CLASIFICACIÓN Y MACHINE LEARNING CON TÉCNICAS DIFUSAS

EVIA 2016

José Antonio Sanz
Universidad Pública de Navarra

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Sistemas de clasificación basados en reglas difusas

Problema de clasificación[Duda01]

- Resolver un problema de clasificación consiste en aprender una regla de decisión que permita determinar la clase de un nuevo objeto dentro de una de las existentes y conocidas

$$D: X(E) \rightarrow C$$

A la regla de decisión, D, se le llama clasificador y debe ser óptima en función de un cierto criterio que determine la calidad del clasificador

El clasificador D se aprende utilizado un conjunto de P ejemplos, $E = \{e_1, \ldots, e_P\}$, llamado conjunto de entrenamiento.

Cada uno de estos P ejemplos, e, está descrito por N valores, uno por cada atributo/variable del sistema

$$X(e) = \{e_1, \ldots, e_N\}$$

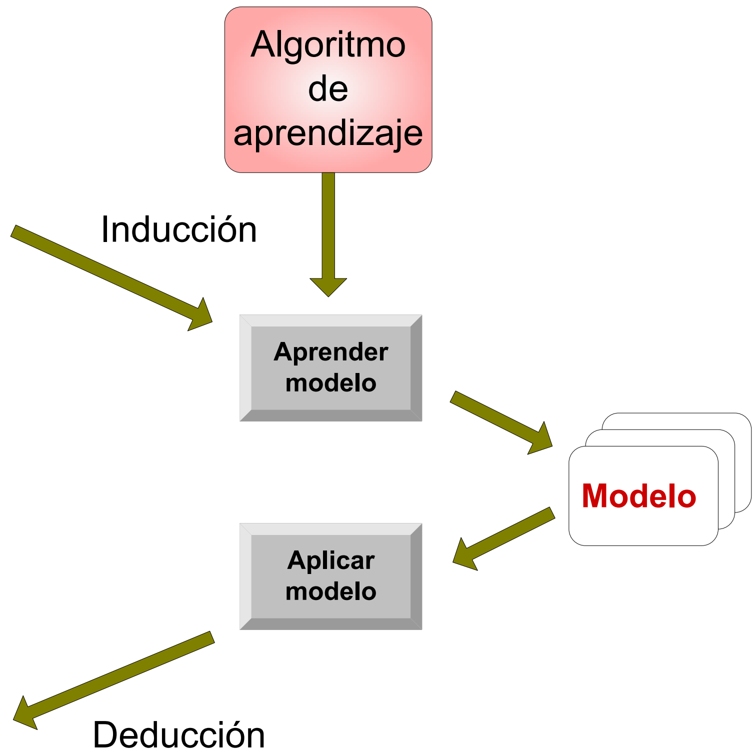
y está etiquetado con una de las M clases existentes: $C = \{C_1, \ldots, C_M\}$

[Duda01] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern classification, Wiley, 2001

Etapas en el proceso de clasificación

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Conjunto de entrenamiento

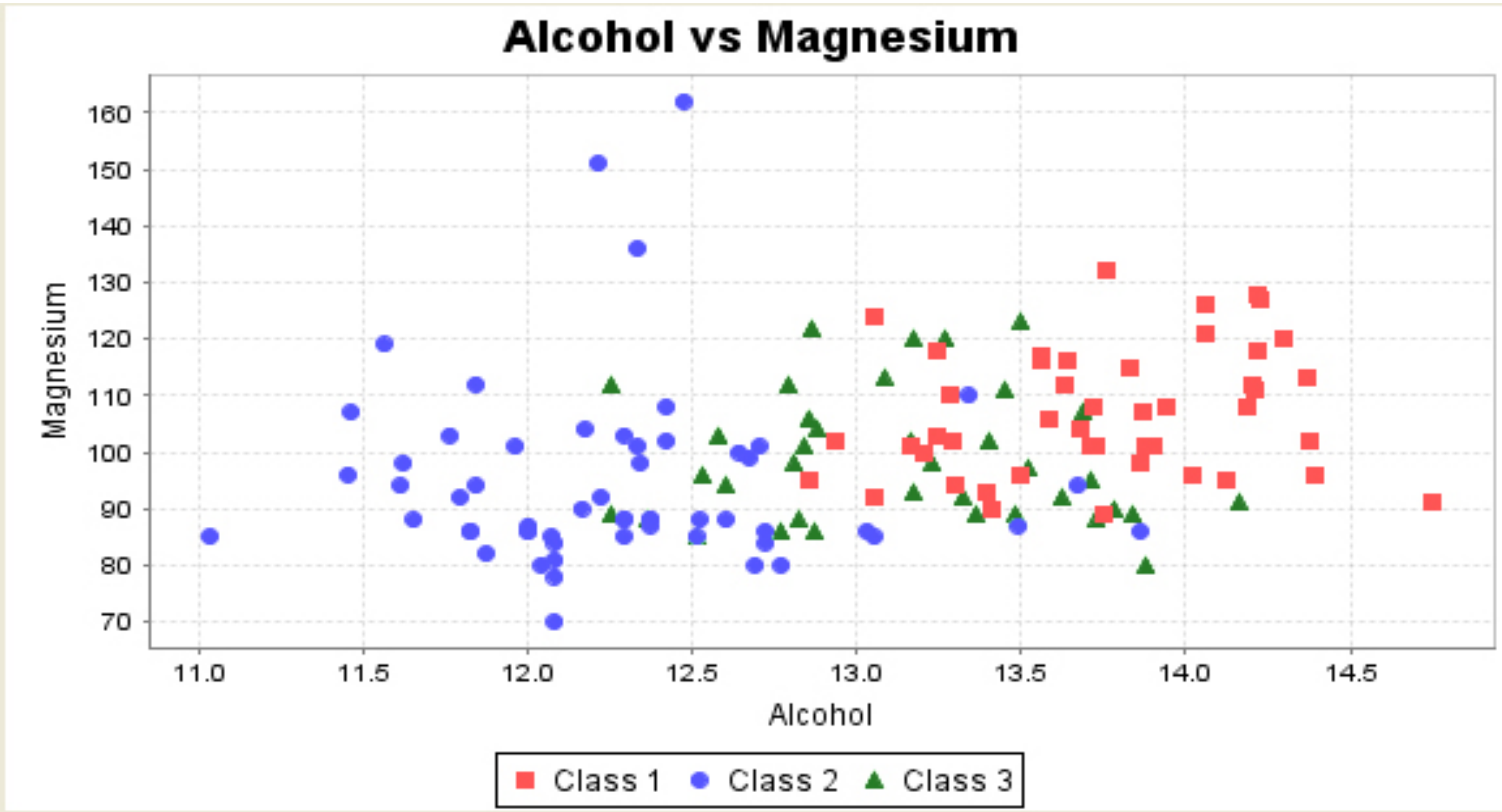


Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Conjunto de prueba

Aplicaciones

- Ejemplo: predecir el tipo de vino en base al alcohol y magnesio



Sistemas de clasificación basados en reglas

- Aprenden un modelo basado en reglas utilizando el conjunto de entrenamiento
  - Clasificador compuesto por un conjunto de reglas
  - Clasifican nuevos ejemplos basándose en el conocimiento aprendido
- Una regla de clasificación está formada por
  - Antecedente: contiene un predicado que se evaluará como verdadero o falso con respecto a cada ejemplo
  - Consecuente: contiene una etiqueta de clase

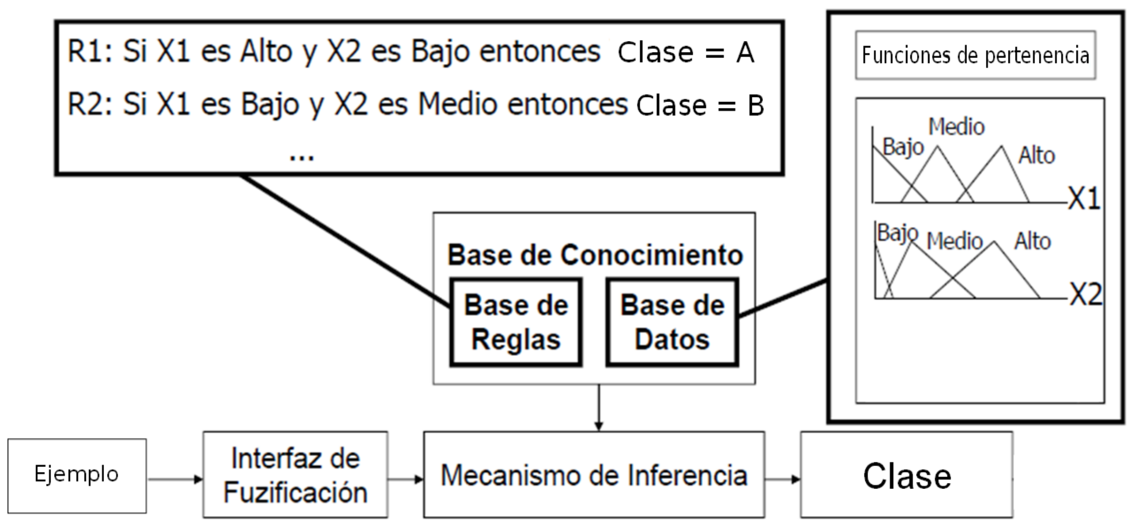
SI gastos > 30.000 Y ingresos < 20.000 ENTONCES Conceder préstamo = NO

Sistemas de clasificación basados en reglas difusas[Ishibuchi05]

- Son sistemas de reglas cuyos antecedentes están formados por predicados que utilizan lógica difusa

SI gastos son altos Y ingresos son bajos ENTONCES Conceder préstamo = NO

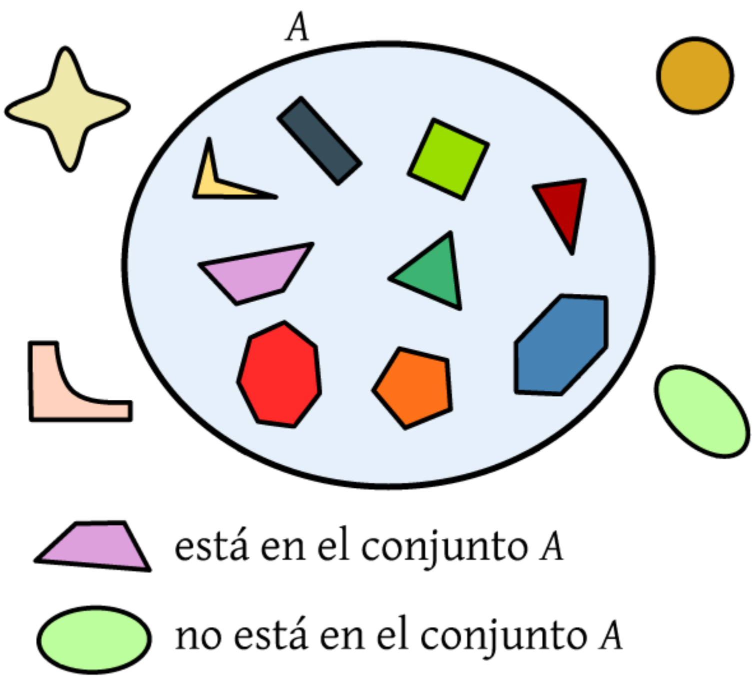
- Esquema general



[Ishibuchi05] Classification and modeling with linguistic information granules. Advanced approaches to linguistic data mining. H. Ishibuchi, T. Nakashima y M. Nii.Springer (2005)

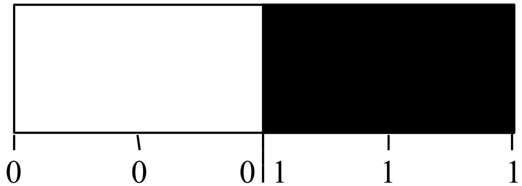
Teoría de conjuntos difusos

- Conjunto: colección de objetos con una/s propiedad/es en común.
  - Cada objeto del conjunto es un elemento.
- Lógica booleana / crisp
  - Los elementos están en el conjunto o no
  - No funciona para términos ambiguos/difusos

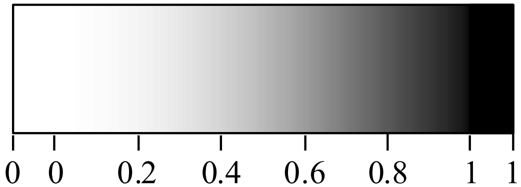


Teoría de conjuntos difusos

- La lógica difusa fue definida en 1965 por Lofti Zadeh
  - Creada para representar matemáticamente la ambigüedad/incertidumbre
  - Ofrece herramientas para tratar la incertidumbre intrínseca en muchos problemas
  - Permite modelar términos lingüísticos: persona alta, persona baja, etc…
    - Provee un modelo interpretable puesto que utiliza el lenguaje natural
  - Basada en que los elementos pertenecen al conjunto con un grado en [0, 1]



(*a*) Lógica booleana.



(*b*) Lógica difusa.

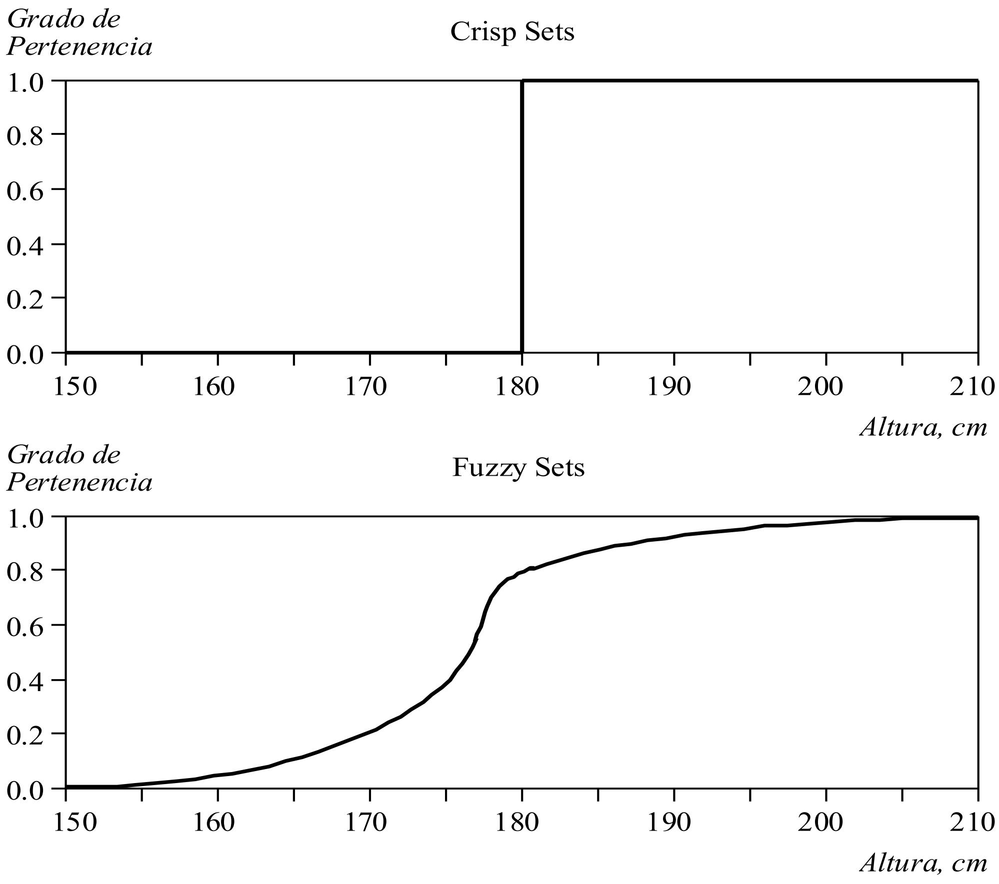
Teoría de conjuntos difusos

- Ejemplo
  - Persona alta
    - Umbral crisp: 180

Name	Height, cm	Degree of Membership	
		Crisp	*Fuzzy*
Chris	208	1	1.00
Mark	205	1	1.00
John	198	1	0.98
Tom	181	1	0.82
David	179	0	0.78
Mike	172	0	0.24
Bob	167	0	0.15
Steven	158	0	0.06
Bill	155	0	0.01
Peter	152	0	0.00

Teoría de conjuntos difusos

- Conjuntos crisp vs. conjuntos difusos



Teoría de conjuntos difusos

- Conjunto crisp

$f_A(x) : X \rightarrow \{0, 1\}$, donde
$$f_A(x) = \begin{cases} 1, \text{si } x \in A \\ 0, \text{si } x \notin A \end{cases}$$

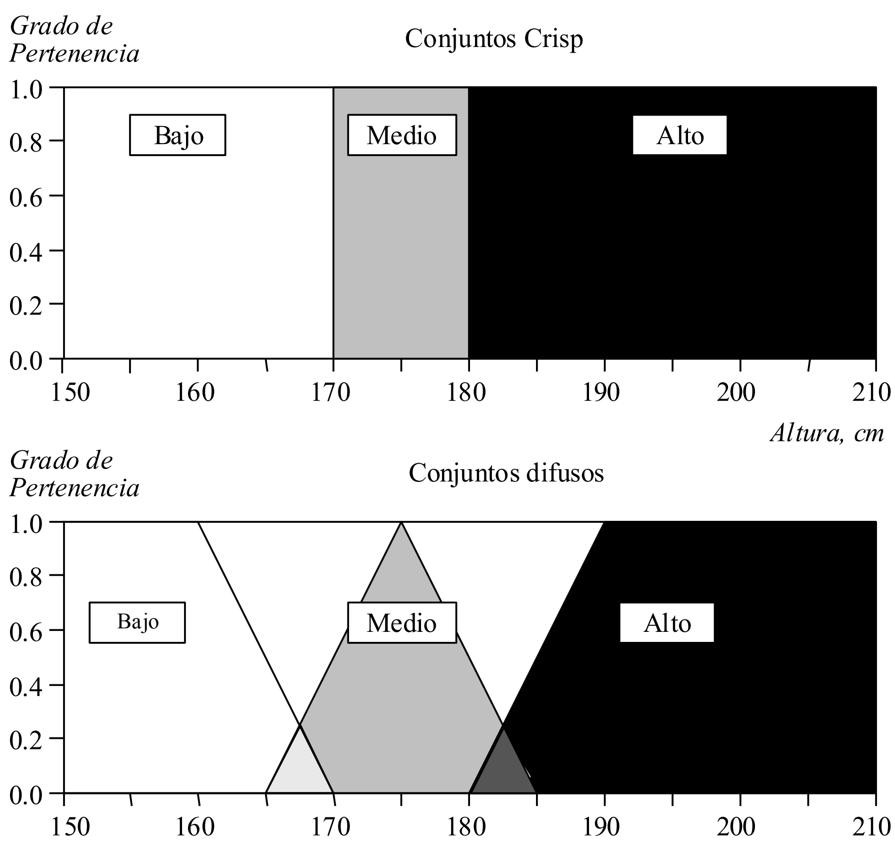
- Conjunto difuso

$\mu_A(x) : X \rightarrow [0, 1]$, donde

$\mu_A(x) = 1$ si x está totalmente en A;
$\mu_A(x) = 0$ si x no está en A;
$0 < \mu_A(x) < 1$ si x está parcialmente en A.

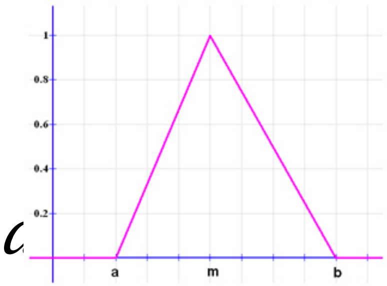
Teoría de conjuntos difusos

- Ejemplo de modelado
  - Variable altura
    - 3 términos (etiquetas lingüísticas)

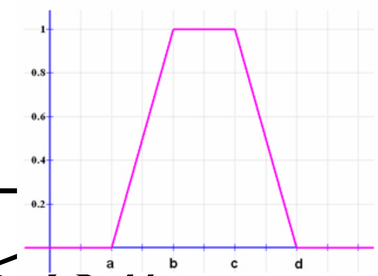


Teoría de conjuntos difusos

- Las funciones de pertenencia modelan a las etiquetas lingüísticas/ términos lingüísticos
- Ejemplos
  - Triangular

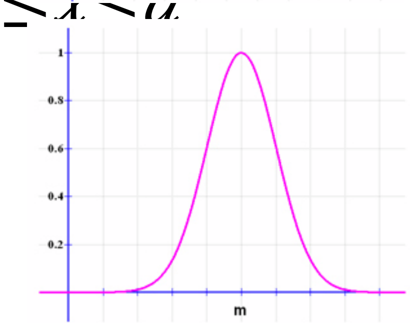
$\mu(x)=\{■0 \& si \& x<a \text{ o } x>b @ x-a/m-$  $m@b$
$-x/b-m \ \& si \& m\le x\le b$

  - Trapezoidal

$\mu(x)=\{■0 \& si \& x<a \text{ o } x>d @ ■x-a/b-$  $/d-c$
$\& ■ si@si@si \ \& ■ a\le x\le b@b\le x\le c@c\le$

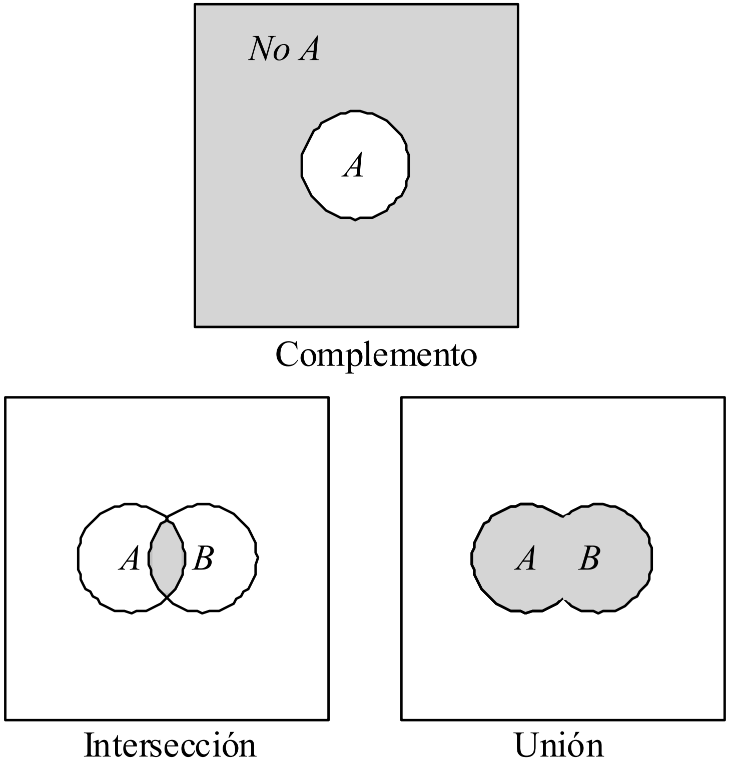
  - Gaussiana

$\mu(x)=e\uparrow-(x-m)\uparrow 2 \ /2\sigma\uparrow 2$



Teoría de conjuntos difusos

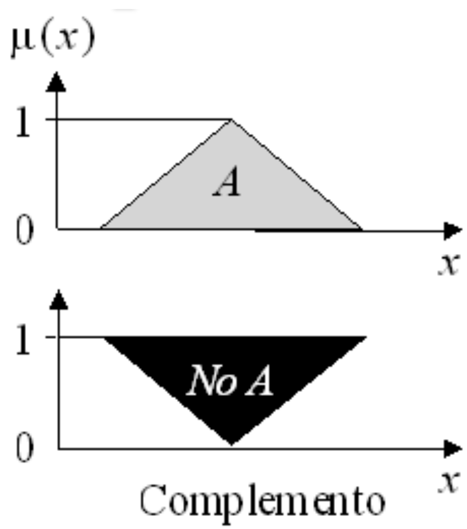
- Operaciones sobre conjuntos difusos



Teoría de conjuntos difusos

- Complemento
  - Lógica booleana: ¿Qué elementos no pertenecen al conjunto?
  - Lógica difusa: ¿Cuánto no pertenecen los elementos al conjunto?

