Representación del Conocimiento y Razonamiento Automático

Departamento de Computación

Universidad de A Coruña

Vicente Moret Bonillo

Senior Member, IEEE

facultade de informática da coruña

Métodos Difusos

- Aspectos generales
- Nomenclatura
- Estructura algebraica
- Operaciones algebraicas
- Representación del conocimiento
- Razonamiento difuso

Métodos Difusos

- Textos Básicos
  - Lofti Zadeh, Fuzzy Sets, Information and Control, vol.8, 1965
  - Watson, Weiss & Donnell, Fuzzy decision analysis, IEEE Trans. Systems, Man & Cybernetics, vol.9, 1979
  - Zadeh, Knowledge representation in fuzzy logic, IEEE Trans. Knowledge & Data Engineering, vol.1, 1989

Métodos Difusos

- Planteamiento general…
  - Criterios lingüísticos para describir los objetos del mundo
  - Ambigüedad
    - Lenguaje
    - Clasificación
    - Taxonomías
    - Jerarquías
    - Razonamiento

Métodos Difusos

- Concepto de “ser vivo”… Definición
  - ¿planta? ¿piedra? ¿virus?
- Cuestiones subjetivas
  - Lo hermoso
  - Esto es hermoso
- Contextos…Ejemplo: Alto
  - Sueco
  - Pigmeo
  - Niño

Conjuntos Difusos

- Lofti Zadeh, 1965
  - Universo N de los números naturales

$$A \subseteq N \,/\, A = naturales\_pares\_menores\_que\_10$$

$$A = \{2,4,6,8\}$$

$$2 \in A$$

$$3 \notin A$$

$$\ldots$$

Conjuntos Difusos

- Referencial U = seres humanos vivos
- B $\subseteq$ U / B = seres humanos vivos morenos y altos…¿B?

$$Conjuntos\_ordinarios$$

$$U, A \subseteq U / \exists \mu_A(x) : U \rightarrow \{0, x \notin A : 1, x \in A\}, \forall x \in U$$

$$Conjuntos\_difusos$$

$$U, A \subseteq U / \exists \mu_A(x) : U \rightarrow [0,1], \forall x \in U$$

$$\mu_A(x) = 1 \rightarrow x \in A$$

$$\mu_A(x) = 0 \rightarrow x \notin A$$

$$0 < \mu_A(x) < 1 \rightarrow (x \in A) \wedge (x \notin A)$$

Conjuntos Difusos

- Criterios para definir la función μ de grado de pertenencia…
  - U = personas vivas
  - $A \subseteq U$ / A = personas vivas jóvenes
  - Criterio para μ
    - ¿edad?
    - ¿cómo?

Conjuntos Difusos

- Aplicación ferroviaria del criterio
  - Joven $\rightarrow$ Inter-Rail
  - No Joven $\rightarrow$ Tarjeta oro de Renfe (3ª edad)

$$\mu_A(x) = 1 : \forall x / Edad(x) \leq 25\_años$$

$$\mu_A(x) = 0 : \forall x / Edad(x) \geq 65\_años$$

$$¿\mu_A(x) : \forall x / 25 < Edad(x) < 65?$$

Conjuntos Difusos

- Criterio lineal para μ:

$$\mu_A(x) = \frac{65 - Edad(x)}{40} : \forall x \,/\, Edad(x) \in [25,65]$$

$$A = Persona\_viva\_joven$$

$$Edad(Juan) = 17 \rightarrow \mu_{joven}(Juan) = 1.00$$

$$Edad(Marisa) = 31 \rightarrow \mu_{joven}(Marisa) = 0.85$$

$$Edad(Blas) = 47 \rightarrow \mu_{joven}(Blas) = 0.45$$

$$Edad(Ana) = 57 \rightarrow \mu_{joven}(Ana) = 0.20$$

$$Edad(Alex) = 73 \rightarrow \mu_{joven}(Alex) = 0.00$$

Conjuntos Difusos

- Cuantificación lingüística (incorrecta)
  - Juan es 1.00 Joven
  - Marisa es 0.85 Joven
  - Blas es 0.45 Joven
  - Ana es 0.20 Joven
  - Alex es 0.00 Joven

Conjuntos Difusos

- Escala semántica (correcta)

$$0.00 = \mu_A(x) = 0.00 \rightarrow No\_es$$

$$0.00 < \mu_A(x) < 0.20 \rightarrow Es\_muy\_poco$$

$$0.20 \leq \mu_A(x) \leq 0.40 \rightarrow Es\_poco$$

$$0.40 < \mu_A(x) < 0.60 \rightarrow Es\_a\lg o$$

$$0.60 \leq \mu_A(x) \leq 0.80 \rightarrow Es\_\mathrm{mod}\,eradamente$$

$$0.80 < \mu_A(x) < 1.00 \rightarrow Es\_bas\tan te$$

$$1.00 = \mu_A(x) = 1.00 \rightarrow Es$$

Conjuntos Difusos

- Traducción

$x \in U$	Etiqueta(μ)	$A \subseteq U$
Juan	Es	Joven
Marisa	Es bastante	Joven
Blas	Es algo	Joven
Ana	Es poco	Joven
Alex	No es	Joven

Conjuntos Difusos

- Comentarios
  - La zona difusa no tiene por qué ser lineal
  - La escala lingüística es arbitraria
  - El número de elementos de la escala es arbitrario
  - Podemos definir conjuntos difusos complementarios
  - Cualquier conjunto es difuminable

Conjuntos Difusos

- Caracterización y nomenclatura de conjuntos ordinarios
  - Implícitamente
    - En función del dominio: Naturales pares
  - Explícitamente
    - Identificando sus elementos: {2, 4, 6, 8}
  - Mediante una función booleana $\forall\ x \in U$

Conjuntos Difusos

- Caracterización y nomenclatura de conjuntos difusos
  - Dado un referencial U, y un subconjunto A de ese referencial U:

$$\forall A \subset U : A\_difuso \leftrightarrow \exists \mu_A(x) : U \rightarrow [0,1] \forall x \in U$$

Conjuntos Difusos

- Dado un referencial U, y sea A $\subset$ U

$$A\_ordinario \rightarrow \mu_A(x) = \{0, x \notin A : 1, x \in A\}$$

$$Naturales\_pares\_menores\_que\_10$$

$$\mu_A(x) = \{\mu_A(1) = 0 + \mu_A(2) = 1 + \mu_A(3) = 0 +$$

$$+ \mu_A(4) = 1 + \mu_A(5) = 0 + \mu_A(6) = 1 + \mu_A(7) = 0 +$$

$$+ \mu_A(8) = 1 + \mu_A(9) = 0 + \mu_A(10) = 0 + \ldots\}$$

$$\mu_A(x) = \{0/1 + 1/2 + 0/3 + 1/4 + 0/5 +$$

$$+ 1/6 + 0/7 + 1/8 + 0/9 + 0/10 + \ldots\}$$

Conjuntos Difusos

- Estructura algebraica
  - Conjunto vacío

$$Universo\_referencial : U$$

$$Z \subset U \,/\, \exists \mu_Z(x) : U \rightarrow [0,1] \forall x \in U$$

$$Z = \varnothing \leftrightarrow \mu_Z(x) = 0 \forall x \in U$$

Conjuntos Difusos

- Identidad

$$Universo\_referencial : U$$

$$A \subset U \,/\, \exists \mu_A(x) : U \rightarrow [0,1] \forall x \in U$$

$$B \subset U \,/\, \exists \mu_B(x) : U \rightarrow [0,1] \forall x \in U$$

$$A = B \leftrightarrow \mu_A(x) = \mu_B(x) : \forall x \in U$$

Conjuntos Difusos

- Complementariedad

$$Universo\_referencial : U$$

$$A \subset U \,/\, \exists \mu_A(x) : U \rightarrow [0,1] \forall x \in U$$

$$\neg A = A' \leftrightarrow \mu_{A'}(x) = 1 - \mu_A(x) : \forall x \in U$$

Conjuntos Difusos

- Inclusión

$$Universo\_referencial : U$$

$$A \subset U / \exists \mu_A(x) : U \rightarrow [0,1] \forall x \in U$$

$$B \subset U / \exists \mu_B(x) : U \rightarrow [0,1] \forall x \in U$$

$$B \subset A \leftrightarrow \mu_B(x) \leq \mu_A(x) : \forall x \in U$$

Conjuntos Difusos

- Ejemplo de inclusión

$$U = \{1,2,3,4,5,6\}$$

$$A \subset U \;/\; A = \{1,2,3,4\}$$

$$B \subset U \;/\; B = \{1,3\}$$

$$\mu_A(x) = \{1/1 + 1/2 + 1/3 + 1/4 + 0/5 + 0/6\}$$

$$\mu_B(x) = \{1/1 + 0/2 + 1/3 + 0/4 + 0/5 + 0/6\}$$

$$1/1 = 1/1 : 1/2 > 0/2 : 1/3 = 1/3$$

$$1/4 > 0/4 : 0/5 = 0/5 : 0/6 = 0/6$$

Conjuntos Difusos

- Unión

$$Universo\_referencial: U$$

$$A \subset U / \exists \mu_A(x): U \rightarrow [0,1] \forall x \in U$$

$$B \subset U / \exists \mu_B(x): U \rightarrow [0,1] \forall x \in U$$

$$C \subset U / \exists \mu_C(x): U \rightarrow [0,1] \forall x \in U$$

$$C = A \cup B \leftrightarrow \mu_C(x) = \max\{\mu_A(x), \mu_B(x)\}: \forall x \in U$$

$$Asociatividad: A \cup (B \cup C) = (A \cup B) \cup C$$

Conjuntos Difusos

- Intersección

$$Universo\_referencial : U$$

$$A \subset U \,/\, \exists \mu_A(x) : U \rightarrow [0,1] \forall x \in U$$

$$B \subset U \,/\, \exists \mu_B(x) : U \rightarrow [0,1] \forall x \in U$$

$$C \subset U \,/\, \exists \mu_C(x) : U \rightarrow [0,1] \forall x \in U$$

$$C = A \cap B \leftrightarrow \mu_C(x) = \min\{\mu_A(x), \mu_B(x)\} : \forall x \in U$$

$$Asociatividad : A \cap (B \cap C) = (A \cap B) \cap C$$

Conjuntos Difusos

- Leyes de DeMorgan

$$Universo\_referencia\ l : U$$

$$A \subset U \ / \ \exists \mu_A(x) : U \to [0,1] \forall x \in U$$

$$B \subset U \ / \ \exists \mu_B(x) : U \to [0,1] \forall x \in U$$

1ª *Ley*

$$\neg(A \cup B) = \neg A \cap \neg B$$

$$\neg[\mu_A(x) \vee \mu_B(x)] = [\neg \mu_A(x) \wedge \neg \mu_B(x)] : \forall x \in U$$

