

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

Departamento Académico de Estadística e Informática



Trabajo Monográfico

**DESCRPCIÓN DE LA METODOLOGÍA DE ANALISIS DE CLUSTER
CON ALGORITMO FUZZY C-MEANS**

Presentado para optar el título de Ingeniero Estadístico e Informático

GIOVANNA FLORES BELLIDO

Modalidad: de Examen Profesional

LIMA - PERÚ

2017

DEDICATORIA

A mis padres con mucho cariño,
por su enseñanza de valorar el esfuerzo
que se realiza para lograr cumplir nuestras metas.

AGRADECIMIENTO

Agradezco a Dios por haberme guiado en momentos difíciles, a mi familia por su comprensión y paciencia a pesar de las adversidades, a mi asesora por la comprensión y el tiempo dedicado en este tiempo para el logro del presente trabajo monográfico y a mis compañeros por el apoyo brindado.

ÍNDICE DE TABLAS

Tabla1: Matriz de Reglas	10
Tabla2 Grado de pertenencia	11
Tabla3 Datos de Resistencia y Rapidez.....	19
Tabla4 Grados de pertenencia a cada cluster	20
Tabla5 Prototipos Finales	21
Tabla6 Asignación de cluster	22

ÍNDICE DE FIGURAS

Figura 1 Matriz de Datos	4
Figura 2 Partición Dura	7
Figura 3 Partición Fuzzy	8
Figura 4 Funciones de pertenencia para las variables	9
Figura 5 Reglas posibles para el sistema de riego	10
Figura 6 Salidas Fuzzy	12
Figura 7 Cluster FCM obtenido	20
Figura8. Anexo Formación de Clueste Kmeans y FCM	37

INDICE

RESUMEN

INTRODUCCIÓN.....	1
CAPÍTULO I. ANÁLISIS DE CLUSTER.....	3
1.1 Métodos numéricos de clasificación.....	3
1.2 Conjunto de datos	4
1.3 La matriz de patrones.....	5
1.1 Clasificación o agrupamiento	5
1.2 Aprendizaje computacional	6
CAPÍTULO II. MODELO FUZZY.....	7
2.1 Lógica fuzzy	7
2.2 Modelado fuzzy	8
CAPÍTULO III.METODOLOGÍA DEL ANÁLISIS DE CLUSTER FUZZY C-MEANS	13
3.1 Fuzzy clustering.....	13
3.2 Fuzzy C-Means (FCM).....	13
3.3 Algoritmo FCM:	16
3.4 Validez del cluster	17
CAPÍTULO IV. CASO DE APLICACIÓN	19
4.1 Aplicación del algoritmo Fuzzy C-Means	19
CAPÍTULO V. CONCLUSIONES	24
CAPÍTULO VI. RECOMENDACIONES.....	25
BIBLIOGRAFÍA.....	26
ANEXO.....	27

RESUMEN

En el presente trabajo se presenta la metodología del algoritmo Fuzzy C-means para el análisis de cluster, el cual fue presentado por Bezdek y Dunn en 1973, la cual combina los métodos basados en la función objetivo con los de la lógica Fuzzy término presentado por Lofty Zadeh en 1960 como medio para modelar la incertidumbre a través de las etapas de fuzzificación, reglas de evaluación y defuzzificación. El algoritmo Fuzzy C-means realiza la formación de cluster a través de una partición suave de los datos, es decir para realizar del reconocimiento de patrones a través del hallazgo de los grados de pertenencia de cada individuo a los diferentes cluster, donde un individuo no tendría pertenecía exclusiva a un solo grupo, sino que un individuo podría tener grados de pertenencia a distintos grupos, a diferencia de otros métodos que realizan la formación de los cluster basados en la lógica binaria o partición dura. Utilizando el software estadístico R se realizó la aplicación del algoritmo Fuzzy C-means sobre datos de jugadores para la formación de cluster a través de rapidez y resistencia.

Key Words: Fuzzy C-means, grado de pertencia, fuzzificación, defuzzificación,

INTRODUCCIÓN

El análisis de clúster es un conjunto de técnicas multivariantes utilizadas para clasificar a un conjunto de individuos en grupos homogéneos, siendo estos grupos desconocidos a priori, estas técnicas multivariantes pertenecen a un tipo de aprendizaje no supervisado puesto que no requieren un aprendizaje previo a partir de los datos.

Las técnicas de clustering son utilizados para organizar los datos en grupos, basados en las similitudes entre los elementos de datos individuales sin conocer la clase perteneciente del individuo, tiene como objetivo modelar la estructura subyacente o la distribución en los datos para aprender más sobre los mismos, ayuda a la obtención de información interesante sobre la estructura de los datos así forma bloques esenciales para futuros modelos.

Una de las técnicas de clustering es el algoritmo Fuzzy C-means, dentro del marco del clustering tiene la capacidad para cuantificar un grado de pertenencia de elementos a grupos detectados que son esenciales cuando lo relacionan con el fenómeno inherente de pertenencia parcial al grupo, esta técnica viene siendo muy utilizada en muchas investigaciones recientemente como son el procesamiento de imágenes, análisis de sistemas de energía, el reconocimiento de patrones. Este algoritmo ha sido reconocido como una poderosa herramienta en el tratamiento de sistemas físicos no tratables debido a la capacidad de integrar información en diferentes fuentes. Los conjuntos Fuzzy sirven como interfaz entre la variable cualitativa, variables cuantitativas, las entradas y salidas de sistemas.

En el presente trabajo monográfico tiene por finalidad el desarrollo de la técnica del algoritmo Fuzzy C-means para la formación de cluster a través de la partición suave de los datos calculando los grados de pertenecía a los cluster o grupos formados, de este modo poder utilizar método para la realización de segmentación en el ámbito laboral.

En el Capítulo 1 se describe el análisis de cluster, la forma del conjunto de datos que se toma para realizar dicho análisis los tipos de datos, para la formación de una matriz de patrones así como la formación de un tipo de aprendizaje a través de los métodos de agrupamiento.

En el capítulo 2 se revisó la lógica Fuzzy, el funcionamiento del sistema Fuzzy así como sus etapas en la que se detalla tres etapas la fuzzydificación, evaluación de reglas y la etapa de defuzzyfificación. En el capítulo 3 se describe el algoritmo Fuzzy C-means.

En el capítulo 4 se realizó la aplicación del método de clustering Fuzzy C-means sobre los datos de atributos de futbolistas en cuanto a rapidez y resistencia para la formación de cluster a partir de estos atributos y agrupar a los futbolistas, se utilizó el software R, para la formación de 3 grupos a través de los grados de pertenencia de cada individuo a cada cluster. Finalmente se obtuvo tres segmentos para los futbolistas: futbolistas de ataque en los cuales se agruparon en este segmento los futbolistas que presentaban rapidez pero poca resistencia, los futbolistas de medio campo en los que se encontrarías los futbolistas rápidos y resistente a la vez, y el último segmento de los futbolistas de defensa conformados por los futbolistas más resistente pero poco rápidos.

CAPÍTULO I. ANÁLISIS DE CLUSTER

1.1 Métodos numéricos de clasificación

Las técnicas numéricas son utilizadas para derivar las clasificaciones originadas largamente en las ciencias naturales como la biología o zoología en un esfuerzo para eliminar taxonomía de esta tradicionalmente naturaleza subjetiva, cuyo objetivo fue proveer estables y objetivas clasificaciones. El objetivo en este sentido es que el análisis de un mismo conjunto de organismos por la misma secuencia de métodos numéricos produce la misma clasificación (Brian S. Everitt *et al.* 2011).

Una clasificación puede representar simplemente un método conveniente para organizar un gran conjunto de datos, comprenderlo fácilmente y obtener información más eficiente. Si los datos pueden ser resumidos por grupos pequeños de objetos, entonces el grupo de clases puede brindar una descripción muy concisa de patrones de similitud y diferencias en los datos (Brian S. Everitt *et al.* 2011).