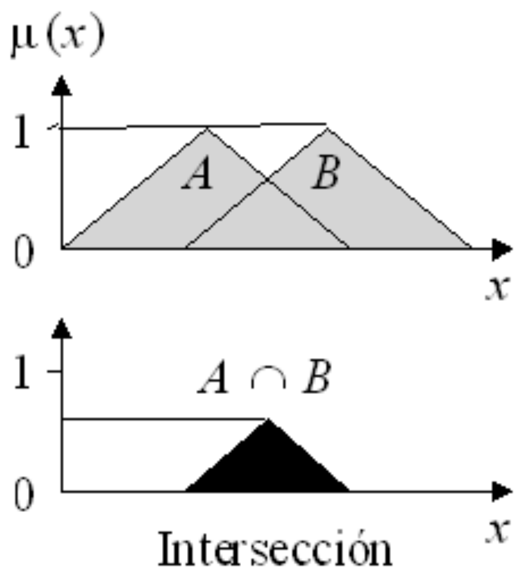
$$\mu_{\neg A}(x) = 1 - \mu_A(x)$$



Teoría de conjuntos difusos

- Intersección
  - Lógica booleana: ¿Pertenece el elemento a ambos conjuntos?
  - Lógica difusa: ¿Cuánto pertenece el elemento a ambos conjuntos?
    - Ejemplos de t-normas son el producto y el mínimo (comúnmente utilizadas)

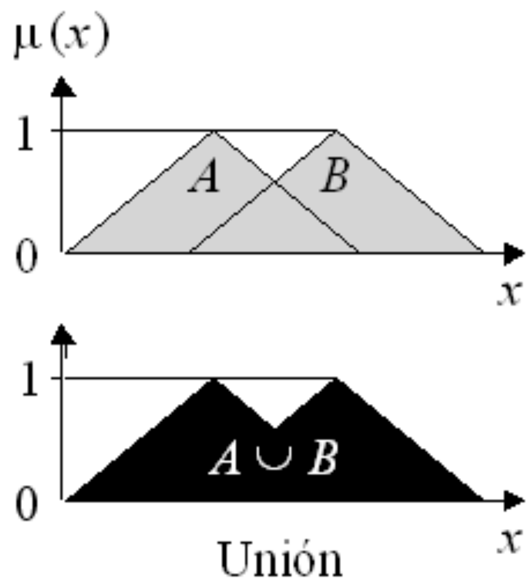
$$\mu_{A \cap B}(x) = T[\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x)$$



Teoría de conjuntos difusos

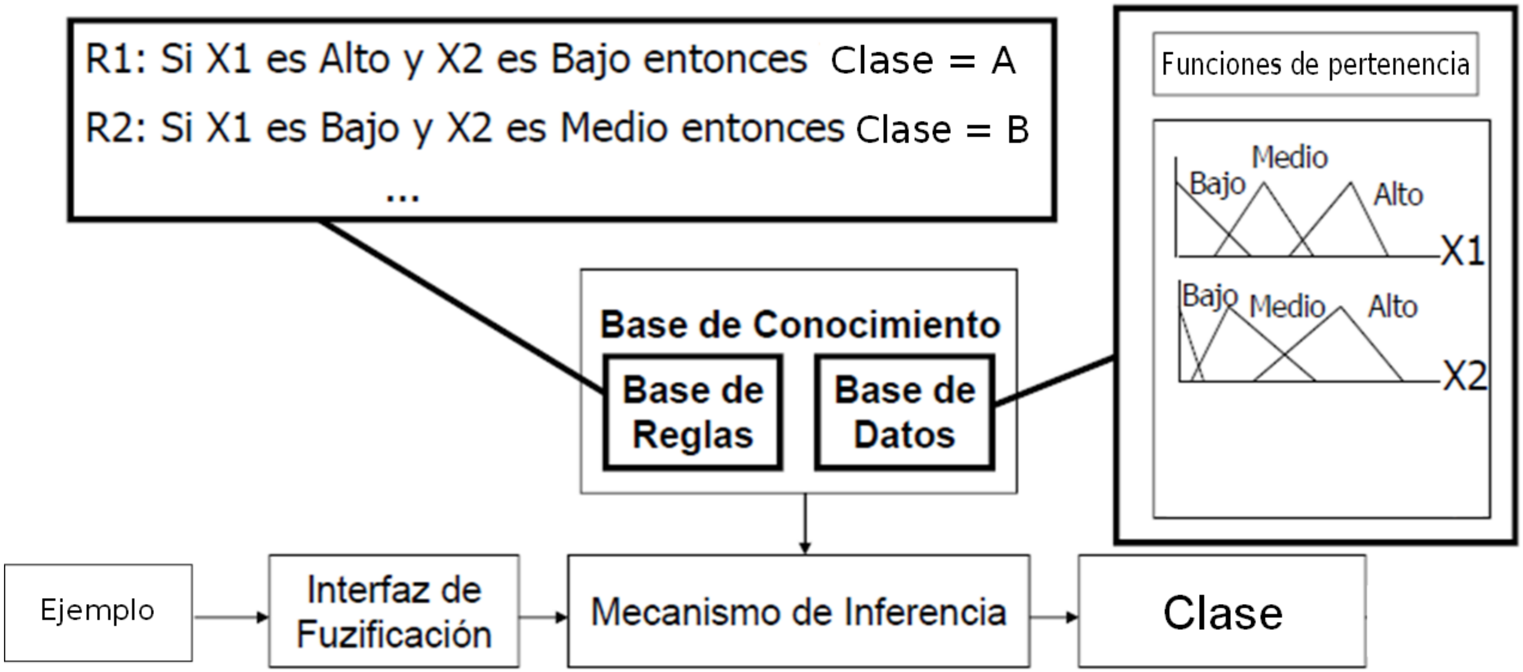
- Unión
  - Lógica booleana: ¿Pertenece el elemento a cualquier conjunto?
  - Lógica difusa: ¿Cuánto pertenece el elemento a cualquier conjunto?
  - Un ejemplo de t-conorma es el máximo (comúnmente utilizada)

$$\mu_{A \cup B}(x) = S[\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x)$$

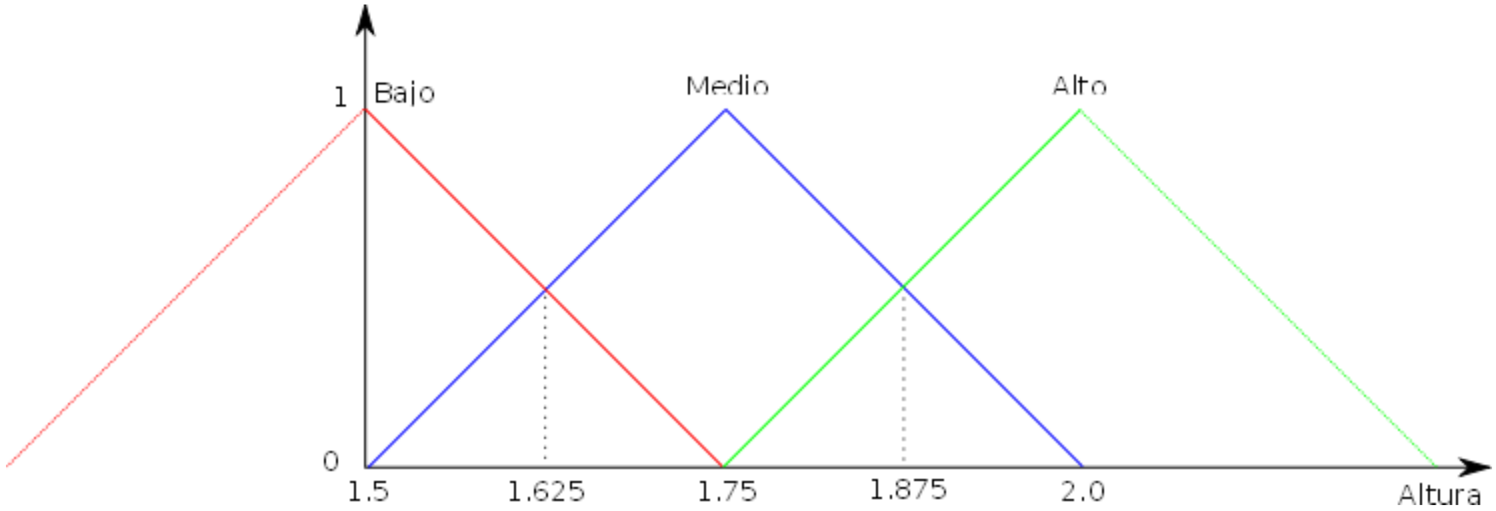


Sistemas de clasificación basados en reglas difusas

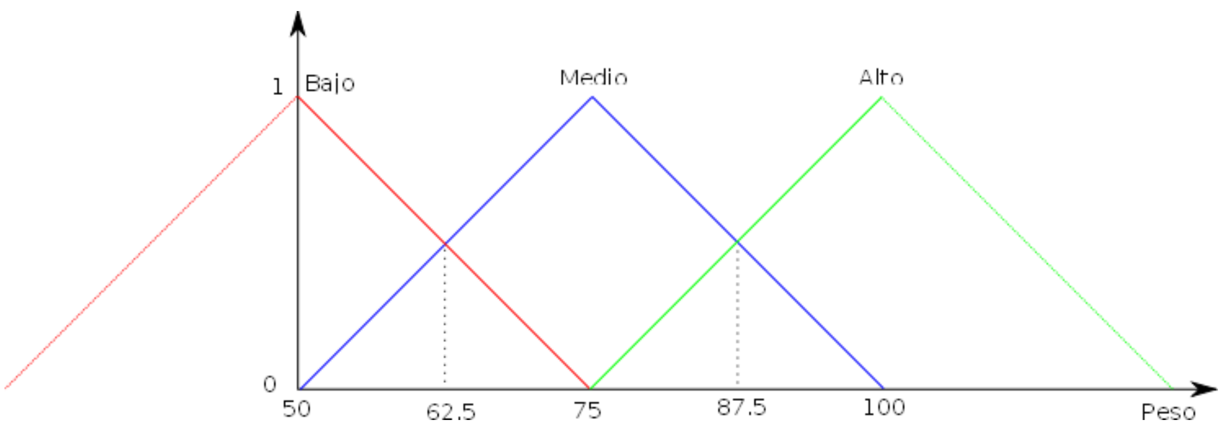
- Esquema general



Reglas difusas

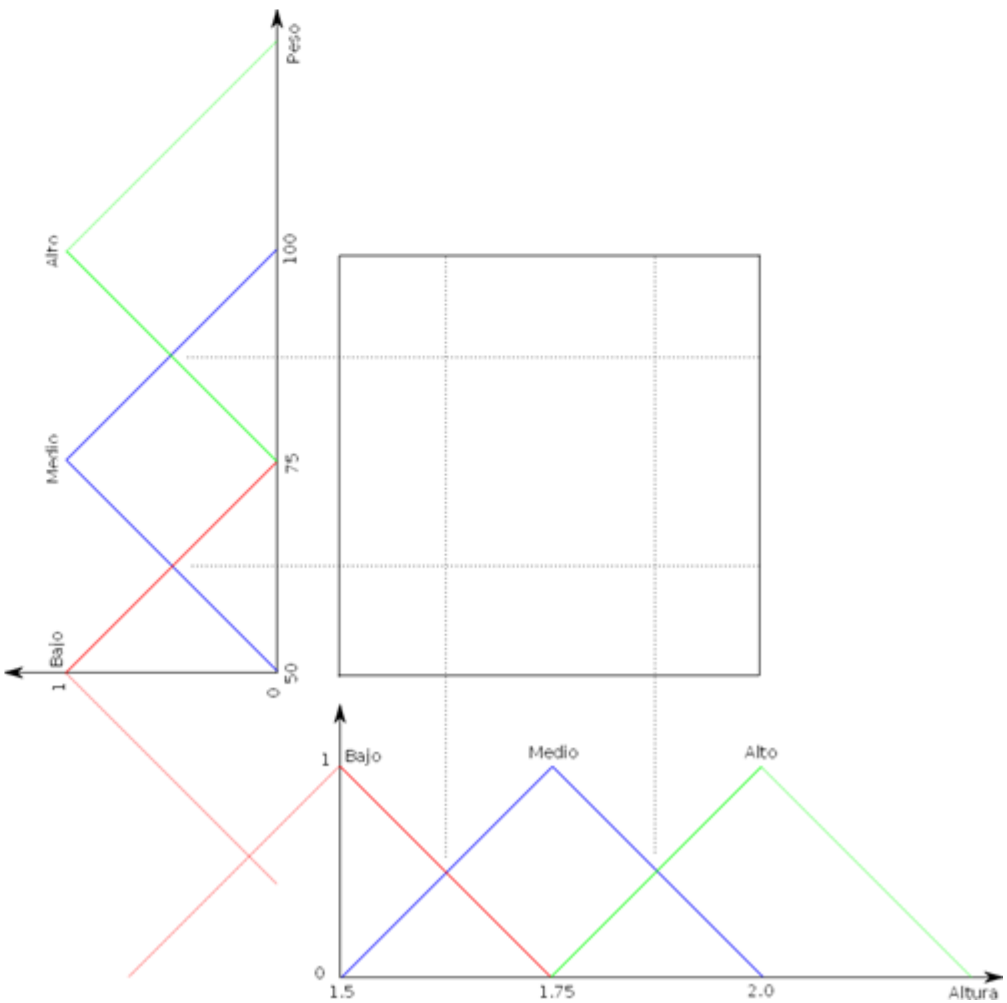


Altura = 1.55 → Mejor función pertenencia: Bajo

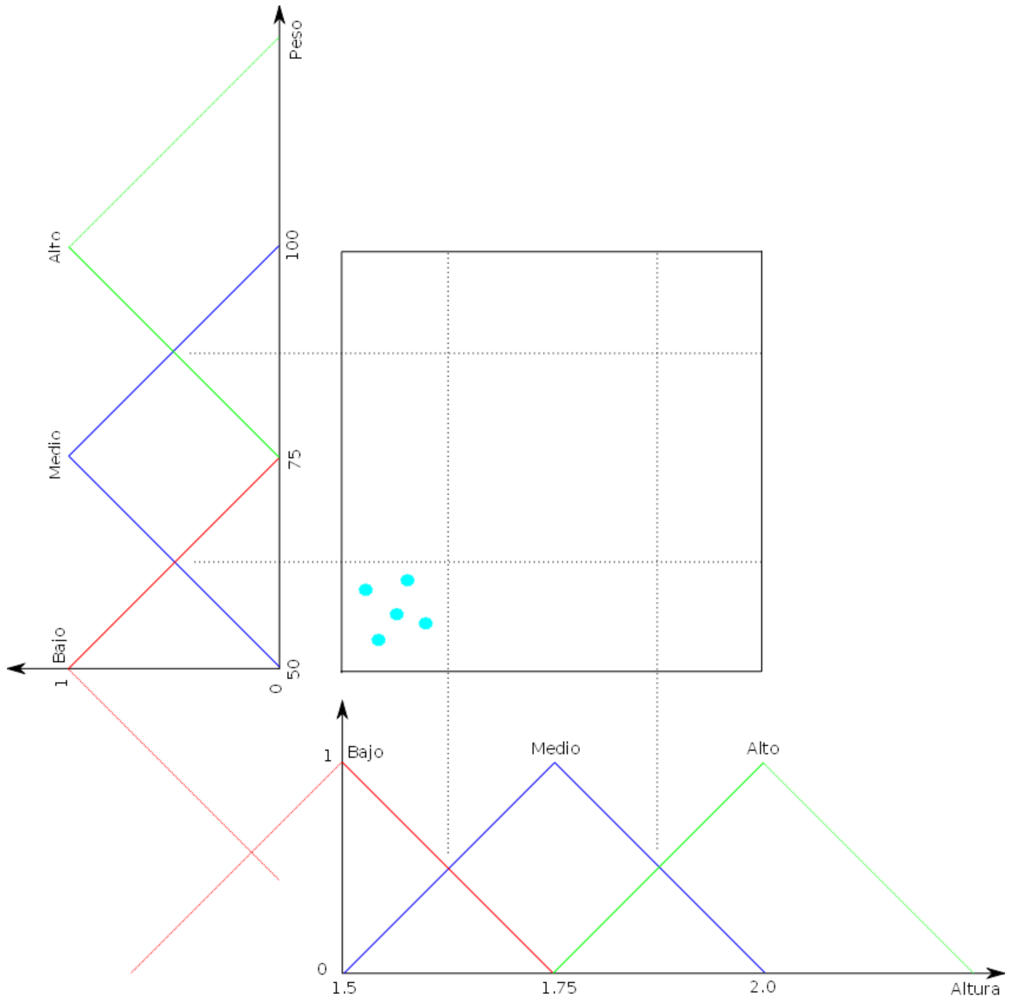


Peso = 55 → Mejor función pertenencia: Bajo

Reglas difusas: particionado del espacio de ejemplos

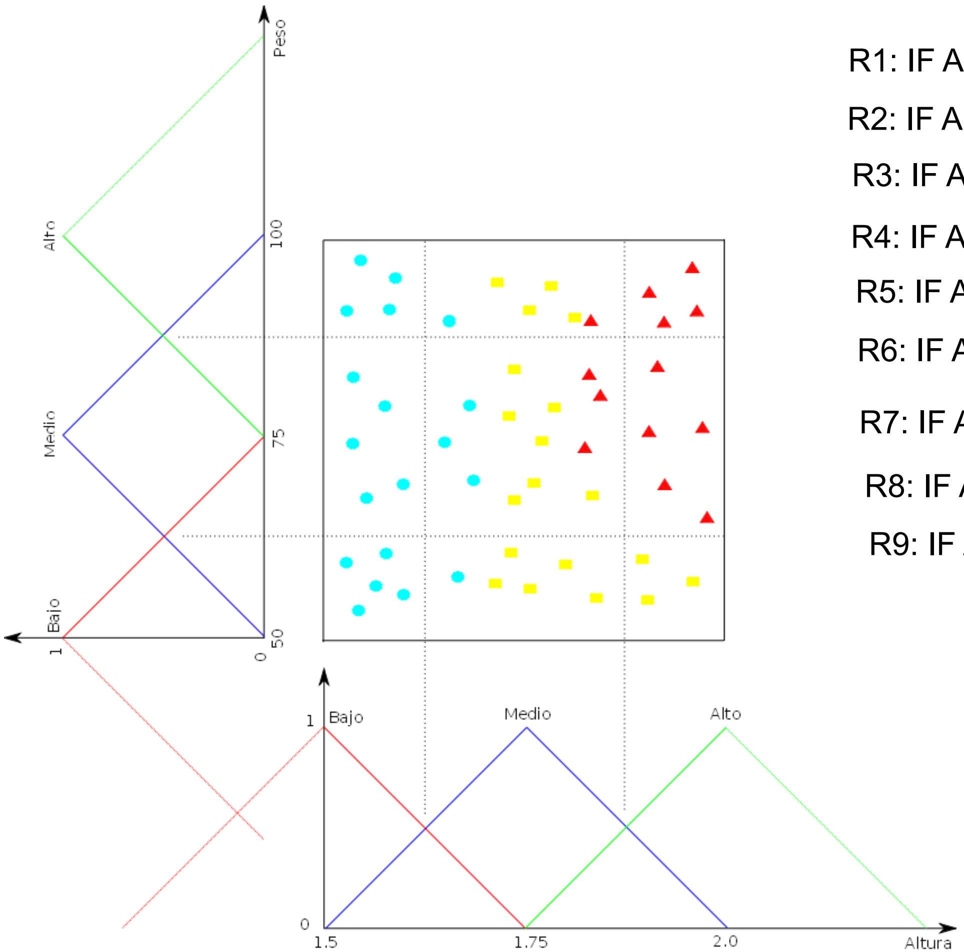


Reglas difusas



R1: IF Altura IS Bajo AND Peso IS Bajo THEN Clase =

Reglas difusas



R1: IF Altura IS Bajo AND Peso IS Bajo THEN Clase =

R2: IF Altura IS Bajo AND Peso IS Medio THEN Clase =

R3: IF Altura IS Bajo AND Peso IS Alto THEN Clase =

R4: IF Altura IS Medio AND Peso IS Bajo THEN Clase =

R5: IF Altura IS Medio AND Peso IS Medio THEN Clase =

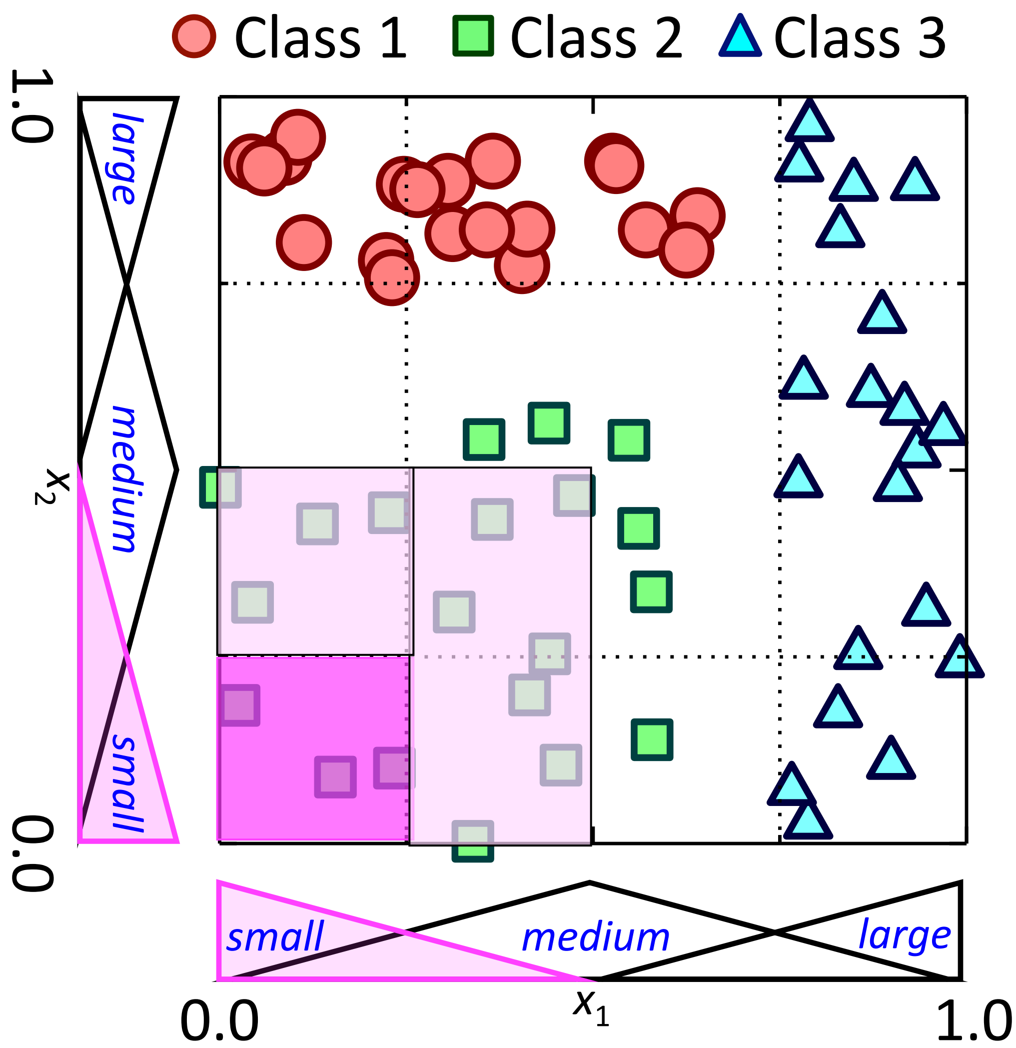
R6: IF Altura IS Medio AND Peso IS Alto THEN Clase =

R7: IF Altura IS Alto AND Peso IS Bajo THEN Clase =

R8: IF Altura IS Alto AND Peso IS Medio THEN Clase =

R9: IF Altura IS Alto AND Peso IS Alto THEN Clase =

Interpretabilidad de las reglas

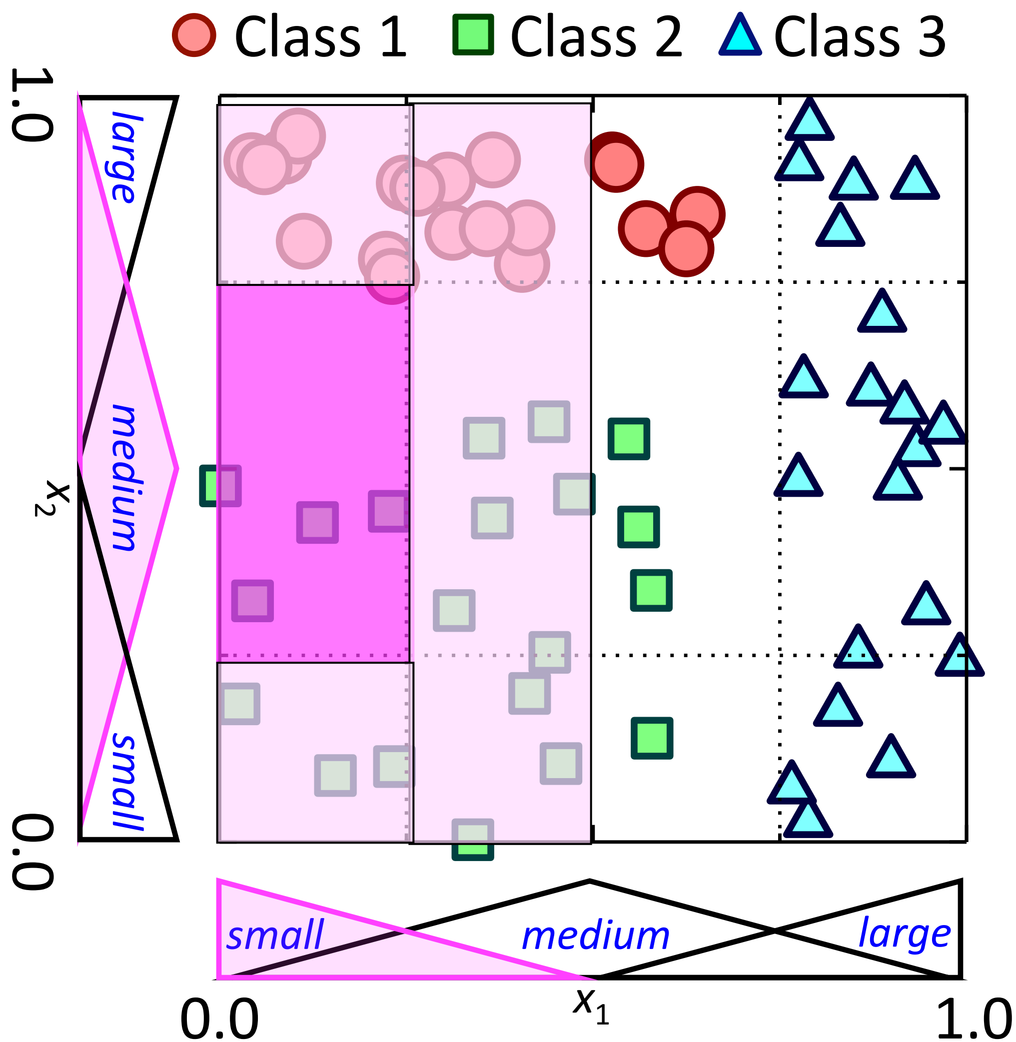


Reglas tipo 1

If x_1 is *small* and x_2 is *small* then Class 2

Interpretabilidad de las reglas

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

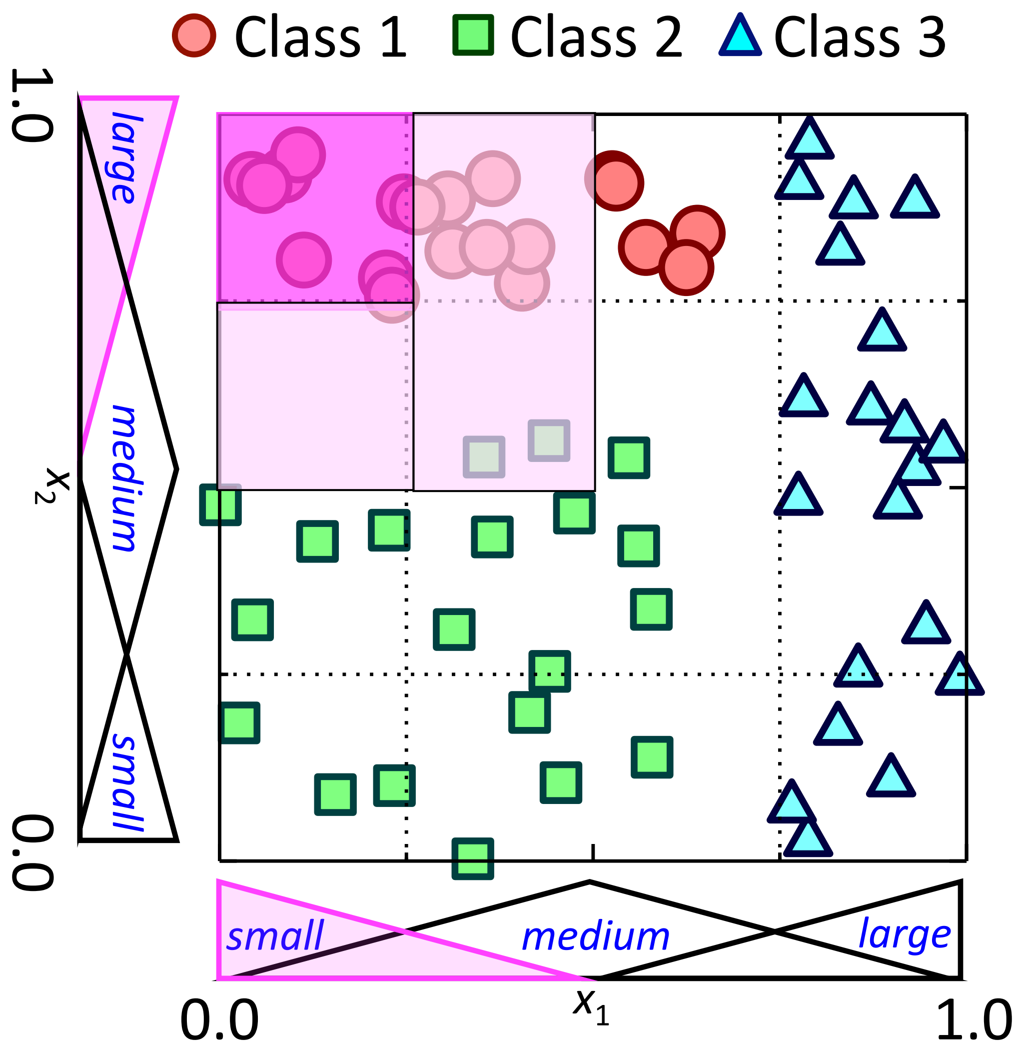


Reglas tipo 1

If x_1 is *small* and x_2 is *small*
then Class 2

If x_1 is *small* and x_2 is *medium*
then Class 2

Interpretabilidad de las reglas



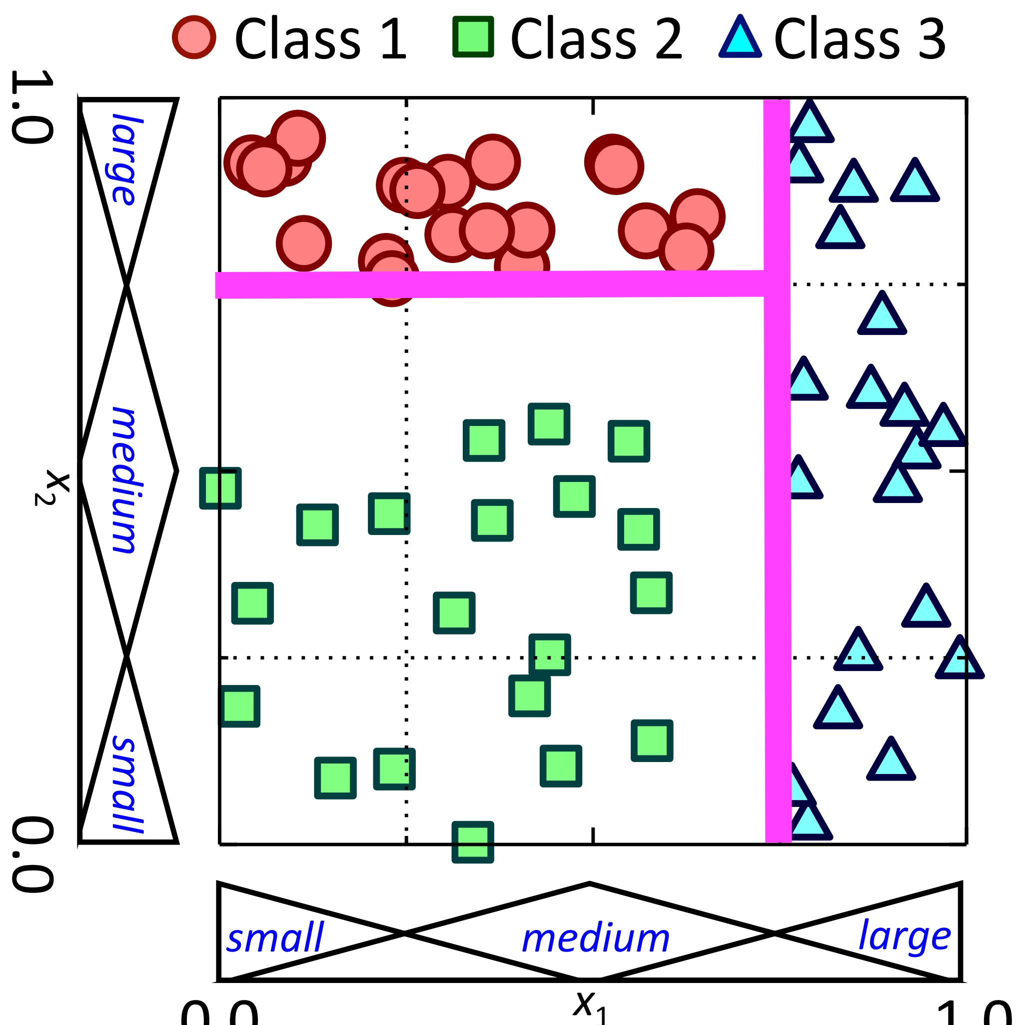
Reglas tipo 1

If x_1 is *small* and x_2 is *small* then Class 2

If x_1 is *small* and x_2 is *medium* then Class 2

If x_1 is *small* and x_2 is *large* then Class 1

Frontera de Clasificación



Reglas tipo 1

If x_1 is *small* and x_2 is *small*
then Class 2

If x_1 is *small* and x_2 is *medium*
then Class 2

If x_1 is *small* and x_2 is *large*
then Class 1

. . .

