [1], [2], [18] and [20].

The foregoing theorem is important in principal components analysis.

Orthogonal vector and matrix: Two vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ are orthogonal

if and only if their inner product

$$\mathbf{x}'\mathbf{y} = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = 0 .$$

If a vector \mathbf{x} is such that

$$\mathbf{x}'\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1x_1 + x_2x_2 + \cdots + x_nx_n = 1$$

Then \mathbf{x} is called a normalized vector. Any vector \mathbf{x} can always be normalized using

$\mathbf{x}_{norm} = \frac{\mathbf{x}}{\sqrt{\mathbf{x}'\mathbf{x}}}$, provided that $\mathbf{x} \neq 0$. Consider a matrix \mathbf{A} whose columns are

mutually orthogonal and normalized, such a matrix is called an orthonormal matrix [20].

Eigenvalue and eigenvectors: Considering any square matrix \mathbf{A} and a scalar λ , if there exists a nonzero vector \mathbf{x} , such that $\mathbf{Ax} = \lambda\mathbf{x}$, then λ is said to be an eigenvalue of \mathbf{A} and \mathbf{x} is said to be its corresponding eigenvector. Finding the eigenvectors of a given matrix, the following relation is used $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0$. The determinant $|\mathbf{A} - \lambda\mathbf{I}|$ is then computed and set equaled to zero, $|\mathbf{A} - \lambda\mathbf{I}| = 0$. Then, find non-trivial eigenvectors [1], [2] and [20].

4.2.1 Spectral Decomposition

Consider a $n \times n$ square matrix \mathbf{A} that has eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Let the diagonal matrix \mathbf{D} of which elements are eigenvalues of $\mathbf{A}_{n \times n}$ and the matrix $\mathbf{P}_{n \times n}$ whose elements are made of corresponding eigenvectors. Then

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}' , \quad (4.2.2)$$

is called the spectral decomposition of the matrix $\mathbf{A}_{n \times n}$.

Square root of a matrix: Consider a positive definite matrix $\mathbf{A}_{n \times n}$ and its spectral decomposition (4.2.1). If a diagonal matrix which elements are all square root of the eigenvalues is used instead of the matrix \mathbf{D} , then the spectral decomposition gives the square root of the matrix $\mathbf{A}_{n \times n}$. This means $\mathbf{A}^{1/2} = \mathbf{P}\mathbf{D}^{1/2}\mathbf{P}'$ [20, 21].

Note: Using the matrices \mathbf{A} , \mathbf{D} and \mathbf{P} , the square and the inverse of the matrix \mathbf{A} can be defined respectively by $\mathbf{A}^2 = \mathbf{P}\mathbf{D}^2\mathbf{P}'$ and $\mathbf{A}^{-1} = \mathbf{P}\mathbf{D}^{-1}\mathbf{P}'$.

4.2.2 Singular Value Decomposition

The concept of spectral decomposition for a square symmetric matrix can be extended to non-square matrices. In this case it is called singular value decomposition. Consider $\mathbf{A}_{n \times p}$, of rank k . then the singular value decomposition of

$(\mathbf{A})_{n \times p}$ is given by $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}'$. Where $\mathbf{D} = \text{diag}(\sqrt{\lambda_1^2}, \sqrt{\lambda_2^2}, \dots, \sqrt{\lambda_k^2})$; with

$(\lambda_1, \lambda_2, \dots, \lambda_k)$ being the nonzero eigenvalues of the matrix $\mathbf{A}\mathbf{A}'$ or $\mathbf{A}'\mathbf{A}$. The k

columns of \mathbf{U} are made of the normalized eigenvectors of the matrix $\mathbf{A}\mathbf{A}'$, whereas

the k columns of \mathbf{V} are made of the normalized eigenvectors of the matrix $\mathbf{A}'\mathbf{A}$ [1] and [2].

4.2.3 Rotation. Let \mathbf{v} be a vector in two dimensional spaces. To rotate \mathbf{v} by an angle θ counter clock wise, \mathbf{v} can be multiplied by the matrix

$$\mathbf{B} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

If $\theta = 90^\circ$ then the rotation will be orthogonal [2, 20].

4.3 Statistical Review

Principal component analysis is a techniques used in dimension reduction. One of the main fields of application of PCA is statistics. Hence, a brief review of some statistical theory is given in the following sub sections [19] and [22].

4.3.1 Covariance Matrix

Consider a random variable X with a probability distribution $f(x)$. The variance of X , is a value that helps to measure the dispersion among the values of the distribution [22].

Considering for instance a sample of n observations of a random variable X , shown as a row vector $\mathbf{x} \rightarrow (x_1, x_2, \dots, x_n)$. The dispersion of this observation can be measured by the computation of the sample variance, using the formula

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (4.3.1)$$

Where \bar{x} is the mean of the observation given in this case by $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$.

It can be shown that the formula (4.3.1) is equivalent to

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}. \quad (4.3.2)$$

In case, a population is considered instead of a sample, the following formula is used to compute the variance of the random variable X .

$$\sigma^2 = E(X - \mu)^2 \quad (4.3.3)$$

It can also be proved that the formula (4.3.3) is equivalent to

$$\sigma^2 = E(X^2) - \mu^2. \quad (4.3.4)$$

Where $E(X^2)$ represented the expectation of the variable X^2 , and μ^2 is the square of the population mean.

Theorem 4.1: Let X be a random variable and c a constant, the variance of the new variable cX is given by [2] and [22].

$$\sigma^2(cX) = c^2\sigma^2(X). \quad (4.3.5)$$

Considering a distribution with two random variables X_1 and X_2 , the measurement of the dispersion between these variables is represented by their covariance. Let us consider a sample of n observations of two random variables $X_1 \rightarrow (x_1, x_2, \dots, x_n)$ and $X_2 \rightarrow (x_1, x_2, \dots, x_n)$. The sample covariance of X_1 and X_2 is given by

$$s_{X_1X_2} = \frac{\sum_{i=1}^n (x_i - \bar{x}_1)(x_i - \bar{x}_2)}{n-1}. \quad (4.3.6)$$

Where \bar{x}_1 and \bar{x}_2 are the means of the random variables X_1 and X_2 respectively.

The formula (4.3.6) is equivalent to the following formula

$$s_{X_1X_2} = \frac{\sum_{i=1}^n x_i x_i - n\bar{x}_1\bar{x}_2}{n-1}. \quad (4.3.7)$$

In case, a population distribution is considered instead of a sample, the following formula is used to compute the covariance among the observation

$$\sigma_{X_1X_2} = E\left[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})\right]. \quad (4.3.8)$$

It is possible to show the equivalence between the formula (4.3.8) and the following formula

$$\sigma_{X_1X_2} = E(X_1X_2) - \mu_{X_1}\mu_{X_2}. \quad (4.3.9)$$

Consider two random variables X_1 and X_2 . Considering also two constants c_1 and c_2 . Then $c_1X_1 + c_2X_2$ is linear combination and its variance-covariance is defined as by

$$\begin{aligned} \sigma^2(c_1X_1 + c_2X_2) &= E\left[(c_1X_1 + c_2X_2) - (c_1\mu_1 + c_2\mu_2)\right]^2 \\ &= E\left[c_1(X_1 - \mu_1) + c_2(X_2 - \mu_2)\right]^2 \\ &= E\left[c_1^2(X_1 - \mu_1)^2 + c_2^2(X_2 - \mu_2)^2 + 2c_1c_2(X_1 - \mu_1)(X_2 - \mu_2)\right] \\ &= c_1^2\sigma_{X_1X_1} + c_2^2\sigma_{X_2X_2} + 2c_1c_2\sigma_{X_1X_2} \end{aligned} \quad (4.3.10)$$