2ª *Ley*

$$\neg(A \cap B) = \neg A \cup \neg B$$

$$\neg[\mu_A(x) \wedge \mu_B(x)] = [\neg \mu_A(x) \vee \neg \mu_B(x)] : \forall x \in U$$

Conjuntos Difusos

- Demostración 1ª ley de DeMorgan (1)

$$Universo\_referencial : U$$

$$A \subset U / \exists \mu_A(x) : U \rightarrow [0,1] \forall x \in U$$

$$B \subset U / \exists \mu_B(x) : U \rightarrow [0,1] \forall x \in U$$

$$\neg(A \cup B) \rightarrow \mu_{\neg(A \cup B)}(x) = 1 - \max\{\mu_A(x), \mu_B(x)\} : \forall x \in U$$

$$\neg A \cap \neg B \rightarrow \mu_{\neg A \cap \neg B}(x) = \min\{1 - \mu_A(x), 1 - \mu_B(x)\} : \forall x \in U$$

Conjuntos Difusos

- Demostración 1ª ley de DeMorgan (2)

$$Casos\_extremos$$

$$\mu_A(x) \geq \mu_B(x) : \forall x \in U$$

$$1 - \max\{\mu_A(x), \mu_B(x)\} = 1 - \mu_A(x) : \forall x \in U$$

$$\min\{1 - \mu_A(x), 1 - \mu_B(x)\} = 1 - \mu_A(x) : \forall x \in U$$

$$\mu_A(x) \leq \mu_B(x) : \forall x \in U$$

$$1 - \max\{\mu_A(x), \mu_B(x)\} = 1 - \mu_B(x) : \forall x \in U$$

$$\min\{1 - \mu_A(x), 1 - \mu_B(x)\} = 1 - \mu_B(x) : \forall x \in U$$

Conjuntos Difusos

- Leyes distributivas

$$Universo\_referencial\_U$$

$$A \subset U / \exists \mu_A(x) \to U : [0,1] \forall x \in U$$

$$B \subset U / \exists \mu_B(x) \to U : [0,1] \forall x \in U$$

$$C \subset U / \exists \mu_C(x) \to U : [0,1] \forall x \in U$$

$$1^a\_ley$$

$$C \cap (A \cup B) = (C \cap A) \cup (C \cap B)$$

$$\mu_C(x) \wedge [\mu_A(x) \vee \mu_B(x)] = [\mu_C(x) \wedge \mu_A(x)] \vee [\mu_C(x) \wedge \mu_B(x)] : \forall x \in U$$

$$2^a\_ley$$

$$C \cup (A \cap B) = (C \cup A) \cap (C \cup B)$$

$$\mu_C(x) \vee [\mu_A(x) \wedge \mu_B(x)] = [\mu_C(x) \vee \mu_A(x)] \wedge [\mu_C(x) \vee \mu_B(x)] : \forall x \in U$$

Conjuntos Difusos

- Los conjuntos difusos no son un álgebra de Boole(1)

$$Ley\_del\_tercero\_excluido$$

$$Universo\_referencial\_U$$

$$Conjuntos\_ordinarios: A \cup \neg A = U$$

$$Conjuntos\_difusos:$$

$$A \subset U / \exists \mu_A(x) \rightarrow U:[0,1] \forall x \in U$$

$$\neg A \rightarrow \mu_{A'}(x) = 1 - \mu_A(x): \forall x \in U$$

$$A \cup \neg A \rightarrow \mu_{A \cup A'}(x) = \max\{\mu_A(x), \mu_{A'}(x)\} = \max\{\mu_A(x), 1 - \mu_A(x)\}: \forall x \in U$$

$$\mu_{A \cup A'}(x) \geq \frac{1}{2}$$

Conjuntos Difusos

- Los conjuntos difusos no son un álgebra de Boole(2)

$$Ley\_de\_no\_contradicción$$

$$Universo\_referencial\_U$$

$$Conjuntos\_ordinarios : A \cap \neg A = \varnothing$$

$$Conjuntos\_difusos :$$

$$A \subset U / \exists \mu_A(x) \to U : [0,1] \forall x \in U$$

$$\neg A \to \mu_{A'}(x) = 1 - \mu_A(x) : \forall x \in U$$

$$A \cap \neg A \to \mu_{A \cap A'}(x) = \min\{\mu_A(x), \mu_{A'}(x)\} = \min\{\mu_A(x), 1 - \mu_A(x)\} : \forall x \in U$$

$$\mu_{A \cap A'}(x) \leq \frac{1}{2}$$

Conjuntos Difusos

- Producto de conjuntos difusos

$$Universo\_referencial\_U$$

$$A \subset U / \exists \mu_A(x) \rightarrow U : [0,1] \forall x \in U$$

$$B \subset U / \exists \mu_B(x) \rightarrow U : [0,1] \forall x \in U$$

$$A \times B \rightarrow \mu_{AB}(x) = \mu_A(x) \times \mu_B(x) : \forall x \in U$$

Conjuntos Difusos

- Relación Producto-Intersección

$$Conjuntos\_ordinarios$$

$$A \times B = A \cap B$$

$$Conjuntos\_difusos$$

$$Universo\_referencial\_U$$

$$A \subset U \;/\; \exists \mu_A(x) \to U : [0,1] \forall x \in U$$

$$B \subset U \;/\; \exists \mu_B(x) \to U : [0,1] \forall x \in U$$

$$A \times B \to \mu_{AB}(x) = \mu_A(x) \times \mu_B(x) : \forall x \in U$$

$$A \cap B \to \mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\} : \forall x \in U$$

$$\mu_{AB}(x) < \min\{\mu_A(x), \mu_B(x)\} : \forall x \in U$$

$$A \times B \subset A \cap B$$

Conjuntos Difusos

- Suma y suma acotada

$$Universo\_referencial\_U$$

$$A \subset U \,/\, \exists \mu_A(x) \rightarrow U : [0,1] \forall x \in U$$

$$B \subset U \,/\, \exists \mu_B(x) \rightarrow U : [0,1] \forall x \in U$$

$$A + B \rightarrow \mu_{A+B}(x) = \mu_A(x) + \mu_B(x) : \forall x \in U$$

$$A \oplus B \rightarrow \mu_{A \oplus B}(x) = \min\{1, \mu_A(x) + \mu_B(x)\} : \forall x \in U$$

Conjuntos Difusos

- Diferencia y diferencia absoluta

$$Universo\_referencial\_U$$

$$A \subset U / \exists \mu_A(x) \rightarrow U : [0,1] \forall x \in U$$

$$B \subset U / \exists \mu_B(x) \rightarrow U : [0,1] \forall x \in U$$

$$A - B \rightarrow \mu_{A-B}(x) = \mu_A(x) - \mu_B(x) : \forall x \in U$$

$$|A - B| \rightarrow \mu_{|A-B|}(x) = |\mu_A(x) - \mu_B(x)| : \forall x \in U$$

Conjuntos Difusos

- Núcleo de un conjunto difuso

$$Universo\_referencial\_U$$

$$A \subset U / \exists \mu_A(x) \to U:[0,1] \forall x \in U$$

$$Núcleo: N_A = \{x \in U / \mu_A(x) = 1\}$$

$$A = Normalizado \leftrightarrow N_A \neq \{\varnothing\}$$

Conjuntos Difusos

- Relación difusa
  - Dado un referencial U, definimos una relación difusa de orden “n” en U, como un conjunto difuso A en el espacio U x U x…x U (n veces), caracterizado por una función de grado de pertenencia del tipo:

$$\mu_A(x1, \ldots, xn) : \forall x \in U$$

Conjuntos Difusos

- Ejemplo
  - Referencial U = {1, 2, 3, 4}
  - Relación difusa de orden 2
  - A = {números aproximadamente iguales}

	U2			
U1	1	2	3	4
1	1.0	0.8	0.2	0.0
2	0.8	1.0	0.8	0.2
3	0.2	0.8	1.0	0.8
4	0.0	0.2	0.8	0.0

Conjuntos Difusos

- Representación y razonamiento
  - Normalmente se tarda alrededor de 45´ en llegar desde Oleiros a Santiago, por la autopista, y con tráfico ligero
  - No es previsible que el paro disminuya en España, al menos de manera drástica, en los próximos meses
  - La mayoría de los expertos opinan que la probabilidad de un terremoto serio en el Caurel es pequeña en un futuro inmediato