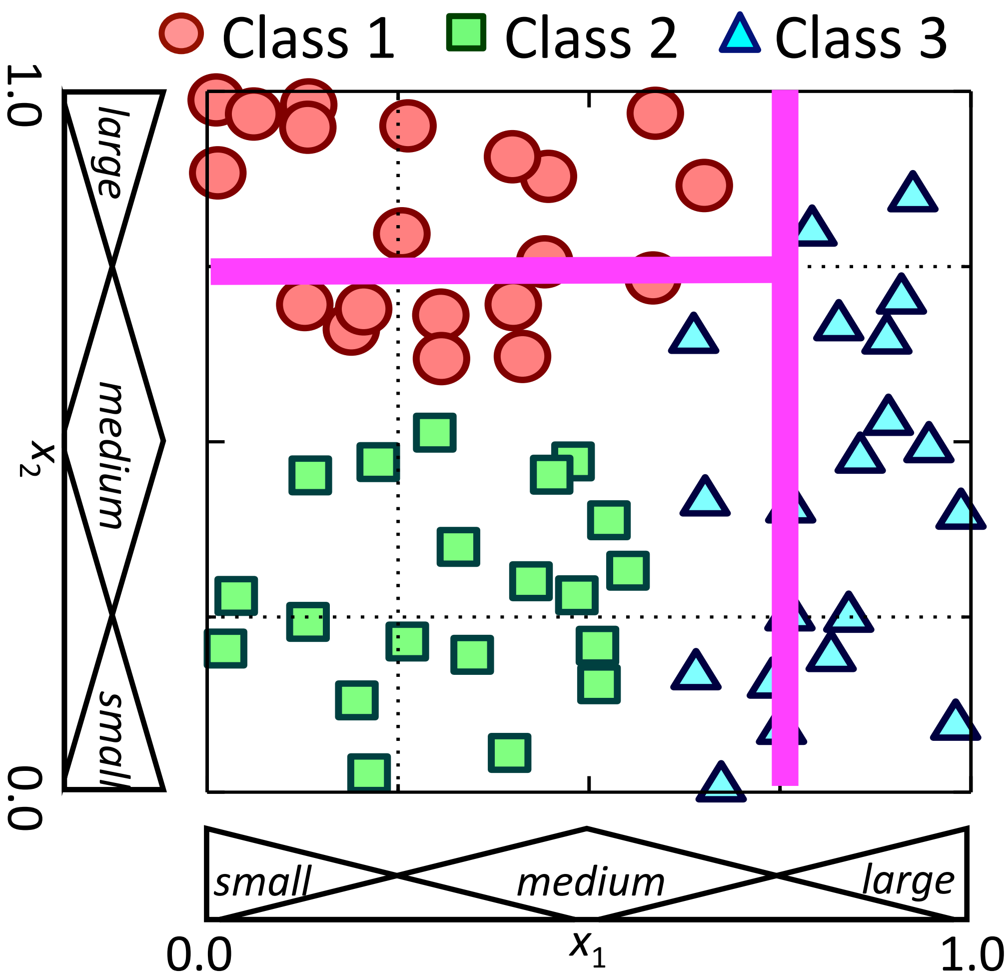
If x_1 is *large* and x_2 is *large*
then Class 3

Interpretabilidad alta

Fácil de entender!

Problema de la forma básica de las reglas

La forma básica no siempre obtiene un buen accuracy



Reglas tipo 1

If x_1 is *small* and x_2 is *small*
then Class 2

If x_1 is *small* and x_2 is *medium*
then Class 2

If x_1 is *small* and x_2 is *large*
then Class 1

. . .

If x_1 is *large* and x_2 is *large*
then Class 3

Interpretabilidad alta

Accuracy bajo

Base de reglas

- Almacena el conjunto de reglas difusas ($R = \{R_1, \ldots, R_L\}$) utilizadas para representar el conocimiento del problema

- Tipos de reglas:
  1. Regla R_j : Si x_1 es A_{j1} y . . . y x_n es A_{jn} entonces Clase = C_j
  2. Regla R_j : Si x_1 es A_{j1} y . . . y x_n es A_{jn} entonces Clase = C_j con RW_j
  3. Regla R_j : Si x_1 es A_{j1} y . . . y x_n es A_{jn} entonces $RW_1, \ldots, RW_M$

- Las reglas del tipo 3 generalizan a las 2 anteriores
  - Si $RW_j = 1$ y $RW_i = 0$ con $j \neq i$, entonces tenemos una regla tipo 1
  - Si $RW_j = RW_j$ y $RW_i = 0$ con $j \neq i$, entonces tenemos una regla tipo 2

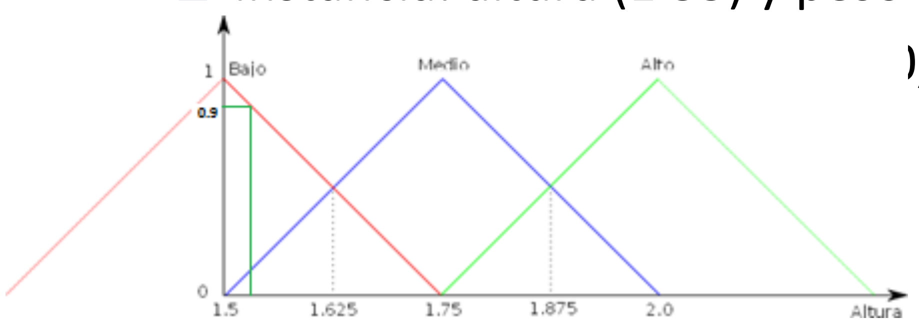
Grado de compatibilidad

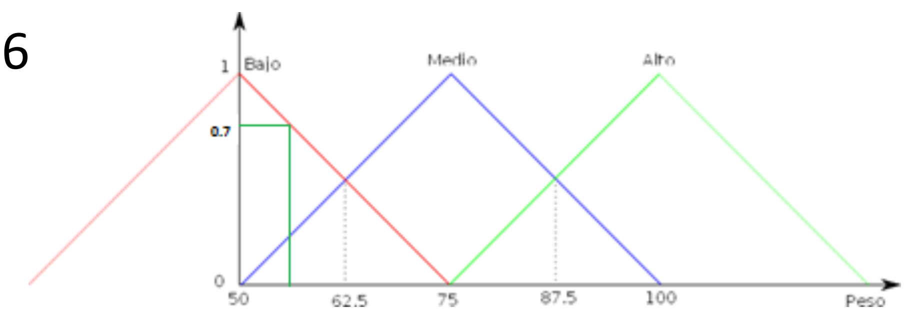
- Mide lo compatible que es un ejemplo con una regla. Sea la regla q y el ejemplo x_p, el grado de compatibilidad es

$$\mu_{A_q}(x_p) = T(\mu_{A_{q1}}(x_{p1}), \ldots, \mu_{A_{qn}}(x_{pn}))$$

- Si Altura es Bajo Y Peso es Bajo Entonces Clase =
  - Instancia: altura (1’55) y peso (55)

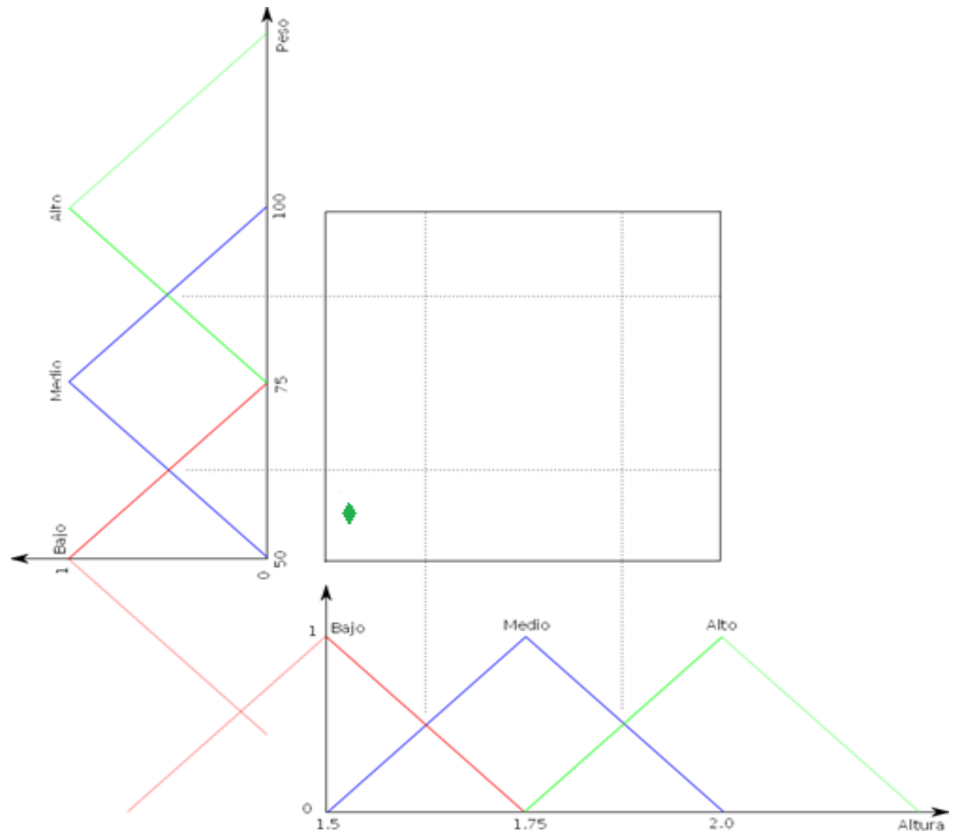
),6





Grado de compatibilidad

- Si Altura es Medio Y Peso es Medio Entonces Clase = ■
  - Instancia: altura (1'55) y peso (55)
  - Compatibilidad: T(0.1, 0.2) = 0.02 → T-norma producto
- Esta instancia nunca será compatible con una regla con un antecedente alto



Peso de las reglas [Ishibuchi01]

- Soporte de la regla q

$$Sop(A_q \rightarrow C_q) = \sum_{x_p \in Class\ C_q} \mu_{A_q}(x_p) \,/\, |N|$$

  - N es el número de ejemplos de entrenamiento

- Confianza de la regla q

$$Conf(A_q \rightarrow C_q) = \sum_{x_p \in Class\ C_q} \mu_{A_q}(x_p) \,/\, \sum_{p=1}^{N} \mu_{A_q}(x_p)$$

[Ishibuchi01] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 4, pp. 506–515, Aug. 2001

Peso de las reglas

- Confianza penalizada con la confianza media del resto de clases

$$Conf_2 = Conf(A_q \rightarrow C_q) - Conf_{media}$$

$$Conf_{media} = 1/M-1 * \sum_{h=1, h\neq C_q}^{M} Conf(A_q \rightarrow C_h)$$

- Confianza penalizada con la segunda mayor confianza

$$Conf_3 = Conf(A_q \rightarrow C_q) - Conf_{segunda}$$

$$Conf_{segunda} = \bigvee_{h=1, h\neq C_q}^{M} Conf(A_q \rightarrow C_h)$$

Peso de las reglas

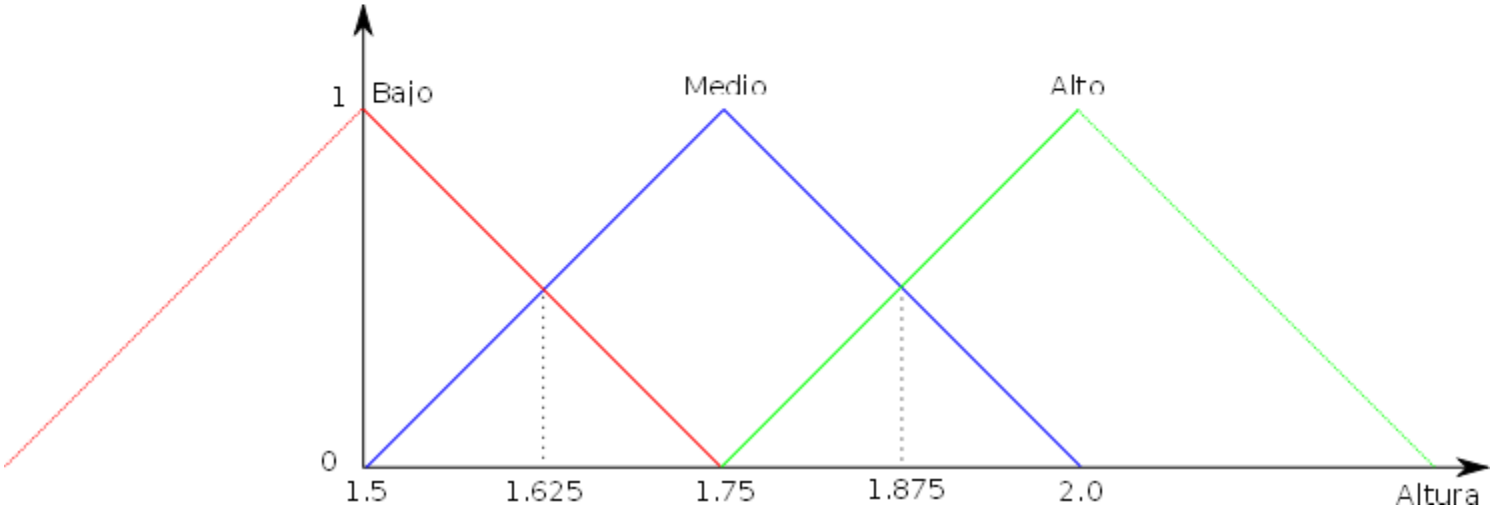
- Confianza penalizada con la suma de la confianza del resto de clases
  - Factor de certeza penalizado

$$Conf_4 = Conf(A_q \rightarrow C_q) - Conf_{suma}$$

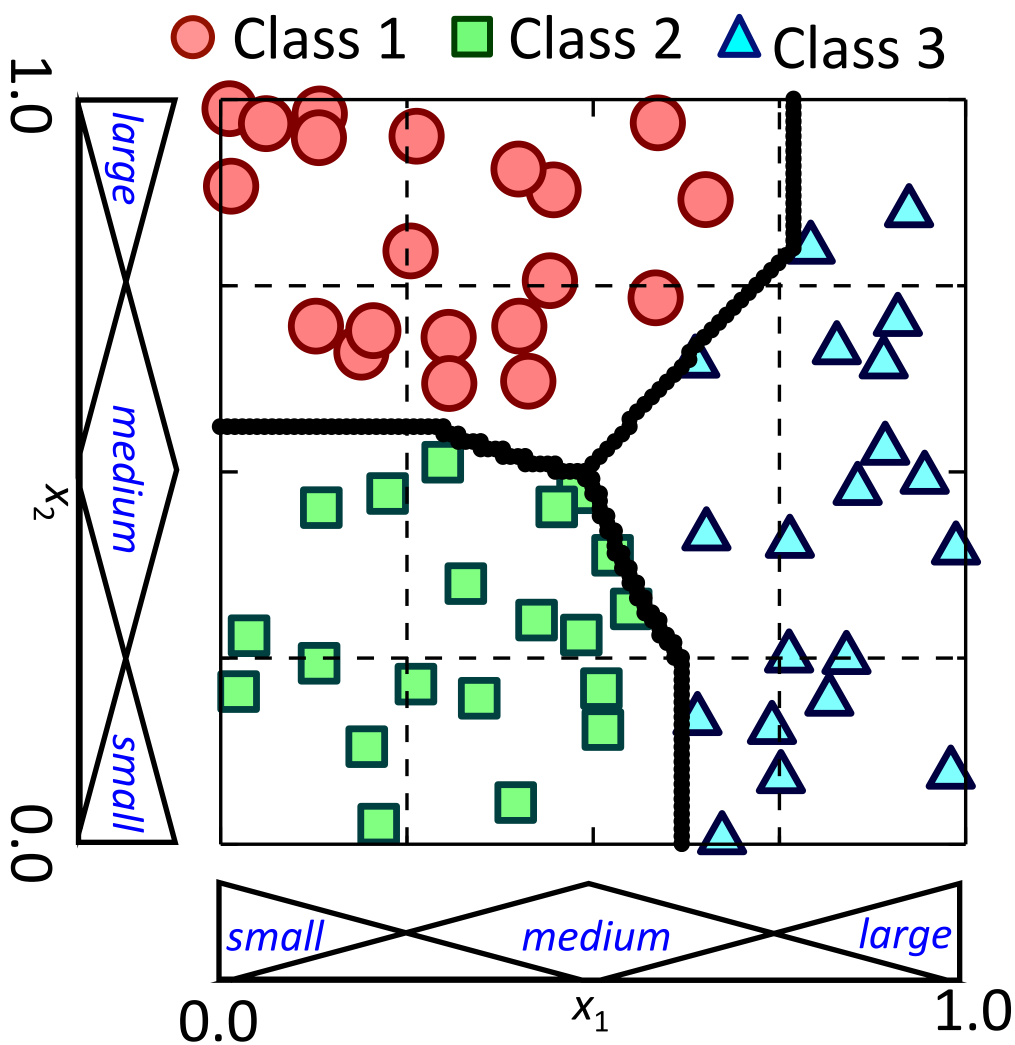
$$Conf_{suma} = \sum_{h=1,\ h \neq C_q}^{M} Conf(A_q \rightarrow C_h)$$

- Observaciones
  - Las 3 confianzas penalizadas son equivalentes si el número de clases es 2
  - En las confianzas penalizadas el resultado puede ser negativo por lo que esa regla no sería generada

Reglas difusas



Frontera de Clasificación



Regla tipo 1

If x_1 is *small* and x_2 is *medium*

then Class 2

Regla tipo 2

If x_1 is *small* and x_2 is *medium*

then Class 2 with 0.158

Algoritmo de generación de reglas de Chi[Chi96]

- Wang y Mendel proponen un método de generación de reglas difusas para problemas de control
- Chi y otros los extienden a problemas de clasificación
- Algoritmo
  1. Calcular las particiones difusas asociadas a las variables de entrada
  2. Generar una regla difusa para cada ejemplo de entrenamiento
     i. Calcular los grados de pertenencia del ejemplo a las diferentes funciones de pertenencia
     ii. Asignar el ejemplo a las funciones de pertenencia que devuelvan el mayor grado de pertenencia para cada variable
     iii. Generar la regla. El antecedente es la conjunción de las funciones de pertenencia y el consecuente es la clase del ejemplo
     iv. Si las reglas son del tipo 2 o 3, calcular el peso de la regla

- Reglas con el mismo antecedente y diferente clase: conservar la de mayor peso

[Chi96] Z. Chi, H. Yan, T. Pham, Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition, World Scientific, 1996

Método de razonamiento difuso[Cordón99]

- Procedimiento de inferencia que usa la información de la base de conocimiento para predecir la clase de un ejemplo no clasificado
  - Clasificación de ejemplos
- Determina la forma de utilizar la información contenida en las reglas

[Cordón99] O. Cordón, M. J. del Jesus, and F. Herrera, "Aproposal on reasoning methods in fuzzy rule-based classification systems," *Int. J. Approx. Reason.*, vol. 20, no. 1, pp. 21–45, 1999

Método de razonamiento difuso

- Sea = {, …, } el ejemplo a clasificar. = {, …, } el conjunto de clases y = {, …, } el conjunto de reglas.
- Los pasos del MRD son:
  - Grado de compatibilidad

$$R^{k}(e)=T(\mu_{A_{k1}}(e_{1}),\ \dots,\ \mu_{A_{kn}}(e_{n})),\quad k=\{1,\ \dots,\ L\}$$

  - Grado de asociación

$$b_{j}^{k}(e)=h(R^{k}(e),\ RW_{j}^{k}),\quad k=\{1,\ \dots,\ L\},\quad j=\{1,\ \dots,\ M\}$$

  - Ponderación

$$B_{j}^{k}(e)=g(b_{j}^{k}(e)),\quad k=\{1,\ \dots,\ L\},\quad j=\{1,\ \dots,\ M\}$$

  - Grado de asociación por clases

$$a_{j}=f(B_{j}^{k}(e)\,|\,B_{j}^{k}(e)>0,\ k=\{1,\ \dots,\ L\})$$

Método de razonamiento difuso

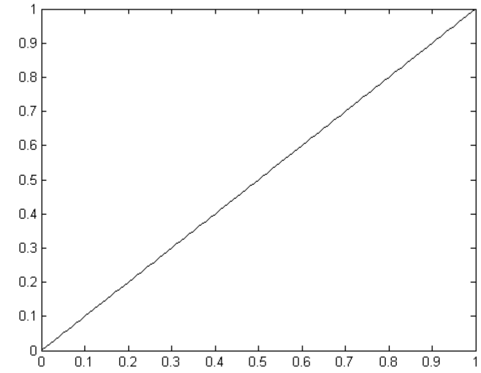
- Grado de compatibilidad
  - T-normas
    - Mínimo
    - Producto

- Grado de asociación
  - Implicaciones
    - Mínimo
    - Producto

Método de razonamiento difuso

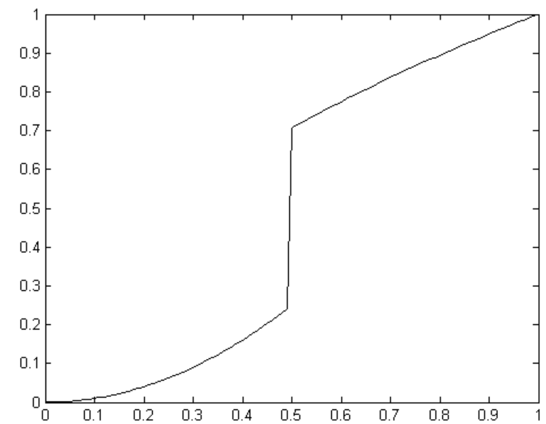
- Función de ponderación
  - Sin ponderar

$$g_1(x) = x \quad \forall x \in [0, 1]$$



  - Penalizar los valores menores a 0.5 y potenciar los mayores

$$g_2(x) = \begin{cases} x^2, & \text{si } x < 0.5 \\ \sqrt{x}, & \text{si } x \geq 0.5 \end{cases}$$



Método de razonamiento difuso

- Grado de asociación por clases

$$a_j = f(B_j^k(e) \mid B_j^k(e) > 0,\ k = \{1, \ldots, L\})$$

  - Funciones de agregación, f, (consideran 1 regla)
    - Máximo (MRD de la regla ganadora)
  - Funciones de agregación, f, (consideran todas las reglas disparadas)
    - Suma aditiva (MRD de la combinación aditiva)
    - Media aritmética
  - Funciones de agregación, f, (seleccionan reglas)
    - OWA