The formula (4.3.10) can be written simply as

$$\sigma^2_{(c_1x_1+c_2x_2)} = c_1^2\sigma_{11} + c_2^2\sigma_{22} + 2c_1c_2\sigma_{12} \quad (4.3.11)$$

Let consider the vectors $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ and $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, then the linear combination can be

written as $c_1X_1 + c_2X_2 = \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{c}'\mathbf{X}$. Furthermore, considering the matrix Σ

defined by $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$, the following relation holds

$$\mathbf{c}'\Sigma\mathbf{c} = \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = c_1^2\sigma_{11} + c_2^2\sigma_{22} + 2c_1c_2\sigma_{12}. \quad (4.3.12)$$

Theorem 4.2: Recalling the formula (4.3.12), the matrix $\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$ is called

variance-covariance matrix of the random variables X_1 and X_2 . The entries σ_{11} and σ_{22} are called variance of the variables X_1 and X_2 respectively. The entrances σ_{12} and σ_{21} are called covariance between the variables X_1 and X_2 [1], [2], [19] and [21].

Remarks:

- σ_{12} measures the dependence between the variables X_1 and X_2 , whereas σ_{21} measures the dependence between the variables X_2 and X_1 . Thus $\sigma_{12} = \sigma_{21}$. This implies that the variance-covariance matrix is a symmetric matrix.
- From the formula (4.3.7), if the random variable X_2 has exactly the same distribution as the random variable X_1 , then $X_1 = X_2$ and

$$\sigma_{X_1X_1} = \frac{\sum_{i=1}^n x_i x_i - n\bar{x}_1 \cdot \bar{x}_1}{n-1} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}_1^2}{n-1} = \sigma_{X_1}^2.$$

Thus $\sigma_{x_1x_1} = \sigma_{x_1}^2$.

- Considering the previous formulas, it is important to mention that the symbol **S** is preferred to denote the sample variance-covariance whereas the symbol σ is used for the population variance-covariance. This notation will be adopted in what will follow.

In multivariate data analysis, it is common to observe simultaneously ($p \geq 3$) variables. In this case to measure the dependence among the variables, computation of a ($p \times p$) covariance matrix is required. Consider n simultaneous observations of p random variables. Then these observations can be represented in matrix form see (4.2.1)

$$\begin{array}{cccc}
 & X_1 & X_2 & \cdots & X_p \\
 \text{observation 1} & x_{11} & x_{12} & \cdots & x_{1p} \\
 \text{observation 2} & x_{21} & x_{22} & \cdots & x_{2p} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 \text{observation n} & x_{n1} & x_{n2} & \cdots & x_{np}
 \end{array}$$

The variance-covariance of these observations is a ($p \times p$) matrix computed and denoted as follow [1] and [2]

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (4.3.13)$$

where the entrances $s_{jj} = s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$. It measures the variance among the variable X_j and it self. For $j \neq k$, the covariance between the variables X_j and X_k is defined and computed as follow

$$s_{jk} = s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k).$$

When a population is considered, the variance-covariance matrix is denoted by

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{23} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \quad (4.3.14)$$

where the entrances σ_{jj} and σ_{jk} have respectively the same meaning with the entries s_{jj} and s_{jk} as given in (4.3.13).

4.3.2 Correlation Matrix

The correlation is a numerical value computed to measure the level of linear correlation between two or more observations from a multivariate data.

Consider two random variables X_1 and X_2 , the correlation coefficient between them is computed and denoted as follow [1], [2], [21] and [22]

$$\rho_{X_1, X_2} = \frac{\sigma_{X_1 X_2}}{\sigma_{X_1} \sigma_{X_2}}. \quad (4.3.15)$$

If there is a perfect linear dependence between the variables X_1 and X_2 , this means $X_1 = aX_2 + b$. Then the correlation coefficient between X_1 and X_2 is $\rho_{X_1, X_2} = 1$ if $a > 0$ and $\rho_{X_1, X_2} = -1$ for $a < 0$. In general for given two random variables X_1 and X_2 , $-1 \leq \rho_{X_1, X_2} \leq 1$.

Consider the multivariate observations of p random variables. The correlation coefficient between the pair of variables can be computed and the correlation coefficient matrix is given as

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}$$

The correlation between a variable and itself is perfect; this is why the diagonal

elements are all equal to 1. $\rho_{jj} = \frac{\sigma_{X_j X_j}}{\sigma_{X_j} \sigma_{X_j}} = \frac{\sigma_{X_j}^2}{\sigma_{X_j} \sigma_{X_j}} = \frac{\sigma_{X_j} \sigma_{X_j}}{\sigma_{X_j} \sigma_{X_j}} = 1$. In the case of n

observation, the correlation matrix will be,

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix} \quad (4.3.16)$$

where $r_{jk} = \frac{s_{X_j X_k}}{s_{X_j} s_{X_k}}$. The correlation coefficient matrix is a symmetric matrix $r_{jk} = r_{kj}$.

4.4 Computation of Principal Components

Given a $n \times p$ data matrix \mathbf{X} , there may be conditions that necessitates the reduction of the size of this matrix. As p represent the number of the variables, it may be desirable to reduce them, without losing the inherent characteristic of the process under study, when p is large. Computations to achieve the required dimension reduction will be explained in the following sub sections [1], [2], [19], [21] and [23].

4.4.1 Theoretical Background

The principle components are linear combination of p random variables X_1, X_2, \dots, X_p . In the computation of PCs, the original coordinate system of the data is rotated to a direction such that the first PC points to the direction with the highest variation in the data, and the last PC point to the direction with the lowest variation in the data.

Computations of the PCs depend of the covariance matrix Σ obtained from the $n \times p$ data matrix. The eigenvalues and corresponding eigenvectors of Σ are $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ and $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$. Since the covariance matrix is symmetric its eigenvectors are orthogonal. Let the square ($p \times p$) matrix \mathbf{E} be made up of the p eigenvectors \mathbf{e}_i ; $i = 1, 2, \dots, p$, then the PC matrix is given by

$$\mathbf{Y} = \mathbf{X}\mathbf{E}' \quad (4.4.1)$$

Or

$$\begin{aligned} Y_1 &= e_{11}X_1 + e_{12}X_2 + \dots + e_{1p}X_p \\ Y_2 &= e_{21}X_1 + e_{22}X_2 + \dots + e_{2p}X_p \\ &\vdots \\ Y_p &= e_{p1}X_1 + e_{p2}X_2 + \dots + e_{pp}X_p \end{aligned} \quad (4.4.2)$$

The variance of the Y matrix is made up of the diagonal element of the resulting matrix given below [1] and [2]

$$\text{Var}(\mathbf{Y}) = \sigma_Y^2 = \mathbf{E}'\Sigma\mathbf{E} \quad (4.4.3)$$

and the covariance of the Y matrix (Σ_Y) is made up of the non-diagonal elements in equation (4.4.3). From here, the i^{th} PC can be written as a linear combination $\mathbf{e}'_i\mathbf{x}$ that maximizes $\text{Var}(\mathbf{e}'_i\mathbf{x})$ subject to $\mathbf{e}'_i\mathbf{e}_i = 1$ and $\text{Cov}(\mathbf{e}'_i\mathbf{x}, \mathbf{e}'_j\mathbf{x}) = 0$, $i \neq j$. Here $\mathbf{x} = (x_1, x_2, \dots, x_p)$, representing p variables the following property is important in PCA and is expressed as

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) \quad (4.4.4)$$

According to equation (4.4.4) the sum of the variances of a random variables is equal to the sum of eigenvalues of the covariance matrix Σ . As a result, the proportion of total variance, explained by the k^{th} PC is given by

$$T = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i}, \lambda = 1, \dots, p \quad (4.4.5)$$

Equation (4.4.5) can be used to determine the number of PCs that can represent high percentage (80% or above) of total variation in the data.