Conjuntos Difusos

- Cuestiones importantes
  - Predicados difusos, cuantificadores difusos, probabilidades difusas
  - El conocimiento derivado del sentido común es léxicamente impreciso
  - El conocimiento derivado del sentido común es no categórico
  - Las lógicas de 1er orden, o las teorías clásicas de la probabilidad no son suficientes para representar el sentido común

Conjuntos Difusos

- Características
  - En lógica difusa, el razonamiento categórico es un caso particular del razonamiento aproximado
  - En lógica difusa todo es cuestión de grado
  - Cualquier sistema lógico puede ser difuminado
  - En lógica difusa el conocimiento se interpreta como una colección de restricciones difusas que operan sobre variables
  - En lógica difusa los procesos inferenciales son propagaciones de las restricciones difusas

Conjuntos Difusos

- Implicación difusa
  - Si x es A Entonces y es B

$$Universo\_referencial\_U$$
$$A \subset U / \exists \mu_A(x): U \to [0,1] \forall x \in U$$
$$Universo\_referencial\_V$$
$$B \subset V / \exists \mu_B(y): V \to [0,1] \forall y \in V$$
$$A \to B \equiv \neg A \oplus B : en\_U \times V$$
$$\mu_{A \to B}(x, y) = \min\{1, 1 - \mu_A(x) + \mu_B(y)\} : x \in U, y \in V$$

Conjuntos Difusos

- Modus Ponens Generalizado:
  - Si x es A entonces y es B
  - x es A*
  - y es B*

$$\mu_{B^*}(y) = \sup_V\{[\neg A \oplus B] \cap A^*\}$$

$$A \subset U : A^* \subset U$$

$$B \subset V : B^* \subset V$$

$$\mu_{B^*}(y) = \sup_V\{\min[\min[1, 1 - \mu_A(x) + \mu_B(y)], \mu_{A^*}(x)]\}$$

$$x \in U : y \in V$$

Conjuntos Difusos

- Ejemplo
  - Sean A ⊂ U, A* ⊂ U, B ⊂ V, B* ⊂ V
  - U = V = {1, 2, 3, 4}

$$\mu_A(x) = \{0.0/1 + 0.6/2 + 1.0/3 + 0.5/4\}$$

$$\mu_{A*}(x) = \{0.0/1 + 0.2/2 + 0.6/3 + 1.0/4\}$$

$$\mu_B(y) = \{0.0/1 + 1.0/2 + 0.6/3 + 0.2/4\}$$

$$x\_es\_A \rightarrow y\_es\_B$$

$$x\_es\_A*$$

Conjuntos Difusos

- Encontrar:
  - $\neg$ A
  - Evaluar $\neg$ A $\oplus$ B en U x V
  - Evaluar [$\neg$ A $\oplus$ B] $\cap$ A* en U x V
  - Encontrar la expresión que caracteriza a B*

Conjuntos Difusos

$$\neg A \to \mu_{\neg A}(x) = \{1.0/1 + 0.4/2 + 0.0/3 + 0.5/4\}$$

$$\neg A \oplus B : U \times V$$

	V			
U	1	2	3	4
1	1.0	1.0	1.0	1.0
2	0.4	1.0	1.0	0.6
3	0.0	1.0	0.6	0.2
4	0.5	1.0	1.0	0.7

Conjuntos Difusos

$$A^* \rightarrow \mu_{A^*}(x) = \{0/1 + 0.2/2 + 0.6/3 + 1/4\}$$

$$(\neg A \oplus B) \cap A^* : U \times V$$

	V			
U	1	2	3	4
1	0.0	0.0	0.0	0.0
2	0.2	0.2	0.2	0.2
3	0.0	0.6	0.6	0.2
4	0.5	1.0	1.0	0.7

Conjuntos Difusos

- Y el resultado final, que caracteriza al conjunto difuso B* es…

$$B^* \to \mu_{B^*}(y) = \sup_V \{(\neg A \oplus B) \cap A^*\} : U \times V$$

$$\mu_{B^*}(y) = \{0.5/1 + 1.0/2 + 1.0/3 + 0.7/4\} : y \in V$$

Conjuntos Difusos

- Diferencias entre razonamiento convencional y razonamiento difuso (1)
  - Modus Ponens Generalizado
  - Certeza
    - Lógica bivalente
      - Cierto o Falso
    - Sistemas multivaluados
      - Elemento de un conjunto finito
      - Un intervalo
      - Un álgebra de Boole

Conjuntos Difusos

- Diferencias entre razonamiento convencional y razonamiento difuso (2)
  - Predicados
    - Sistemas bivalentes
      - Predicados categóricos
    - Sistemas difusos
      - Predicados difusos
  - Modificadores
    - Sistemas clásicos : NOT
    - Sistemas difusos : MUY, MÁS O MENOS, BASTANTE,…

Conjuntos Difusos

- Diferencias entre razonamiento convencional y razonamiento difuso (3)
  - Cuantificadores
    - Sistemas clásicos: $\forall$, $\exists$
    - Sistemas difusos: pocos, bastantes, la mayoría,…
  - Probabilidades
    - Sistemas clásicos: probabilidad numérica
    - Sistemas difusos: etiquetas lingüísticas (plausible,…)

Conjuntos Difusos

- Diferencias entre razonamiento convencional y razonamiento difuso (4)
  - Posibilidades
    - A diferencia de lo que ocurre con los sistemas lógicos clásicos, el concepto de posibilidad en los sistemas difusos no es bivalente. Al igual que sucede con las probabilidades, las posibilidades pueden ser tratadas como variables lingüísticas que adoptan valores del tipo: casi imposible, bastante posible,…

Conjuntos Difusos

- Modos de Razonamiento (1)
  - Razonamiento categórico
    - Utiliza declaraciones difusas, pero no emplea ni cuantificadores difusos ni probabilidades difusas
    - Ejemplo
      - María es una chica delgada
      - María es una chica muy inteligente
      - ____
      - María es una chica delgada y muy inteligente

Conjuntos Difusos

- Modos de Razonamiento (2)
  - Razonamiento Silogístico
    - Produce inferencias con premisas que incorporan cuantificadores difusos
    - Ejemplo
      - La mayoría de los suecos son rubios
      - La mayoría de los suecos rubios son altos
      - ____________________________________
      - $(\text{La mayoría})^2$ de los suecos son rubios y altos
      - $(\text{La mayoría})^2$ es el cuadrado de (La mayoría) en aritmética difusa

Conjuntos Difusos

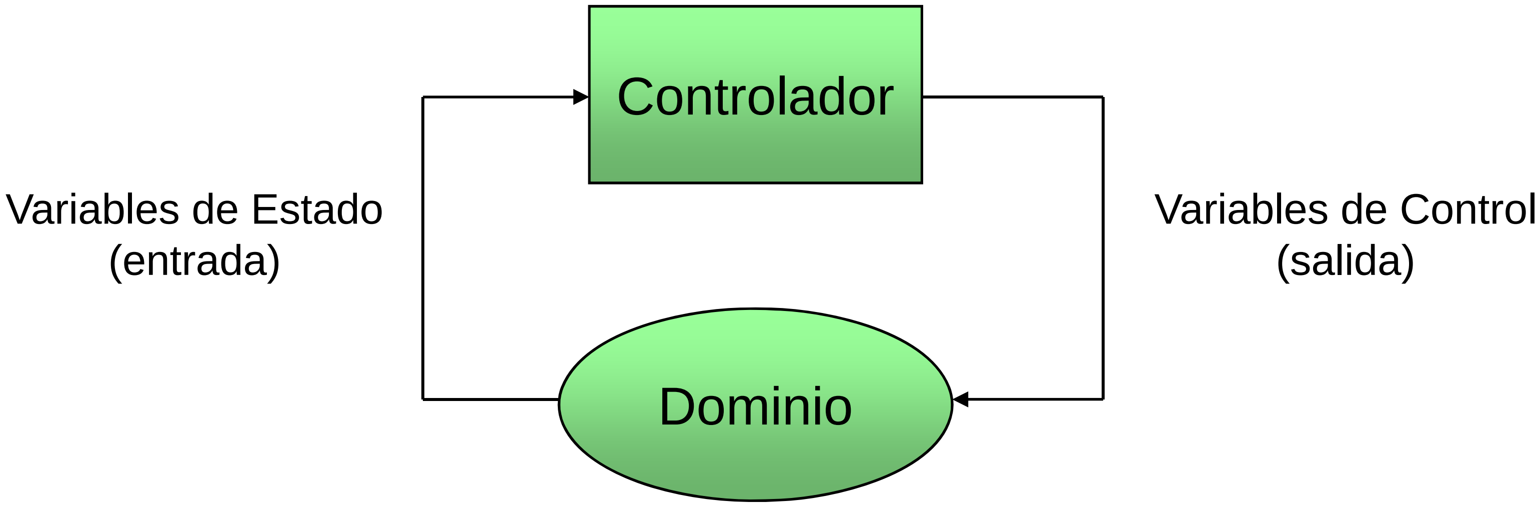
- Modos de Razonamiento (3)
  - Razonamiento Cualitativo
    - Se define como un modo de razonamiento en el cual las relaciones entrada/salida de un sistema se representan por medio de una colección de reglas difusas de tipo IF-THEN, en las que los antecedentes y los consecuentes incluyen variables lingüísticas
    - Este tipo de razonamiento es el empleado habitualmente en las aplicaciones de la lógica difusa al análisis de sistemas y al control de procesos

Control Difuso

- Control
  - Cuando hablamos de control nos referimos a la regulación de características que pueden ser físicas, mecánicas, numéricas, etc.
  - Ejemplos de características a controlar serían: la temperatura, la corriente eléctrica, el fluido de gases o líquidos, medidas financieras como el nivel de caja óptimo, el tiempo de espera en un semáforo, etc.
- Controlador
  - Un controlador es un mecanismo que lee una serie de variables de estado de un determinado dominio, en base a estas entradas decide cómo actuar sobre el dominio modificando una serie de variables de control
  - Ej. Un termostato lee la temperatura ambiente, cuando esta cae por debajo de un mínimo enciende la caldera, cuando sobrepasa un máximo apaga la caldera

Control Difuso

- Partes de un controlador



Control Difuso

- Partes de un controlador difuso
  - Codificación (Fuzzificación)
    - Se encarga de leer y codificar las variables de entrada
    - Estas variables de entrada serán datos numéricos no difusos (crisp) que es necesario convertir a datos difusos para ser tratados por el controlador difuso
    - Por ejemplo, si leemos que la temperatura es de 25º C esto se puede codificar como que la temperatura pertenece al conjunto "temperatura alta", con un grado de pertenencia del 0,75.