A estos métodos numéricos aplicados, se les llama de muchos nombres, de acuerdo al área de aplicación Taxonomía numérica es usada en biología, en psicología *Q análisis*, en inteligencia artificial *Reconocimiento de patrones no supervisado*, en marketing siempre se habla sobre la segmentación entre ellos análisis de cluster el cual busca descubrir grupos en los datos (Brian S. Everitt *et al.* 2011).

En la mayoría de aplicaciones de análisis de cluster una partición de los datos es buscado, en dicha partición cada objeto pertenece a un simple cluster y el conjunto completo de cluster contiene a todos objetos individualmente, adicionalmente es una herramienta para realizar la exploración de datos y debe ser suplementado por técnicas de visualización de datos. La visualización más directa es un Figura de dos dimensiones que muestren los objetos a ser agrupados como puntos. Los datos multivariados nosiempre pueden ser reproducidos en dos dimensiones entonces es de mucha ayuda revisar un algoritmo de clasificación (Brian S. Everitt *et al.* 2011)

1.2 Conjunto de datos

Un conjunto de datos multivariados se define como la matriz de datos X que contienen los valores de las variables que describen a cada objeto a ser agrupado.

Figura1. Matriz de Datos

$$X = \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \dots \\ \vec{x}_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,p} \\ x_{2,1} & x_{2,2} & \dots & x_{2,p} \\ \dots & \dots & \dots & \dots \\ x_{n,1} & x_{n,2} & \dots & x_{n,p} \end{pmatrix}$$

Fuente: Tomado de "Tutorial sobre Fuzzy-c-Means e Fuzzy Learning Vector Quantization Volumen 19 (Thiago Rocha" Sarajane Marquez Péres 2012),

Donde x_j es un vector de p coordenadas y n el número de elementos del conjunto de datos, cada vector representa una observación de ese conjunto y cada coordenada de ese vector representa un atributo de esa observación.

Un conjunto de datos X reside en un espacio R^p , este espacio está referenciado para los algoritmos de análisis de datos como *espacio de datos*, *espacio de entrada* o *espacio vectorial*. Generalmente esos datos presentan un tipo de organización que puede ser explicada por medio de particiones de conjunto de datos o estudio de esas particiones mediante un análisis de datos. Las particiones pueden ser caracterizadas como clases o subgrupos de datos comúnmente llamados clusters (Tutorial sobre Fuzzy-c-Means e Fuzzy Learning Vector).

Las variables en la matriz X puede ser una mixtura de valores continuos, ordinales y/o categóricos y algunas veces contener valores perdidos, esta mixtura puede complicar la agrupación de objetos y en algunas aplicaciones las filas de la matriz puede contener medidas repetidas de la misma variable pero bajo diferentes condiciones o diferentes tiempos o en un número de posiciones espaciales.

1.3 La matriz de patrones

Si cada objeto en un conjunto de n objetos puede ser representado por un conjunto de d dimensiones (atributos o score) entonces cada objeto es representado por un patrón o un vector de d lugares, el conjunto mismo es visto como una matriz de patrones, cada fila de esta matriz describe un patrón y cada columna denota una información o medida, el conjunto de los valores de esta información para una señal es un patrón, nosotros requerimos que esa información sea medida para todos los patrones. (Anil K.Jain, *et al.* 1988).

Las d informaciones son usualmente graficadas como un conjunto ortogonal de x , los n patrones son entonces puntos encerrados en un espacio de d dimensiones llamado espacio de patrones. (Anil K.Jain, *et al.* 1988).

Un cluster puede ser visualizado como una colección de patrones los cuales son cercanos el uno al otro. Al mencionar a un algoritmo de agrupamiento es para identificar como es su agrupamiento natural en espacios de muchas dimensiones, nuestra percepción visual es limitada a tres dimensiones, el real beneficio del análisis de cluster es la organización de datos multidimensionales donde la percepción visual falla (Anil K.Jain, *et al.* 1988).

1.4 Clasificación o agrupamiento

(Tomado de Tutorial sobre Fuzzy-c-Means e Fuzzy Learning Vector). El termino de clase debe ser usado cuando existe información sobre cuantas y cuáles son las particiones presentes en el subconjunto de datos, ver que datos pertenece a que partición. Sea un conjunto de vectores de datos de entrada \vec{X} pertenece a \mathbb{R}^p , para un conjunto C finito de rótulos donde el término de C es modelado en términos de una función:

$$F : \mathbb{R}^p \times W \rightarrow C$$

Donde W es un espacio de parámetros ajustados por medio de un algoritmo de aprendizaje supervisado.

El termino grupo debe ser usado cuando no existe cualquier información sobre cómo está organizado los datos, en nuestro caso el agrupamiento tiene por objetivo estudiar las relaciones, de similaridad entre las observaciones, determinando las observaciones que determinarán el grupo. Los grupos son formados de manera que maximizan la similaridad entre los elementos de un grupo (similaridad intra grupos) y minimizar la similaridad entre

grupos diferentes (similaridades entre grupos). Formalmente dado un conjunto de entrada \vec{X} pertenece R^p , se encuentra una función

$$g : R^p \times W \rightarrow C$$

Donde W es un vector de parámetros ajustados por medio de un algoritmo de aprendizaje supervisado o no supervisado que determina c -clases o grupos en X

$$C = C_1, \dots, C_c (c \leq n)$$

Tal que:

1. $C_i \neq \phi, i = 1, \dots, c;$
2. $\bigcup_{i=1}^c C_i = X$
3. $C_i \cap C_j = \phi, i, j = 1, \dots, c$

Asumiendo una clasificación o agrupamiento clásico. ($i \neq j$) (Tutorial sobre Fuzzy-c-Means e Fuzzy Learning Vector).

1.5 Aprendizaje computacional

En las teorías de la inteligencia artificial existen diferentes tipos de aprendizaje computacional que pueden ser implementados por diferentes técnicas que en nuestro interés será desarrollada la modalidad no supervisada, implementados por algoritmos de optimización (*c-mean, fuzzy c-means*).

Este aprendizaje intuitivo, patrones y reglas son determinados a partir de un conjunto de datos, en la modalidad no supervisada los parámetros de una función son ajustadas con la base de una maximización de medidas de calidad aplicadas a los resultados obtenidas con el uso de tal función.

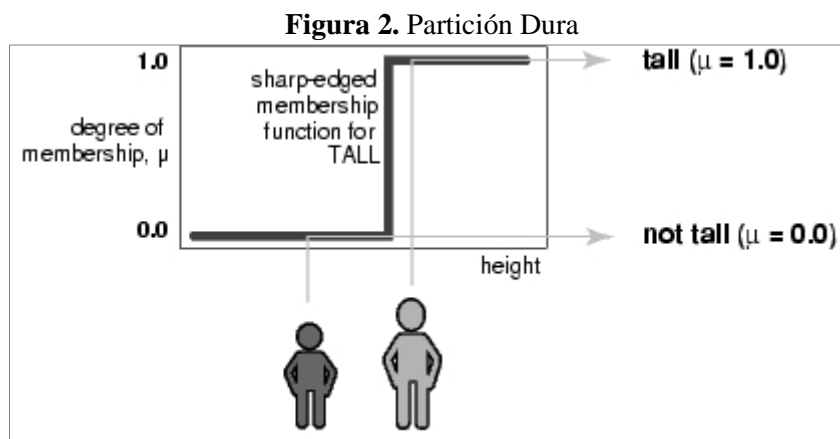
CAPÍTULO II. MODELO FUZZY

2.1 Lógica fuzzy

El termino Lógica Fuzzy fue presentada por Lofty Zadeh en los Años 60, como un medio para modelar la incertidumbre del lenguaje natural, Zadeh dice “no se debe considerar la teoría Fuzzy como un simple teoría, sino que se debería considerar el proceso de Fuzzyficación como una metodología para generalizar cualquier teoría desde su versión ordinaria o discreta a una nueva versión continua o Fuzzy”.