Another method to determine the number of PCs is the scree plot of the eigenvalue [1]. Since the eigenvalues are in decreasing order, the graph forms an elbow. Generally the point where the elbow is observed is taken as the number of PCs necessary to represent the majority of variation in the data.

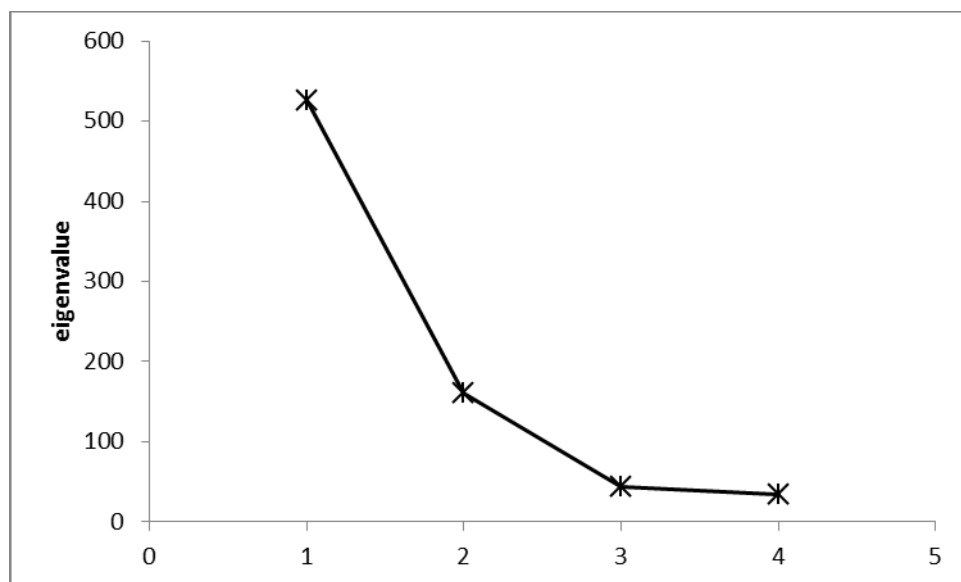


Figure 4: Eigenvalues of example 1 forming an elbow.

The PCs are obtained using the eigenvectors of the covariance matrix

$(Y_1 = \mathbf{e}'_1 \mathbf{x}, \dots, Y_p = \mathbf{e}'_p \mathbf{x})$. The correlation between PCs and the random variables is a useful tool. They show the relationship between PCs and the constituent variables.

They are given by

$$\rho_{Y_i, X_k} = \frac{e_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \quad i, k = 1, \dots, p \quad (4.4.6)$$

Example: Consider the following dataset, representing the grades of 10 IT students in the following courses. Java programming (X_1), Database management (X_2), Introduction to statistics (X_3) and English communication (X_4). The data related with 4 variables is given in table 4.1.

Table 4.1: Student grades for 4 different subjects.

	X_1	X_2	X_3	X_4
1	90	75	65	55
2	70	80	60	50
3	85	60	50	65
4	70	70	55	50
5	85	60	70	60
6	45	55	55	50
7	40	35	35	35
8	60	45	50	40
9	40	65	70	35
10	60	65	65	60

Compute the PCs using both covariance and correlation matrices.

Using the data from the table 4.4, the covariance matrix, its eigenvalues and corresponding eigenvectors are computed and given below.

Covariance matrix

$$\mathbf{S} = \begin{pmatrix} 352.50 & 131.11 & 65.28 & 152.78 \\ 131.11 & 182.22 & 102.78 & 63.89 \\ 65.28 & 102.78 & 118.06 & 38.89 \\ 152.78 & 63.89 & 38.89 & 111.11 \end{pmatrix}.$$

Eigenvalues

$$\lambda_1 = 526.09, \lambda_2 = 160.80, \lambda_3 = 43.48, \lambda_4 = 33.52.$$

Eigenvectors

$$\mathbf{e}_1 = \begin{pmatrix} 0.7675 \\ 0.4433 \\ 0.2703 \\ 0.3761 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0.4735 \\ -0.6208 \\ -0.5944 \\ 0.1926 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} -0.1793 \\ -0.5359 \\ 0.6002 \\ 0.5661 \end{pmatrix} \text{ and } \mathbf{e}_4 = \begin{pmatrix} 0.3932 \\ -0.3619 \\ 0.4619 \\ -0.7078 \end{pmatrix}.$$

The principal components are the linear combination of the random variables where the coefficients are the elements of the eigenvectors

$$\begin{cases} Y_1 = 0.7675X_1 + 0.4433X_2 + 0.2703X_3 + 0.3761X_4 \\ Y_2 = 0.4735X_1 - 0.6208X_2 - 0.5944X_3 + 0.1926X_4 \\ Y_3 = -0.1793X_1 - 0.5359X_2 + 0.6002X_3 + 0.5661X_4 \\ Y_4 = 0.3932X_1 - 0.3619X_2 + 0.4619X_3 - 0.7078X_4 \end{cases}.$$

The correlation matrix, its eigenvalues and corresponding eigenvectors computed from the same data are given as follows.

Correlation matrix

$$\mathbf{R} = \begin{pmatrix} 1.00 & 0.52 & 0.32 & 0.78 \\ 0.52 & 1.00 & 0.70 & 0.45 \\ 0.32 & 0.70 & 1.00 & 0.34 \\ 0.77 & 0.45 & 0.34 & 1.00 \end{pmatrix}.$$

Eigenvalues

$$\lambda_1 = 2.56, \lambda_2 = 0.94, \lambda_3 = 0.30, \lambda_4 = 0.20 .$$

Eigenvectors

$$\mathbf{e}_1 = \begin{pmatrix} 0.5175 \\ 0.5216 \\ 0.4518 \\ 0.5060 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0.4665 \\ -0.3960 \\ -0.6229 \\ 0.4874 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} -0.2756 \\ -0.6434 \\ 0.5547 \\ 0.4499 \end{pmatrix} \text{ and } \mathbf{e}_4 = \begin{pmatrix} 0.6623 \\ -0.3964 \\ 0.3166 \\ -0.5514 \end{pmatrix}.$$

The principal components obtained are

$$\begin{cases} Y_1 = 0.5175X_1 + 0.5216X_2 + 0.4518X_3 + 0.5060X_4 \\ Y_2 = 0.4665X_1 - 0.3960X_2 - 0.6229X_3 + 0.4874X_4 \\ Y_3 = -0.2756X_1 - 0.6434X_2 + 0.5547X_3 + 0.4499X_4 \\ Y_4 = 0.6623X_1 - 0.3964X_2 + 0.3166X_3 - 0.5514X_4 \end{cases}.$$

Based on the computed PCs the percentage of variation represented by each PC are given table (4.2), where T is given in (4.4.5)

Table 4.2: Eigenvalues obtained from the covariance matrix, and corresponding percentage of variation they represent.