Control Difuso

- Partes de un controlador difuso
  - Base de conocimiento
    - Contiene el conocimiento asociado al dominio de la aplicación y los objetivos del control
    - Está formado por un conjunto de reglas difusas de la forma:

      Si (X1 es A1) y (X2 es A2) y … (Xn es An) entonces (Y es B)

      Donde Xi son variables de estado, Y es una variable de control y los Ai y B son conjuntos difusos.
    - Ejemplo: Si la temperatura es alta la potencia de la caldera tiene que ser baja

Control Difuso

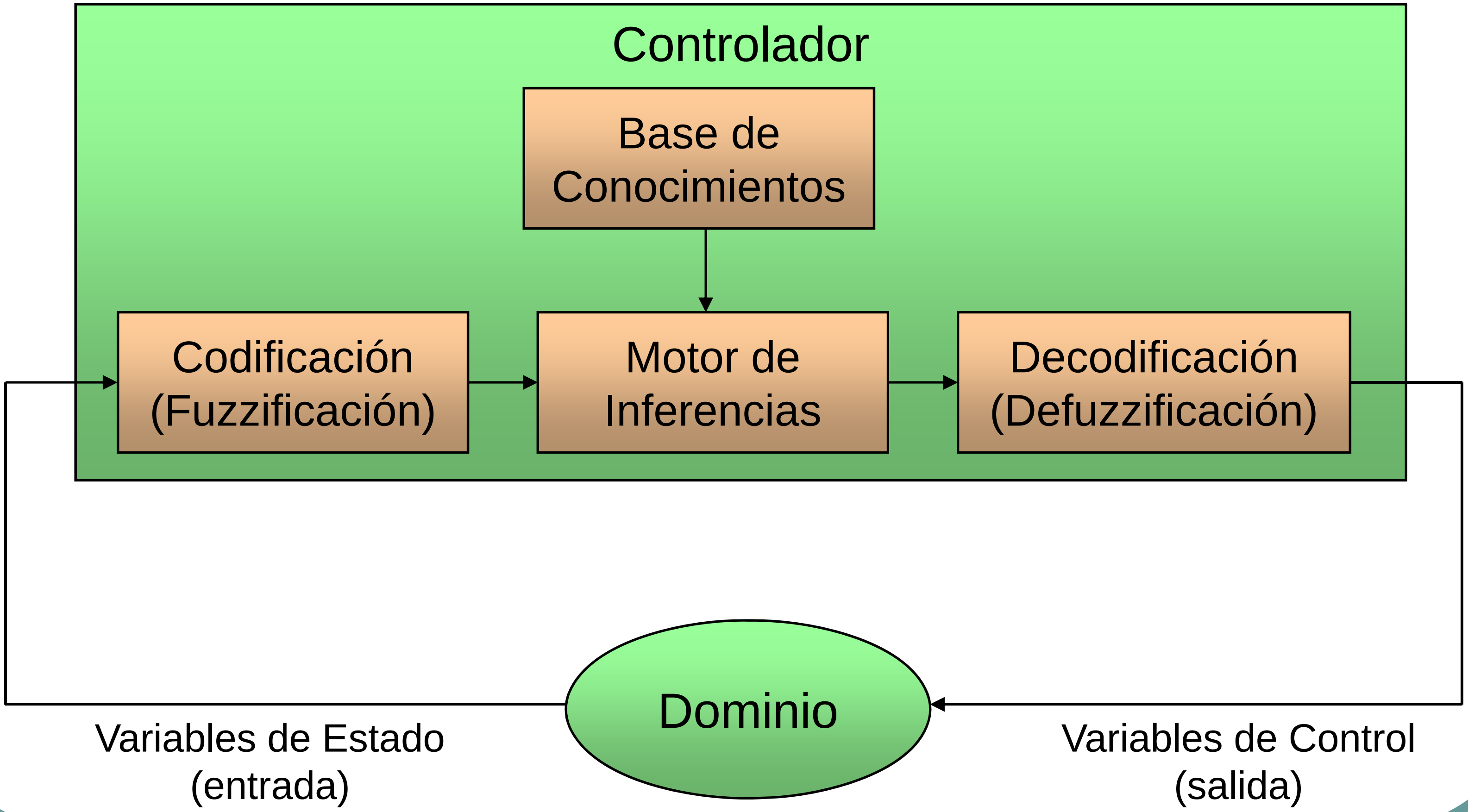
- Partes de un controlador difuso
  - Motor de inferencias
    - Es el núcleo del controlador difuso
    - Infiere las acciones de control utilizando las reglas de la base de conocimientos y simulando el proceso de decisión humano mediante la implicación difusa y las reglas de inferencia de la lógica difusa

Control Difuso

- Partes de un controlador difuso
  - Decodificación (Defuzzificación)
    - La salida de un controlador difuso es una variable difusa que es necesario convertir en una variable concreta (crisp) del dominio de discurso
    - El proceso de decodificación genera una acción no difusa a partir da la acción difusa resultante del sistema de inferencia
    - Ejemplo, el resultado puede ser: potencia de la caldera media con un grado de pertenencia del 0,25 y potencia de la caldera baja con un grado de pertenencia del 0,5. La decodificación convertirá este conjunto difuso en el valor exacto al que hay que poner la potencia de la caldera

Control Difuso

- Partes de un controlador difuso



Control Difuso

- Implicación clásica
  - La implicación en lógica clásica tiene la siguiente tabla de verdad

A	B	$A \rightarrow B$
0	0	1
0	1	1
1	0	0
1	1	1

- Ejemplo

Llueve	Nubes	Si Llueve $\rightarrow$ Hay Nubes	Explicación
0	0	1	**Si no llueve ni hay nubes no se contradice la implicación**
0	1	1	**Si no llueve pero hay nubes no se contradice la implicación**
1	0	0	**Si llueve y no hay nubes la implicación es falsa**
1	1	1	**Si llueve y hay nubes la implicación se cumple**

Control Difuso

- Implicación en control difuso
  - La implicación puede interpretarse como una implicación “local” y no “global” como en lógica clásica
  - Quiere decir que la implicación se cumple cuando el antecedente y el precedente son ciertos
  - En lógica difusa significa que el valor de verdad de la función de implicación será alto si el valor de verdad del antecendente y el consecuente es alto.

Abrir válvula	Salir gas	Abrir válvula $\rightarrow$ Salir gas	Explicación
0	**0**	**0**	**Situación indeterminada**
0	**1**	**0**	**Situación indeterminada**
1	**0**	**0**	**No se cumple la implicación**
1	**1**	**1**	**La implicación se cumple**

Control Difuso

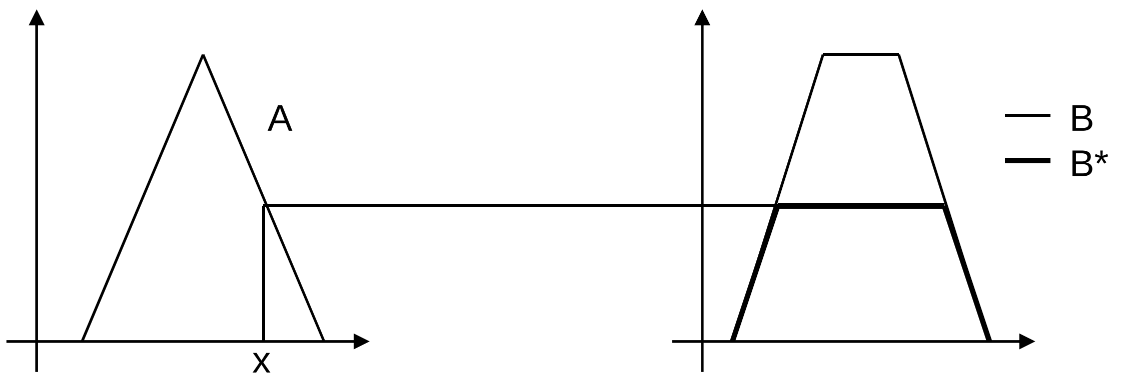
- Implicación en control difuso (cont.)
  - La implicación se convierte en una regla
    IF A THEN B ELSE NADA
  - La función de implicación definida de esta forma es igual a la función de conjunción
    - $f_{\rightarrow}(a,b) = f_{\wedge}(a,b)$
  - Este tipo de implicación se conoce como implicación de Mamdani ya que fue el primer autor que la propuso en un trabajo de control difuso
  - Para representar la conjunción puede utilizarse cualquier norma aunque lo habitual es emplear la norma estándar del mínimo
    - $f_{\rightarrow}(a,b) = f_{\wedge}(a,b) = \min(a,b)$

