

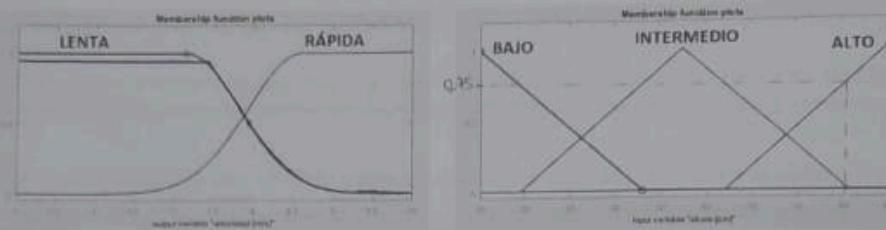


NOMBRE y DNI: Lucas José Muñoz DNI 42624508

PROBLEMA 1:

Dado un objeto que tiene una altura de 90 cm y un peso de 1000 g, mostrar la forma del conjunto difuso de la salida del siguiente sistema de inferencia difusa con una sola regla, antes de la defuzzificación. Utilizar los operadores [producto] para la conjunción, [suma probabilística] para la disyunción y el operador [mínimo] para la implicación.

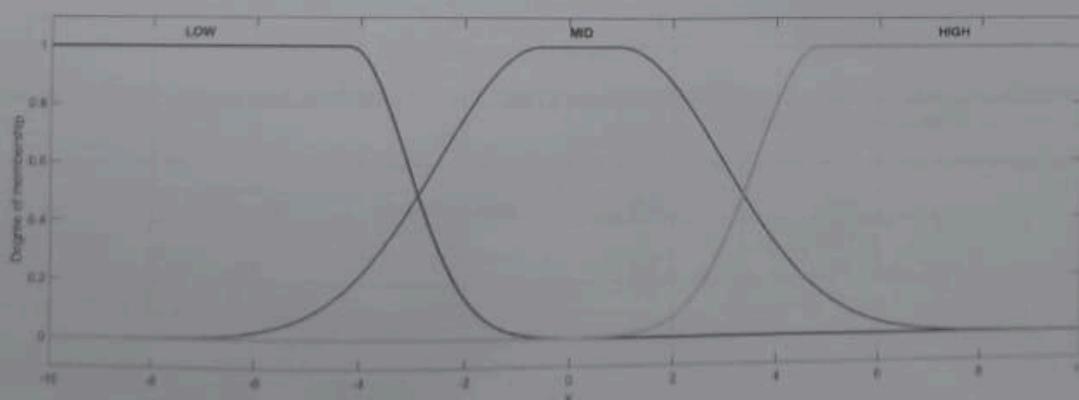
Si el objeto es alto o es pesado ENTONCES la velocidad es lenta.