Covariance matrix					
	λ_1	λ_2	λ_3	λ_4	Total
	526.09	160.80	43.48	33.52	763.89
T	0.69	0.21	0.06	0.04	1
Cum. T	0.69	0.90	0.96	1	

Table 4.3: Eigenvalues obtained from the correlation matrix, and corresponding percentage of variation they represent.

Correlation matrix					
	λ_1	λ_2	λ_3	λ_4	Total
	2.56	0.94	0.30	0.20	4
T	0.64	0.235	0.075	0.05	1
Cum. T	0.64	0.875	0.95	1	

It is observed that the first eigenvalue obtained from the covariance matrix represent 90% of total variation in the data. Respective eigenvalue of the correlation matrix represents 87.5% of total variation. Since the two percentages are very close, it indicates that for each variable the variance of the row data and the standardized data is compatible.

Linear correlation between the PCs and involved variables is given in equation (4.4.6). Using this equation $\rho_{Y_1, X_i}, \rho_{Y_2, X_i}, \rho_{Y_1, Z_i}$ and ρ_{Y_2, Z_i} are computed and presented in Table (4.3).

Table 4.4: Correlation coefficient values between the PCs and random variables, obtained from the covariance and correlation matrices.

	X_i	X_1	X_2	X_3	X_4
S matrix	ρ_{Y_1, X_i}	0.9376	0.7532	0.5705	0.8183
	ρ_{Y_2, X_i}	0.3197	-0.5831	-0.6936	0.2316
		Z_1	Z_2	Z_3	Z_4
R matrix	ρ_{Y_1, Z_i}	0.8280	0.8345	0.7228	0.8096
	ρ_{Y_2, Z_i}	0.4522	-0.3839	-0.6039	0.4725

Coefficients of the PCs are eigenvalues. They indicate the contribution of each variable to the particular PC. The linear correlation coefficient between a PC and its constituent variables shows the degree of linear correlation. It must be noted that the variable with the highest contribution to the PC, may not also have the highest correlation with the PC. This is clearly visible from table (4.4). For example in the first PC obtained from the covariance matrix, the highest contributing variable is X_1 also has the highest correlation with Y_1 . On the other hand, the second highest contributing variable X_2 has the third highest correlation with Y_1 .

Chapter 5

FUZZY LOGIC AND PRINCIPAL COMPONENTS ANALYSIS

The idea of principal components analysis has been extended to be adapted as a solution to various problems. Indeed, many methods with the roots in principal components analysis (PCA) exist. Probabilistic principal components analysis (PPCA), weighted principal components analysis (WPCA), fuzzy principal components analysis (FPCA) can be named just to mention a few. In this chapter, the association of fuzzy algebra and PCA will be discussed theoretically and some of its application will be presented [6], [23] and [24].

5.1 Fuzzy Principal Components Analysis

The fuzzy principal component analysis (FPCA) enables the fuzzification of the raw data matrix that helps to diminish the influence of possible outliers.

5.1.1 Fuzzification Procedure and its Algorithms.

To determine the structure of the data, the fuzzy clustering tool is designed to be adequate. The general procedure of the fuzzy clustering algorithm which is associated to an objective function is defined as follow. Consider a finite set of feature vector in \mathbb{R}^p . This means $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\} \subset \mathbb{R}^p$. Then \mathbf{X} is a data matrix with its vectors defined as: $\mathbf{x}_1 = [x_{11}, x_{12}, \dots, x_{1p}]$, $\mathbf{x}_2 = [x_{21}, x_{22}, \dots, x_{2p}]$, ..., $\mathbf{x}_p = [x_{n1}, x_{n2}, \dots, x_{np}]$. Explicitly \mathbf{X} can be written as,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}.$$

This is actually the view of a data set made of n simultaneous observations of p random variables. The entrance \mathbf{x}_{jk} is defined by $\mathbf{x}_{jk} = [x_{j1}, x_{j2}, \dots, x_{jp}]'$. Let consider now a s -tuple of supports or prototypes $\mathbf{\Lambda} = (\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_s)$ each of them characterizing one of the s defined clusters. Using all what mentioned, a partitioning of \mathbf{X} into s clusters is computed by the minimization of the objective function [5] and [24].

$$J(\mathbf{A}, \mathbf{\Lambda}) = \sum_{i=1}^s \sum_{j=1}^n (A_i(x_j))^m d^2(x_j, L_i)$$

where $\mathbf{A} = \{A_1, A_2, \dots, A_s\}$ being the fuzzy partition. $A_i(x_j) \in [0, 1]$ is the degree of membership of the feature point x_j to the fuzzy cluster A_i . The integer $m > 1$ is the index of fuzziness. $d(x_j, L_i)$ is the distance between the prototype of cluster A_i and the feature point x_j . The mentioned distance may be the Euclidean distance if L_i is defined as elements in \mathbb{R}^p . The clustering algorithm is mainly based on the prototype choice and the defined distance. In this study, the linear prototype denoted $L(\mathbf{u}, \mathbf{v})$ is chosen to characterize the fuzzy set. With the given main direction \mathbf{u} such that $\|\mathbf{u}\| = 1$ and \mathbf{v} is the center of the class. The line from the linear prototype is also called first PC of the set, with its direction given by the association \mathbf{u} of to the largest eigenvalue λ_1 . The covariance matrix computed from the fuzzification procedure is [24] and [25]

$$C_{ij} = \frac{\sum_{k=1}^p A(x_k)^m \cdot (x_{ki} - \bar{x}_i) \cdot (x_{kj} - \bar{x}_j)}{\sum_{k=1}^p A(x_k)^m}, \quad i, j = 1, \dots, n. \quad (5.1.1)$$

Equation (5.1.1) and what preceded, mean that the characterization of the fuzzy set A is done by the linear prototype principal component 1 (PC1) computed based on the fuzzy covariance matrix. At this point, the aim is to determine the membership degree called α such a way to make PC1 fit best to the data \mathbf{X} . The following algorithm is used to determine the fuzzy memberships α .

Algorithm 1: Computation of fuzzy memberships [6], [8], [13], [24], and [26].

Step 1: Set $A(x) = 1, \forall x \in \mathbf{X}$.

Step 2: Compute the prototype $L(u, v)$, with \mathbf{u} being the eigenvector computed from the largest eigenvalue, itself computed from the fuzzy covariance matrix defined by equation (5.1). Compute v the center of the fuzzy cluster A by

$$v = \frac{\sum_{j=1}^n A(x_j)^m \cdot x_j}{\sum_{j=1}^n A(x_j)^m}.$$

Step 3: Compute the new degrees of fuzzy membership $A(x_j)$ as follow

$$A(x_j) = \frac{\frac{\alpha}{1-\alpha}}{\frac{\alpha}{1-\alpha} + \left(d^2(x_j, L)\right)^{\frac{1}{m-1}}}.$$

Step 4: Compare the new fuzzy set to the old fuzzy set. If they are close enough, then stop and accept this as the new fuzzy set, otherwise, go back to step 2 and reiterate the algorithm.