Control Difuso

- Ejemplo de implicación difusa
  - Modus ponens generalizado

    IF x es A THEN y es B
    x es A*
    ----
    y es B*

  - Representación gráfica del operador de implicación de Mamdani



Control Difuso

- Reglas difusas de la base de conocimientos

Rule: 1

IF	x is A3
OR	y is B1
THEN	z is C1

Rule: 1

IF	project_funding is adequate
OR	project_staffing is small
THEN	risk is low

Rule: 2

IF	x is A2
AND	y is B2
THEN	z is C2

Rule: 2

IF	project_funding is marginal
AND	project_staffing is large
THEN	risk is normal

Rule: 3

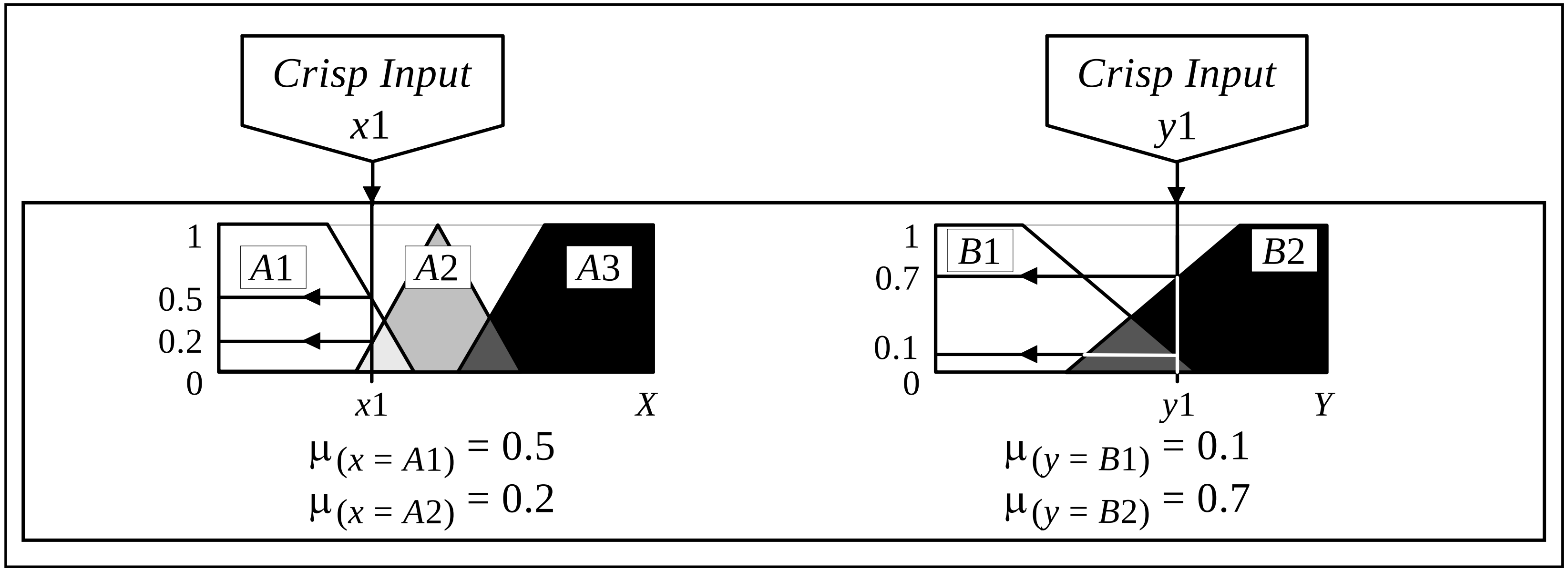
IF	x is A1
THEN	z is C3

Rule: 3

IF	project_funding is inadequate
THEN	risk is high

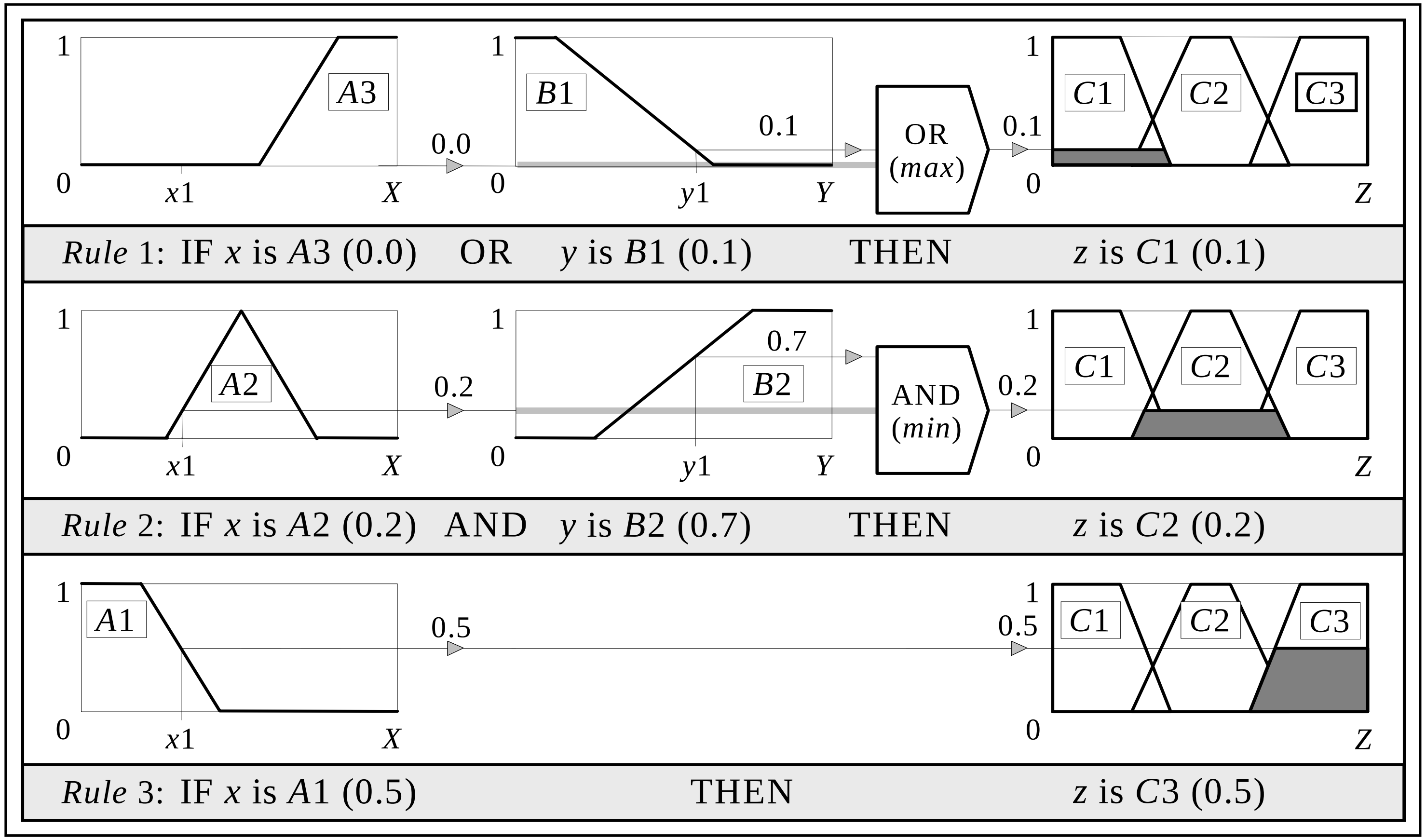
Control Difuso

- Codificación
  - Variable x1 en el referencial {A1, A2, A3}
  - Variable y1 en el referencial {B1. B2}



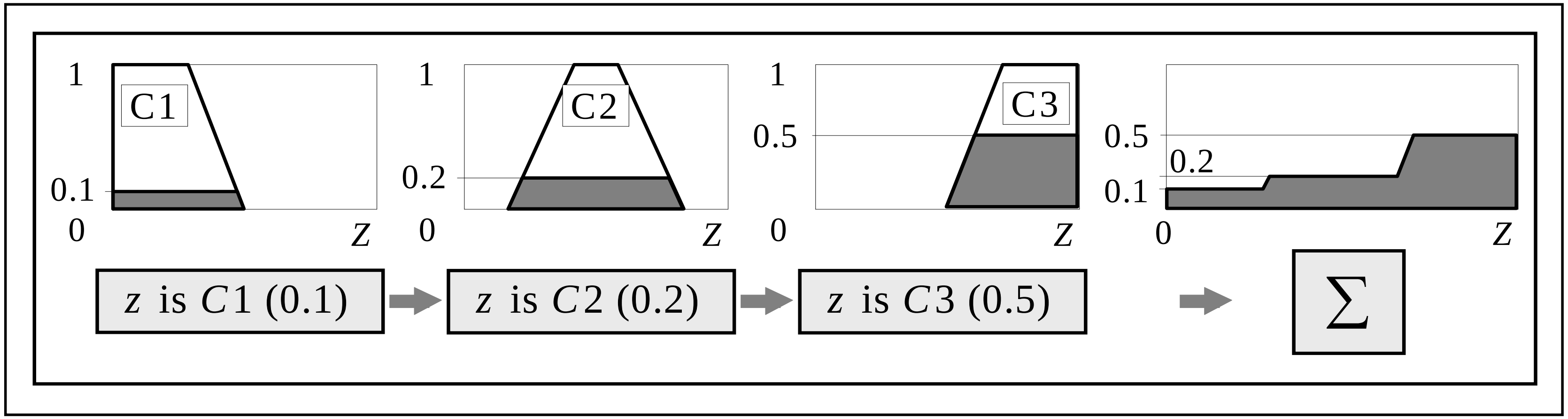
Control Difuso

- Razonamiento (evaluación de reglas)