CLASIFICACIÓN Y MACHINE LEARNING CON TÉCNICAS DIFUSAS

Mikel Galar
Universidad Pública de Navarra

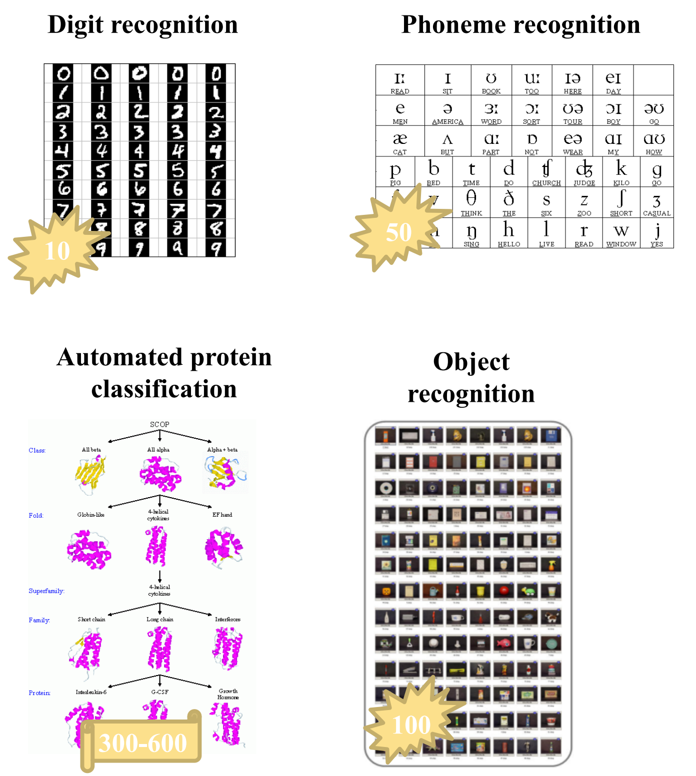
EVIA 2016

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Ensembles basados en descomposición

Introducción

- **Clasificación**
  - **Dos tipos** de problemas
    - **Binarios**
      - Diagnóstico médico (si / no)
    - **Multi-clase**
      - Reconocimiento de dígitos (1, 2,…)
  - Problemas **binarios** generalmente **más sencillos**
  - Algunos clasificadores no soportan múltiples clases
    - SVM, PDFC…

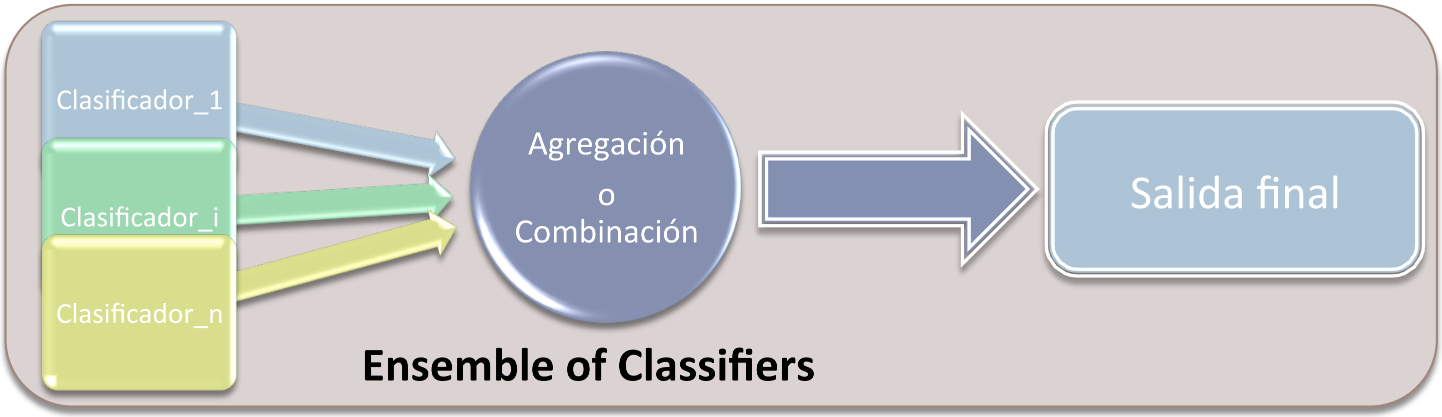


Ensembles basados en descomposición

- **Objetivo**
  - Afrontar problemas multi-clase con clasificadores binarios
- **Funcionamiento**
  - Descomponer un problema multi-clase
    - En **problemas binarios** (más sencillos de resolver)
  - Aprender **un clasificador para cada subproblema**
  - Para clasificar una **nueva instancia**
    - **Agregar las salidas** de los clasificadores base

Ensembles basados en descomposición

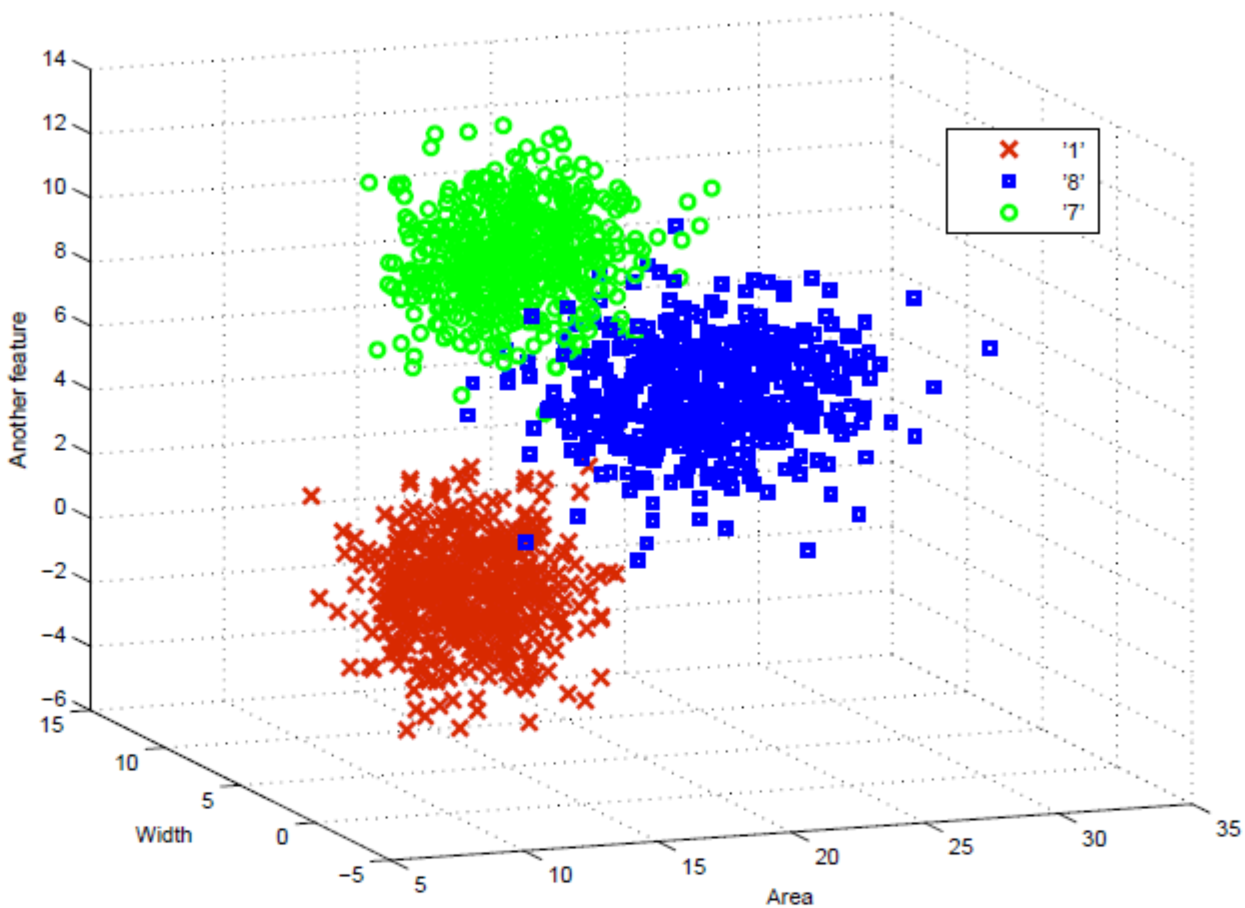
- **Paradigma divide y vencerás**[Gal11]



[Gal11] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera. An overview of ensemble methods for binary classifiers in multiclass problems: Experimental study on one-vs-one and one-vs-all schemes. Pattern Recognition, 44(8):1761–1776, 2011

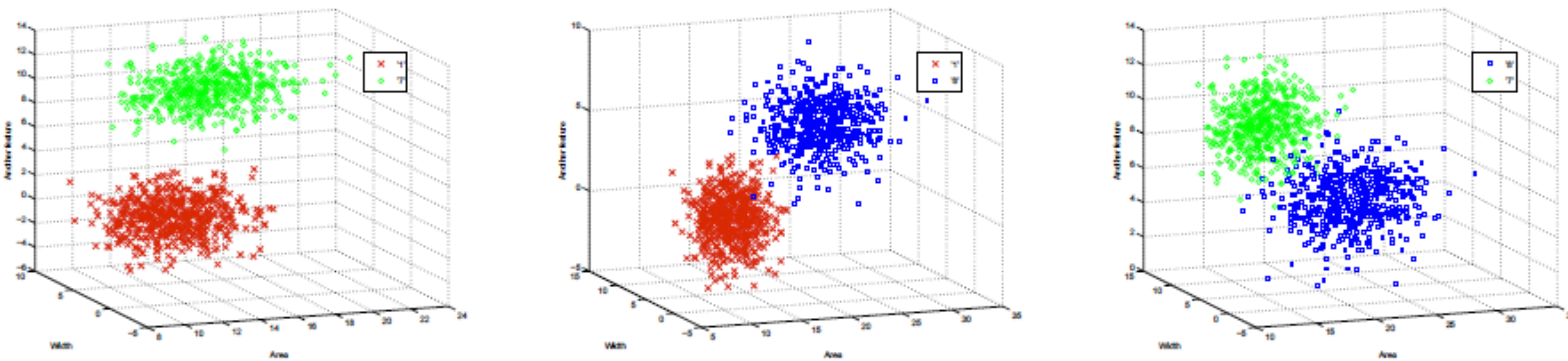
Ensembles basados en descomposición

- **Ejemplo de problema multi-clase**



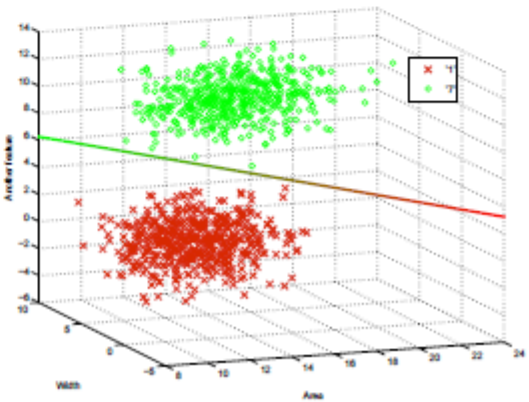
One-vs-One (OVO)

- **Divide el problema multi-clase en tantos problemas como pares de clases**

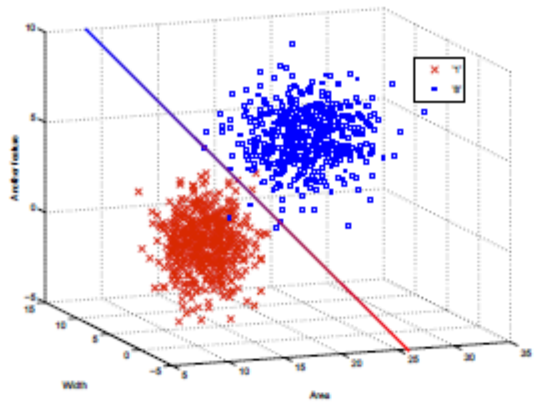


(a) '1' vs. '7' (b) '1' vs. '8' (c) '8' vs. '7'

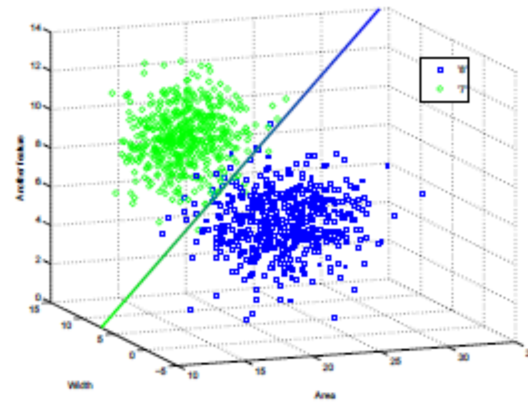
One-vs-One (OVO)



(a) '1' vs. '7'



(b) '1' vs. '8'



(c) '8' vs. '7'

One-vs-One (OVO)

- **Representación de las salidas**
  - **Matriz de votos**
  - $r_{ij} \in [0,1]$ confianza a favor de la clase i frente a la j
  - $r_{ji} = 1 - r_{ij}$

$$R = \begin{pmatrix} - & r_{12} & \cdots & r_{1m} \\ r_{21} & - & \cdots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \cdots & - \end{pmatrix}$$

  - Con la mayoría de clasificadores difusos
    - **Necesario normalizar**

$$\hat{r}_{ij} = \begin{cases} \dfrac{r_{ij}}{r_{ij} + r_{ji}} & \text{if } r_{ij} \neq 0 \text{ or } r_{ji} \neq 0 \\ 0.5 & \text{if } r_{ij} = r_{ji} = 0 \end{cases}$$

One-vs-One (OVO)

- **Agregaciones más comunes**
  - **Voto (VOTE)**[Friedman96]
    - Cada clasificador vota por la clase predicha

$$Class = \arg \max_{i=1,\dots,m} \sum_{1 \le j \ne i \le m} s_{ij},$$

    - donde s_{ij} es 1 si $r_{ij} > r_{ji}$ y 0 en otro caso
  - **Voto ponderado (WV)**
    - Cada clasificador vota con cierta confianza a cada clase

$$Class = \arg \max_{i=1,\dots,m} \sum_{1 \le j \ne i \le m} r_{ij}$$

  - En ambos casos, la clase con mayor confianza es la que se predice

[Friedman96] J. H. Friedman, Another approach to polychotomous classification, Tech. rep., Department of Statistics, Stanford University (1996).

One-vs-One (OVO)

- **Agregaciones basadas en la teoría difusa**
  - **Learning Valued Preference for Classification (LVPC)** [Hüllermeier08, Huhn09]
    - Matriz de votos = Relación de preferencia fuzzy
    - Se **descompone en tres relaciones** diferentes
      - Preferencia estricta $$\begin{aligned} P_{ij} &= r_{ij} - min\{r_{ij}, r_{ji}\} \\ P_{ji} &= r_{ji} - min\{r_{ij}, r_{ji}\} \end{aligned}$$
      - Conflicto $C_{ij} = min\{r_{ij}, r_{ji}\}$
      - Ignorancia $I_{ij} = 1 - max\{r_{ij}, r_{ji}\}$
    - La decisión final se basa en un **voto usando las tres relaciones**

$$Class = arg \max_{i=1,\dots,m} \sum_{1 \leq j \neq i \leq m} P_{ij} + \frac{1}{2} C_{ij} + \frac{N_i}{N_i + N_j} I_{ij}$$

[Hüllermeier08] E. Hüllermeier and K. Brinker. Learning valued preference structures for solving classification problems. Fuzzy Sets and Systems, 159(18):2337–2352, 2008.

One-vs-One (OVO)

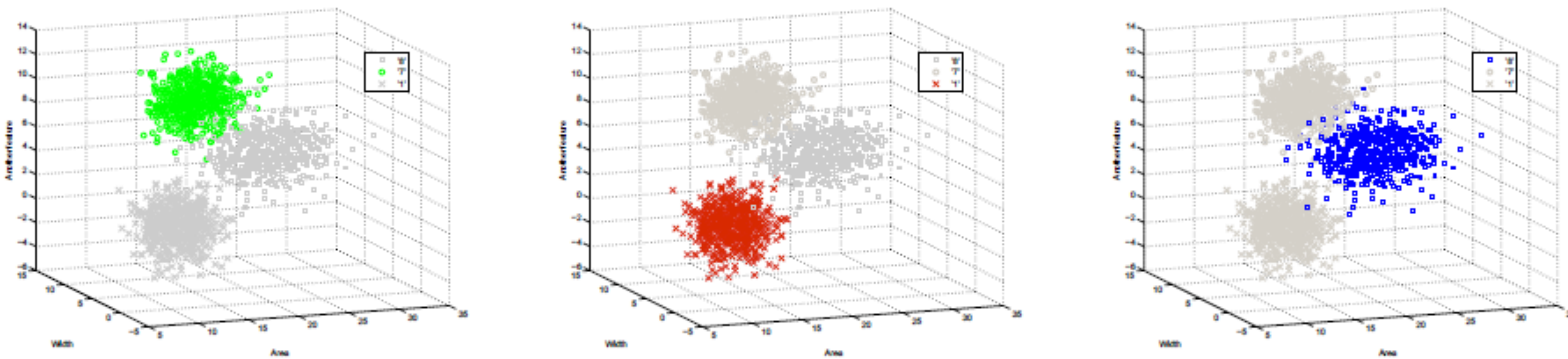
- **Agregaciones basadas en la teoría difusa**
  - **Non-Dominance Criterion (ND)** [Fernandez10]
    - Toma de decisión y modelado de la preferencia [Orlovsky78]
    - Matriz de votos = Relación de preferencia fuzzy
      - $r_{ji} = 1 - r_{ij}$, sino → normalizar $\bar{r}_{ij} = \frac{r_{ij}}{r_{ij} + r_{ji}}$
      - **Calcular el elemento con máximo grado de no dominancia**
        - Construir la relación de preferencia estricta $r'_{ij} = \begin{cases} \bar{r}_{ij} - \bar{r}_{ji}, & \text{when } \bar{r}_{ij} > \bar{r}_{ji} \\ 0, & \text{otherwise.} \end{cases}$
        - Calcular el grado de no dominancia $ND_i = 1 - \sup_{j \in C}[r'_{ji}]$
          - *El grado en el que la clase i está dominada por ninguna de las clases restantes*
        - Output $Class = arg \max_{i=1,\ldots,m} \{ND_i\}$

[Fernandez10] A. Fernández, M. Calderón, E. Barrenechea, H. Bustince, F. Herrera, Solving mult-class problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations, Fuzzy Sets and System *161:23 (2010) 3064-3080,*

[Orlovsky78] S. A. Orlovsky, Decision-making with a fuzzy preference relation, Fuzzy Sets and Systems 1 (3) (1978) 155–167.

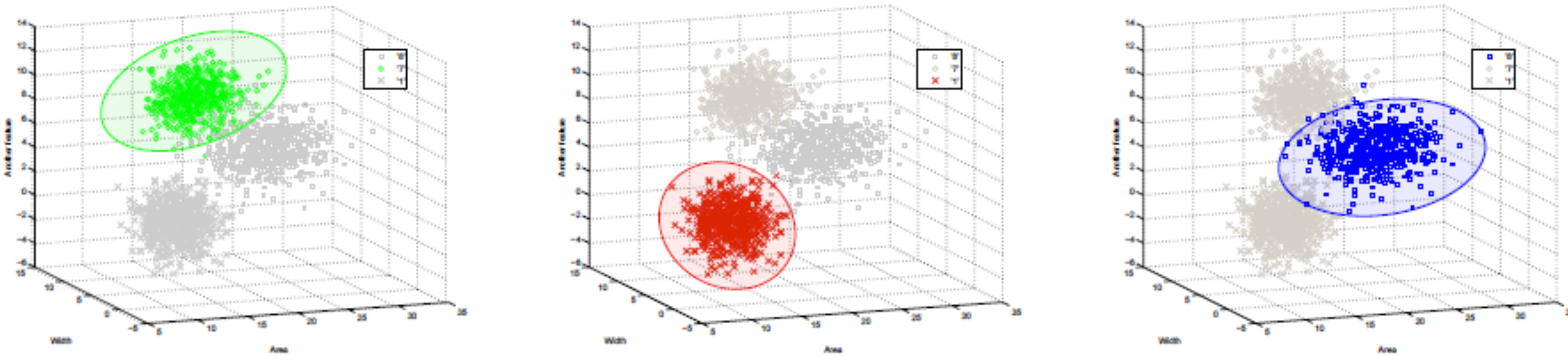
One-vs-All (OVA)

- **Divide el problema multi-clase en tantos problemas como clases**



(a) '7' vs. '1' and '8' (b) '1' vs. '7' and '8' (c) '8' vs. '1' and '7'

One-vs-All (OVA)



(a) '7' vs. '1' and '8'   (b) '1' vs. '7' and '8'   (c) '8' vs. '1' and '7'

One-vs-All (OVA)

- **Representación de las salidas**
  - **Vector de votos**
  - $r_i \in [0,1]$ confianza a favor de la clase i

$$R = (r_1, r_2, \ldots, r_i, \ldots, r_m)$$

  - **Normalización**
    - Respecto al vector $\overline{R}$ de confianzas para la clase negativa

$$\hat{r}_i = \frac{r_i}{r_i + \overline{r}_i}$$

  - **Agregación**
    - **Máximo**
      - Clase de salida = clase con mayor confianza

OVO vs. OVA

- **Ventajas OVO**
  - Problemas **más sencillos**
  - Problemas **más pequeños**
  - Computacionalmente **más rápido**
  - Generalmente, **más preciso**
- **Desventajas OVO**
  - **Región no clasificable** (Voto)
  - Clasificadores **no competentes**
  - **Crecimiento cuadrático** en el número de clasificadores

OVO vs. OVA

- **Ventajas OVA**
  - Utiliza **todos los ejemplos**
    - **No hay** clasificadores **no competentes**
  - **Agregaciones más simples**
  - **Crecimiento lineal** en el número de clasificadores
- **Desventajas OVA**
  - Problemas **no balanceados**
  - Problemas **más complejos**
  - Computacionalmente **más costoso**

OVO vs. OVA con clasificadores difusos

Marco experimental

Table 2
Summary of the features of the datasets used in the experimental study.

Id.	Dataset	#Ex.	#Atts.	#Num.	#Nom.	#Class.
aut	autos	159	25	15	10	6
bal	balance	625	4	4	0	3
cle	cleveland	297	13	13	0	5
con	contraceptive	1473	9	6	3	3
eco	ecoli	336	7	7	0	8
gla	glass	214	9	9	0	7
hay	hayes-roth	132	4	4	0	3
iri	iris	150	4	4	0	3
new	newthyroid	215	5	5	0	3
pag	pageblocks	548	10	10	0	5
pen	penbased	1100	16	16	0	10
sat	satimage	643	36	36	0	7
seg	segment	2310	19	19	0	7
shu	shuttle	2175	9	9	0	5
tae	tae	151	5	3	2	3
thy	thyroid	720	21	21	0	3
veh	vehicle	846	18	18	0	4
vow	vowel	990	13	13	0	11
win	wine	178	13	13	0	3
yea	yeast	1484	8	8	0	10

Table 3
Setup of the methods parameters.

Algorithm	Parameters
CHI	Num. of linguistic labels per variable: 3
	Rule weight: certainty factor
SLAVE	Num. of linguistic labels per variable: 5
	Number of individuals: 100
	Mutation probability: 0.01
	Max. iterations without change: 500
FURIA	Num. of optimizations: 2
	Num. of folds: 3
FARC-HD	Num. of linguistic labels per variable: 5
	Minimum support: 0.05
	Minimum confidence: 0.8
	Maximum depth: 3
	Parameter k: 2
	Evaluations: 20,000
	Number of individuals: 50
	α parameter: 0.02
	Bits per gen: 30
	Rule weight: certainty factor

OVO vs. OVA con clasificadores difusos

Resultados experimentales

Table 4
Average accuracy rate obtained in testing by each method.

	CHI	SLAVE	FURIA	FARC-HD
Baseline	**75.07**	**76.71**	**80.56**	**80.37**
OVA	**73.76**	**69.91**	80.39	79.92
OVO^{ND}	**77.10**	**77.41**	81.97	81.45
OVO^{VOTE}	**77.90**	**77.73**	82.37	81.52
OVO^{LVPC}	**77.93**	**71.72**	**82.52**	**79.77**
OVO^{WV}	**78.11**	**72.23**	**82.61**	80.19
OVO^{WinWV}	**78.19**	**76.07**	82.44	81.50

OVO vs. OVA con clasificadores difusos

Resultados experimentales

- **Número de reglas / Número de antecedentes**

Table 9
Average number of rules and antecedents.