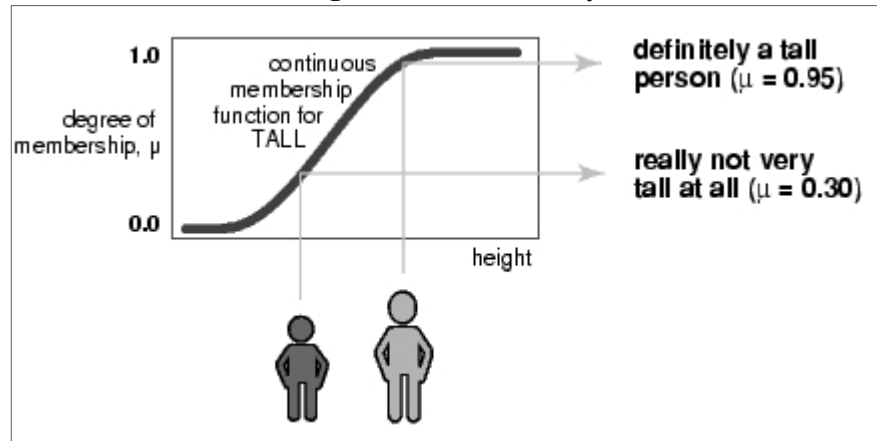
La lógica booleana es conocida como la más precisa, a pesar de sus ventajas de exactitud, esta lógica booleana presenta una desventaja, que es el no poder reproducir los patrones del pensamiento humano, mientras que la lógica Fuzzy está diseñada para reaccionar a cambios continuos, de la variable a ser controlada y no está restringida a dos únicos valores de 0 y 1 (J. C. Bezdek (1981).

Por ejemplo la siguiente Figura 1 describe la pertenencia al conjunto de los altos presenta una función nítida que funciona muy bien con operaciones binarias, pero no funciona bien describiendo el mundo real. Por el contrario en la gráfica 2, muestra una aproximación de los conjuntos Fuzzy que estaría definido por medio de una función continua, que muestran una mejor representación de la característica ser alto de una persona.



Fuente: Tomado de Presentación Lógica Fuzzy Aplicada a Sistemas de Diagnostico (Eletronorte, Isabela Monico, Brasil 2014)

Figura 3 Partición Fuzzy



Fuente: Tomado de Presentación Lógica Fuzzy Aplicada a Sistemas de Diagnostico (Eletronorte, Isabela Monico, Brasil 2014)

La Función de pertenencia define el conjunto difuso para los posibles valores de altura (eje horizontal) y en el eje vertical los grados de pertenencia entre 0 y 1. De tal manera que el grado de pertenencia de la primera persona es 0.3 y por tanto no es muy alto, mientras que el segundo tienen un grado de pertenencia 0.95 definitivamente alto.

Un conjunto Fuzzy permite a sus elementos tener un grado de pertenencia, si el valor 1 es asignado a los elementos que están completamente en el conjunto y 0 a los que están completamente fuera, entonces los objetos que están parcialmente en el conjunto tendrán un valor de pertenencia estrictamente entre 0 y 1.

2.2 Modelado fuzzy

El proceso de este modelo se puede presentar en las siguientes etapas:

Etapas 1: Fuzzificación

Es la etapa donde se calcula el grado de pertenencia que podría tener una entrada abrupta a una o varias funciones de pertenencia de una variable el resultado se denomina entrada Fuzzy.

Primero se debe determinar las funciones de pertenencia de la variable, luego comparar las entradas abruptas con la función de pertenencia correspondiente para producir los valores de entrada Fuzzy.

Ejemplo:

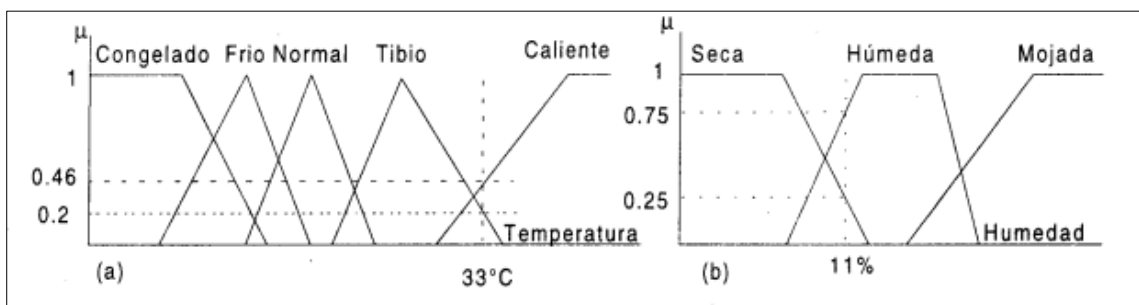
En un sistema de riego donde el control Fuzzy tendría por objetivo de controlar el tiempo de riego dependiendo de la temperatura del aire y de la humedad de suelo.

Definimos las etiquetas de las variables:

- Temperatura : Congelado, Frio, Normal, Tibio, Caliente
- Humedad del suelo: Seca, Humeda, Mojada

Seguido definimos para cada una de las etiquetas las funciones de pertenencia correspondientes (Ver Figura 3)

Figura 4. Funciones de pertenencia para las variables



Fuente: Tomado de Tutorial de Logica Fuzzy (Facultad de Electronica de la UNMSM 2000 Ing.Mg Guillermo Tejada Muños)

Las entradas Fuzzy para la temperatura de 33C y 11% de humedad en el Figura 3 serán:

- La temperatura es tibia con un grado de verdad de 0.2
- La temperatura es caliente con grado de verdad de 0.46
- La Humedad es Seca con un grado de verdad de 0.25.
- La Humedad del suelo es húmeda con un grado de verdad de 0.75

Etapa 2: Reglas de Evaluación

Esta etapa es la etapa de inferencia Fuzzy, el controlador Fuzzy usa reglas lingüísticas sobre los resultados que fueron generados en la etapa de Fuzzyficación Las reglas Fuzzy son sentencias SI-THEN que describen la acción a ser tomada en respuesta a varias entradas Fuzzy.

“Si Antecedente 1 y Antecedente 2... Entonces Consecuente1 y consecuente 2...”

Formato del antecedente:

Variable de entrada =Etiqueta [La temperatura (variable) es caliente (etiqueta)]

Formato del consecuente:

Variable de salida =Etiqueta [la duración del regadío(variable) es prolongada(etiqueta)]

Las reglas siguen el sentido común del comportamiento del sistema y son escritos en términos de las etiquetas de las funciones de pertenencia.

Para el ejemplo anterior las reglas pueden ser representadas en la matriz siguiente

Figura 5. Reglas posibles para el sistema de riego

A N T E C E D E N T E 2	H U M E D A D	ANTECEDENTE 1					
		TEMPERATURA					
			Congelado	Frío	Normal	Tibio	Caliente
		Mojada	Corto	Corto	Corto	Corto	Corto
		Húmeda	Corto	Medio	Medio	Medio	Medio
		Seca	Prolongado	Prolongado	Prolongado	Prolongado	Prolongado

Fuente: Tomado de Tutorial de Logica Fuzzy (Facultad de Electronica de la UNMSM 2000 Ing.Mg Gullermo Tejada Muños)

La temperatura y la humedad son las variables de entrada el tiempo de regadío es la variable de salida. Se observa que son 15 las posibles reglas para el sistema de riego.

Siguiendo el ejemplo anterior resultaron de la etapa de Fuzzyficación cuatro entradas Fuzzy entonces para esa 4 entradas solo será posible cuatro reglas dela matriz.