Knowing that the membership degree is a value between 0 and 1, a second algorithm can be defined, using the previous one to determine the best membership degree, computed from the previous algorithm. This membership degree should be such that, it maximizes the first eigenvalue.

Algorithm 2: Computation of the best membership degree α [6], [13], [17] and [24].

Step 1: Set $\alpha_0 = 0$ and $\lambda_0 = 0$.

Step 2: Set the first membership value to be used ($\alpha = \text{initial value}$).

Step 3: Using Algorithm 1 with initial α value and compute the optimal degree of fuzzy membership $A(x)$.

Step 4: Use α and $A(x)$ computed in step 3, to compute the fuzzy covariance matrix C as given in equation (5.1.1), then choose its largest eigenvector.

Step 5: If $\lambda > \lambda_0$ then $\alpha_0 = \alpha$ and $\lambda_0 = \lambda$.

Step 6: Compute $\alpha = \alpha + \text{increment}$. If $\alpha < 1$ then reiterate the procedure from Step 3; else stop the procedure and return the current α_0 as optimal α .

After the definition of the algorithms “Computation of fuzzy memberships” and “Computation of the best membership degree α ” given by Algorithm 1 and Algorithm 2 respectively, one can write a final algorithm that helps to compute the fuzzy principal component analysis as follow.

Algorithm 3: Fuzzy principal component analysis [6], [13], [17], [24] and [26].

Step 1: Compute the optimal α using Algorithm 2.

Step 2: Using Algorithm 1 to compute the optimal fuzzy membership degrees using the α computed in step 1.

Step 3: Use the result (fuzzy membership degrees) computed in Step 2, to compute the fuzzy covariance matrix \mathbf{C} .

The classical procedure of computing PCs through eigenvalues and eigenvectors can now be carried out on the obtained \mathbf{C} matrix to find the fuzzy principal components.

5.1.2 Application of PCA Concepts to Fuzzy Data.

Example 1: Body mass index (BMI). Consider the following data set given in Table 5.1 representing 21 observations of 4 random variables. Each random variable represent the membership value from the BMI data where the fuzzy interval for BMI values is assume to be between 19 and 21, for the categories *thin* and *normal*. Based on the assumption that membership functions are linear, different fuzzy membership functions are used to compute membership values. According to fuzzy interval widths used functions are

$$\begin{aligned}
 \mu_{x_1} &= 0.5x - 9.5 & 19 - 21 \\
 \mu_{x_2} &= 0.25x - 4.5 & 18 - 22 \\
 \mu_{x_3} &= 0.167x - 2.83 & 17 - 23 \\
 \mu_{x_4} &= 0.125x - 2. & 16 - 24
 \end{aligned} \tag{5.1.2}$$

Here, X values used in each equation, are the BMI values between 19 and 21 with increments 0.1. Therefore, from each function in equation (5.1.2), 21 membership values were computed. Computed values form a data matrix of 21×4 . First column in the matrix represents fuzzy membership values computed for μ_{x_1} . Similarly, columns 2, 3 and 4 are computed from respective fuzzy membership functions for μ_{x_2} , μ_{x_3} , and μ_{x_4} (5.1.2). This data given in the table (5.1).

Table 5.1: Fuzzy membership data obtained from BMI values.

X_1	X_2	X_3	X_4
0	0.0125	0.020057	0.01875
0.05	0.0625	0.070058	0.0125
0.1	0.1125	0.120059	0.11875
0.15	0.1625	0.17006	0.16875
0.2	0.2125	0.220061	0.21875
0.25	0.2625	0.270062	0.26875
0.3	0.3125	0.318934	0.31875
0.35	0.3625	0.370064	0.36875
0.4	0.4125	0.420065	0.41875
0.45	0.4625	0.470066	0.46875
0.5	0.5125	0.520067	0.51875
0.55	0.5625	0.570068	0.56875
0.6	0.6125	0.620069	0.61875
0.65	0.6625	0.67007	0.66875
0.7	0.7125	0.720071	0.71875
0.75	0.7625	0.770072	0.76875
0.8	0.8125	0.820073	0.81875
0.85	0.8625	0.870074	0.86875
0.9	0.9125	0.920075	0.91875
0.95	0.9625	0.970076	0.96875
1	1	1	1

The z scores for each variable are computed and checked for normality through the standard normal quintile – quintile plot. Figure (5.1) shows the quintile – quintile plot for X_1 . The fit of the point to a near perfect straight line is an indication of normality.

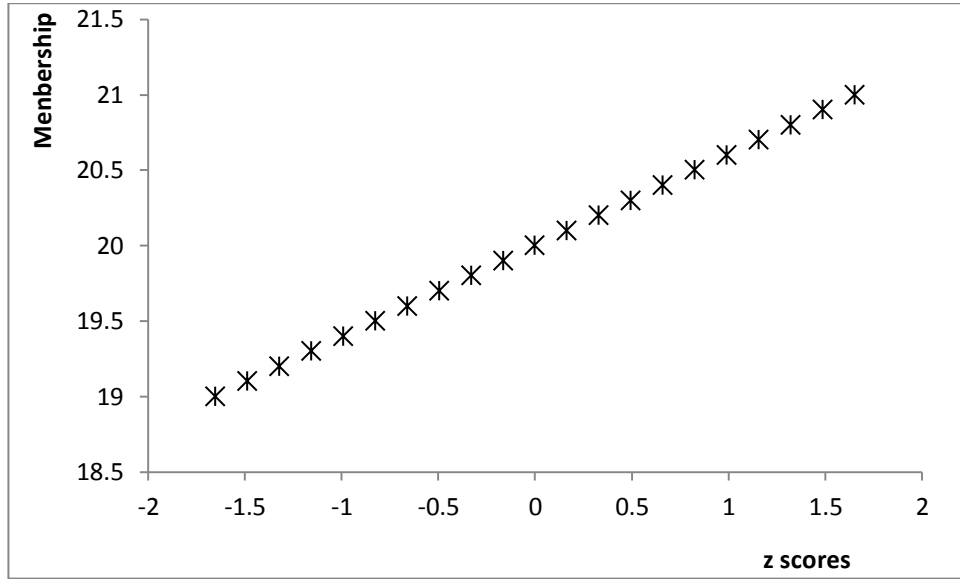


Figure 5.1: Standard normal quintile – quintile plot for X_1 .

The data in Table 5.1 will be used in PCA

Principal component analysis based on the covariance matrix

The covariance matrix computed from the BMI data in Table 5.1, is given below.

$$\mathbf{S} = \begin{pmatrix} 0.0963 & 0.0959 & 0.0958 & 0.0970 \\ 0.0959 & 0.0956 & 0.0955 & 0.0967 \\ 0.0958 & 0.0955 & 0.0953 & 0.0966 \\ 0.0970 & 0.0967 & 0.0966 & 0.0980 \end{pmatrix}$$

Eigenvalues and corresponding eigenvectors are

$$\mathbf{L}_s = \begin{pmatrix} 0.0000 & 0 & 0 & 0 \\ 0 & 0.0000 & 0 & 0 \\ 0 & 0 & 0.0001 & 0 \\ 0 & 0 & 0 & 0.3851 \end{pmatrix}, \mathbf{E}_s = \begin{pmatrix} 0.3073 & -0.7346 & 0.3406 & 0.4999 \\ -0.8093 & 0.1312 & 0.2819 & 0.4983 \\ 0.5006 & 0.6633 & 0.2490 & 0.4974 \\ 0.0013 & -0.0557 & -0.8617 & 0.5043 \end{pmatrix}$$

It is observed from the diagonal \mathbf{L} matrix, $\ell_1 = 0.3851$ is very large compared with the remaining eigenvalues. This indicates that only the first PC is enough to represent the fuzzy membership values, for the BMI process. This is given as

$$Y_1 = 0.4999X_1 + 0.4983X_2 + 0.4974X_3 + 0.5043X_4 .$$

Principal component analysis based on the correlation matrix.