		Avg. rules	Avg. antecedents
CHI	Baseline	170.43	12.40
	OVA	155.49	12.40
	OVO	86.94	12.40
SLAVE	Baseline	18.47	4.33
	OVA	3.78	2.36
	OVO	4.14	2.51
FURIA	Baseline	16.54	2.76
	OVA	7.95	2.05
	OVO	4.50	1.58
FARC-HD	Baseline	32.67	2.34
	OVA	13.03	1.76
	OVO	8.55	1.61

CLASIFICACIÓN Y MACHINE LEARNING CON TÉCNICAS DIFUSAS

EVIA 2016

Alberto Fernández Hilario
Universidad de Granada

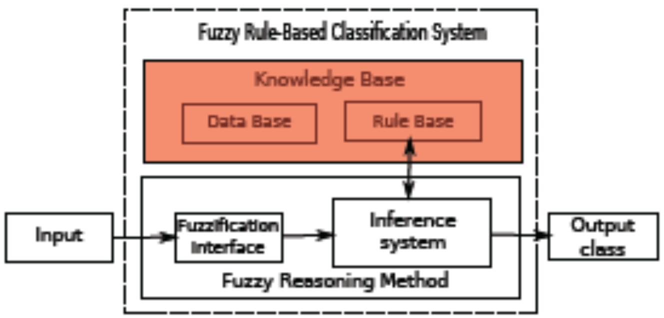
Modelos y escenarios avanzados para clasificación con sistemas difusos

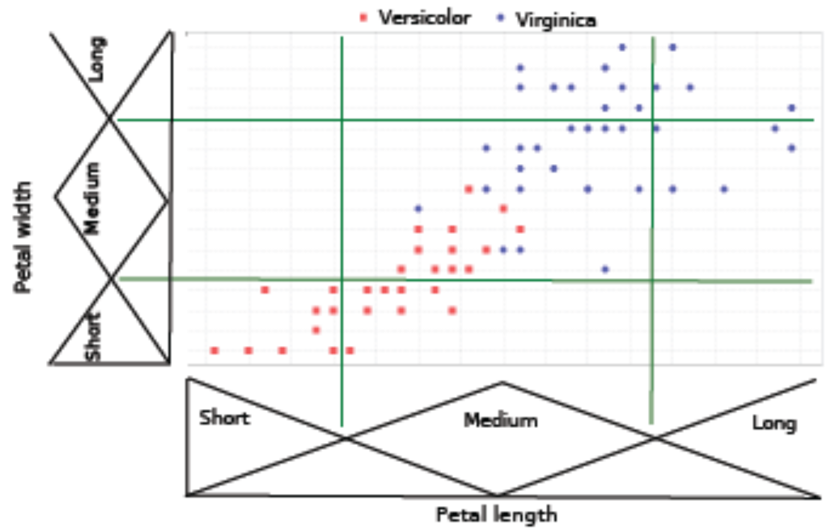
Índice

Índice

1. Interval-valued Fuzzy Rule Based Systems
    1. How FRBCS address difficult problems
    2. Improving the performance of FRBCS using Interval-Valued Fuzzy Sets
    3. Case study with IVTURS
2. Classification with Imbalanced Data
3. Fuzzy Systems in Big Data and Data Science

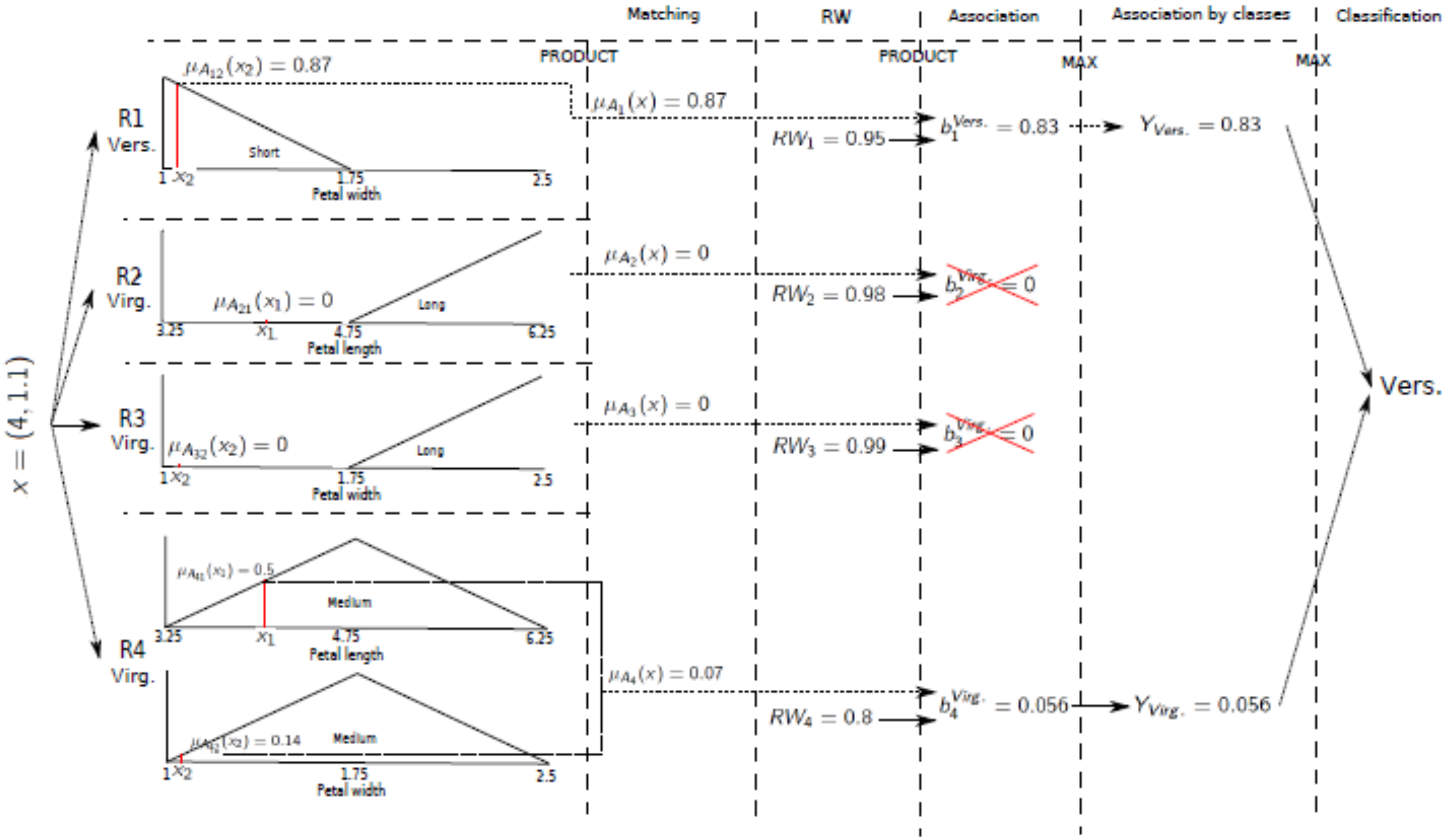
How FRBCS address difficult problems





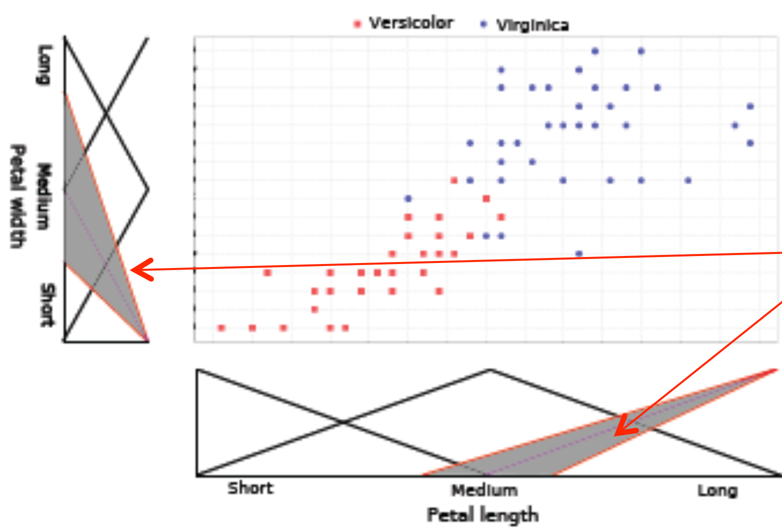
R_1: If *Width* is *Short* then *Class* = *Versicolor*
R_2: If *Length* is *Long* then *Class* = *Virginica*
R_3: If *Width* is *Long* then *Class* = *Virginica*
R_4: If *Length* is *Medium* and *Width* is *Medium* then *Class* = *Virginica*

How FRBCS address difficult problems



How FRBCS address difficult problems

- Two problems of FRBCSs:
  - Difficulty in the choice of membership functions.
    - Degree of ignorance (unknowledge).

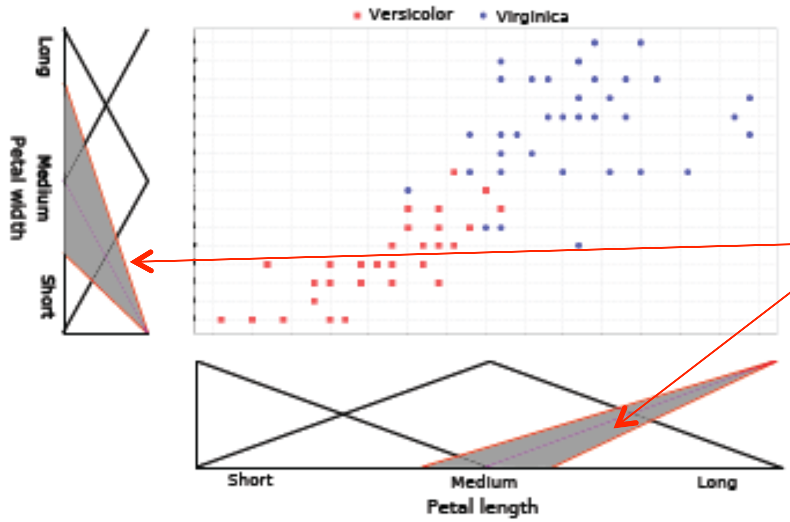


Degree of ignorance on the exact Membership degree.
It may be useful to consider and model it to improve the FRBCS

  - Suboptimal performance of the initial system.
    - Standard initial values (suboptimal) of the system' parameters.

How FRBCS address difficult problems

- Two problems of FRBCSs:
  - Difficulty in the choice of membership functions.
    - Degree of ignorance (unknowledge).



We can consider the interval length to represent and model the ignorance.

  - Suboptimal performance of the initial system.
    - Standard initial values (suboptimal) of the system' parameters.

Índice

1. Interval-valued Fuzzy Rule Based Systems
   1. How FRBCS address difficult problems
   2. Improving the performance of FRBCS using Interval-Valued Fuzzy Sets
   3. Case study with IVTURS
2. Classification with Imbalanced Data
3. Fuzzy Systems in Big Data and Data Science

Improving the performance of FRBCS using Interval-Valued Fuzzy Sets

- **Interval-valued fuzzy sets**: to model de degree of ignorance related to the definition of the membership functions.
- **Genetic algorithms**: to optimize the values of the model's parameters.
- With this aim, it is necessary to:
  - Define methods for constructing interval-valued fuzzy sets.
  - Extend the fuzzy reasoning methods.
  - Design new genetic tuning approaches.

Interval-Valued Fuzzy Sets (IVFS)

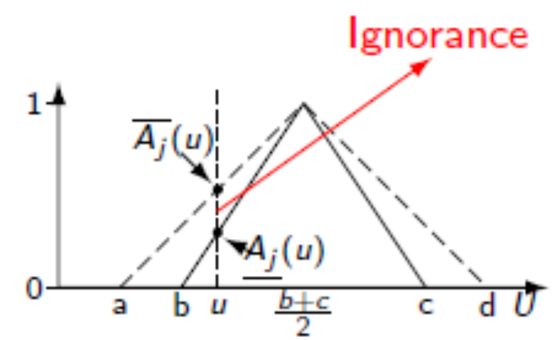
Definition

An interval-valued fuzzy set A in the universe $U \neq \emptyset$ is a mapping $A : U \to L([0,1])$, such that

$$A(u) = [\underline{A}(u), \overline{A}(u)] \in L([0,1]), \text{ for all } u \in U.$$

$$L([0,1]) = \{\mathrm{x} = [\underline{x}, \overline{x}] | (\underline{x}, \overline{x}) \in [0,1]^2 \text{ and } \underline{x} \leq \overline{x}\}$$

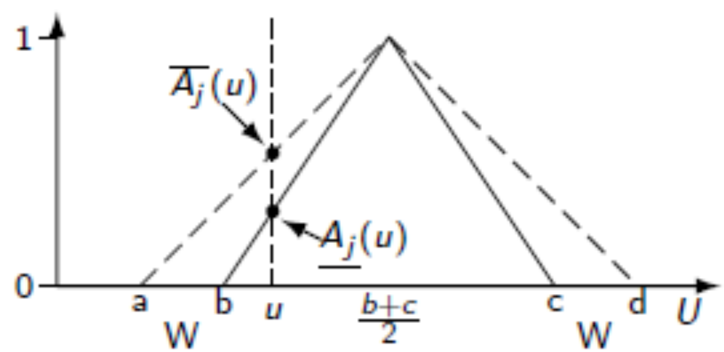
Interval length: $\mathcal{L} = \overline{A}(u) - \underline{A}(u)$



Interval-valued fuzzy set

IVFS: Construction Method

- Lower bound: initial fuzzy set.
- Upper bound construction:
  - Centered in the lower bound.
  - Symmetrical in both sides.
  - Amplitude of the support: 50% greatter than that of the lower bound.

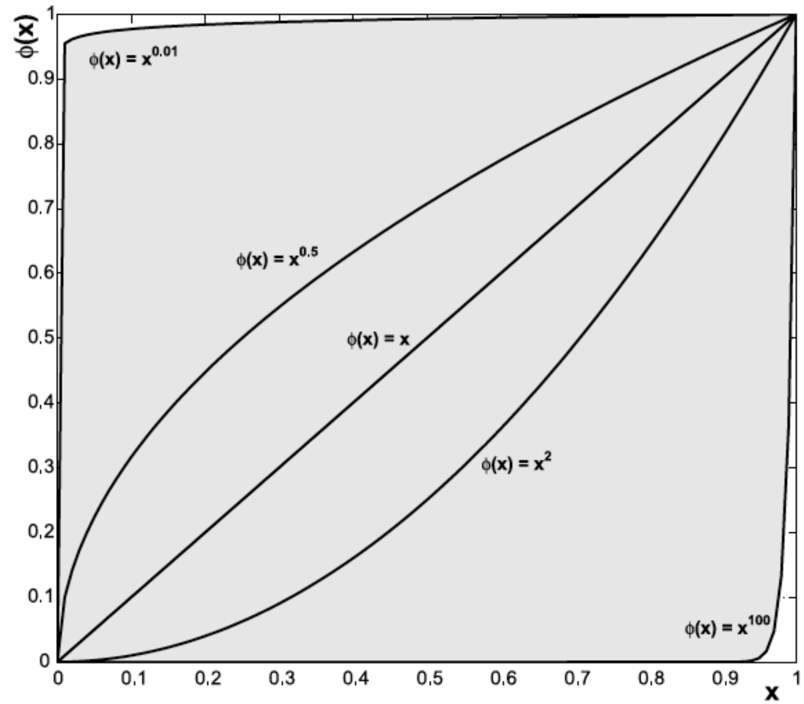


W: parameter that determines the amplitude.
Initial construction $W = 0.5$

IVFS: Restricted Equivalence Functions

- IV-REFs are used to quantify the equivalence degree between two intervals
- We use a construction method based on automorphisms

$$IV\text{-}REF(\mathrm{x},\mathrm{y}) = [T(\phi_1^{-1}(1-|\phi_2(\underline{x})-\phi_2(\underline{y})|),\phi_1^{-1}(1-|\phi_2(\overline{x})-\phi_2(\overline{y})|))$$
$$S(\phi_1^{-1}(1-|\phi_2(\underline{x})-\phi_2(\underline{y})|),\phi_1^{-1}(1-|\phi_2(\overline{x})-\phi_2(\overline{y})|))].$$



Example 3: Taking $\phi_1(x) = x, \phi_2(x) = x$, we obtain the following IV-REF:

$$IV\text{-}REF(\mathrm{x},\mathrm{y}) = [T(1-|\underline{x}-\underline{y}|, 1-|\overline{x}-\overline{y}|)$$
$$S(1-|\underline{x}-\underline{y}|, 1-|\overline{x}-\overline{y}|)].$$

References: A. Jurio, M. Pagola, D. Paternain, C. Lopez-Molina, P. Melo-Pinto: Interval-valued restricted equivalence functions applied on Clustering Techniques. Proceedings of IFSA 2009 , pp. 831-836 (2009)

M. Galar, J. Fernandez, G. Beliakov, H. Bustince "Interval-Valued Fuzzy Sets Applied to Stereo Matching of Color Images" IEEE Trans. Image Processing, vol. 20(7), 1939-1949, 2011.

IVFS: Extension of the FRM

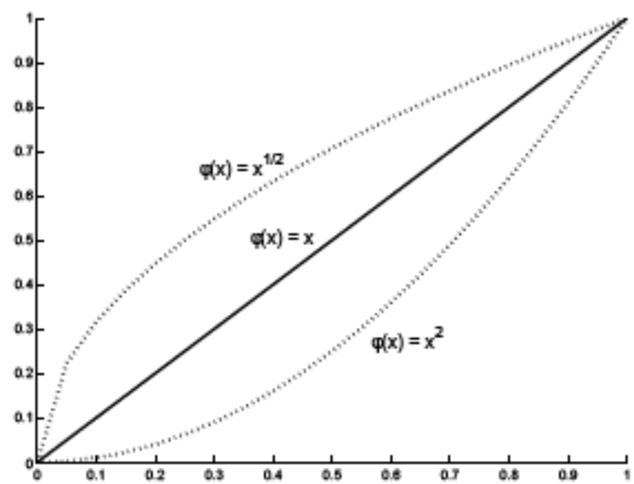
1. *Interval matching degree:*

$$[\underline{A_j}(x_p), \overline{A_j}(x_p)] = \mathbf{T}_{T_1,T_2}(IV\text{-}REF([\underline{A_{j1}}(x_{p1}), \overline{A_{j1}}(x_{p1})], [1,1]), \ldots,$$
$$IV\text{-}REF([\underline{A_{jn}}(x_{pn}), \overline{A_{jn}}(x_{pn})], [1,1])), \qquad j = 1, \ldots, L$$

where $IV\text{-}REF(\mathbf{x}, \mathbf{y}) = [T(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})), S((REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})))]$

with $REF(x, y) = \phi_1^{-1}(1 - |\phi_2(x) - \phi_2(y)|)$ and $\phi_1(x) = x^a$, $\phi_2(x) = x^b$

The IV-REF of each attribute
is defined by
the values of a and b.



The goal is to show how equivalent are the interval membership degrees of the antencedent of the rules to the ideal interval membership [1,1].

IVFS: Extension of the FRM

2. *Interval association degree:*

$$[\underline{b_j^k}, \overline{b_j^k}] = h([\underline{A_j}(x_p), \overline{A_j}(x_p)], [\underline{RW_j^k}, \overline{RW_j^k}]) \qquad k = 1, \ldots, M, \quad j = 1, \ldots, L.$$

$$[\underline{RW_j}, \overline{RW_j}] = \frac{\sum_{x_p \in ClassC_j} A_j(x_p) - \sum_{x_p \notin ClassC_j} A_j(x_p)}{\sum_{p=1}^{m} A_j(x_p)}$$

where

$$A_j(x_p) = [\underline{A_j}(x_p), \overline{A_j}(x_p)]$$

3. *Interval association degree by classes.*

$$[\underline{Y_k}, \overline{Y_k}] = f([\underline{b_j^k}, \overline{b_j^k}],\ j = 1, \ldots, L \text{ y } [\underline{b_j^k}, \overline{b_j^k}] > 0_L), \qquad k = 1, \ldots, M.$$

4. *Classification.*

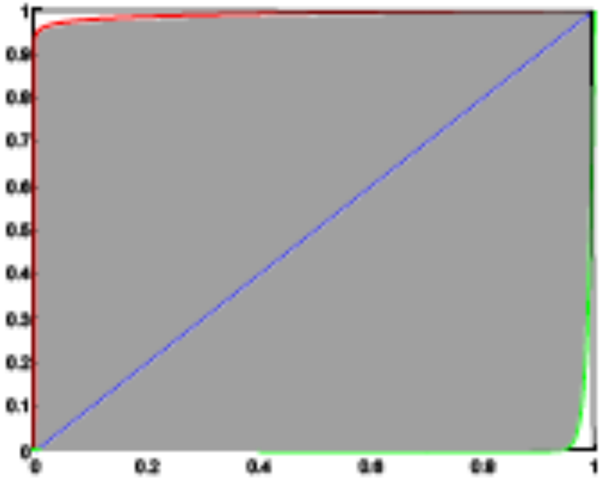
$$F([\underline{Y_1}, \overline{Y_1}], ..., [\underline{Y_M}, \overline{Y_M}]) = \arg \max_{k=1,\ldots,M}([\underline{Y_k}, \overline{Y_k}])$$

In step4 we use the order given by Yager and Xu.

Genetic Tuning of the IV-REF

- **Aim**: to improve the initial performance of the FRBCS.
- **Methodology**: to learn the values of the parameters a and b used in the IV-REF construction related to each attribute.
- a and b take values within the interval $[0.01, 1.99]$:
  - Values of a and b used in the initial construction of the IV-REFs: 1.0.
- **Coding scheme**:

$$C_E = (a_1, b_1, a_2, b_2, ..., a_n, b_n)$$



The selection of a suitable IV-REF to measure the equivalence degree in each variable could lead to an improvement on the FRBCS' behaviour.

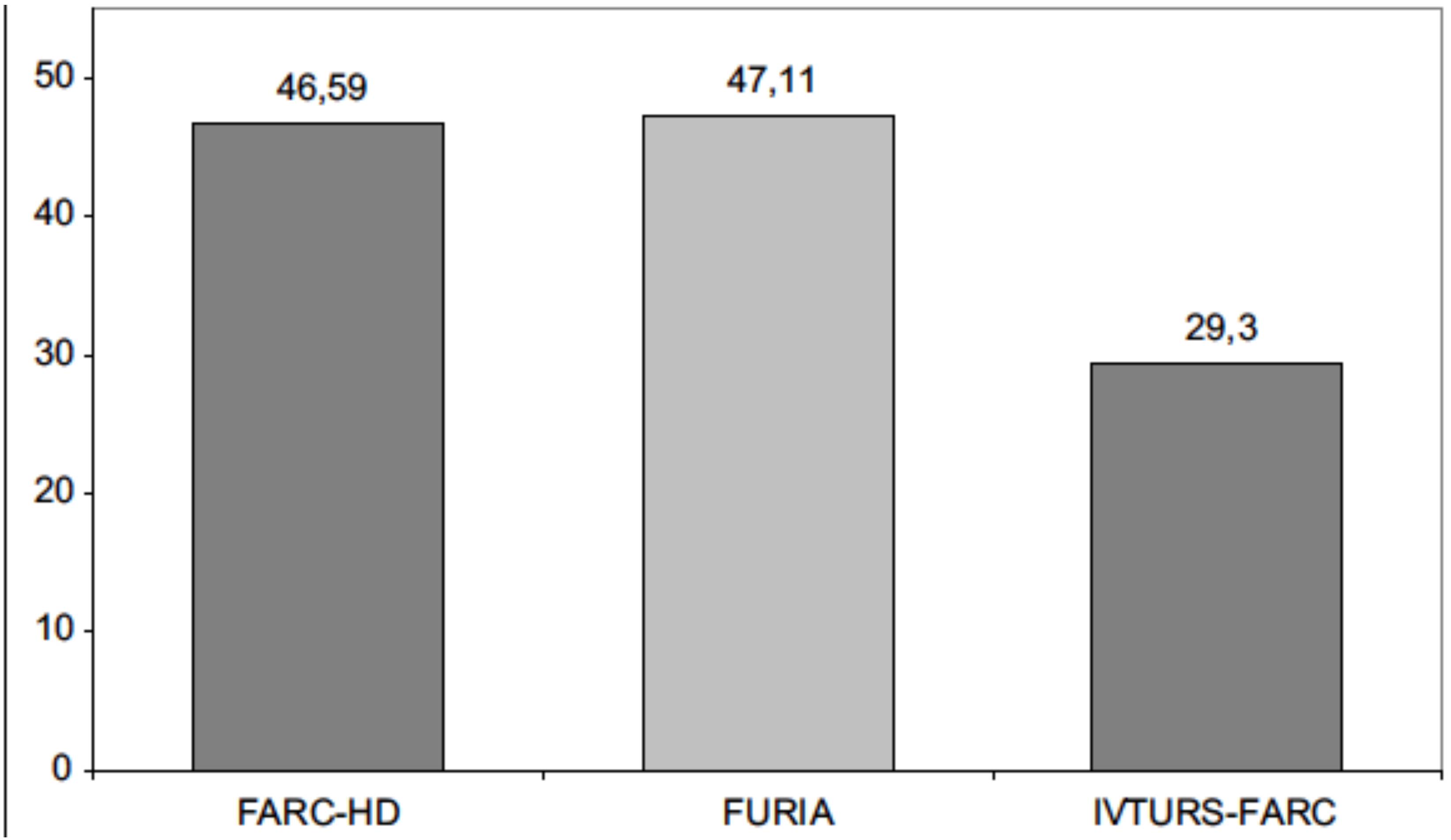
Índice

Experimental Study

Data Set	FS_T_E Tr.	FS_T_E Tst	FS_T_E+R Tr.	FS_T_E+R Tst	IVFS_T_E Tr.	IVFS_T_E Tst	IVFS_T_WI Tr.	IVFS_T_WI Tst	IVFS_T_E+WI Tr.	IVFS_T_E+WI Tst	FARC-HD Tr.	FARC-HD Tst	FURIA Tr.	FURIA Tst	IVTURS-FARC Tr.	IVTURS-FARC Tst
aus	88.44	**86.38**	90.65	84.93	88.77	85.07	83.22	83.48	88.84	85.51	90.43	85.51	88.99	86.09	90.04	85.80
bal	84.68	80.96	91.56	82.40	86.92	82.08	88.56	82.40	88.12	82.88	92.12	**87.36**	88.84	83.68	91.84	85.76
cle	83.50	58.92	88.30	54.54	80.81	57.58	80.81	57.25	81.82	59.59	89.56	57.92	62.37	56.57	85.44	**59.60**
con	56.96	53.02	61.80	53.90	57.04	52.68	57.16	52.68	57.37	52.55	62.80	52.68	56.81	**54.17**	59.69	53.36
crx	89.01	85.45	91.58	86.83	89.40	86.53	84.07	82.85	89.47	**87.75**	91.65	86.53	89.70	86.37	91.42	87.14
der	99.44	92.75	100.00	90.49	99.65	**94.69**	98.39	**94.69**	99.65	94.42	100.00	89.94	98.88	93.86	99.86	94.42
eco	88.17	78.60	91.07	**80.07**	86.24	76.20	85.49	77.69	87.43	77.39	92.11	**80.07**	89.66	80.06	89.06	78.58
ger	81.95	71.10	86.98	**73.80**	81.48	71.90	78.25	71.00	81.65	70.70	87.70	71.60	76.90	73.30	85.35	73.10
hab	78.68	**73.84**	82.19	72.19	77.61	72.22	74.43	73.20	78.92	73.51	81.53	71.22	75.57	72.55	80.72	72.85
hay	86.36	76.41	91.28	79.46	86.36	80.23	81.26	77.18	86.36	80.23	91.28	80.20	88.44	**81.00**	91.28	80.23
hea	91.94	**88.15**	94.44	86.67	92.31	86.30	90.09	86.30	92.22	85.19	94.63	84.44	89.72	78.15	93.61	**88.15**
ion	98.29	91.18	98.58	90.60	98.36	91.17	96.15	**92.33**	98.58	90.33	98.50	90.32	97.65	88.91	98.72	89.75
iri	98.17	96.00	98.50	94.67	98.17	**96.67**	97.83	**96.67**	98.33	96.00	98.50	94.00	98.50	94.00	98.17	96.00
mag	82.14	79.13	83.43	80.33	81.32	78.81	78.25	76.18	81.65	78.97	84.46	80.49	84.03	**80.65**	82.28	79.76
new	95.47	92.09	97.91	94.42	96.63	92.09	95.23	93.02	98.14	93.49	98.95	**95.35**	99.07	94.88	98.84	**95.35**
pag	96.44	94.52	96.85	94.89	96.03	94.71	94.66	93.42	96.26	94.52	96.94	94.34	99.54	**95.25**	96.85	95.07
pen	96.09	91.82	97.68	92.00	93.68	91.09	94.45	91.16	94.23	90.18	98.34	**92.64**	98.98	92.45	95.50	92.18
pim	79.56	75.51	82.29	**76.17**	78.71	74.99	75.39	72.00	79.46	74.35	83.66	74.08	79.17	**76.17**	80.57	75.90
sah	79.17	**72.28**	82.74	69.70	78.52	69.28	77.22	70.99	79.06	68.40	83.49	70.77	74.84	70.33	79.55	70.99
spe	91.20	77.51	92.60	77.87	89.42	**80.52**	85.68	77.13	90.08	80.51	92.42	78.64	95.70	77.88	90.45	**80.52**
tae	70.04	52.37	75.50	**54.43**	70.37	53.68	67.55	50.39	72.19	54.34	77.15	48.41	54.63	47.08	73.18	50.34
tit	77.06	77.06	79.07	**78.87**	77.65	77.65	77.65	77.65	77.65	77.65	79.07	**78.87**	78.46	78.51	79.07	**78.87**
two	96.05	89.19	96.76	90.54	95.91	**93.24**	95.71	91.49	96.39	92.97	97.84	89.32	99.46	88.11	96.35	92.30
veh	75.30	64.66	78.69	66.90	70.77	65.26	70.30	62.89	71.57	63.48	80.38	68.44	79.34	**70.21**	72.81	67.38
win	99.44	93.24	99.86	95.49	99.30	96.08	98.88	**97.76**	99.44	94.95	99.86	96.62	99.58	93.78	99.30	97.19
wiR	62.59	59.54	64.93	**59.66**	60.94	58.91	58.30	56.35	61.60	58.97	56.22	53.96	57.12	51.29	62.01	58.28
wis	97.80	96.49	98.61	**96.78**	97.88	96.34	96.12	95.90	98.13	**96.78**	98.76	96.63	98.83	96.63	98.50	96.49
Mean	86.07	79.56	88.66	79.95	85.56	79.85	83.75	79.04	86.10	79.84	88.83	79.64	85.21	79.33	87.42	**80.57**

J. Sanz, A. Fernandez, H. Bustince, F. Herrera. IVTURS: a linguistic fuzzy rule-based classification system based on a new Interval-Valued fuzzy reasoning method with TUning and Rule Selection. IEEE-TFS 21:3 (2013) 399-411

Experimental Study



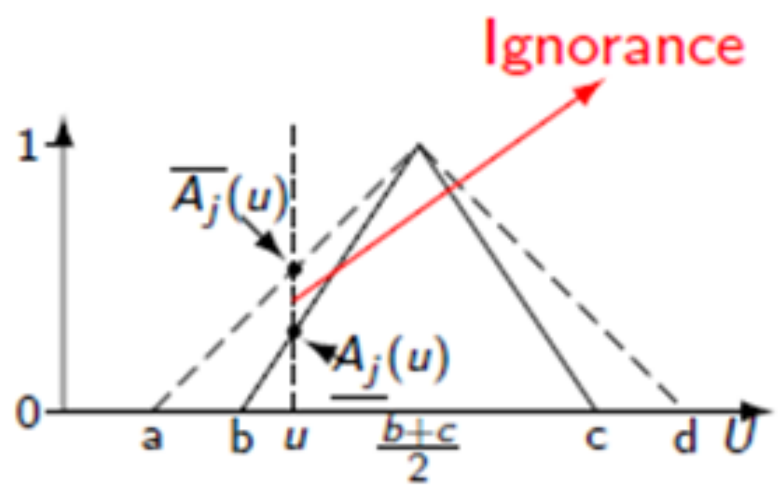
J. Sanz, A. Fernandez, H. Bustince, F. Herrera. IVTURS: a linguistic fuzzy rule-based classification system based on a new Interval-Valued fuzzy reasoning method with TUning and Rule Selection. IEEE-TFS 21:3 (2013) 399-411

Summary

- Methodology to face difficult classification problems.
- The ignorance degree is modeled by means of IVFS.
- IVFS are built by means of REFs
- Fuzzy reasoning methods are extended in a natural way to work with IVFS
- There is a positive cooperation between the tuning approaches.
- The model obtains high quality results that can be compared with the state of the art in the topic.