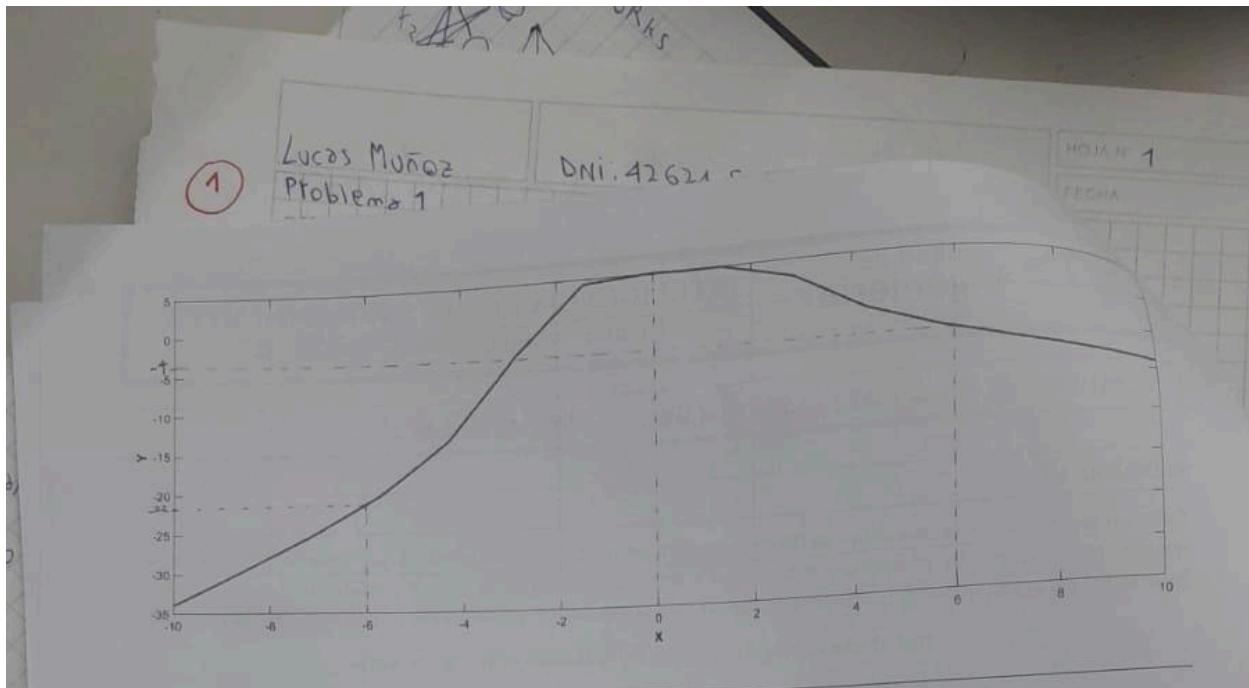


PROBLEMA 2:

Se tienen las siguientes funciones de pertenencia en un sistema de inferencia difusa de Sugeno de una entrada X y una salida Y . Se da también la salida que se obtiene de este sistema.

Escriba todas las reglas de este sistema. Explique claramente cómo las deduce y, si le es útil, recuerde que la ecuación de una recta de pendiente m que pasa por el punto (x_0, y_0) es $y - y_0 = m(x - x_0)$.





PROBLEMA 3:

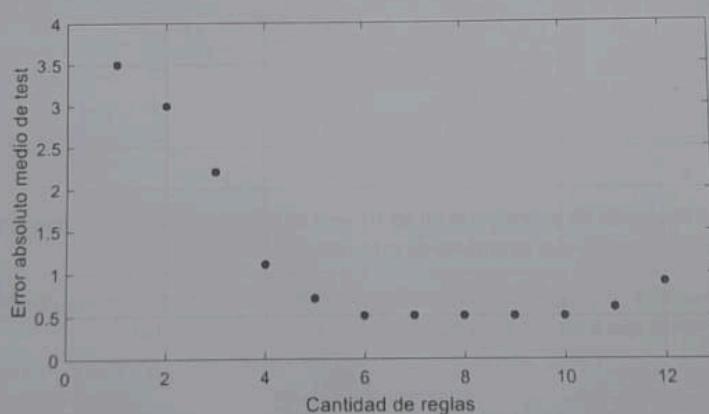
Se intenta diseñar un modelo de Sugeno para estimar un valor de alarma de enfermedad renal entre 0 y 10 (KDA, kidney disease alarm), dadas dos variables: la tasa de filtración glomerular (GFR, glomerular filtration rate) y la cantidad de albúmina en sangre (ALB).

Para generar el modelo se dispone de 160 casos en los que se conoce GFR, ALB y su respectivo valor KDA.

Se ha decidido utilizar el 5% para datos de test y el resto para datos de "entrenamiento" (generación del modelo). Se utiliza el error absoluto medio para evaluar el comportamiento del modelo. Se utiliza el método de hold-out para la evaluación con 20 iteraciones.

a) Indique las dimensiones de las matrices que contienen los datos para entrenamiento y los datos para test.

b) Se varió la cantidad de reglas de diversos modelos y se observa lo siguiente:



Seleccione la cantidad de reglas que estima adecuada para elegir uno de los modelos. Explique cuál puede ser la causa del aumento del error en la última parte.

c) Calcule el error para los siguientes 3 datos, según lo que estimó el modelo elegido:

GFR	ALB	KDA estimado	KDA real
110	24	1.2	0.5
12	52	9.7	9.4
55	31	2.2	3.0

0.8

PROBLEMA 4:

Una empresa de venta online de celulares requiere que se obtenga un agrupamiento de datos para analizar el grado de satisfacción de sus clientes y posteriormente tomar decisiones para incrementar sus ventas. Quieren dividirse los datos en dos grupos.

Le proporciona un Dataset de tamaño 5000 x 4 donde tiene la información recolectada sobre la marca de celular comprado (Categórica, XIAOMI – SAMSUNG – APPLE), el número de IMEI que identifica únicamente el equipo, si fue devuelto o no luego de una semana (Dicotómica) y el grado de satisfacción sus clientes (Entero en el intervalo 1-5, donde 5 representa el mayor grado de satisfacción).