Tabla1. Matriz de Reglas

REGLA1	IF la temperatura es caliente (0.46) AND la humedad es seca (0.25) THEN la duración es prolongada.	0.25
REGLA2	IF la temperatura es tibia(0.2) AND la humedad es húmeda (0.75) THEN la duración es media	0.20
REGLA3	IF la temperatura es tibia (0.2) AND la humedad es seca THEN la duración es prolongada(0.25)	0.20
REGLA4	IF la temperatura es caliente (0.46) AND la humedad es húmeda THEN la duración es media (0.75).	0.46

*Fuente: Tomado de Tutorial de Logica Fuzzy (Facultad de Electronica de la UNMSM
2000 Ing.Mg Gullermo Tejada Muños)*

Ahora se tiene que determinar las salidas Fuzzy comparando las fuerzas de todas las reglas, a pesar que hay dos reglas que muestran la misma salida la salida Fuzzy será determinada por la máxima fuerza en este caso serán:

- La duración de regadío es prolongada con un grado de verdad de 0.25
- La duración de regadío es media con un grado de verdad de 0.46

El método de la regla utilizada en ese ejemplo es MIN MAX, ya que toma el mínimo de los antecedentes para determinar la fuerza de cada regla y toma la más fuerte para cada consecuente, los cuales determinan la salida Fuzzy.

Etapa 3: Defuzzificación

En esta etapa el objetivo final se basa en encontrar las salidas abruptas, de manera que cada salida Fuzzy generada en la etapa de evaluación modificara su respectiva función de pertenencia de salida. Los métodos más comunes de la defuzzyficación son:

Método de la media ponderada en el cual se calcula el promedio entre los valores de salida que se obtendrían para cada uno de los conjuntos Fuzzy ponderados por el peso de la correspondiente regla o grado de pertenencia al subconjunto.

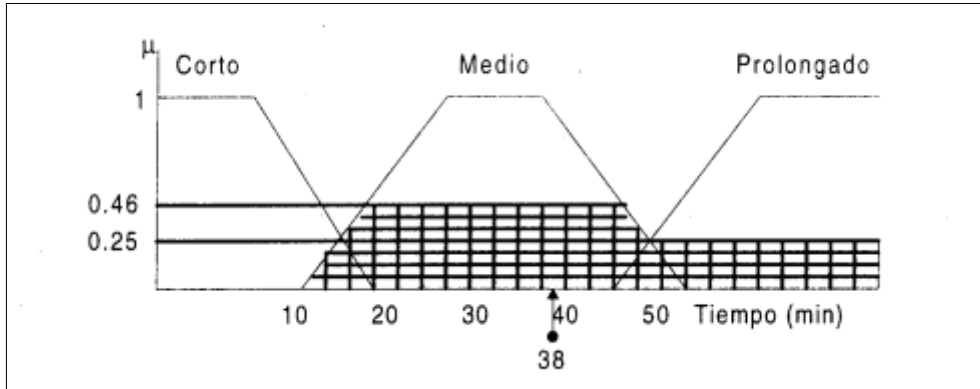
Método de centro de masa (Centro de gravedad o centroide) se determina el centro de gravedad del conjunto de salida, resultante de la unión de las contribuciones de todas las reglas válidas. El valor de la salida es la abscisa del punto de centro de gravedad

Siguiendo con el ejemplo anterior las etiquetas de las funciones de salida hacen referencia a la duración de regadío en el terreno. En este método cada función de pertenencia de salida es cortadas en el nivel indicado por su respectiva salida Fuzzy. Luego las funciones de pertenencia cortadas resultantes son luego combinadas para calcular su centro de gravedad. Las salidas Fuzzy que fueron calculadas en la etapa anterior se muestran a continuación.

Tabla2. Grado de pertenencia

Corto	Medio	Prolongado
0	0.46	0.25

Figura 6. Salidas Fuzzy



Fuente: Tomado de Tutorial de Logica Fuzzy (Facultad de Electronica de la UNMSM 2000 Ing.Mg Gullermo Tejada Muños)

Luego se debe encontrar el punto de balance del centro de gravedad, del área sombreada. Este valor representa la defuzzyficación.

$$Salida\ abrupta = \frac{\sum_i^n y_i \mu(y_i)}{\sum_i^n \mu(y_i)}$$

Finalmente de la formula se obtendrá una salida abrupta de 38 minutos, que equivale al tiempo de duración de regadío para una humedad de 11% y una temperatura de 33C

CAPÍTULO III. METODOLOGÍA DEL ANÁLISIS DE CLUSTER FUZZY C-MEANS

3.1 Fuzzy clustering

En 1973, Bezdek y Dunn presentaron un método de clustering que combinaba los conceptos de los métodos basados en función objetivo con los de la lógica Fuzzy.

De esta manera, un patrón podría tener grados de pertenencia a los distintos subgrupos resultantes en vez de solamente poseer una pertenencia discreta (0 o 1). La literatura en torno a este concepto y sus diversas aplicaciones es extensa. El algoritmo fue introducido por Raspín y más tarde ampliado por Dunn y Bezdek.

Los algoritmos basados en clasificación Fuzzy son ampliamente utilizados en aplicaciones de comprensión, organización y clasificación de información.

En la actualidad se viene realizando investigaciones sobre la base del algoritmo FCM, en especial para la aplicación de reconocimiento de patrones para imágenes dentro del ámbito médico, y geográfico, de la misma manera para minería de datos con Fuzzy Miner y en la BIGDATA adaptando el método Fuzzy a diferentes métodos para obtener resultados robustos.

3.2 Fuzzy C-Means (FCM)

El objetivo de este método es particionar estos n patrones en c cluster, estas particiones se representan a través de una matriz de pertenencia U , donde $[u_{ik}]$ representa el grado de pertenencia del patrón k al i -ésimo cluster y por consiguiente un dato específico puede pertenecer parcialmente a más de un cluster, realizando una partición suave del conjunto de datos.

Sea $X = \{x_1, \dots, x_n\} \subset R^s$ un conjunto de datos (los reales) p -dimensional cada uno, se dice que una partición

$$P = \{c1, c2, \dots, cc\}$$

$$\text{Donde } U = [u_{ik}] \in R^{cn}$$

c es una partición suave de X si y solo si se cumple que

$$u_{ik} \in [0,1] \quad 1 \leq i \leq c; \quad 1 \leq k \leq n \quad (1a)$$

$$\sum_{i=1}^c u_{ik} = 1; \quad 1 \leq k \leq n, \quad (1b)$$

$$\sum_{k=1}^n u_{ik} > 0; \quad 1 \leq i \leq c, \quad (1c)$$

El conjunto de todas las matrices en R^{cn} satisfaciendo (1a) es denotado por M_{fcn} una matriz $U \in M_{fcn}$ puede ser usado para describir la estructura del cluster de X interpretando u_{ik} como el grado de pertenencia de x_k para el cluster i , mientras $u_{ik} = 0.95$ representa una fuerte asociación de x_k para el cluster i un $u_{ik} = 0.01$ representa una muy débil.

- En (1a) vemos como el grado de pertenencia de un objeto k a un cluster i deber estar ente 0 y 1
- En (1b) vemos como la suma de los grados de pertenencia de un objeto k a los distintos grupos ha de ser igual a 1, al cumplir esta condición mas es llamada una partición suave restringida.
- En (1c) vemos como la suma de todos los grados de pertenencia en un cluster tiene que ser mayor a 0 y menor que n es decir no podemos tener grupos vacíos ni un cluster con todos los elementos.

Tomamos a M_{cn} como el subconjunto de M_{fcn} que contiene solo matrices con todos los u_{ik} en $\{0, 1\}$, es exactamente el conjunto de no degenerado crisp (o convencional) c - particiones de X .

Otra información útil acerca de la subestructura de cluster puede ser identificada para prototipos (o centros de cluster) $v = (v_1, \dots, v_c)^T \in R^{cs}$ donde v_i es el prototipo para clase i , $1 \leq i \leq c$, $v_i \in R^s$. "Buenas" particiones U de X y representativos (v_i para i clases) puede ser definido por considerando minimización de una de la familia de c -means objetivos funcionales $J_m : (M_{fcn} \times R^{cs}) \rightarrow R$ definido por

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m \|x_k - v_i\|^2, \quad (2)$$

Donde m es el factor difuso que indica cuanto queremos que se solapen los grupos y tiene que cumplir que $1 \leq m \leq \infty$ y $\|\cdot\|$ es alguna norma inducida sobre R^s . Esta aproximación fue primero dado por $m=2$ en Dun (1973) y entonces generalizado para el rango de valores de m en Bezdek (1973).