Summary

- The proposal allows us to manage the **lack of knowledge (ignorance)** associated with the **data intrinsic characteristics**
  - Application on different scenarios: overlapping, borderline, and so on.



Interval-valued fuzzy set

Índice

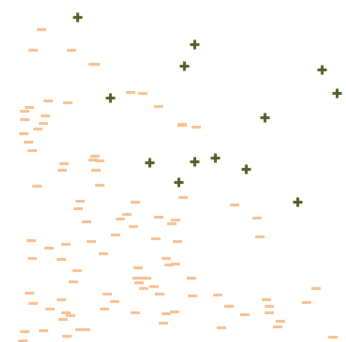
Classification with Imbalanced Data

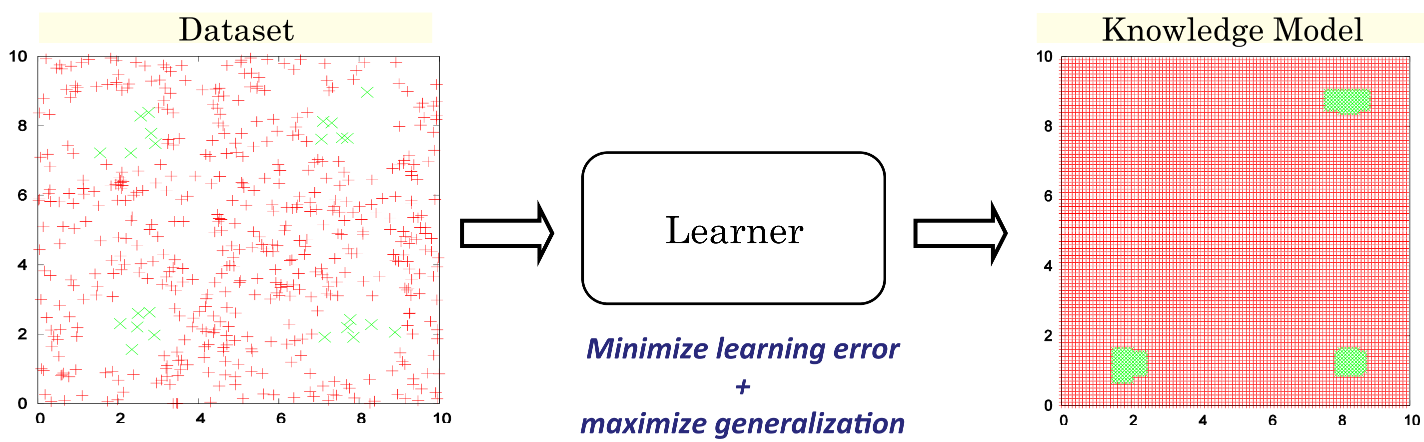
- Real application areas in engineering characterised by having a very different distribution of examples among their classes.
- Intrinsic to the problem or due to limitations during the data collection process.
- Some examples:
  - Medical diagnosis / Fraud detection
  - Objects identification / Bioinformatics
- **Problem of imbalanced data-sets**: it sets a handicap for the correct identification of the different concepts to be learnt.
- Positive class often represents the concept of the highest interest for the problem, whereas the negative class represents counter-examples.

Properties of imbalanced problems

Why learning from imbalanced data-sets might be difficult?

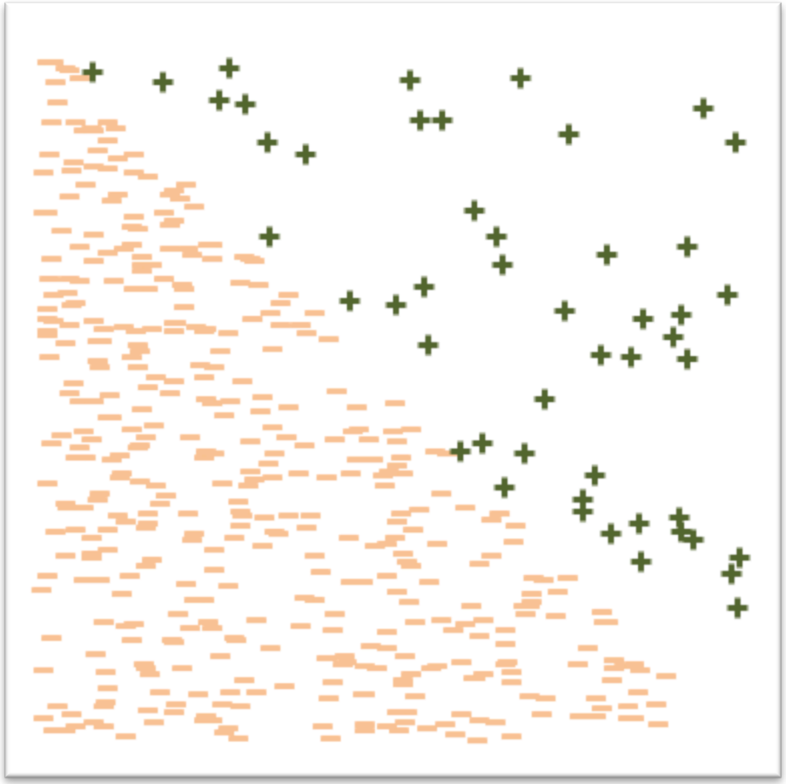
1. Search process guided by global error rates.
2. Classification rules over the positive class are highly specialized.
3. Classifiers tend to ignore small classes concentrating on classifying large ones accurately





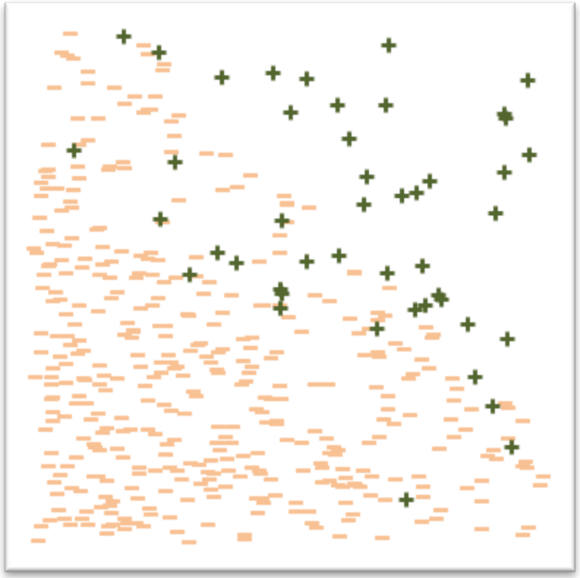
Properties of imbalanced problems

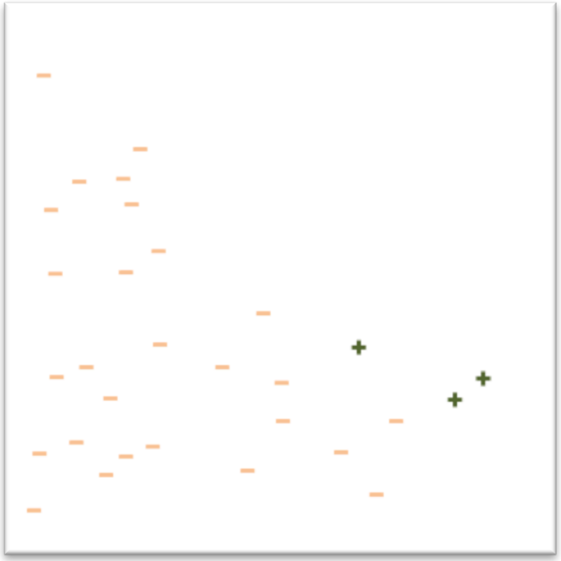
- Skewed class distribution:
  - Measured by the fraction between majority and minority samples
  - Imbalance ratio (IR)
- **Intrinsic Data Characteristics**
  - Not only imbalance hinders classification performance
  - IR ≈ 9

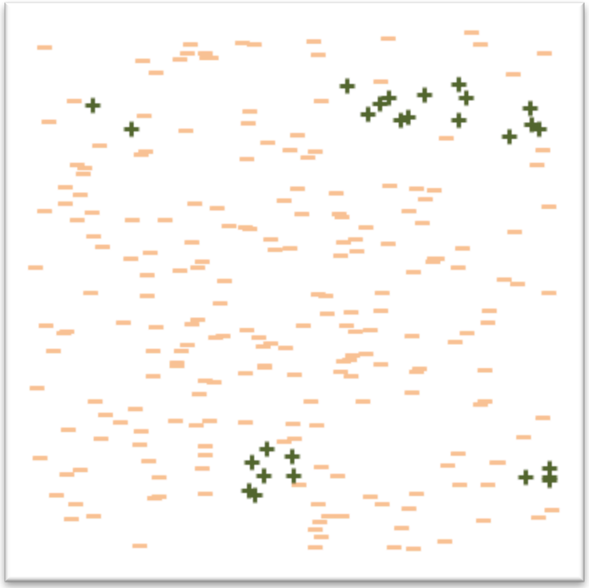


V. López, A. Fernandez, S. García, V. Palade, F. Herrera, An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using Data Intrinsic Characteristics. Information Sciences 250 (2013) 113-141

Properties of imbalanced problems







V. López, A. Fernandez, S. García, V. Palade, F. Herrera, An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using Data Intrinsic Characteristics. Information Sciences 250 (2013) 113-141

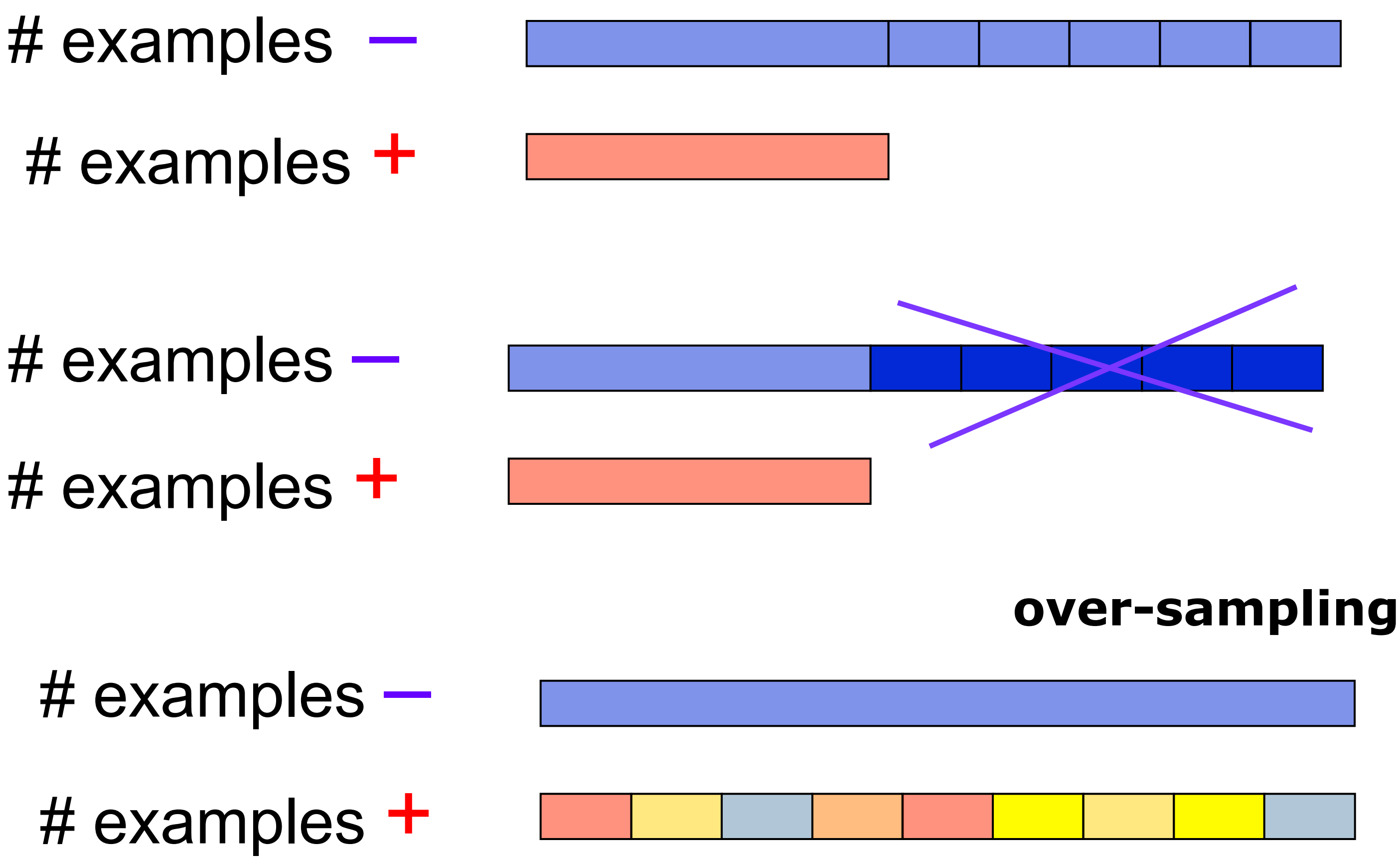
Índice

Addressing imbalanced data-sets

- Strategies to deal with imbalanced datasets:
  1. **Data level**: instance preprocessing (re-sampling).
  2. **Algorithm level**: change the behaviour of the algorithm itself.
  3. **Cost-Sensitive learning:** considers the varying costs of different misclassification types, given by a cost matrix. Mix of strategies

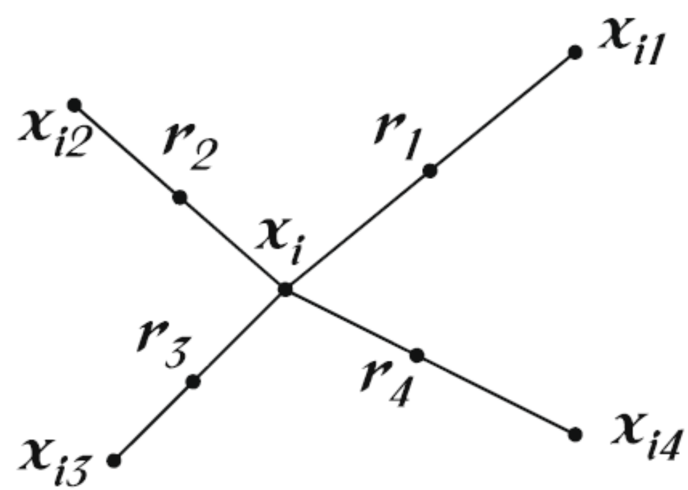
Preprocessing algorithms

- Undersampling vs oversampling



Preprocessing algorithms: SMOTE

- Synthetic Minority Over-sampling TEchnique and SMOTE related approaches:
  - **Generation** of new minority class examples
  - Interpolation among several minority class instances that lie together
  - Some drawbacks:
    - Problem of **over-generalization**
    - Lack of **flexibility**



For each minority sample
- Find its k-nearest minority neighbours
- Randomly select j neighbours
- Randomly generate synthetic samples along the lines joining the minority sample selected and its j neighbours

(j depends on the amount of oversampling desired)

N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16 (2002) 321-357

Preprocessing + FRBCS

- Good synergy among both techniques*
- Many Evolutionary Fuzzy Systems (EFS) in conjunction with preprocessing (SMOTE):
  - Adaptive inference system
  - Linguistic adjustment based on 2-tuples
  - Analysis for low quality data
  - A priori learning of the Data Base: feature selection + granularity
  - Study of the behavior of EFS with dataset shift

* A. Fernandez, S. García, M.J. del Jesus, F. Herrera. A Study of the Behaviour of Linguistic Fuzzy Rule Based Classification Systems in the Framework of Imbalanced Data Sets. Fuzzy Sets and Systems, 159:18 (2008) 2378-2398

Fuzzy algorithm level approaches

- Fuzzy Logic And Genetic algorithms for Imbalanced Datasets (FLAGID):
  - To create ad hoc membership functions for the positive class
- Hierarchical fuzzy rule based classification system:
  - Different granularity levels: higher for borderline instances
- IVTURS for imbalanced classification:
  - Rule weight rescaling method regarding the positive class

Cost-sensitive learning

- Errors made on minority class examples higher than those of the majority class in computing training error.
- Needs a cost matrix, which encodes misclassification penalty.
- Consider the cost-matrix throughout the building of the model for achieving the lowest cost.
- **Direct methods:**
  - Introduce and use misclassification costs into the learning algorithms.
- **Meta-learning:**
  - Preprocessing mechanism for training data or post-processing of the output. The original learning algorithm is not modified:
    - **Sampling:** assigning instance weights
    - **Thresholding** based on the Bayes decision theory: assign instances to class with minimum expected cost.

C. Elkan, The foundations of cost-sensitive learning, in: Proceedings of the 17th International Joint Conference on Artificial Intelligence, 2001, pp. 973–978

Cost-sensitive fuzzy learning

- Include weights during the learning stage:
  - Evaluation function of genetic rule learning procedure
- Acting on the computation of the rule weight
  - Values are “weighted” regarding class distribution
- Modifying the inference process
  - Consider the compatibility degree of the example and the fuzzy label together with the cost associated to that example

Índice

GP-COACH-H: A Hierarchical Genetic Programming Fuzzy Rule-Based Classification System with Rule Selection and Tuning

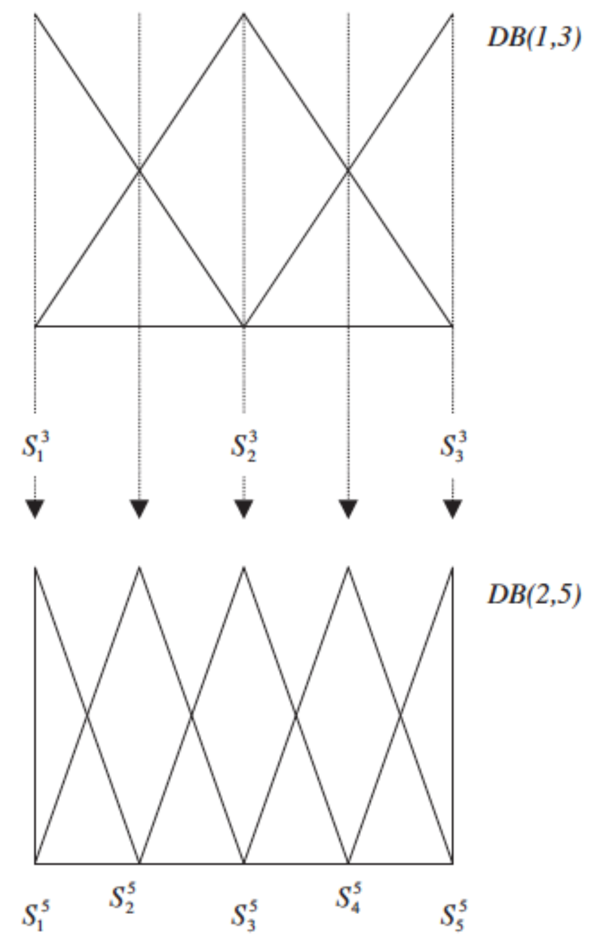
- This proposal combines several strategies that are able to obtain a good synergy between:
  - **Data preprocessing** modifies the input data to ease the learning process
  - A h**ierarchical linguistic classifier** extends the definition of the KB to model complex search spaces such as imbalanced data with hard data intrinsic characteristics.
- **Genetic tuning** of KB parameters to enhance the final model by a better characterization of the classes of the dataset

Rationale for Hierarchical KB

- A standard KB structure has problems for modeling complex problems: Imbalanced data with hard data intrinsic characteristics
  - Lack of flexibility because of the rigid partitioning of the input space (linguistic labels).
  - The size of the RB directly depends on the number of variables and linguistic terms in the system
  - Obtaining an accurate FRBS requires a significant granularity amount.
    - Number of rules to rise significantly
    - Interpretability is lost.
    - Overfitting on the training data

Hierarchical Multigranularity

- Flexibilization of the KB to become a Hierarchical KB.
- Linguistic variables of the linguistic rules could take values from fuzzy partitions with different granularity levels.
- *Example: DB^1* becomes a linguistic partition in *DB^2*.



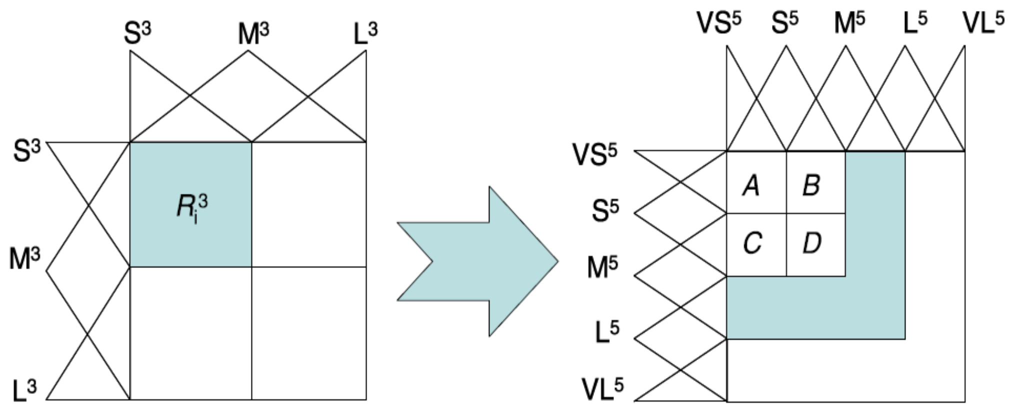
Example of the expansion process

$DB(1,3) = \{S^3, M^3, L^3\},$
$DB(2,5) = \{VS^5, S^5, M^5, L^5, VL^5\},$

R_i^3 : IF x_1 *is* S_{i1}^3 AND x_2 *is* S_{i2}^3 THEN $Class = C$ with RW_i,

$I(S_{i1}^3) = \{VS^5, S^5\}, \quad I(S_{i2}^3) = \{VS^5, S^5\},$
$I(R_i^3) = I(S_{i1}^3) \times I(S_{i2}^3).$

$LRG(I(R_i^3), \quad X(R_i^3)) = \left\{R_{i1}^5, R_{i2}^5, R_{i3}^5, R_{i4}^5\right\}$

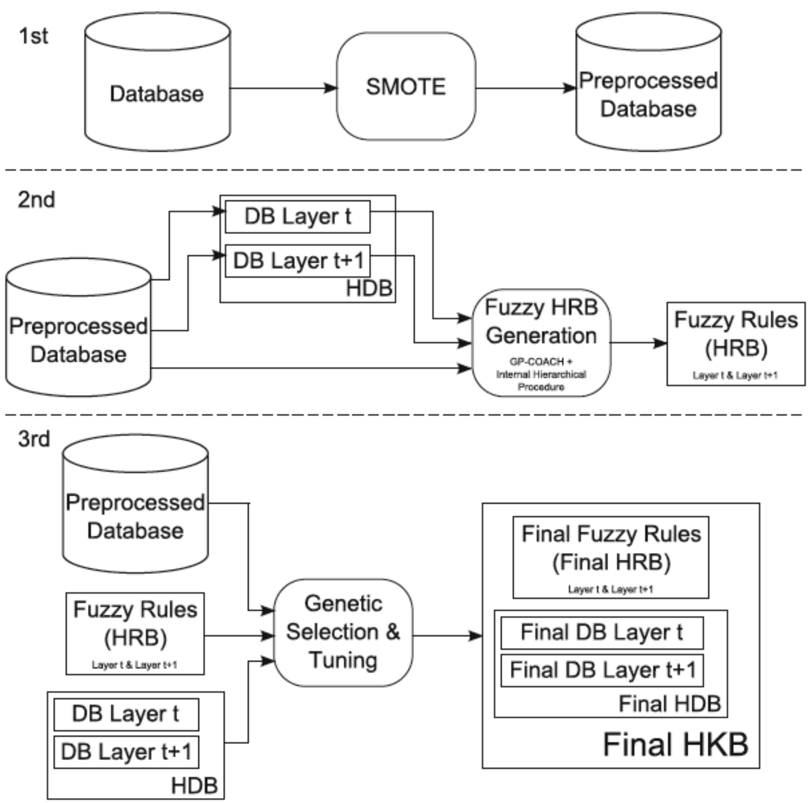


$R_i^3 = \text{IF } x_1 \text{ is } S^3 \text{ AND } x_2 \text{ is } S^3 \text{ THEN } Class = C \text{ with } RW_i$

$R_{i1}^5 = \text{IF } x_1 \text{ is VS}^5 \text{ AND } x_2 \text{ is VS}^5 \text{ THEN } Class = C \text{ with } RW_{i1}$
$R_{i2}^5 = \text{IF } x_1 \text{ is VS}^5 \text{ AND } x_2 \text{ is } S^5 \text{ THEN } Class = C \text{ with } RW_{i2}$
$R_{i3}^5 = \text{IF } x_1 \text{ is } S^5 \text{ AND } x_2 \text{ is VS}^5 \text{ THEN } Class = C \text{ with } RW_{i3}$
$R_{i4}^5 = \text{IF } x_1 \text{ is } S^5 \text{ AND } x_2 \text{ is } S^5 \text{ THEN } Class = C \text{ with } RW_{i4}$

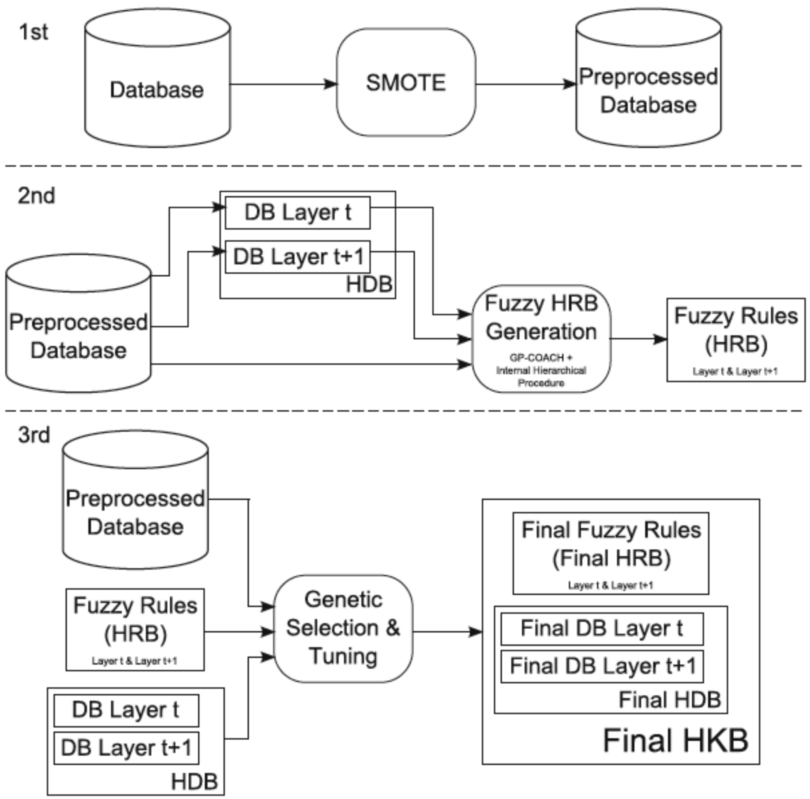
Flowchart of GP-COACH-H

1. **Data preprocessing**: SMOTE algorithm to address skewed data distribution
2. **HKB generation process** with an EFS (GP-COACH algorithm):
   - Identification of *good and bad* rules in the current population. Bad rules are replaced by new high granularity rules.
   - New evaluation function to consider different granularity levels.
   - New constraints over the crossover operator: only applied to rules of the same hierarchy.