Correlation matrix computed from the data given in Table 5.1 is

$$\mathbf{R} = \begin{pmatrix} 1.0000 & 1.0000 & 0.9999 & 0.9992 \\ 1.0000 & 1.0000 & 1.0000 & 0.9993 \\ 0.9999 & 1.0000 & 1.0000 & 0.9993 \\ 0.9992 & 0.9993 & 0.9993 & 1.0000 \end{pmatrix}.$$

Eigenvalues and corresponding eigenvectors computed from the \mathbf{R} matrix are

$$\mathbf{L}_R = \begin{pmatrix} 0.0000 & 0 & 0 & 0 \\ 0 & 0.0001 & 0 & 0 \\ 0 & 0 & 0.0011 & 0 \\ 0 & 0 & 0 & 3.9988 \end{pmatrix}, \mathbf{E}_R = \begin{pmatrix} 0.3083 & -0.7355 & 0.3376 & 0.5000 \\ -0.8094 & 0.1290 & 0.2796 & 0.5001 \\ 0.4998 & 0.6628 & 0.2468 & 0.5000 \\ 0.0013 & -0.0563 & -0.8643 & 0.4999 \end{pmatrix}$$

Similar to the covariance matrix case, the first eigenvalue is dominating. Therefore, the first PC is enough to represent the fuzzy membership process.

The first PC can be written as

$$Y_1 = 0.5000X_1 + 0.5001X_2 + 0.5000X_3 + 0.4999X_4.$$

First PC obtained from the \mathbf{S} and \mathbf{R} matrices are the same, with all coefficients are either 0.5 or extremely close to 0.5. This is an unusual situation indicating extremely high correlation between the variables, which is evident from the \mathbf{R} matrix. They are all practically equal to 1. Reason behind this being the linear membership functions used in the generation of fuzzy membership values given in Table (5.1). To overcome this handicap, a new data set generated using a different approach, given in Example 2.

Example 2: Referring to example 2 of chapter 3, the main ingredient sugar in determining the sweetness of the cake is divided into 4 categories. An interval

around the boundary values between the categories is defined as a fuzzy interval in transition from one category to another.

Five different fuzzy membership functions are used in determining the membership values in the fuzzy interval. Decreasing membership values in a fuzzy interval obtained from each function are taken as the realizations of a random variable. Hence, an association is established between the membership values and the random variables. This is particularly true for each membership function, as there are infinitely many possible sets of membership values obtainable from each function, depending on its parameters.

The following fuzzy membership functions are used in generating the membership values [8]. For each fuzzy interval as the membership values to one category decrease, for the other category increase. In this example only the decreasing membership values for one fuzzy interval are considered. This is based on the assumption that all fuzzy intervals are of the same width.

1- Linear membership function. $\mu_{X_1}(x; a, b) = aX + b$. The coefficients a and b are computed from the lower and upper boundaries of the fuzzy interval. Obtained membership values are assigned to random variable X_1 .

2- Z shaped membership function. Formally it is defined as

$$\mu_{X_2}(x; a) = \begin{cases} 1 & x \leq a_1 \\ 1 - 2 \left(\frac{x - a_1}{a_2 - a_1} \right)^2 & a_1 < x \leq \frac{a_1 + a_2}{2} \\ 2 \left(\frac{x - a_2}{a_2 - a_1} \right)^2 & \frac{a_1 + a_2}{2} < x \leq a_2 \\ 0 & x > a_2 \end{cases} \quad (5.1.3)$$

Suitable values are assigned to the constants a_1 and a_2 , and membership values are computed. These set of membership values are assigned to the random variable X_2 .

3- Cauchy membership function. This function is given by

$$\mu_{X_3}(x; a, b, c) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad (5.1.4)$$

Parameter b is usually positive. If $b > 0$ curve becomes concave up bell shaped. This function is the generalized version of the Cauchy distribution. Membership values from this function are assigned to X_3 .

4- Gaussian membership function.

$$\mu_{X_4}(x; c, \sigma) = e^{-0.5\left(\frac{x-c}{\sigma}\right)^2} \quad (5.1.5)$$

Gaussian membership function is completely determined by the parameters c and σ . c is the central value and σ is the measure of width for the function.

Membership values computed are assigned to X_4 .

5- Sigmoidal membership function.

$$\mu_{X_5}(x; a, c) = \frac{1}{1 + \exp[-a(x-c)]} \quad (5.1.6)$$

The sigmoidal membership function produces a tail to the right or left, depending on the value of a . X_5 represents the membership values obtained from this function.

Application of these 5 membership functions to the sugar content in a cake example, for the fuzzy interval 80 – 120 grams and the decreasing part of the membership values from 1 to 0, produced the results given in Figure (5.2).

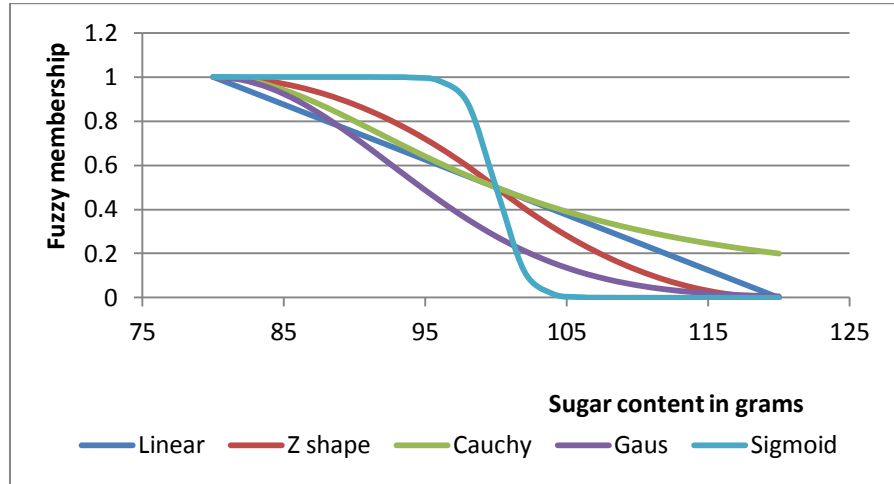


Figure 5.2: Fuzzy membership values obtained from 5 different membership functions

The same data is also given in Table (5.2), and are used in dimension reduction analysis.

Table 5.2: Fuzzy membership values from different membership functions assigned to random variables.