Para $m > 1$, Bezdek (1973) otorga el seguimiento necesario de las condiciones para un mínimo (U^*, v^*) de $J_m(U, v)$ sobre $M_{fcn} \times R^{cs}$.

$$(v_i^*) = \frac{\sum_{k=1}^n (u_{ik}^*)^m x_k}{\sum_{k=1}^n (u_{ik}^*)^m} \quad \text{Para todo } \sum_{k=1}^n (u_{ik}^*)^m x_k \quad (3a)$$

Y para cada k dado que $d_{ik}^* = \|x_k - v_i^*\|^2 > 0 \forall i$, entonces

$$u_{ik}^* = \left[\sum_{k=1}^n a_{ijk}^* \right]^{-1} \quad \text{Para todo } i$$

Donde

$$a_{ijk}^* = \left[d_{ik}^* / d_{jk}^* \right]^{1/(m-1)}$$

Pero si k es como que $d_{ik}^* = \|x_k - v_i^*\|^2 = 0$ por algún i , entonces $u_{ik}^* \forall i$ es algún número no negativo que satisface.

$$\sum_{i=1}^c u_{ik}^* = 1 \quad \text{y} \quad u_{ik}^* = 0 \quad \text{si} \quad d_{ik}^* \neq 0. \quad (3b2)$$

El algoritmo de FCM consiste de interacciones alternando entre ecuaciones (3a) y (3b). El proceso inicia con un valor inicial para las particiones \mathbf{U} o un inicial valor para el vector de prototipos \mathbf{v} , y está continuando hasta sucesivas interacciones de la matriz de diferentes particiones; hasta parar con el primer U^{r+1} mientras que $\|U^{r+1} - U^r\| < \varepsilon$, donde ε es un pequeño número positivo.

La convergencia numérica que resulta siguiendo el comportamiento de la secuencia $\{U^r\}$ y $\{v^r\}$, mientras la teoría refiere como mínimo bueno de (2) actualmente representa la subestructura del cluster de una población bajo la asunción de probabilidad estadística.

3.3 Algoritmo FCM:

Considerando a U^r como el valor de la matriz de partición y v^r como el valor de la matriz de prototipos en la iteración r-ésima del método se presenta el siguiente algoritmo:

- 1) Fijar c y m , escoger una norma métrica.
- 2) Inicializar $U^0 \in M_{fc}$ y v^0 aleatoriamente.
- 3) Si $\|U^{r+1} - U^r\| < \varepsilon$ entonces terminar.
- 4) Sino

Actualizar U^r con v^{r-1} y la ecuación 3a

Actualizar v^r con U^r y la ecuación 3a

En el paso 2 se demuestra la convergencia del algoritmo a un mínimo local de la función objetivo.

Algunas definiciones:

Partición suave especial: es aquella partición en la que la suma de los grados de pertenencia de un punto específico es igual a 1.

Parámetro c : El parámetro de c es el número de cluster esperados, es uno de los parámetros del algoritmo FCM difícil de establecer

Parámetro m : El parámetro m indica un grado de fuzziness, es un cociente que controla el grado en que las pertenencias están compartidas entre los cluster Fuzzy en X . Un mayor valor implica una mayor característica Fuzzy, valores típicos [2, 2.5]

La matriz U : Es una matriz que contiene las funciones de pertenencia con sus valores iniciales asignados.

Distancia utilizada $\|\cdot\|$: Norma inducida de un producto interno, es un espacio de p tuplas de números reales (Norma euclidiana, Norma Diagonal, Norma de Mahalanobis).

ε : Umbral de ϵ que opera como criterio de partici3n para los centroides de los cluster, valor t3pico ser3a 0.01.

3.4 Validez del cluster

Las medidas de validez de una partici3n suave restringida son de tres categor3as

a. Medidas basadas en el grado de pertenencia

Estas medidas calculan ciertas propiedades de las funciones de pertenencia en una partici3n suave restringida, una de estas medidas es el coeficiente de partici3n introducido por Jim Bezdek en 1973, este coeficiente mide el grado de fuzzi3n (fuzziness) del cluster. Entre m3s difusos son los clusters, peor es la partici3n. La f3rmula para esta medida es

El coeficiente de la partici3n es igual a uno en caso que la partici3n sea dura y en el caso que la partici3n sea mas difusa posible es igual a $1/c$. por tanto es mas deseable que el coeficiente de la partici3n este lo mas cercano a 1 y de esta manera tener una buena partici3n. Las medidas de validez basadas en el grado de pertenencia como el coeficiente de partici3n pueden ser usadas para comparar particiones alternativas con igual n3mero de cluster pero tienen la limitaci3n de mejorar a medida que se aumenta la cantidad de cluster, tampoco consideran propiedades de la partici3n como la separaci3n entre cluster, por estos motivos se desarrollaron las medidas geom3tricas.

$$V_{pc} = \frac{1}{n} \sum_{i=1}^c \sum_{k=1}^n (u_{ik}^*)$$

b. Medidas basadas en la geometr3a

Entre m3s compactos y separados est3n los cluster mejor es una partici3n, una medida que considera estos espacios fue introducida por X. Xie y G. Beni, su f3rmula es:

$$V_{XB} = \left(\frac{\sum \sigma_1}{n} \right) \frac{1}{d_{\min}^2}$$

$$\sigma_i = \sum_j (u_{ik}^*) d_{ik}$$

c. Medidas basadas en el desempeño

Estas medidas evalúan una partición con base a su desempeño respecto a un objetivo predefinido, por ejemplo el mínimo error en una prueba de clasificación.

CAPÍTULO IV. CASO DE APLICACIÓN

4.1 Aplicación del algoritmo Fuzzy C-Means

(Tomado de Papper Técnicas de lógica difusa aplicada a la minería de datos)

En el artículo mencionado anteriormente se evalúa la rapidez y la resistencia de 11 futbolistas. Un valor cercano a uno significa que el jugador es bastante rápido o resistente según sea el caso, un valor cercano a 0 muestra que el jugador es lento o poco resistente, se escogerá la norma de euclidiana.

Con los datos de la aplicación presentada en el papper antes mencionado, se replicó mediante una muestra con reemplazo de tamaño 100.

Tabla3. Datos de Resistencia y Rapidez

JUGADOR	RAPIDEZ	RESISTENCIA
1	0.58	0.33
2	0.9	0.11
3	0.68	0.17
4	0.11	0.44
5	0.47	0.81
6	0.24	0.83
7	0.09	0.18
8	0.82	0.11
9	0.65	0.5
.		
.		
.		
100	0.98	0.24

Fuente: Elaboración Propia

Se procede a inicializar los parámetros iniciales

$$C = 3, m = 2$$

Se calcula los grados de pertenencia a través de la norma elegida, para el caso se elige la norma euclidiana, con el cual se procede a calcular las distancias de los datos a los prototipos iniciales obteniendo de esta manera la matriz de distancias. Matriz de distancias (Para las primera 11 observaciones)

$$d_{ik}^* = \|x_k - v_i^*\|^2 = 0$$

$$d_{ik} = \begin{bmatrix} 0.1733 & 0.0773 \\ 0.6421 & 0.1621 \\ 0.3393 & 0.1233 \\ 0.0117 & 0.4797 \\ 0.169 & 0.205 \\ 0.1105 & 0.4225 \\ 0.1145 & 0.6065 \\ 0.5365 & 0.1525 \\ 0.2025 & 0.0225 \\ 0.029 & 0.521 \\ 0.676 & 0.1 \end{bmatrix}$$

Con la matriz de distancias obtenía se procede a calcular los grados de pertenencia de cada individuo a cada cluster.