Flowchart of GP-COACH-H

3. Genetic tuning of the HKB parameters:
   - Selection of rules seeking a good cooperation
   - Tuning the existing hierarchical DBs following a 2-tuples linguistic representation.
   - An unique genetic procedure using the CHC evolutionary algorithm to profit from the synergy that these optimizations can achieve.



Case study: Initial DB

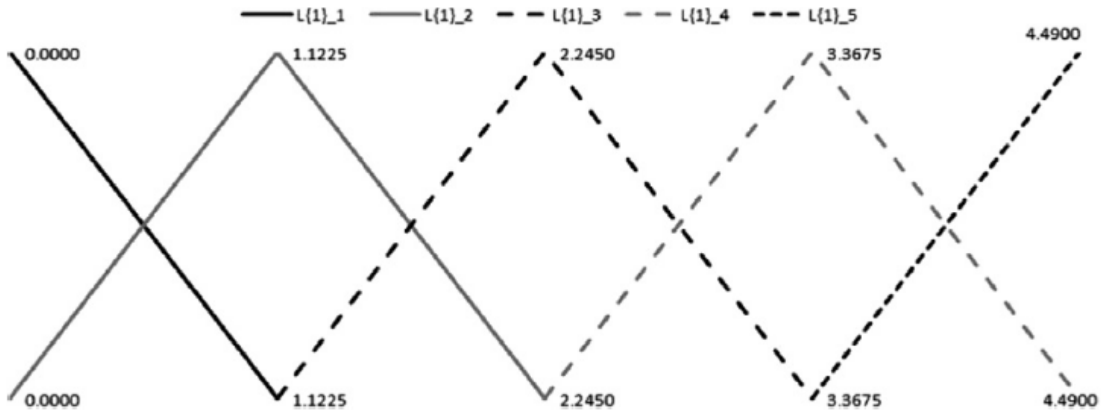
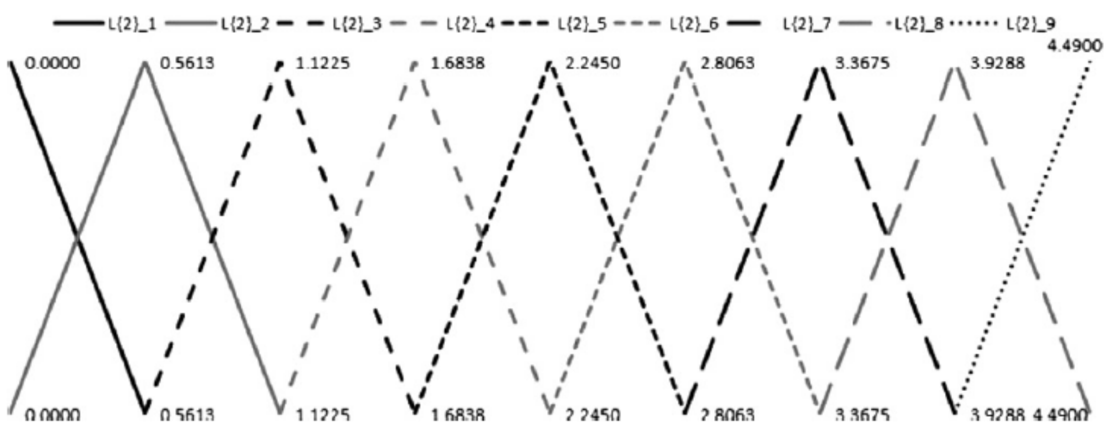


Fig. 12. Database Layer 1 with 5 labels, M_g attribute.



Case study: Initial RB


```
@Number of rules: 13

1{1}: IF Mg IS (L{1}_1 OR L{1}_2 OR L{1}_3) THEN negative with RW: .9969
2{1}: IF Ca IS (L{1}_1 OR L{1}_4 OR L{1}_5) THEN negative with RW: .9999
3{1}: IF RI IS (L{1}_4 OR L{1}_5) THEN negative with RW: 1.0
4{1}: IF Al IS (L{1}_3 OR L{1}_4 OR L{1}_5) AND Fe IS (L{1}_3 OR L{1}_5) THEN negative with RW: 1.0
5{1}: IF Al IS L{1}_3 AND Si IS (L{1}_1 OR L{1}_2) AND Fe IS (L{1}_2 OR L{1}_5) THEN negative with RW: 1.0
6{1}: IF Na IS (L{1}_1 OR L{1}_2 OR L{1}_5) AND Si IS (L{1}_1 OR L{1}_2 OR L{1}_5) THEN negative with RW: .9969
7{1}: IF RI IS (L{1}_3 OR L{1}_4 OR L{1}_5) AND Al IS L{1}_3 AND Fe IS (L{1}_2 OR L{1}_5) THEN negative with RW:
      .9904
8{1}: IF RI IS (L{1}_3 OR L{1}_4 OR L{1}_5) AND Na IS (L{1}_1 OR L{1}_2 OR L{1}_5) AND Fe IS (L{1}_2 OR L{1}_5)
      THEN negative with RW: .9590
9{2}: IF RI IS L{1}_3 AND Na IS L{1}_4 AND Mg IS L{1}_4 AND Al IS L{1}_2 AND Si IS L{1}_3 AND K IS L{1}_1 AND
      Ca IS L{1}_2 AND Ba IS L{1}_1 AND Fe IS L{1}_2 THEN positive with RW: .8156
10{2}: IF RI IS L{1}_2 AND Na IS L{1}_3 AND Mg IS L{1}_4 AND Al IS L{1}_2 AND Si IS L{1}_3 AND K IS L{1}_1 AND
       Ca IS L{1}_2 AND Ba IS L{1}_1 AND Fe IS L{1}_1 THEN positive with RW: .6675
11{2}: IF RI IS L{2}_3 AND Na IS L{2}_5 AND Mg IS L{2}_7 AND Al IS L{2}_2 AND Si IS L{2}_6 AND K IS L{2}_1 AND
       Ca IS L{2}_3 AND Ba IS L{2}_1 AND Fe IS L{2}_1 THEN positive with RW: .9654
12{2}: IF RI IS L{2}_4 AND Na IS L{2}_5 AND Mg IS L{2}_7 AND Al IS L{2}_2 AND Si IS L{2}_5 AND K IS L{2}_1 AND
       Ca IS L{2}_4 AND Ba IS L{2}_1 AND Fe IS L{2}_2 THEN positive with RW: .7443
13{2}: IF RI IS L{2}_5 AND Na IS L{2}_7 AND Mg IS L{2}_1 AND Al IS L{2}_3 AND Si IS L{2}_5 AND K IS L{2}_1 AND
       Ca IS L{2}_6 AND Ba IS L{2}_1 AND Fe IS L{2}_2 THEN negative with RW: 1.0
```

Case study: Final DB

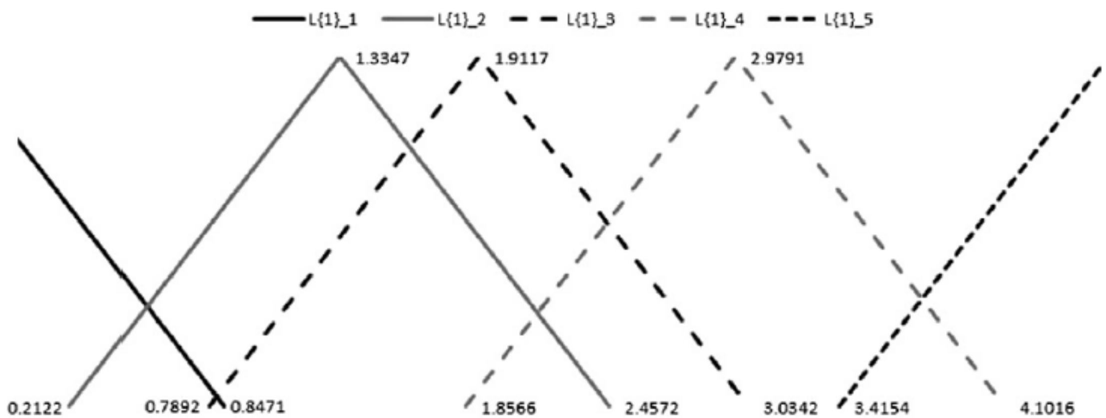
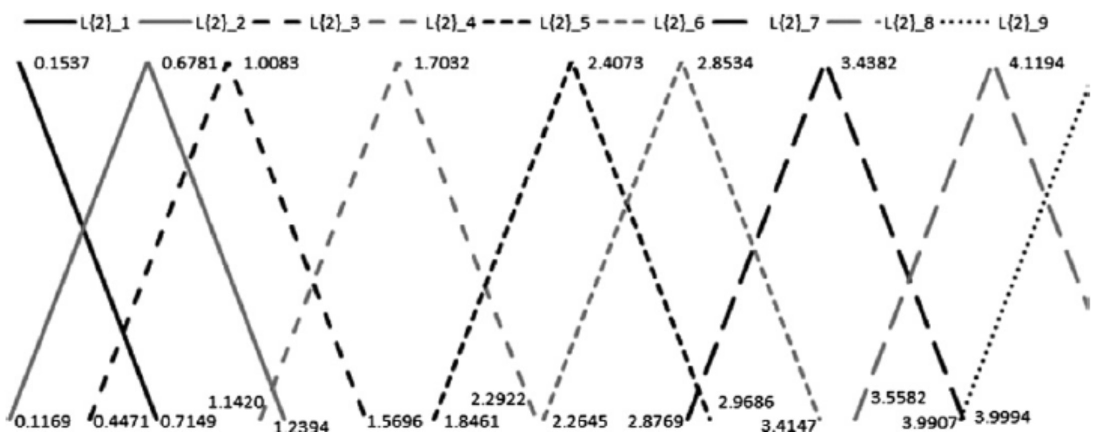


Fig. 15. Final database Layer 1 with 5 labels, M_g attribute.



Case Study: Final RB and results


```
@Number of rules: 10

1{1}: IF Mg IS (L{1}_1 OR L{1}_2 OR L{1}_3) THEN negative with RW: 1.0
2{1}: IF Ca IS (L{1}_1 OR L{1}_4 OR L{1}_5) THEN negative with RW: .9871
3{1}: IF RI IS (L{1}_4 OR L{1}_5) THEN negative with RW: .9981
4{1}: IF Al IS (L{1}_3 OR L{1}_4 OR L{1}_5) AND Fe IS (L{1}_3 OR L{1}_5) THEN negative with RW: .9892
5{1}: IF Al IS L{1}_3 AND Si IS (L{1}_1 OR L{1}_2) AND Fe IS (L{1}_2 OR L{1}_5) THEN negative with RW: .9902
6{1}: IF Na IS (L{1}_1 OR L{1}_2 OR L{1}_5) AND Si IS (L{1}_1 OR L{1}_2 OR L{1}_5) THEN negative with RW: 1.0
7{1}: IF RI IS (L{1}_3 OR L{1}_4 OR L{1}_5) AND Na IS (L{1}_1 OR L{1}_2 OR L{1}_5) AND Fe IS (L{1}_2 OR L{1}_5)
      THEN negative with RW: .9461
8{2}: IF RI IS L{1}_3 AND Na IS L{1}_4 AND Mg IS L{1}_4 AND Al IS L{1}_2 AND Si IS L{1}_3 AND K IS L{1}_1 AND Ca
      IS L{1}_2 AND Ba IS L{1}_1 AND Fe IS L{1}_2 THEN positive with RW: .6544
9{2}: IF RI IS L{1}_2 AND Na IS L{1}_3 AND Mg IS L{1}_4 AND Al IS L{1}_2 AND Si IS L{1}_3 AND K IS L{1}_1 AND Ca
      IS L{1}_2 AND Ba IS L{1}_1 AND Fe IS L{1}_1 THEN positive with RW: .6719
10{2}: IF RI IS L{2}_3 AND Na IS L{2}_5 AND Mg IS L{2}_7 AND Al IS L{2}_2 AND Si IS L{2}_6 AND K IS L{2}_1 AND Ca
       IS L{2}_3 AND Ba IS L{2}_1 AND Fe IS L{2}_1 THEN positive with RW: .9561
```

Data-set	GM_{tr}	GM_{tst}
GP-COACH-5	.8763 ± .0307	.7897 ± .1212
GP-COACH-9	.9056 ± .0267	.7845 ± .1334
HFRBCS(Chi)	.9331 ± .0117	.7901 ± .1325
GP-COACH-H	.9576 ± .0121	**.8175 ± .1193**
C4.5	.9549 ± .0180	.7848 ± .1452

Class Imbalance: Data-sets & algorithms

KEEL Data Mining Tool:
It includes algorithms and data set partitions

http://www.keel.es

KEEL-dataset
Data set repository

Class Imbalance: Data-sets & algorithms

- 145 data sets included:
  - 100 for 2 classes,
  - 15 for multiple classes
  - 30 for noise and borderline.

KEEL-dataset

Data set repository

Imbalanced data sets

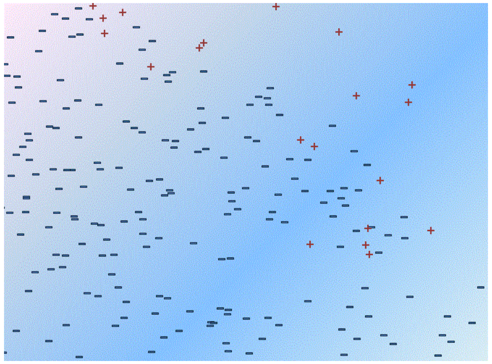
We divide our Imbalanced data sets into the following sections:

- Imbalance ratio between 1.5 and 9
- Imbalance ratio higher than 9 - Part I
- Imbalance ratio higher than 9 - Part II
- Imbalance ratio higher than 9 - Part III
- Multiple class imbalanced problems
- Noisy and Borderline Examples

- Preprocessed partitions are also available.

Thematic Website

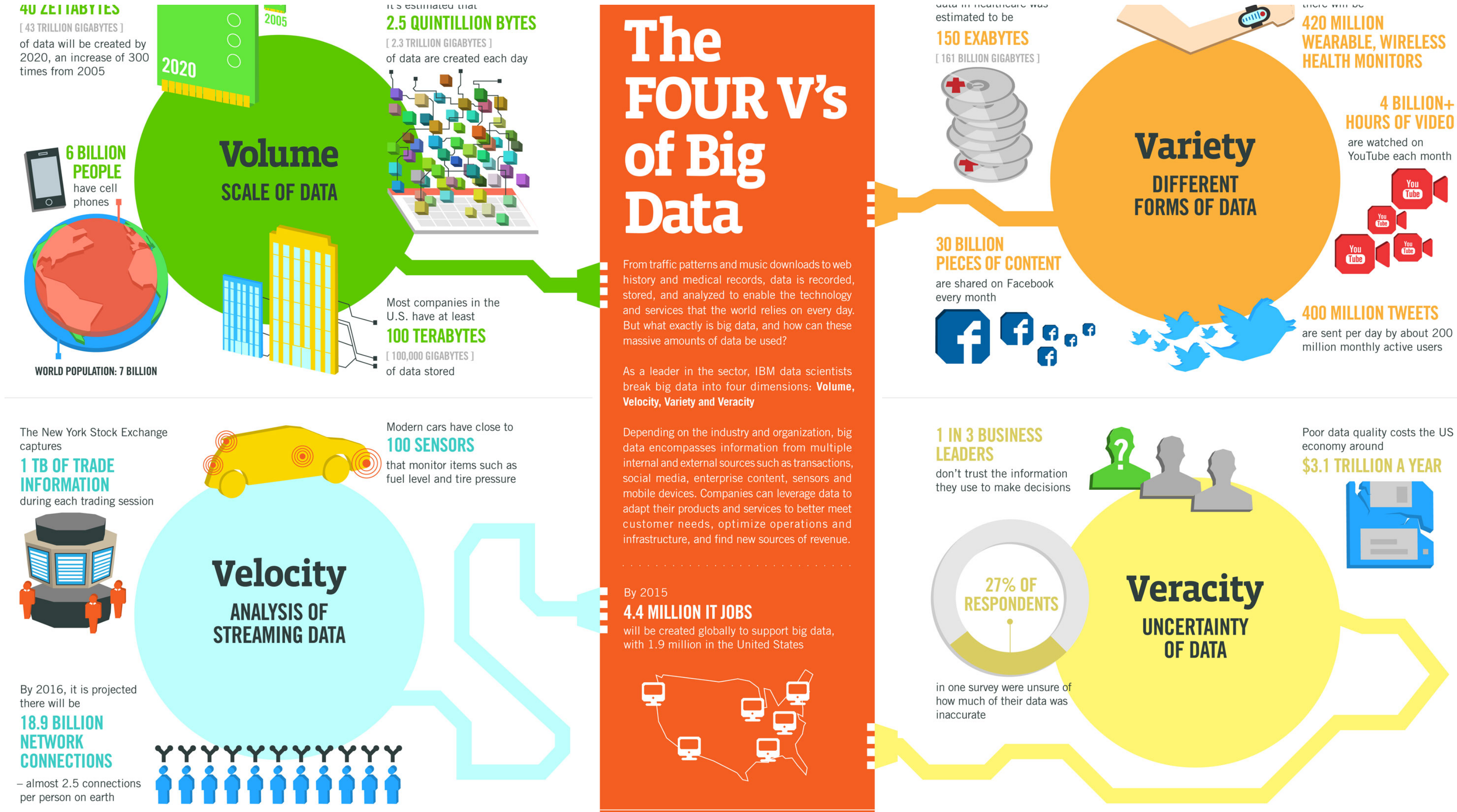
- http://sci2s.ugr.es/imbalanced/



- The web is organized according to the following summary:
  - Introduction to Classification with Imbalanced Datasets
  - Imbalanced Datasets in Classification
    - The problem of imbalanced datasets
    - Evaluation in imbalanced domains
  - Problems related with intrinsic data characteristics
  - Addressing Classification with imbalanced data: preprocessing, cost-sensitive learning and ensemble techniques
    - Preprocessing imbalanced datasets: resampling techniques
    - Cost-sensitive learning
    - Ensemble methods
    - Experimental Results
  - **Software, Algorithm Implementations and Dataset Repository**
  - **Literature review**
    - **Papers of interest: surveys and highlighted studies**
    - **SCI2S Related Approaches**

Índice

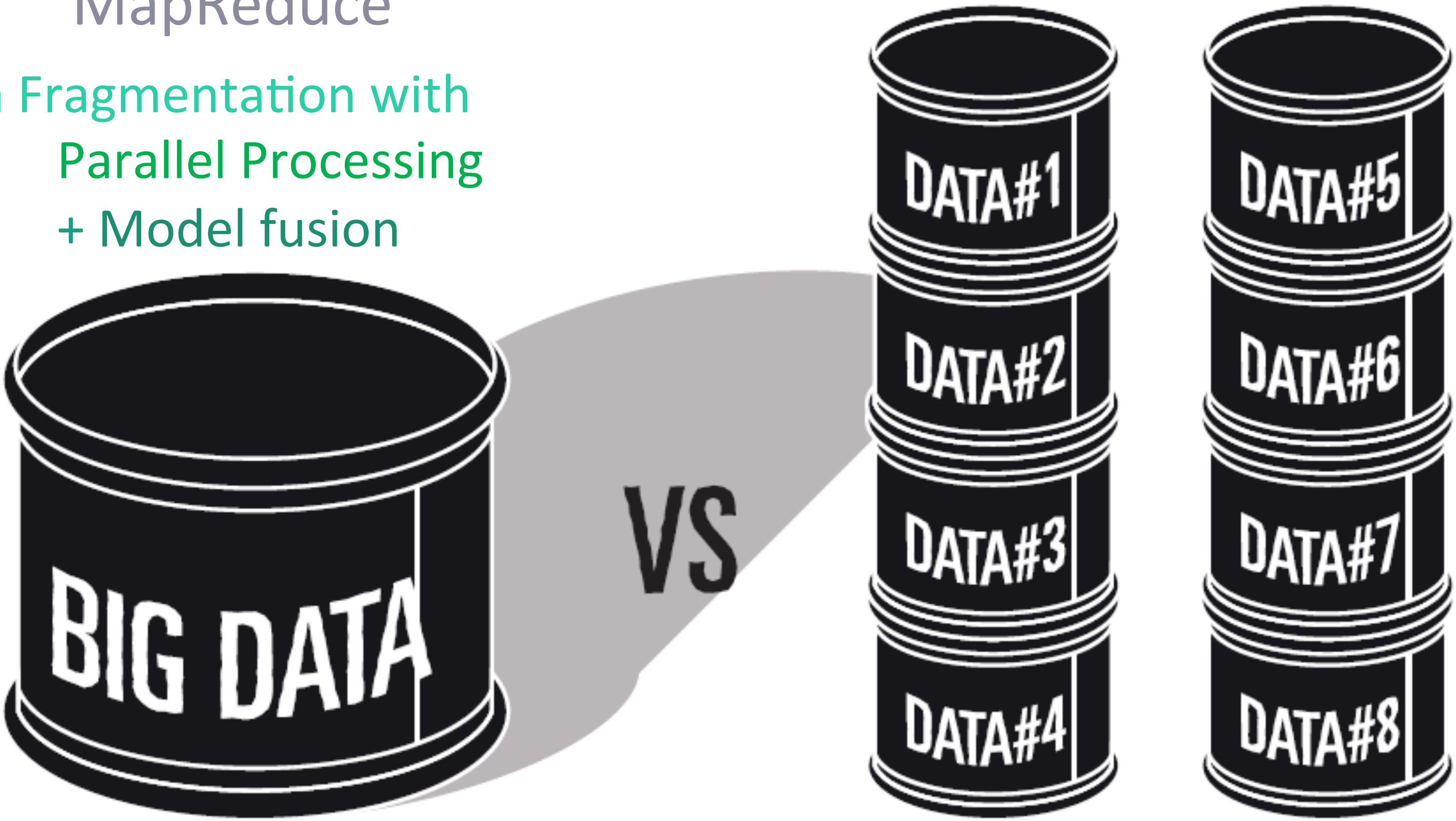
Sistemas Difusos en Big Data Analytics



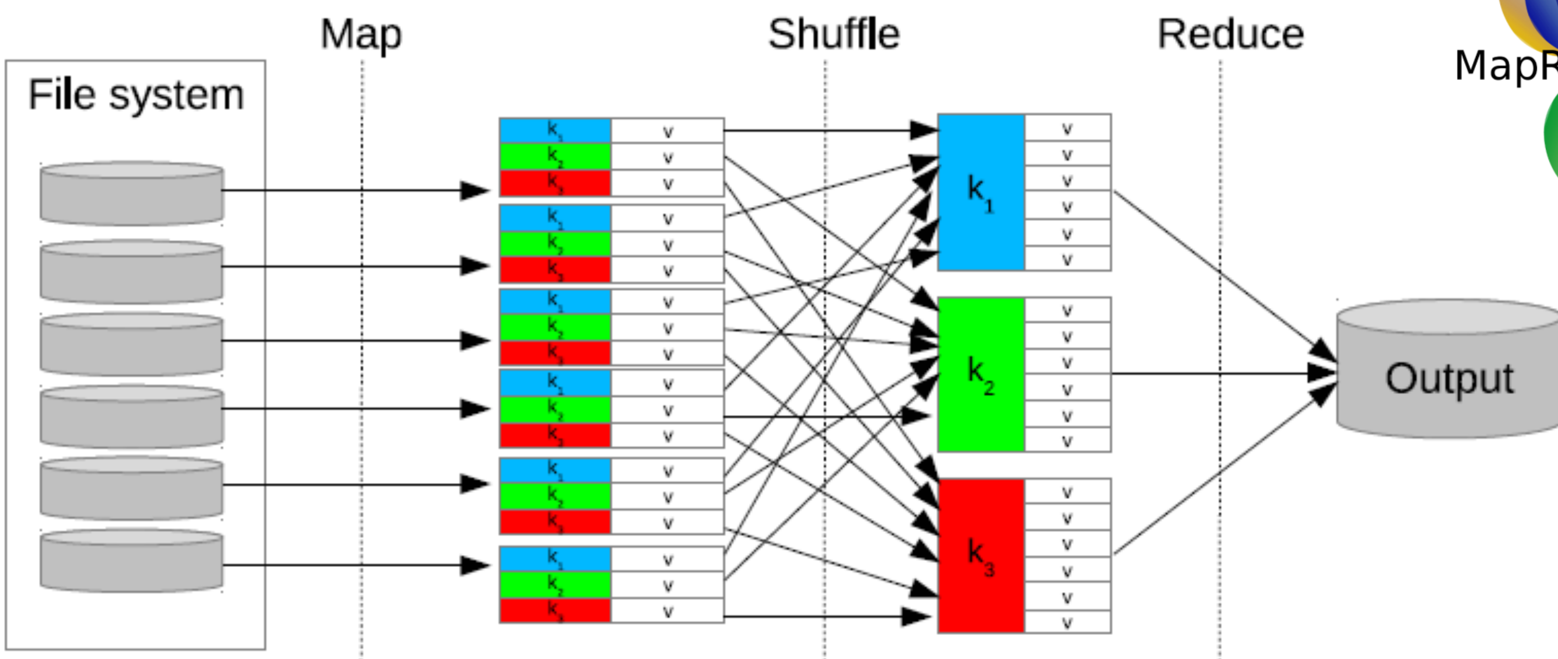
A. Fernandez et al. Big Data with Cloud Computing: An Insight on the Computing Environment, MapReduce and Programming Frameworks. WIREs DMKD 4:5 (2014) 380-409. doi: 10.1002/widm.1134

MapReduce

Data Fragmentation with
Parallel Processing
+ Model fusion



- Flujo de datos en MapReduce: transparente para el programador.