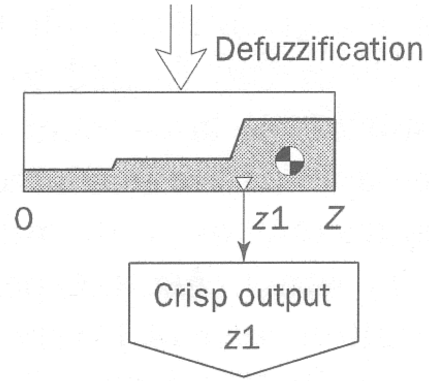
Control Difuso

- Razonamiento
  (Agregación de los consecuentes de las reglas)



Control Difuso

- Decodificación
  - Variable Z en el referencial {C1, C2, C3} se convierte a un valor numérico
  - Principal método de decodificación: Centro de Gravedad (COG) o método del centroide
    - Encuentra el punto en el que una línea vertical cortaría el conjunto difuso resultante en dos, cada parte con la misma "masa"



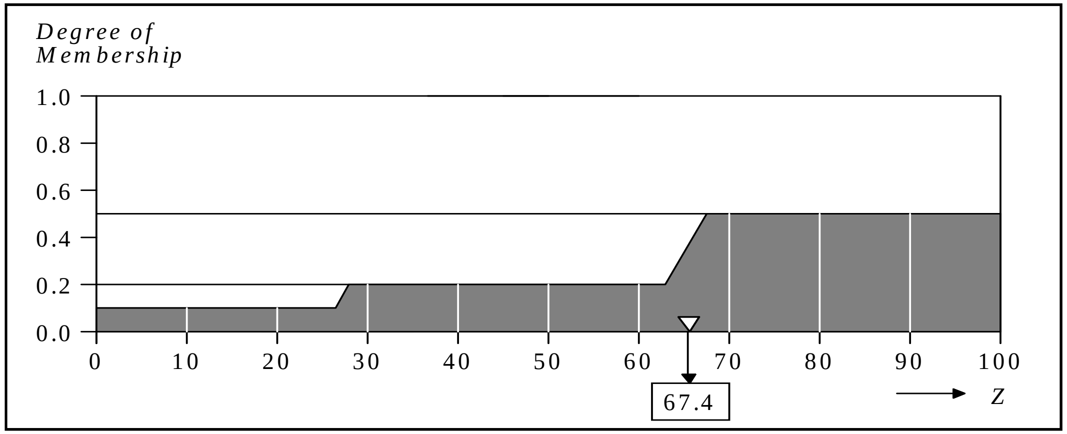
$$COG = \frac{\int_a^b \mu_A(x)\, x\, dx}{\int_a^b \mu_A(x)\, dx}$$

Control Difuso

- Decodificación (cont.)
  - El método COG implica la realización de una integral.
  - Puede obtenerse una aproximación razonable muestreando la función de pertenencia y hallando el centro de gravedad según esas muestras (no es necesaria la realización de una integral)
  - Ejemplo:

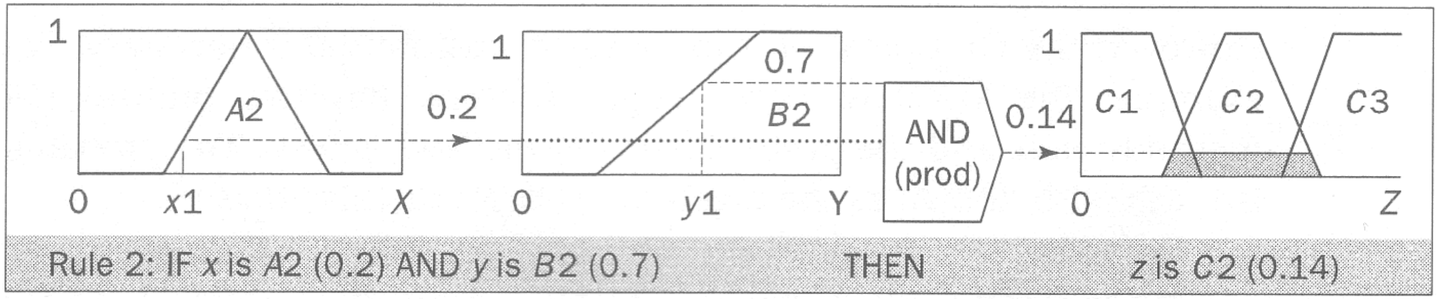
$$COG = \frac{(0+10+20)\times 0.1 + (30+40+50+60)\times 0.2 + (70+80+90+100)\times 0.5}{0.1+0.1+0.1+0.2+0.2+0.2+0.2+0.5+0.5+0.5+0.5} = 67.4$$

$$\mathrm{COG} = \frac{\sum_{x=a}^{b} \mu_A(x)\,x}{\sum_{x=a}^{b} \mu_A(x)}$$



Control Difuso

- Alternativas al razonamiento
  - Pueden utilizarse otras normas para la conjunción y otras conormas para la disyunción
  - Ejemplo: Utilización de la norma producto para representar el AND difuso



The AND *product* fuzzy operation

Control Difuso

- Alternativas al razonamiento (cont.)

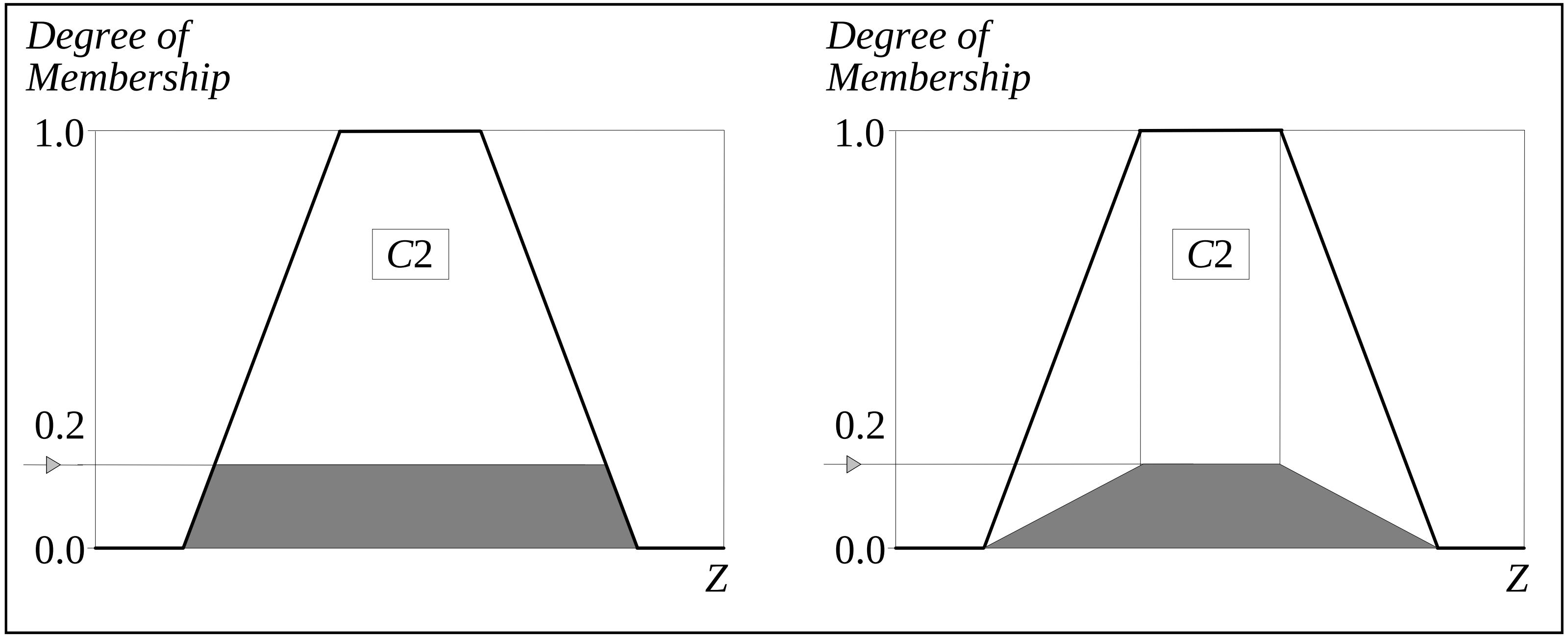
  - La utilización de la norma mínimo o producto en la definición de la implicación también da lugar a variaciones ($f_{\rightarrow} = f_{\wedge}$).