Tabla4. Grados de pertenencia a cada cluster

Individuo	Cluster 1	Cluster 2	Cluster 3
1	0.31639851	0.11524808	0.5683534
2	0.08122006	0.16588419	0.75289574
3	0.07630475	0.8739776	0.04971765
4	0.08017652	0.05100601	0.86881747
5	0.91484839	0.04917449	0.03597713
6	0.12540004	0.61562905	0.25897091
7	0.12709324	0.32550715	0.54739961
8	0.43970873	0.13154441	0.42874686
9	0.45402672	0.12980208	0.41617121
10	0.10096147	0.84188648	0.05715205
11	0.17981887	0.27031692	0.54986421

Fuente: Elaboración propia

En la tabla4 se presentan la pertenencia de los individuos a cada cluster, como se mencionó en capítulos anteriores la partición realizada por el algoritmo FCM es una partición suave en la cual no asigna una cluster definitivo a cada individuo, sino más bien cada individuo tienen un nivel de pertenencia a un cluster.

Según la cercanía que tiene los grados de pertenencia a 1 se asignará al cluster correspondiente.

Actualización de los prototipos después de las iteraciones realizadas

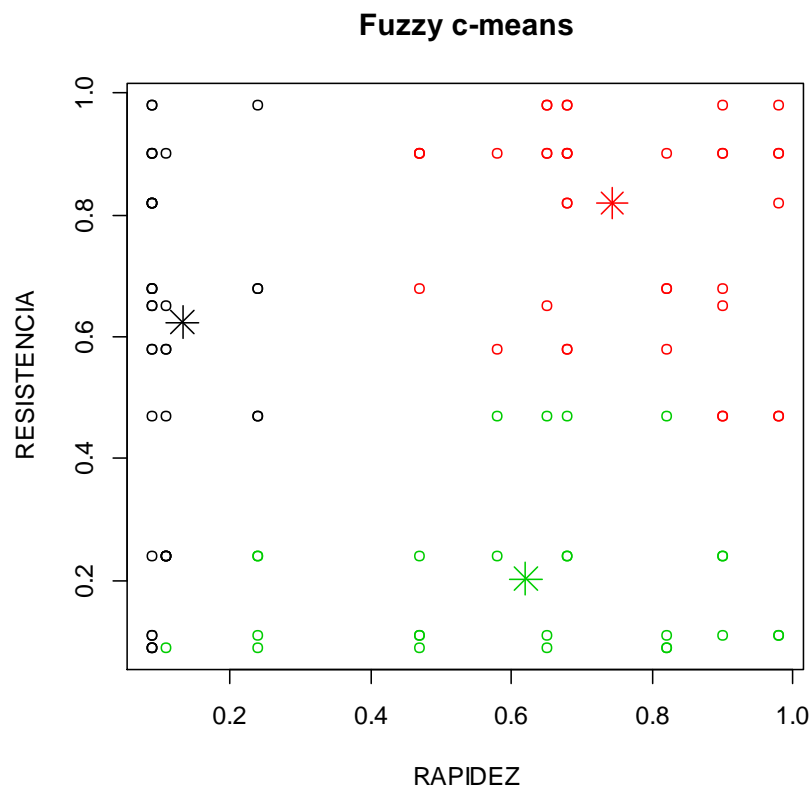
$$(v_i^*) = \frac{\sum_{k=1}^n (u_{ik}^*)^m x_k}{\sum_{k=1}^n (u_{ik}^*)^m}$$

Tabla5. Prototipos Finales

Agrupación	RAPIDEZ	RESISTENCIA
Cluster 1	0.1326445	0.6239699
Cluster 2	0.7433209	0.8196052
Cluster 3	0.6206765	0.2025221

Fuente: Elaboración Propia

Figura 7. Cluster FCM obtenidos



En el Figura 7 podemos observar los tres prototipos obtenidos después de las iteraciones, el primer prototipo se encuentra junto al primer cluster1 que se podría definir como los futbolistas menos rápidos pero con mayor resistencia, con el segundo prototipo estaría formando el cluster de los futbolistas más rápidos y resistentes (grupo optimo), y el prototipo

tres encontrarían con la cercanía de los individuos para el tercer cluster conformados por los futbolistas con mayor resistencia y pero no muy rápidos.

Definición de Perfil Final

Cluster 1: futbolistas de ataque

Cluster 2: futbolistas de medio campo

Cluster 3: futbolistas de defensa

Agrupación de las 20 primeras observaciones según el grado de pertenecía

Tabla 6. Asignación de cluster

Individuo	Cluster	Grado de Pertenencia
1	3	0.5683534
2	3	0.7528957
3	2	0.8739776
4	3	0.8688175
5	1	0.9148484
6	2	0.615629
7	3	0.5473996
8	1	0.4397087
9	1	0.4540267
10	2	0.8418865
11	3	0.5498642
12	3	0.8572652
13	2	0.4697945
14	1	0.8657974
15	1	0.8657974
16	3	0.5774499
17	2	0.4697945
18	1	0.7958123
19	3	0.8618083
20	1	0.4397087

Fuente: Elaboración propia

En la tabla 6 muestra la asignación del futbolista 1 al Cluster 3 al grupo de futbolistas de defensa, alcanzando un grado de pertenencia a este cluster de 0.56 por presentar este grado como el más cercano a uno, no lo define totalmente como un futbolista. Si observamos este futbolista también tiene grados de pertenencia a los otros cluster como son: posee un grado de pertenencia al grupo de los futbolistas de medio campo de 0.11 y a los futbolistas de

ataque de 0.31. Es decir que este futbolista es un futbolista no tan rápido y que no posee una alta resistencia esto se refleja en el grado de pertenencia de cluster 2 porque tiene un bajo grado de pertenencia a este grupo.

En este caso se presenta la partición suave que realiza el algoritmo FCM a los individuos para agruparlos, donde se agruparon a los futbolistas según los atributos de rapidez y resistencia.

CAPÍTULO V. CONCLUSIONES

1. El algoritmo Fuzzy cmeans es un algoritmo que realiza una partición suave de a través del grado de pertenencia a cada cluster en la formación de los mismos, de tal forma que un individuo podría pertenecer a más de un cluster a diferencia de la partición dura o convencional.
2. Este algoritmo es un método adicional con el que se puede realizar los agrupamientos bajo el aprendizaje no supervisado utilizando la lógica Fuzzy.
3. En la agrupación hallada con FCM muestra la cantidad de individuos en cada cluster asignada según el grado de pertenencia al mismo con una partición suave de los datos, Kmeans estaría mostrando una partición más dura de los mismos datos.
4. El caso se tuvo como objetivo presentar la partición suave que realiza este algoritmo para la formación de grupos según el grado de pertenencia, así mismo de mostrar que un individuo no pertenece exclusivamente a un solo grupo.
5. Este algoritmo también presenta sus desventajas, como son el número de cluster a generar se debe ingresar a priori, además para datos reales se necesitaría de una agrupación a priori para comprobar la calidad de los cluster formados.

CAPÍTULO VI. RECOMENDACIONES

1. Se recomendaría realizar comparaciones con otros métodos no supervisados para obtener la una tasa de clasificación sobre datos reales con una clase conocida a priori.
2. En la actualidad se vienen desarrollando variaciones de este algoritmo teniendo como base al método Fuzzy c-means para el reconocimiento de patrones, en especial para el reconocimiento de imágenes De la misma forma a entrado en la dirección de machine learning a trabajar de manera conjunta con otros métodos como redes neuronales y algoritmos genéticos, en el que se puede seguir investigando.

BIBLIOGRAFÍA

BRIAN S. EVERITT . SABINE LANDAU MORVEN LEESE . DANIEL STAHL King's College London, UK Cluster Analise 5th Edition.

J. C. BEZDEK (1981): "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York

ZADEH, L. A. Fuzzy sets. Information and Control 8, 3 (1965), 338–353.