Ficheros de entrada
Particionamiento en bloques

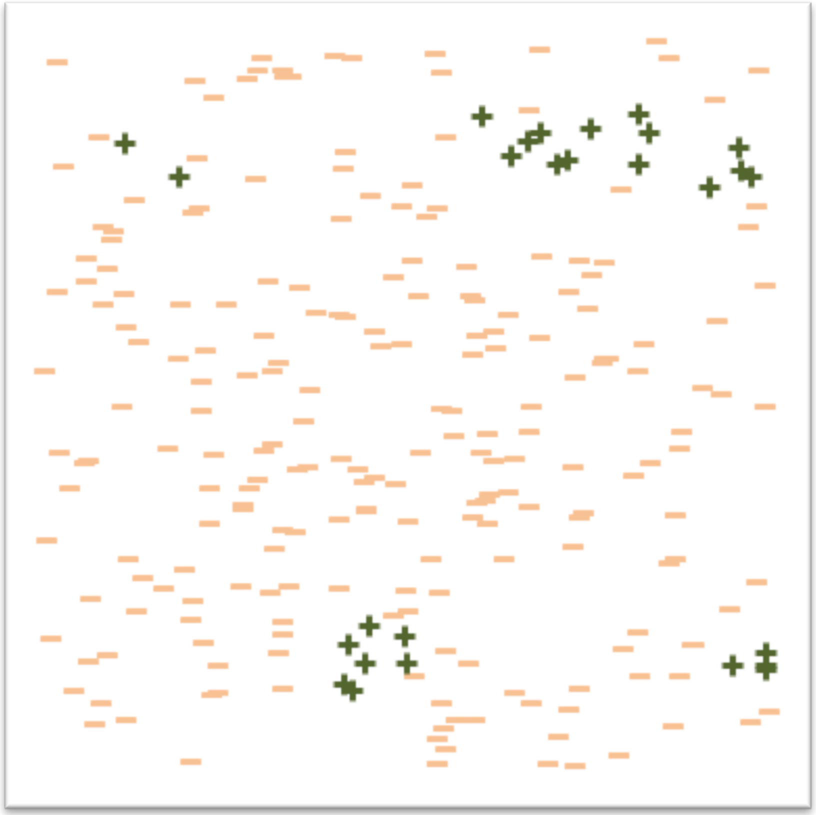
Ficheros Intermedios

Ficheros salida

Índice

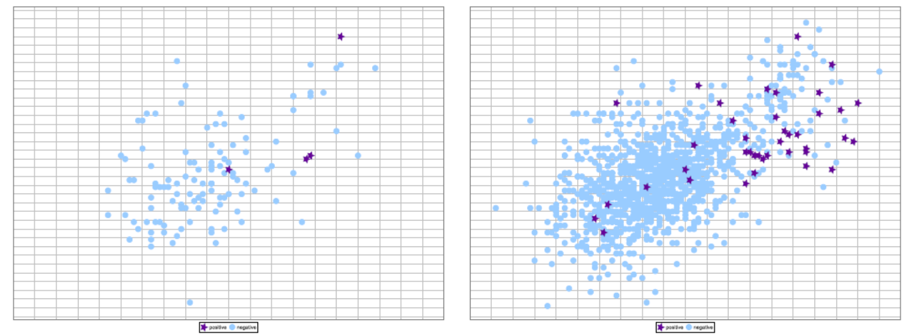
Sistemas Difusos en Big Data Analytics

- Potencial presencia de “small disjuncts”.
- Zonas que necesitan reglas muy específicas para representarlas.
- Consecuencia directa de la división de datos en el proceso Map.

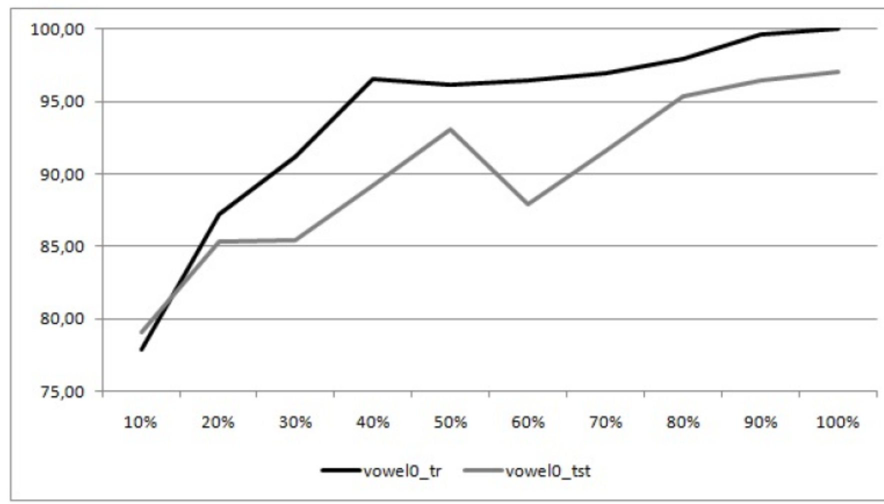


A. Fernandez, C.J. Carmona, M.J. del Jesus, F. Herrera. A View on Fuzzy Systems for Big Data: Progress and Opportunities. International Journal of Computational Intelligence Systems, 9:1 (2016), 69-80

- Falta de datos y densidad en las particiones de entrenamiento.
- Conecta con el problema de “Small Disjuncts”
- Localidad de los datos que implica sobreaprendizaje



(a) 10 % of training instances (b) 100 % of training instances

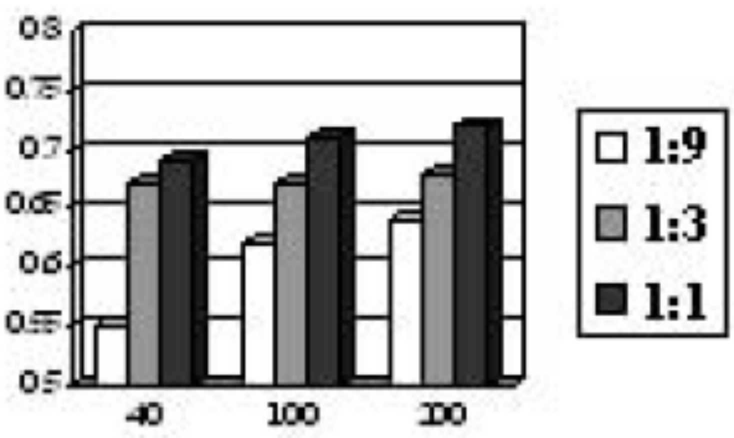


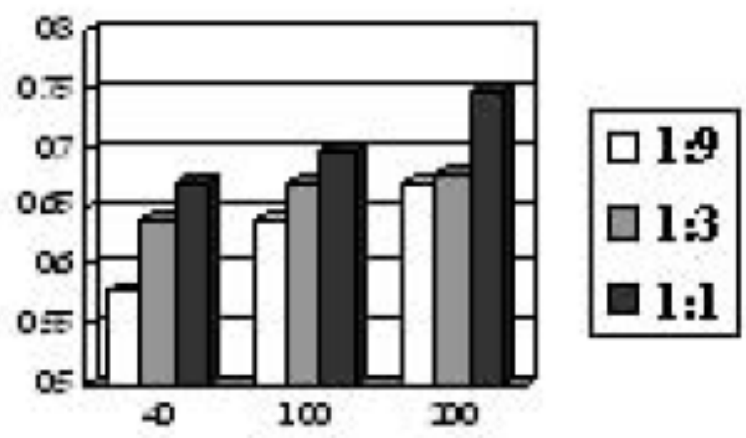
A. Fernandez, C.J. Carmona, M.J. del Jesus, F. Herrera. A View on Fuzzy Systems for Big Data: Progress and Opportunities. International Journal of Computational Intelligence Systems, 9:1 (2016), 69-80

Falta de datos:

Pima data sets, C4.5, Backpropagation

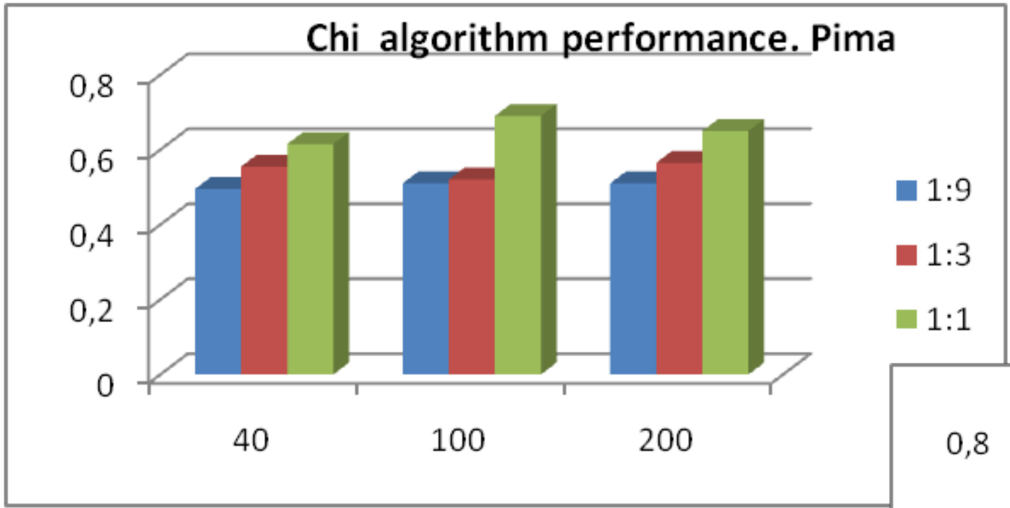
El mal comportamiento inicial de los clasificadores es reparado cuando aumenta el tamaño del conjunto de datos.

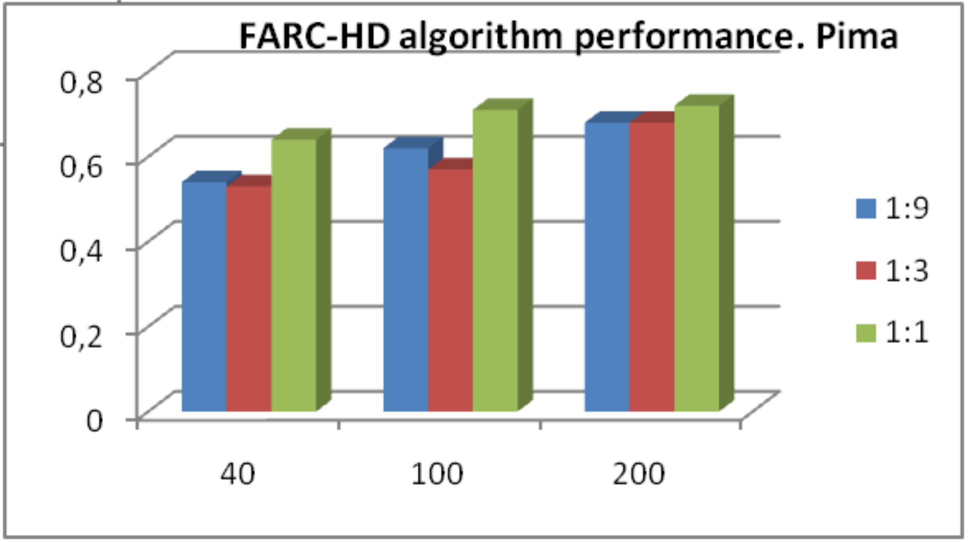




T. Jo, N. Japkowicz. Class imbalances versus small disjuncts. SIGKDD Explorations 6:1 (2004) 40-49

Ejemplo de Falta de datos. Comportamiento de sistemas difusos.





¿Robustez ante la falta datos? Problema abierto que hay que estudiar a fondo.

Big Data: Imprecisión e incertidumbre

- La imprecisión y la incertidumbre es inherente a Big Data debido a:
  - **Origen heterogéneo de los datos**
  - **Variedad de tipo de datos**
  - **Datos Imperfectos/incompletos/ densidad baja de datos**
- Los sistemas difusos pueden manejar:
  - **Incertidumbre**
  - **Imprecisión**
  - **Densidad baja de datos, falta de datos/ fragmentación de los datos**

- El uso de variables lingüísticas ofrece grandes ventajas en el marco de trabajo de Big Data
  - Permiten simplicidad y flexibilidad en el modelado de las soluciones.
  - Gestión intrínseca de la variedad (heterogeneidad), veracidad, y no completitud/imperfección de los datos
  - El solapamiento de etiquetas difusas facilita la cobertura del espacio del problema: muy importante en procesamiento MapReduce (small disjuncts!)

A. Fernandez, C.J. Carmona, M.J. del Jesus, F. Herrera. A View on Fuzzy Systems for Big Data: Progress and Opportunities. International Journal of Computational Intelligence Systems, 9:1 (2016), 69-80

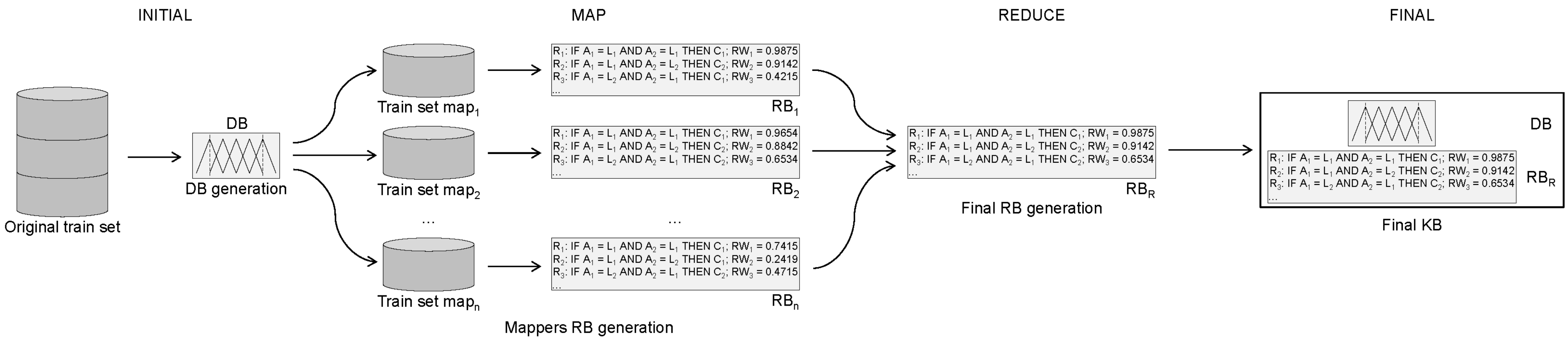
Índice

Chi-FRBCS (extensión: Wang & Mendel model)

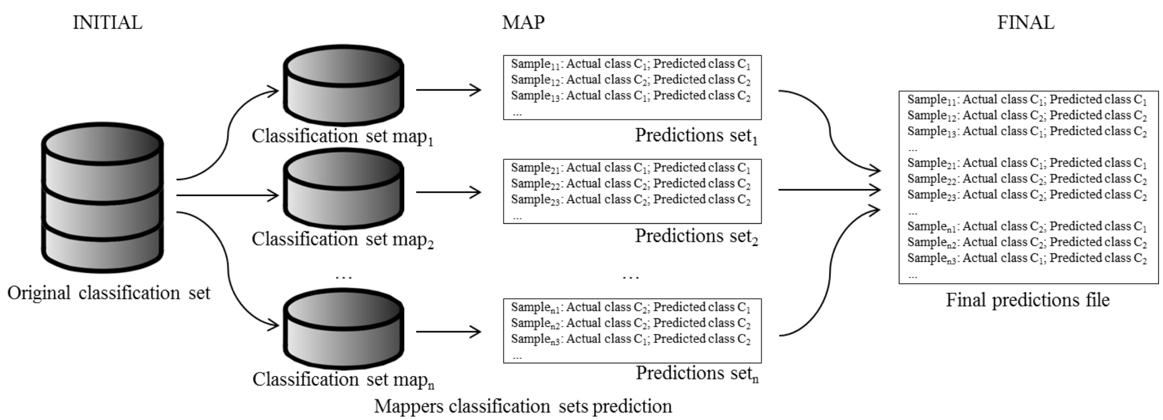
- “**Rule R_j:** IF x_1 IS A^1_j AND … AND x_n IS A^n_j THEN Class = C_j with RW_j”
- Builds the fuzzy partition using equally distributed triangular membership functions
- Builds the RB creating a fuzzy rule associated to each example
- Rules with the same antecedent may be created:
  - Same consequent → Delete duplicated rules
  - Different consequent → Preserve highest weight rule

Z. Chi, H. Yan and T. Pham, **Fuzzy algorithms with applications to image processing and pattern recognition**, World Scientific, 1996.

Sistemas Difusos en Big Data Analytics



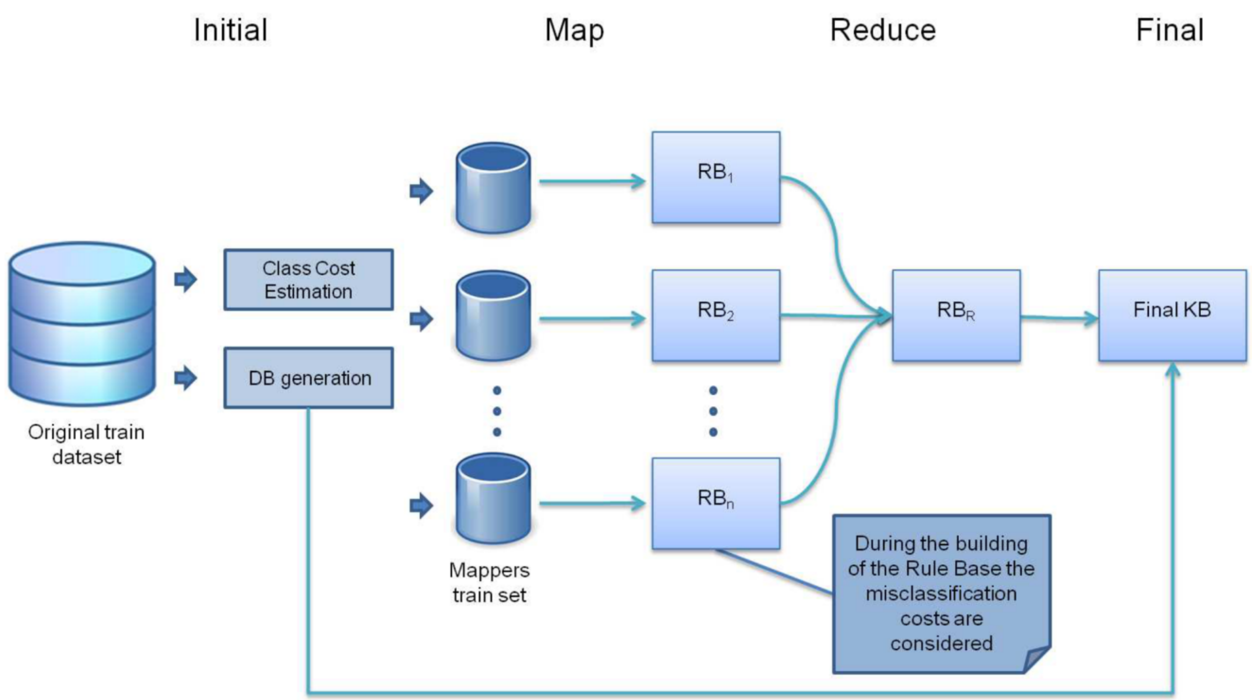
Flujo de cómo se gestiona la **construcción de la BC** en Chi-FRBCS-BigData



Flujo de cómo se gestiona la **clasificación** en Chi-FRBCS-BigData

S. Río, et al. A MapReduce Approach to Address Big Data Classification Problems Based on the Fusion of Linguistic Fuzzy Rules. International Journal of Computational Intelligence Systems 8:3 (2015) 422-437.

- Chi-FRBCS-BigDataCS: **Algoritmo imbalanced BigData**



V. López, S. Río, J.M. Benítez, F. Herrera. *Cost-Sensitive Linguistic Fuzzy Rule Based Classification Systems under the MapReduce Framework for Imbalanced Big Data*. Fuzzy Sets and Systems 258 (2015) 5-38.

Construyendo la RB con Chi-FRBCS-BigData-Max

R_1: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_1; $RW_1 = 0.8743$
R_2: IF $A_1 = L_2$ AND $A_2 = L_2$ THEN C_2; $RW_2 = 0.9142$
...

RB_1

R_1: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_2; $RW_3 = 0.9254$
R_2: IF $A_1 = L_1$ AND $A_2 = L_2$ THEN C_2; $RW_2 = 0.8842$
...

RB_2

R_1: IF $A_1 = L_2$ AND $A_2 = L_1$ THEN C_2; $RW_3 = 0.6534$
R_2: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_1; $RW_1 = 0.7142$
...

RB_3

R_1: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_2; $RW_1 = 0.2143$
R_2: IF $A_1 = L_3$ AND $A_2 = L_2$ THEN C_2; $RW_3 = 0.4715$
...

RB_4

...

R_1: IF $A_1 = L_2$ AND $A_2 = L_3$ THEN C_2; $RW_3 = 0.7784$
R_2: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_1; $RW_2 = 0.8215$
...

RB_n

REDUCE

R_1: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_1; $RW_1 = 0.9254$
R_2: IF $A_1 = L_2$ AND $A_2 = L_2$ THEN C_2; $RW_2 = 0.9142$
R_3: IF $A_1 = L_1$ AND $A_2 = L_2$ THEN C_2; $RW_2 = 0.8842$
R_4: IF $A_1 = L_2$ AND $A_2 = L_1$ THEN C_2; $RW_3 = 0.6534$
R_5: IF $A_1 = L_3$ AND $A_2 = L_2$ THEN C_2; $RW_3 = 0.4715$
R_6: IF $A_1 = L_2$ AND $A_2 = L_3$ THEN C_2; $RW_3 = 0.7784$
...

RB_R

Final RB generation

RB_1, R_1, C_1, RW = 0.8743
RB_2, R_1, **C_2**, **RW = 0.9254**
RB_3, R_2, C_1, RW = 0.7142
RB_4, R_1, C_2, RW = 0.2143
RB_5, R_2, C_1, RW = 0.8215

Construyendo la RB con Chi-FRBCS-BigData-Ave

R_1: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_1; $RW_1 = 0.8743$
R_2: IF $A_1 = L_2$ AND $A_2 = L_2$ THEN C_2; $RW_2 = 0.9142$
…

RB_1

R_1: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_2; $RW_3 = 0.9254$
R_2: IF $A_1 = L_1$ AND $A_2 = L_2$ THEN C_2; $RW_2 = 0.8842$
…

RB_2

R_1: IF $A_1 = L_2$ AND $A_2 = L_1$ THEN C_2; $RW_3 = 0.6534$
R_2: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_1; $RW_1 = 0.7142$
…

RB_3

R_1: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_2; $RW_1 = 0.2143$
R_2: IF $A_1 = L_3$ AND $A_2 = L_2$ THEN C_2; $RW_3 = 0.4715$
…

RB_4

…

R_1: IF $A_1 = L_2$ AND $A_2 = L_3$ THEN C_2; $RW_3 = 0.7784$
R_2: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_1; $RW_2 = 0.8215$
…

RB_n

REDUCE

R_1: IF $A_1 = L_1$ AND $A_2 = L_1$ THEN C_1; $RW_1 = 0.8033$
R_2: IF $A_1 = L_2$ AND $A_2 = L_2$ THEN C_2; $RW_2 = 0.9142$
R_3: IF $A_1 = L_1$ AND $A_2 = L_2$ THEN C_2; $RW_2 = 0.8842$
R_4: IF $A_1 = L_2$ AND $A_2 = L_1$ THEN C_2; $RW_3 = 0.6534$
R_5: IF $A_1 = L_3$ AND $A_2 = L_2$ THEN C_2; $RW_3 = 0.4715$
R_6: IF $A_1 = L_2$ AND $A_2 = L_3$ THEN C_2; $RW_3 = 0.7784$
…

RB_R

Final RB generation

$RB_1, R_1, C_1, RW = 0.8743$
$RB_2, R_1, C_2, RW = 0.9254$
$RB_3, R_2, C_1, RW = 0.7142$
$RB_4, R_1, C_2, RW = 0.2143$
$RB_5, R_2, C_1, RW = 0.8215$

→

RC_1, **C_1, $RW_{ave} = 0.8033$**
$RC_2, C_2, RW_{ave} = 0.5699$

Análisis de comportamiento: Precisión

¡Buena precisión con MapReduce!

Datasets	8 maps					
	Chi-FRBCS		Chi-BigData-Max		Chi-BigData-Ave	
	Acc_{tr}	Acc_{tst}	Acc_{tr}	Acc_{tst}	Acc_{tr}	Acc_{tst}
Poker_0_vs_1	63.72	**61.77**	62.93	60.74	63.12	60.91
Covtype_2_vs_1	74.65	74.57	74.69	**74.63**	74.66	74.61
Census	96.52	86.06	97.12	**93.89**	97.12	93.86
Fars_Fatal_Inj_vs_No_Inj	99.66	89.26	97.01	95.07	97.18	**95.25**
Average	83.64	77.92	82.94	81.08	83.02	**81.16**

Datasets	#Ex.	#Atts.	Selected classes	#Samples per class
RLCP	5749132	2	(FALSE; TRUE)	(5728201; 20931)
Kddcup_DOS_vs_normal	4856151	41	(DOS; normal)	(3883370; 972781)
Poker_0_vs_1	946799	10	(0; 1)	(513702; 433097)
Covtype_2_vs_1	495141	54	(2; 1)	(283301; 211840)
Census	141544	41	(-_50000.; 50000+.)	(133430; 8114)
Fars_Fatal_Inj_vs_No_Inj	62123	29	(Fatal_Inj; No_Inj)	(42116; 20007)

3 etiquetas partición
T-norm Prod.
Peso W: PCF
10fcv

Análisis de comportamiento: Tiempo escalable

Datasets		8 maps	
	Chi-FRBCS	Chi-BigData-Max	Chi-BigData-Ave
	Runtime (s)	Runtime (s)	Runtime (s)
Census	38655.60	**1102.45**	1343.92
Covtype_2_vs_1	86247.70	**2482.09**	2512.16
Fars_Fatal_Inj_vs_No_Inj	8056.60	**241.96**	311.95
Poker_0_vs_1	114355.80	**5672.80**	7682.19
Average	61828.93	**2374.82**	2962.56

Sistemas Difusos en Big Data Analytics

- **Características a destacar**: Resultados prometedores
  - Es factible procesar grandes volúmenes de datos
  - Se puede mantener una buena precisión
  - Respuesta rápida (incrementando el número de Maps)

- **Retos y estudios a realizar:** Muchos interrogantes
  - ¿Cómo diseñar modelos no lingüísticos?
  - Análisis "ensembles vs fusion de rules"

    ¿Son útiles los ensembles lingüísticos? Los ensembles de árboles (Random Forest) son modelos importantes, permiten conocer la importancia de las variables, precisión, …
  - Análisis de los "small disjuncts" vía preprocesamiento y granularidad de las particiones lingüísticas.
  - ¿Cómo abordar problemas de alta dimensionalidad?

Gracias por su atención

- ¿Preguntas?

CLASIFICACIÓN Y MACHINE LEARNING CON TÉCNICAS DIFUSAS

José Antonio Sanz y Mikel Galar
Universidad Pública de Navarra

Alberto Fernández Hilario
Universidad de Granada

EVIA 2016