- a) Determine los hiperparámetros a utilizar y elija las variables adecuadamente para aplicar el algoritmo de clustering Hard-C-means.
 - b) Indique dimensiones de la matriz de pertenencias obtenida y de la matriz de centros de clusters. Esquematice una matriz de pertenencia posible para los tres primeros datos. ¿Cómo difiere si hubiera elegido Fuzzy C Means?
 - c) Elija una medida de calidad de modelo y describa brevemente su utilización para evaluar el agrupamiento obtenido.
 - d) Proponga un cromosoma y una función de costo/fitness para la optimización de K y los centros iniciales con Algoritmos Genéticos.
-

Lucas Muñoz

DNI: 42624 508

HOJA N° 1

9

1

Problema 1

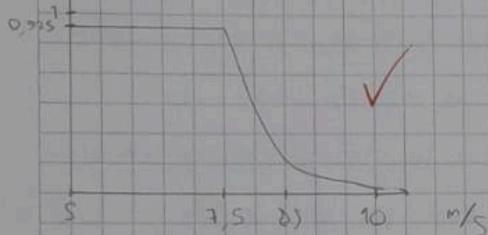
operación OR

 $N(1000)$ OR $N(500)$

$$0,75 + 0,7 - 0,75 \cdot 0,7 = \checkmark$$

$$1,45 - 0,525 = 0,925 \checkmark$$

concepto difuso desolido



Problema 3

0.8

a- La matriz de entrenamiento tendrá 152 datos por lo tanto 152 filas que es el 95% de los datos totales. Dado que hay solo dos variables de entrada y se debe tener un target ya que es un modelo de sugerencia este tendrá 3 columnas. Lo que queda una matriz de 152×3 ✓

En el caso del test serán del 5% de los datos, es decir, el resto. Una matriz de 8×3 . ✓

b- La cantidad adecuada de reglas es la que tenga menor error absoluto medio. En este caso vemos en el gráfico que entre 6 y 10 se tiene el mismo valor y el mismo por lo tanto elegiría cualquiera de estos. En mi caso elijo 6 ya que son menos reglas a evaluar por lo que se tendrá una optimización de tiempo. ✓

Lo causó porque cada una aumenta el error en la última parte es debido a que puede suceder un over-training. Se complejiza tanto el modelo que pierde la capacidad de predecir la interpolación de datos. ✓

c.	KDA estimado	KDA real	error
	1,2	0,5	0,7
	9,4	9,4	0,5
	2,2	3,0	0,8

y el mae?

Problema 4:

a) Un hiperparametro del Hard-C-means es K , la cantidad de grupos que se quiere formar, en este caso la empresa quiere dividir los datos en dos grupos, por lo tanto usaria $K=2$.

Otros hiperparametros podrian ser los centros de cluster iniciales. En este caso no pude observar la distribucion de datos por lo que los asigno de forma aleatoria, o como la empresa le interesa dividir los grupos en grados de satisfaccion, pondria como primer centro de cluster un dato que tengo satisfaccion '1' y otro que tengo satisfaccion '5'. uno
0.15/1

b) La matriz de pertenencia tendra la dimension 2×5000 donde las filas serian los dos cluster y los columnas los datos totales. La matriz de centros de clusters indica donde estan estos centros la dimension de esto sera 2×4 , dos porque son dos clusters y 4 porque son 4 dimensiones, 4 variables de entrada.

Una matriz de pertenencia posible:

c. $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix}$ Si hubiera elegido Fuzzy CMeans esta matriz tendría valores entre el intervalo $[0,1]$, esto significaria en cuanto pertenece cada cluster cada dato.

c. Un indice para medir la calidad del modelo es silhouette que al aplicarlo a cada cluster obtendremos un numero entre -1 y 1 donde -1 es malo y 1 es bueno, esto nos evalua la cohesión y separación del agrupamiento obtenido. 0.25
1

Lucas Muñoz

DNI. 42624508

HOJA DE 2

FECHA

d) optimizar $[k, c_1, c_2, \dots]$

w centros iniciales

Función costo puede ser igualmente silhouette siendo así

$$SC = \frac{1}{N} \sum_{i=1}^N S(x_i) \rightarrow S(x_i) = \frac{b(x_i) - a(x_i)}{\max(a(x_i), b(x_i))} \rightarrow a \text{ en numerador}$$

$$\frac{0.25}{1}$$

Se buscaba que este índice sea lo más cercano a 1 posible.

Un cromosoma puede ser $k=2$ $c_1 = [\dots, \text{grado de satisfacción}=7]$ $c_2 = [\dots, \text{grado de satisfacción}=5]$

TOTAL: 0.9 / 1

0.8

Problema 2

Reglas del sistema

R1 Si X es Low entonces $Y = m_1(X+6) - 22$

R2 Si X es Mid entonces $Y = m_2(X-0) + 5$

R3 Si X es High entonces $Y = m_3(X-6) - 4$

Para obtener las reglas obtuve los valores para cuando la variable de entrada cumple al máximo con la condición Low, Mid y High, en este caso $(-6, 0, 6)$ y luego obtuve su salida con el gráfico X vs Y $(-22, 5, -4)$ por lo tanto obtuve los

reglas en base a rectas de pendientes m_i que pasan por los puntos (x_0, y_0) antes mencionados. ✓

* faltan las pendientes m_i

* ok la idea del método