X_1 : Lin	X_2 : Z shape	X_3 : Cauchy	X_4 : Gauss	X_5 : Sigmoid
1	1	1	1	1.00000000
0.95	0.995	0.990099	0.987282	0.99999998
0.9	0.98	0.961538	0.950089	0.99999989
0.85	0.955	0.917431	0.891188	0.99999917
0.8	0.92	0.862069	0.81481	0.99999386
0.75	0.875	0.8	0.726149	0.99995460
0.7	0.82	0.735294	0.630779	0.99966465
0.65	0.755	0.671141	0.534085	0.99752738
0.6	0.68	0.609756	0.440784	0.98201379
0.55	0.595	0.552486	0.354588	0.88079708
0.5	0.5	0.5	0.278037	0.50000000
0.45	0.405	0.452489	0.212503	0.11920292
0.4	0.32	0.409836	0.15831	0.01798621
0.35	0.245	0.371747	0.114957	0.00247262
0.3	0.18	0.337838	0.081366	0.00033535
0.25	0.125	0.307692	0.056135	0.00004540
0.2	0.08	0.280899	0.037749	0.00000614
0.15	0.045	0.257069	0.024743	0.00000083
0.1	0.02	0.235849	0.015809	0.00000011
0.05	0.005	0.21692	0.009845	0.00000002
0	0	0.2	0.005976	0.00000000

Processing the data from Table 5.2 for dimension reduction

Initially the covariance (**S**) and correlation (**R**) matrices from the data are to be computed to help determine the relationship between the variables. These are computed as below

$$\mathbf{S} = \begin{pmatrix} 0.0963 & 0.1169 & 0.0858 & 0.1111 & 0.1367 \\ 0.1169 & 0.1458 & 0.1055 & 0.1374 & 0.1772 \\ 0.0858 & 0.1055 & 0.0785 & 0.1032 & 0.1249 \\ 0.1111 & 0.1374 & 0.1032 & 0.1369 & 0.1641 \\ 0.1367 & 0.1772 & 0.1249 & 0.1641 & 0.2374 \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} 1.0000 & 0.9865 & 0.9872 & 0.9680 & 0.9041 \\ 0.9865 & 1.0000 & 0.9857 & 0.9724 & 0.9523 \\ 0.9872 & 0.9857 & 1.0000 & 0.9955 & 0.9150 \\ 0.9680 & 0.9724 & 0.9955 & 1.0000 & 0.9103 \\ 0.9041 & 0.9523 & 0.9150 & 0.9103 & 1.00003 \end{pmatrix}$$

While correlations between the variables are very high, the lowest one is between the membership values obtained from the linear and sigmoidal membership functions. Eigenvalues (l) and eigenvectors (\mathbf{e}) obtained from **S** and **R** matrices are given below

$$\mathbf{L}_S = \begin{pmatrix} 0.0000 & 0 & 0 & 0 & 0 \\ 0 & 0.0005 & 0 & 0 & 0 \\ 0 & 0 & 0.0045 & 0 & 0 \\ 0 & 0 & 0 & 0.0231 & 0 \\ 0 & 0 & 0 & 0 & 0.6668 \end{pmatrix}; \mathbf{L}_R = \begin{pmatrix} 4.8321 & 0 & 0 & 0 & 0 \\ 0 & 0.1264 & 0 & 0 & 0 \\ 0 & 0 & 0.0372 & 0 & 0 \\ 0 & 0 & 0 & 0.0042 & 0 \\ 0 & 0 & 0 & 0 & 0.0001 \end{pmatrix}$$

Eigenvalues from **S** and **R** matrices both indicate that one PC is enough to represent the whole process. This is because the first eigenvalue obtained from **S** accounts for 95.5% of total variation in the data, and respective value from the **R** matrix is 96.6% from equation 4.4.5. Also the graph on Figure (5.3) shows an elbow at the end of the

first eigenvalue, and after the second eigenvalue tails off, supporting the sufficiency of one PC to represent the whole process.

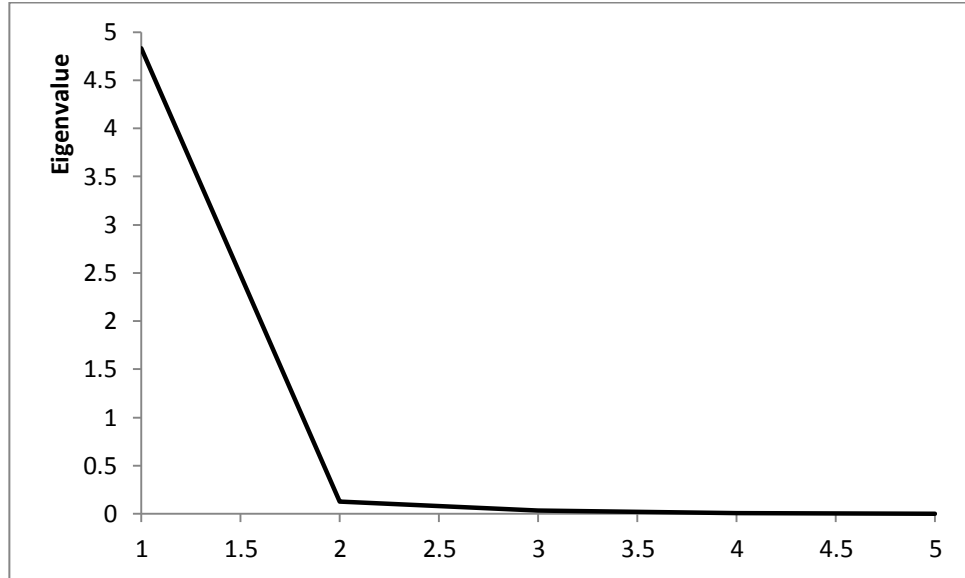


Figure 5.3: Eigenvalues of example 2 forming and elbow

$$\mathbf{E}_S = \begin{pmatrix} 0.2409 & -0.5994 & 0.5717 & 0.3436 & 0.3712 \\ 0.0738 & 0.7728 & 0.4103 & 0.1108 & 0.4655 \\ -0.8758 & -0.0938 & -0.1316 & 0.3042 & 0.3382 \\ 0.4114 & 0.0048 & -0.6925 & 0.3930 & 0.4436 \\ -0.0174 & -0.1862 & -0.0891 & -0.7891 & 0.5783 \end{pmatrix}$$

$$\mathbf{E}_R = \begin{pmatrix} -0.4487 & -0.3060 & 0.6219 & 0.5062 & -0.2491 \\ -0.4533 & 0.0538 & 0.3299 & -0.8209 & -0.0942 \\ -0.4521 & -0.2924 & -0.1908 & 0.0598 & 0.8186 \\ -0.4487 & -0.2829 & -0.6785 & 0.0149 & -0.5081 \\ -0.4330 & 0.8591 & -0.0875 & 0.2570 & 0.0285 \end{pmatrix}$$

Coefficients of the PC obtained from the covariance matrix are the values of the eigenvector that corresponds to the largest eigenvalue of the covariance matrix. Then the PC is

$$Y_1 = 0.3712X_1 + 0.4655X_2 + 0.3382X_3 + 0.4436X_4 + 0.5783X_5.$$

PC obtained from the covariance matrix shows that X_5 has the highest and X_3 the lowest contribution to the linear combination. Similarly the PC from the correlation matrix is

$$Y_1 = -0.4487X_1 - 0.4533X_2 - 0.4521X_3 - 0.4487X_4 - 0.4330X_5.$$

Coefficients of this PC are almost equal to each other. Hence, carry no useful information about the contribution of the variables to the PC. This is because the correlation values between variables are very high. Never the less compared with example 1, they do not exhibit perfectness.