GUILLERMO TEJADA MUÑOZ, Tutorial de Logica Fuzzy, San Marcos (2000) Lima Perú.

UNIVERSIDAD TECNOLÓGICA DE PEREIRA. ISSN 0122-1701 técnicas de lógica difusa aplicada a la minería de datos.

THIAGO ROCHA, SARAJANE MARQUEZ PÉRES 2012, "Tutorial sobre Fuzzy-c-Means e Fuzzy Learning Vector Quantization: Volumen 19

KRINIDIS, S. E CHATZIS, V. A robust Fuzzy local information c-means clustering algorithm. IEEE Transactions on Image Processing 19, 5 (May 2010), 1328 –1337.

ISABELA MONICO, Presentación Lógica Fuzzy Aplicada a Sistemas de Diagnostico (Eletronorte, Brasil 2014)

ALGORITHMS FOR FUZZY CLUSTERING_ METHODS IN C-MEANS CLUSTERING (Studies in Fuzziness and Soft Computing 229) Sadaaki Miyamoto, Hidetomo Ichihashi, Katsuhiko Honda (auth.)

ZEYNEL CEBECI1, FIGEN YILDIZ2 Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures, 2015

CARLOS GONZÁLEZ MORCILLO Lógica Difusa Una introducción práctica Técnicas de Softcomputing , 2011

ANEXO

Comandos con Aplicación R

Paquete fclust

```
DATR=c(0.58,      0.90, 0.68, 0.11, 0.47, 0.24, 0.09, 0.82, 0.65, 0.09,
        0.98)
SRA=sample(DATR,100,T)
DATRE=c(0.33,      0.11, 0.17, 0.44, 0.81, 0.83, 0.18, 0.11, 0.50, 0.63,
        0.24)
SRE=sample(DATR,100,T)
x=data.frame(RAPIDEZ=c(SRA),RESISTENCIA=c(SRE))
```

```
clustx=FKM(x,k=3,m=2,maxit=200)
clustx
clustx
$U
      Clus 1      Clus 2      Clus 3
1  0.316398513 0.115248082 0.568353404
2  0.081220064 0.165884193 0.752895744
3  0.076304749 0.873977604 0.049717647
4  0.080176515 0.051006014 0.868817470
5  0.914848386 0.049174489 0.035977125
6  0.125400041 0.615629048 0.258970911
7  0.127093242 0.325507150 0.547399608
8  0.439708732 0.131544413 0.428746855
9  0.454026718 0.129802076 0.416171206
10 0.100961473 0.841886481 0.057152046
11 0.179818868 0.270316919 0.549864212
12 0.059284807 0.083449985 0.857265208

$H
      RAPIDEZ RESISTENCIA
Clus 1 0.1326445  0.6239699
```

Clus 2 0.7433209 0.8196052

Clus 3 0.6206765 0.2025221

\$clus

	Cluster	Membership degree
1	3	0.5683534
2	3	0.7528957
3	2	0.8739776
4	3	0.8688175
5	1	0.9148484
6	2	0.6156290
7	3	0.5473996
8	1	0.4397087
9	1	0.4540267
10	2	0.8418865
11	3	0.5498642
12	3	0.8572652
13	2	0.4697945
14	1	0.8657974
15	1	0.8657974
16	3	0.5774499
17	2	0.4697945
18	1	0.7958123
19	3	0.8618083
20	1	0.4397087
21	2	0.8939320
22	2	0.9530687
23	3	0.5550958
24	2	0.4584686
25	2	0.4502615
26	2	0.9300101
27	3	0.7528957
28	1	0.6996841
29	1	0.9894533
30	1	0.6996841
31	1	0.5665760

32	3	0.8719105
33	3	0.4610859
34	2	0.9077961
35	1	0.9148484
36	3	0.9830539
37	1	0.8935294
38	2	0.8939320
39	1	0.5665760
40	1	0.4540267
41	3	0.8796771
42	2	0.7541428
43	1	0.8657974
44	2	0.6302770
45	2	0.5001325
46	1	0.9946774
47	3	0.8572652
48	1	0.7746803
49	1	0.9836762
50	1	0.7802698
51	1	0.8938418
52	2	0.6302770
53	3	0.8796771
54	1	0.9795878
55	2	0.8710010
56	2	0.9300101
57	1	0.5803766
58	1	0.9894533
59	1	0.9887279
60	3	0.9751353
61	1	0.4397087
62	1	0.7958123
63	1	0.9836762
64	2	0.9530687
65	3	0.9651550
66	2	0.9560653
67	2	0.8473375

68	1	0.9795878
69	2	0.8710010
70	2	0.6302770
71	2	0.8741872
72	3	0.7213899
73	1	0.5665760
74	1	0.8657974
75	2	0.9077961
76	1	0.7802698
77	1	0.9887279
78	2	0.8739776
79	2	0.9782033
80	2	0.9530687
81	3	0.4610859
82	3	0.9751353
83	2	0.4502615
84	1	0.5952166
85	1	0.7802698
86	2	0.7800081
87	2	0.6156290
88	3	0.4778065
89	2	0.8473375
90	2	0.8208136
91	3	0.9528788
92	2	0.8447324
93	1	0.9795878
94	1	0.5665760
95	2	0.6753361
96	3	0.7213899
97	2	0.8276813
98	3	0.4443856
99	2	0.9782033
100	3	0.7902285
\$value		
Start 1		

4.461796

\$cput

Start 1

0.33

\$iter

Start 1

200

\$k

Number of clusters

3

\$m

Parameter of fuzziness

2

\$stand

Standardization (1=Yes, 0=No)

0

\$Xca

RAPIDEZ RESISTENCIA

[1,]	0.24	0.11
[2,]	0.90	0.24
[3,]	0.65	0.98
[4,]	0.47	0.24
[5,]	0.24	0.68
[6,]	0.68	0.58
[7,]	0.68	0.47
[8,]	0.09	0.09
[9,]	0.09	0.11
[10,]	0.58	0.90
[11,]	0.58	0.47
[12,]	0.82	0.09

[13,]	0.98	0.47
[14,]	0.09	0.82
[15,]	0.09	0.82
[16,]	0.24	0.09
[17,]	0.98	0.47
[18,]	0.24	0.47
[19,]	0.82	0.11
[20,]	0.09	0.09
[21,]	0.68	0.98
[22,]	0.68	0.90
[23,]	0.65	0.47
[24,]	0.47	0.68
[25,]	0.90	0.47
[26,]	0.65	0.90
[27,]	0.90	0.24
[28,]	0.09	0.98
[29,]	0.09	0.65
[30,]	0.09	0.98
[31,]	0.11	0.24
[32,]	0.47	0.09
[33,]	0.24	0.24
[34,]	0.90	0.90
[35,]	0.24	0.68
[36,]	0.58	0.24
[37,]	0.11	0.47
[38,]	0.68	0.98
[39,]	0.11	0.24
[40,]	0.09	0.11
[41,]	0.47	0.11
[42,]	0.65	0.65
[43,]	0.09	0.82
[44,]	0.47	0.90
[45,]	0.58	0.58
[46,]	0.11	0.65
[47,]	0.82	0.09
[48,]	0.11	0.90

[49,]	0.09	0.58
[50,]	0.09	0.90
[51,]	0.09	0.47
[52,]	0.47	0.90
[53,]	0.47	0.11
[54,]	0.09	0.68
[55,]	0.82	0.68
[56,]	0.65	0.90
[57,]	0.09	0.24
[58,]	0.09	0.65
[59,]	0.11	0.58
[60,]	0.68	0.24
[61,]	0.09	0.09
[62,]	0.24	0.47
[63,]	0.09	0.58
[64,]	0.68	0.90
[65,]	0.65	0.11
[66,]	0.82	0.90
[67,]	0.98	0.90
[68,]	0.09	0.68
[69,]	0.82	0.68
[70,]	0.47	0.90
[71,]	0.90	0.98
[72,]	0.98	0.11
[73,]	0.11	0.24
[74,]	0.09	0.82
[75,]	0.90	0.90
[76,]	0.09	0.90
[77,]	0.11	0.58
[78,]	0.65	0.98
[79,]	0.68	0.82
[80,]	0.68	0.90
[81,]	0.24	0.24
[82,]	0.68	0.24
[83,]	0.90	0.47
[84,]	0.24	0.98