  - Corte (implicación de Mamdani)
    - La evaluación de la implicación se hace mediante la aplicación de la función mínimo entre el antecedente de la regla y la función de pertenencia del resultado
    - La operación es poco compleja de realizar y el resultado es fácil de defuzzificar
    - Sin embargo al hacer un corte se pierde la verdadera forma del conjunto difuso

  - Escalamiento (implicación de Larsen)
    - La evaluación de la implicación se hace mediante la aplicación de la función producto entre el antecedente de la regla y la función de pertenencia del resultado
    - La operación es más compleja de realizar que el mínimo
    - Pero el resultado pierde menos información ya que lo que hace función de pertenencia del resultado para que tenga la misma fo l pero sin que sobrepase el valor del antecedente

Control Difuso

- Alternativas al razonamiento (cont.)
  - Representación gráfica del Corte y Escalamiento



EJEMPLO CONJUNTOS DIFUSOS

CASO PRÁCTICO

Baloncesto

Entre el pueblo de Torquemada de la Inquisición y el de Carnaval de la Juerga hay una gran rivalidad en cuestiones de baloncesto, hasta el punto de que el alcalde de la villa de Carnaval ha destinado una cantidad importante de dinero (hasta 1.000.000 euros) para reforzar al equipo. El entrenador ha sido instruido para seleccionar al mejor candidato, y –de paso- debe tratar de ahorrar dinero. Así, se pagará más al mejor aspirante, siempre y cuando reúna unos requisitos mínimos. El entrenador llegó a la conclusión de que lo que necesita el equipo es un hombre alto, que tenga –además- buenos porcentajes en la canasta. Lo que no sabe es cómo ahorrar dinero. Por otra parte, no tiene claro qué es un hombre alto con buenos porcentajes. Da la casualidad de que su hijo está estudiando inteligencia artificial, y le sugiere el empleo de los modelos difusos para resolver su problema. Después de estudiar mucho se hace el siguiente planteamiento:

Baloncesto

- Porcentaje (sobre 20 tiros)

Muy_malo	→	(1/0 – 1/3 – 0/7)
Malo	→	(0/3 – 1/5 – 1/7 – 0/9)
Regular	→	(0/7 – 1/10 – 0/13)
Bueno	→	(0/11 – 1/13 – 1/15 – 0/17)
Muy_bueno	→	(0/15 – 1/17 – 1/20)

Baloncesto

- Estatura (en centímetros)

  - Muy_bajo → (1/150 – 1/170 – 0/180)
  - Bajo → (0/170 – 1/175 – 1/180 – 0/185)
  - Normal → (0/180 – 1/190 – 0/195)
  - Alto → (0/190 – 1/195 – 1/205 – 0/210)
  - Muy_alto → (0/200 – 1/210 – 1/250)

Baloncesto

- **Como tiene que pagar al candidato en función de su valía, y de los requisitos mínimos, establece una distribución de 0 a 100 (que luego multiplicada por 10.000 le da la cantidad en euros de la ficha), con el fin de clasificar a los aspirantes. Según este criterio la distribución es la siguiente:**

Baloncesto

- **Candidato**
  - **Malo** → **(1/0)**
  - **Normal** → **(1/0 – 1/25 – 0/40)**
  - **Bueno** → **(0/25 – 1/40 – 1/60 – 0/75)**
  - **Muy_bueno** → **(0/60 – 1/75 – 1/100)**

- **Evidentemente no fichará a candidatos “malos”… ¡¡¡Se trata de ganar al equipo de Torquemada de la Inquisición!!!**

Baloncesto

- A continuación define las siguientes reglas difusas:

  R1: SI: (1) <ESTATURA = NORMAL> Y <PORCENTAJE = BUENO>
  O:(2) <ESTATURA = NORMAL> Y <PORCENTAJE = MUY_BUENO>
  O:(3) <ESTATURA = ALTO> Y <PORCENTAJE = REGULAR>
  O:(4) <ESTATURA = MUY_ALTO> Y <PORCENTAJE = REGULAR>

- EL CANDIDATO ES NORMAL

Baloncesto

- R2:SI:(1) <ESTATURA = ALTO> Y <PORCENTAJE = BUENO>
  O:(2) <ESTATURA = MUY_ALTO> Y <PORCENTAJE = BUENO>

- EL CANDIDATO ES BUENO

Baloncesto

- R3: SI:(1) <ESTATURA = ALTO> Y <PORCENTAJE = MUY_BUENO>
- O:(2) <ESTATURA = MUY_ALTO> Y <PORCENTAJE = BUENO>

- EL CANDIDATO ES MUY_BUENO

Baloncesto

- R4: SI: (1) <PORCENTAJE = MUY_MALO>
- O: (2) <PORCENTAJE = MALO>

- EL CANDIDATO ES MALO

Baloncesto

- R5: SI: (1) <ESTATURA = MUY_BAJO>
- O: (2) <ESTATURA = BAJO>
- EL CANDIDATO ES MALO

Baloncesto

- R6: SI: (1) <ESTATURA = NORMAL>
- Y (2) <PORCENTAJE = REGULAR>
- → EL CANDIDATO ES MALO

Baloncesto

- Una vez que ha comprobado que su modelo es correcto, el entrenador hace pública una preselección, a la que se presentan los siguientes candidatos:

Elías	198 cm	16/20
Blas	193 cm	17/20
Luis	188 cm	17/20
Juan	203 cm	15/20
Raúl	176 cm	18/20
Cholo	186 cm	18/20

- ¿A quién tratará de fichar el entrenador? ¿Cuánto dinero ofrecerá para contratarle?

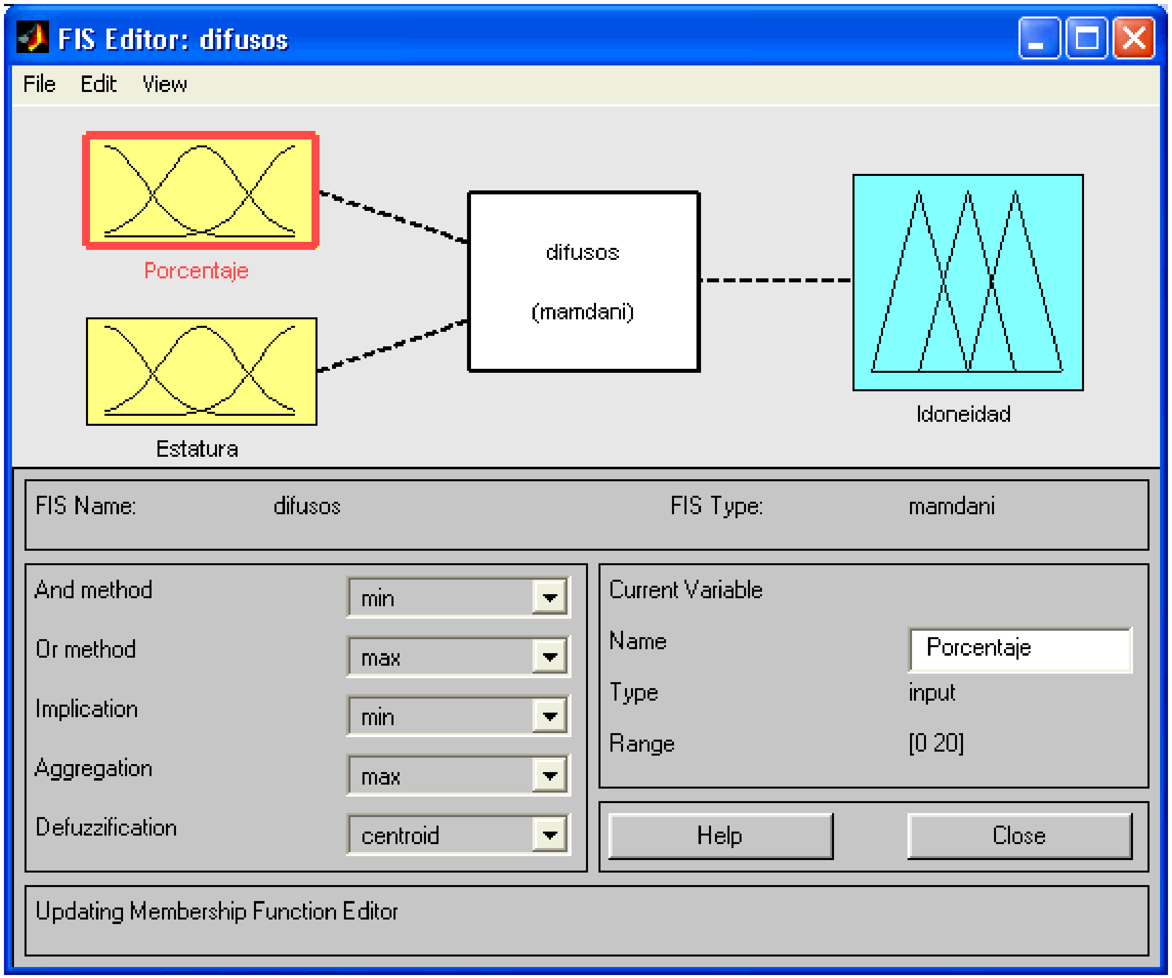
Ejemplo de Conjuntos Difusos utilizando Matlab