Table 5.3: Correlation coefficient values between the PCs and random variables, obtained from the covariance and correlation matrices.

		X_1	X_2	X_3	X_4	X_5
S matrix	ρ_{Y_1, X_i}	0.9767	0.9955	0.9857	0.9790	0.9692
		Z_1	Z_2	Z_3	Z_4	Z_5
R matrix	ρ_{Y_1, Z_i}	-0.9863	-0.9964	-0.9938	-0.9863	-0.9518

The correlation between the variables and the PC is not directly related with the level of contribution. For example in the PC obtained from the **S** matrix the highest contributing variable X_5 , turns out to have the lowest correlation with the PC.

On the other hand, the correlation between the standardized variables Z_i and Y_1 are again almost all the same, confirming the previous comments.

Chapter 6

CONCLUSION

Fuzzy set theory provides a means to extract useful information where large volumes of data is concerned or generated. Areas where data is generated in vast volumes include business and banking, engineering, health science. Other sectors are rapidly growing in terms of data generation. There are situations where large data has to be divided into categories. Boundaries of such classifications are sometimes difficult to delineate as to whether a data value is a member of one class or the other. Fuzzy logic together with mathematical concepts provides a solution to the problem by introducing the *fuzzy interval* and *fuzzy membership* concepts. With the use of a *fuzzy membership function* a data value can be assigned with partial membership to neighboring classes, within the fuzzy interval between these classes $\mu_{x_1}(a)$ and $\mu_{x_2}(a)$. The fuzzy membership values assigned to a point must add up to 1, i.e. $\mu_{x_1}(a) + \mu_{x_2}(a) = 1$.

In this thesis work, this very basic concept is built on to generate different fuzzy membership multivariate data sets as explained under chapters 3 and 5 with examples.

Generated multivariate data sets are then processed through the PCA concepts to assess the possibility of dimension reduction. PCA is a methodology that examines

the multivariate data via the covariance and/or the correlation matrices obtained from the data.

Example 1 under section 5.1.2, 4 sets of fuzzy membership data values were generated assuming linear membership functions. Each function is obtained from a different fuzzy membership interval. Using the same type of membership function resulted in extremely high correlation between the data sets, each represented by a different variable. As expected, the PCA could not distinguish between the variables as it is observed from equal coefficients being applied to every variable in the PC. Further, it is seen that almost all variation in the data was attributed to the first PC, rendering the remaining 3 PCs useless. Hence, becomes clear that application of PCA to such data is not appropriate.

In example 2 of section 5.2.1, 5 different fuzzy membership functions were employed using a fixed fuzzy interval. Generated data sets are assigned to 5 variables. High correlation between the variables still evident, but not to the extent to hamper the application of PCA concepts. The first PC was sufficient to represent the process generated by 5 variables. However, the coefficients attached to each variable showing its contribution to the PC were different as expected.

REFERENCES

- [1] Rencher, A. C. (2002). *Methods of Multivariate Analysis*. Second Edition. John Wiley & Sons.
- [2] Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Sixth Edition. Pearson, Prentice Hall.
- [3] Chen, G., & Trung Tat Pham, T. T. (2001). *Introduction to fuzzy sets, fuzzy logic, and fuzzy control system*. ISBN 0-8493-1658-8.
- [4] Bouhouche, S., Lahreche, M., Moussaoui, A., & Bast, J. (2007). Quality Monitoring Using Principal Component Analysis and Fuzzy Logic Application in Continuous Casting Process. *American Journal of Applied Sciences*, 4 (9): 637-644
- [5] Dumitrescu, D., Sarbu, C., & Pop, H. F. (1994). *A Fuzzy Divisive Hierarchical Clustering Algorithm for the Optimal Choice of Sets of Solvent Systems*, Anal. Lett. 24, 1031-1054.
- [6] e-book. Intelligent control techniques in mechatronics. http://www.ro.feri.uni-mb.si/predmeti/int_reg/Predavanja/Eng/4.Fuzzy%20logic/_03.html.
- [7] Dadone, P. (2001). *Design Optimization of Fuzzy Logic Systems*. PhD Thesis, Blacksburg: Virginia Polytechnic Institute and State University.

- [8] Gil-Lafuente, A. M. (2005). *Fuzzy Logic in Financial Analysis*. Berlin Heidelberg: Springer.
- [9] Leondes, C. T. (1998). *Fuzzy Logic and Expert Systems Applications*. San Diego.
- [10] Zadah, L. (1978). *Fuzzy sets as a basis of possibility*. Fuzzy Sets Systems. Volume 1, pp 3 – 28.
- [11] Ross, T. J. (1995). *Fuzzy Logic with Engineering Applications*. McGraw-Hill.
- [12] Passino, K. M., & Yurkovich (1998). *Fuzzy Control*. Addison Wesley.
- [13] Ilich, M. S., & Jain, L. C. (2006). *Innovations in Fuzzy Clustering*. Theory and Applications. Springer. ISBN-10 3-540-34356-3.
- [14] Tzeng, G. H., & Huang, J. H. (2014). *Fuzzy Multiple Objective Decision Making*. Taylor & Francis Group, CRC Press.
- [15] Deb, K. (2001). *Multi Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons.
- [16] Zadeh, L. A. (1968). fuzzy sets. *information and control* , 338-353.
- [17] Sato-Ilic, M., & Jain, L. C. (2006). *Innovations in Fuzzy Clustering Theory and Applications*. Berlin : Springer.

- [18] Jolliffe, I. (2002). *Principal Component Analysis*. Second Edition. New York: Springer.
- [19] Neil, H. T. (2002). *Applied Multivariate Analysis*. New York: Springer.
- [20] Stephen, H. F., Arnold, J. I. & Lawrence, E. S. (2002). *Linear Algebra*. Upper Saddle River: Pearson.
- [21] Hardle, W. K., & Simar, L. (2003, 2007, 2012). *Applied Multivariate Statistical Analysis*. Berlin Heidelberg: Springer.
- [22] Ronald, E. W., Raymond, H. M., Sharon, L. M., & Keying, Y. (2012). *Probability & Statistics for Engineers & Scientists*. Boston: Pearson.
- [23] Yoshio, T. (2014). *Constrained Principal Component Analysis and Related Techniques*. London: CRC Press.
- [24] Pop, H. F. (2001). *Principal Components Analysis Based on a Fuzzy Sets Approach*. Studia Univ. Babeş-Bolyai, Informatica, Volume XLVI, No. 2.
- [25] Chen, F. -C., Tzeng, Y.-F., Hsu, M.-H., & Chen, W. R. (2010). Combining Taguchi Method, Principal Component Analysis and Fuzzy Logic to the Tolerance Design of a Dual-purpose Six-bar Mechanism. *Transactions of the Canadian Society for Mechanical Engineering*, Vol. 34, No. 2.

- [26] Vidal, R., Ma, Y., & Sastry, S. S. (2006). *Generalized Principal Component Analysis Modeling & Segmentation of Multivariate Mixed Data*. California.
- [27] Girshich, M. A. (1939). On the Sampling Theory of Determinantal Equations. *Annals of Mathematical Statistics* 10; 203-224.
- [28] Jeffers, J. N. R. (1967). Two case studies in the Application of Principal Component Analysis. *Applied Statistics*, 16, 225-236.