[85,]	0.09	0.90
[86,]	0.90	0.65
[87,]	0.68	0.58
[88,]	0.82	0.47
[89,]	0.98	0.90
[90,]	0.90	0.68
[91,]	0.65	0.09
[92,]	0.98	0.82
[93,]	0.09	0.68
[94,]	0.11	0.24
[95,]	0.82	0.58
[96,]	0.98	0.11
[97,]	0.98	0.98
[98,]	0.11	0.09
[99,]	0.68	0.82
[100,]	0.90	0.11

\$X

RAPIDEZ RESISTENCIA

[1,]	0.24	0.11
[2,]	0.90	0.24
[3,]	0.65	0.98
[4,]	0.47	0.24
[5,]	0.24	0.68
[6,]	0.68	0.58
[7,]	0.68	0.47
[8,]	0.09	0.09
[9,]	0.09	0.11
[10,]	0.58	0.90
[11,]	0.58	0.47
[12,]	0.82	0.09
[13,]	0.98	0.47
[14,]	0.09	0.82
[15,]	0.09	0.82
[16,]	0.24	0.09
[17,]	0.98	0.47

[18,]	0.24	0.47
[19,]	0.82	0.11
[20,]	0.09	0.09
[21,]	0.68	0.98
[22,]	0.68	0.90
[23,]	0.65	0.47
[24,]	0.47	0.68
[25,]	0.90	0.47
[26,]	0.65	0.90
[27,]	0.90	0.24
[28,]	0.09	0.98
[29,]	0.09	0.65
[30,]	0.09	0.98
[31,]	0.11	0.24
[32,]	0.47	0.09
[33,]	0.24	0.24
[34,]	0.90	0.90
[35,]	0.24	0.68
[36,]	0.58	0.24
[37,]	0.11	0.47
[38,]	0.68	0.98
[39,]	0.11	0.24
[40,]	0.09	0.11
[41,]	0.47	0.11
[42,]	0.65	0.65
[43,]	0.09	0.82
[44,]	0.47	0.90
[45,]	0.58	0.58
[46,]	0.11	0.65
[47,]	0.82	0.09
[48,]	0.11	0.90
[49,]	0.09	0.58
[50,]	0.09	0.90
[51,]	0.09	0.47
[52,]	0.47	0.90
[53,]	0.47	0.11

[54,]	0.09	0.68
[55,]	0.82	0.68
[56,]	0.65	0.90
[57,]	0.09	0.24
[58,]	0.09	0.65
[59,]	0.11	0.58
[60,]	0.68	0.24
[61,]	0.09	0.09
[62,]	0.24	0.47
[63,]	0.09	0.58
[64,]	0.68	0.90
[65,]	0.65	0.11
[66,]	0.82	0.90
[67,]	0.98	0.90
[68,]	0.09	0.68
[69,]	0.82	0.68
[70,]	0.47	0.90
[71,]	0.90	0.98
[72,]	0.98	0.11
[73,]	0.11	0.24
[74,]	0.09	0.82
[75,]	0.90	0.90
[76,]	0.09	0.90
[77,]	0.11	0.58
[78,]	0.65	0.98
[79,]	0.68	0.82
[80,]	0.68	0.90
[81,]	0.24	0.24
[82,]	0.68	0.24
[83,]	0.90	0.47
[84,]	0.24	0.98
[85,]	0.09	0.90
[86,]	0.90	0.65
[87,]	0.68	0.58
[88,]	0.82	0.47
[89,]	0.98	0.90

```
[90,] 0.90 0.68
[91,] 0.65 0.09
[92,] 0.98 0.82
[93,] 0.09 0.68
[94,] 0.11 0.24
[95,] 0.82 0.58
[96,] 0.98 0.11
[97,] 0.98 0.98
[98,] 0.11 0.09
[99,] 0.68 0.82
[100,] 0.90 0.11
```

```
$call
```

```
FKM(X = x, k = 3, m = 2, maxit = 200)
```

```
attr(,"class")
```

```
[1] "fclust"
```

Aplicación de función Kmeans.

```
o1<- kmeans(x, centers = 3, nstart = 20)
```

```
o1
```

Agrupamiento por Kmeans y FCM

Figura8. Anexo Formación de Cluste Kmeans y FCM

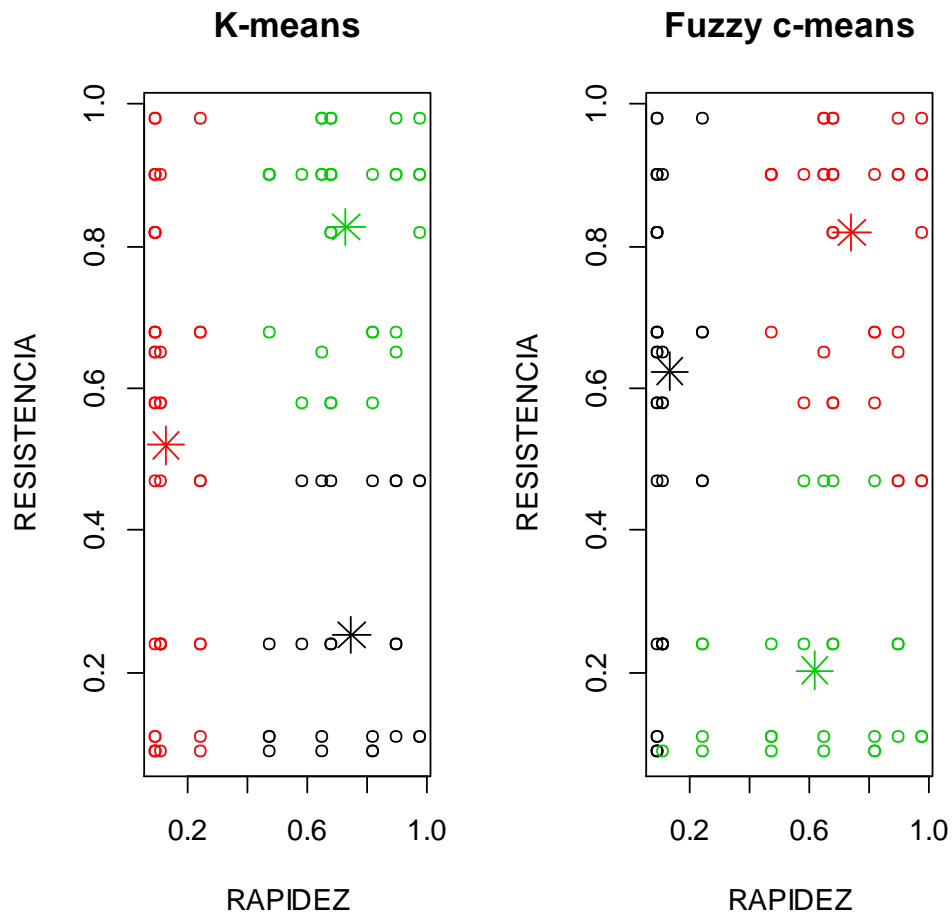


Tabla 7. Anexo Comparativo Agrupación de Individuos

Cluster	Fuzzy Cmeans	Kmeans
1	37	25
2	37	42
3	26	33

No se obtienen similar cantidad de individuos agrupados en cada cluster, y los centroides hallados difieren en ambos métodos, el método Kmeans en la grafica muestra la agrupación mejor definida visualmente, pero la agrupación obtenida con el método FCM se observa algunos individuos clasificados en otro cluster en especial entre el cluster 1 y el cluster 2.