Inteligencia Artificial

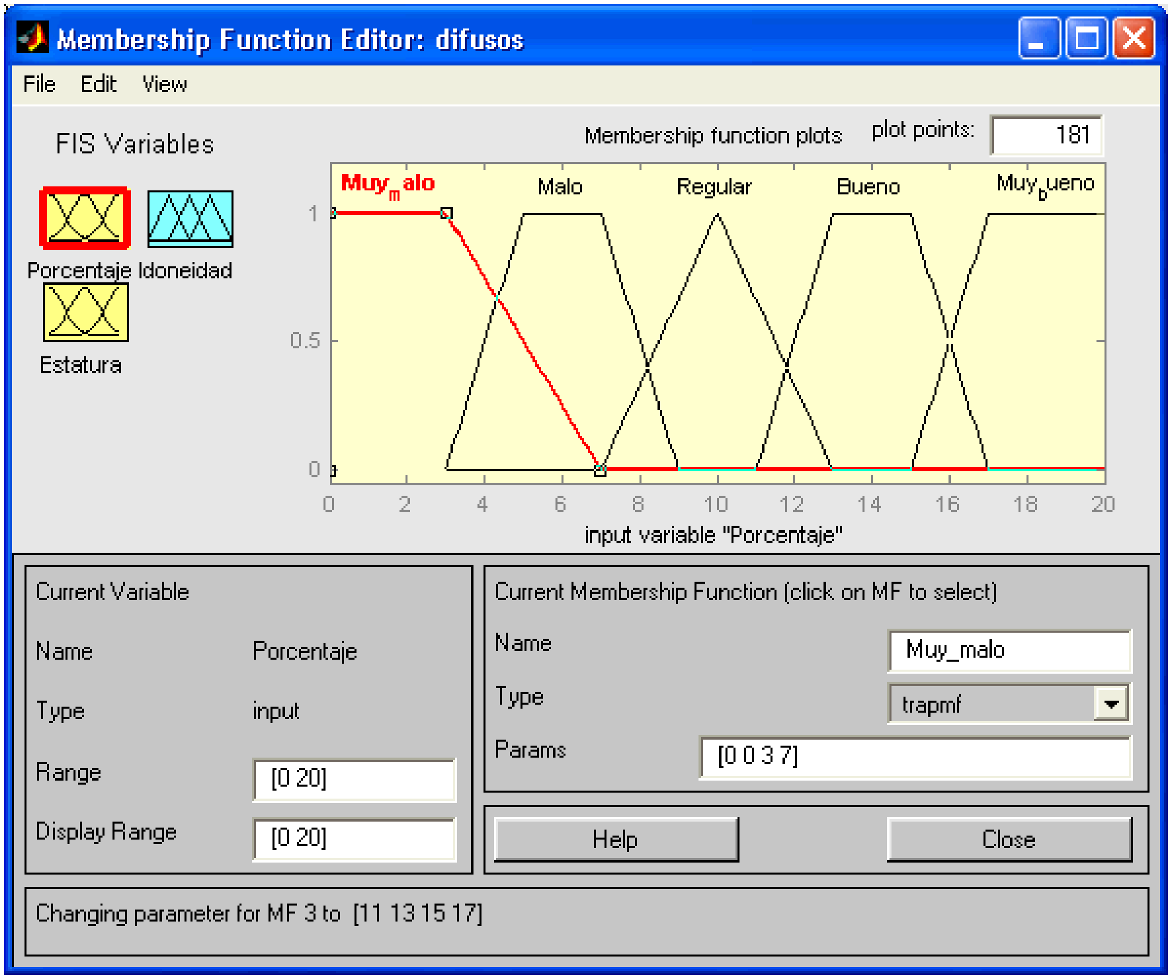
Facultad de Informática

UNIVERSIDADE DA CORUÑA

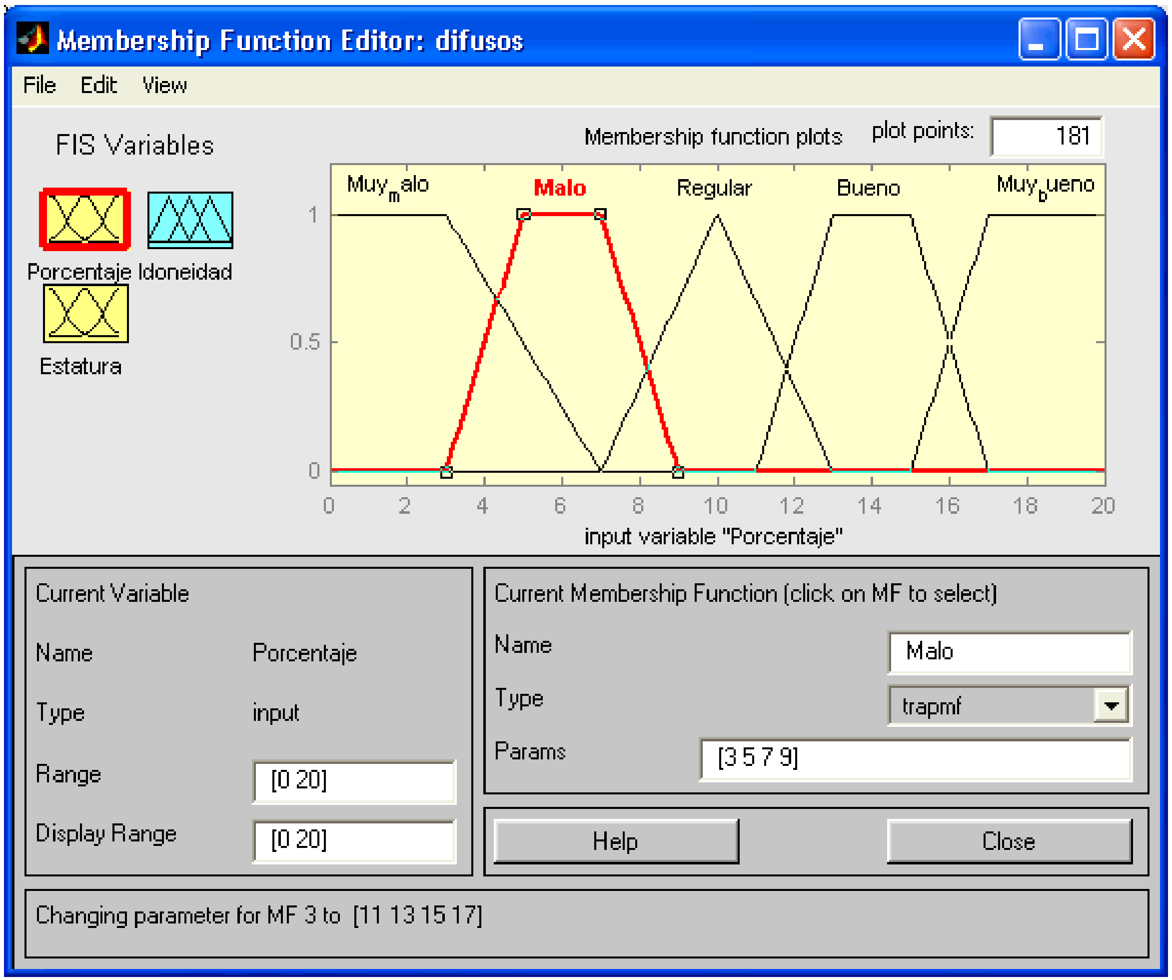
Sistema difuso



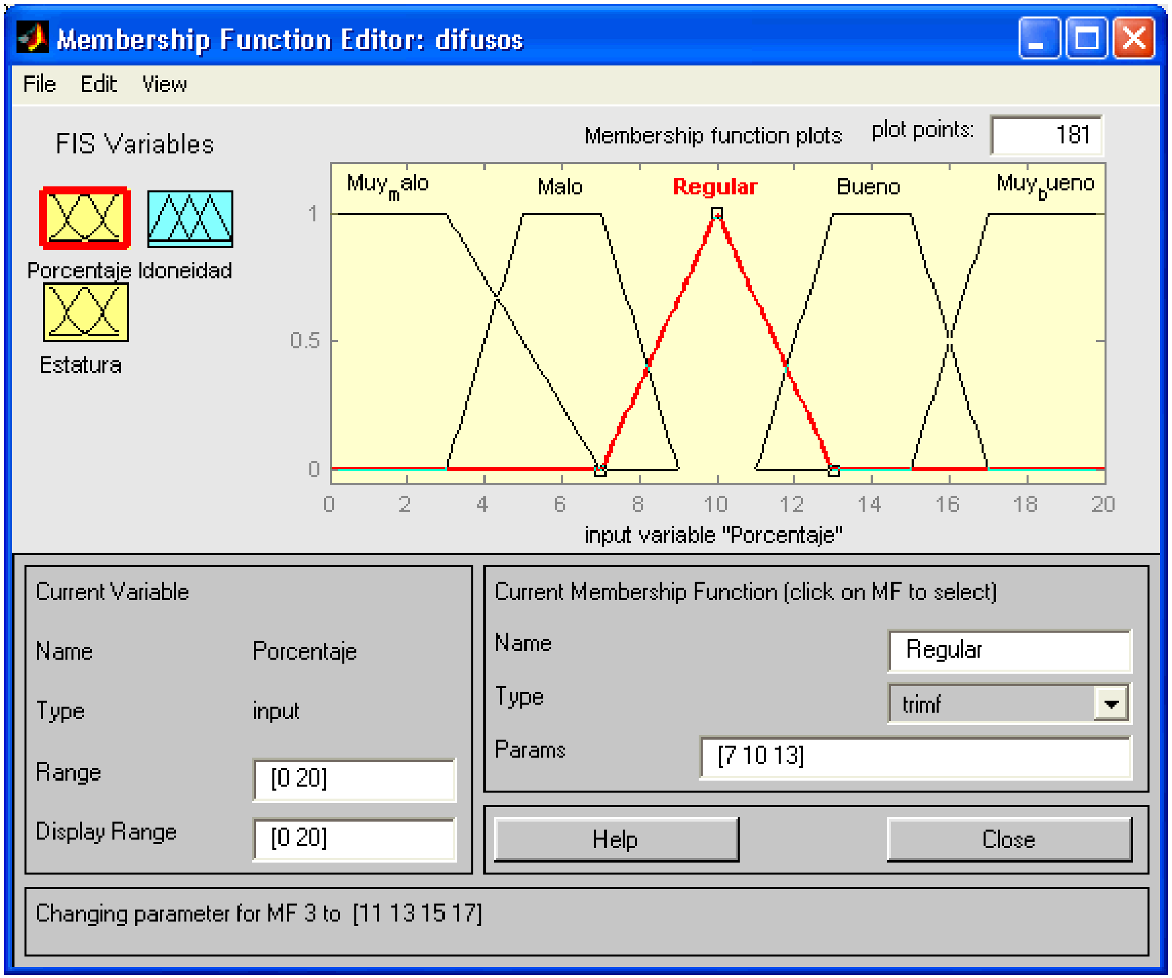
Variable “Porcentaje”



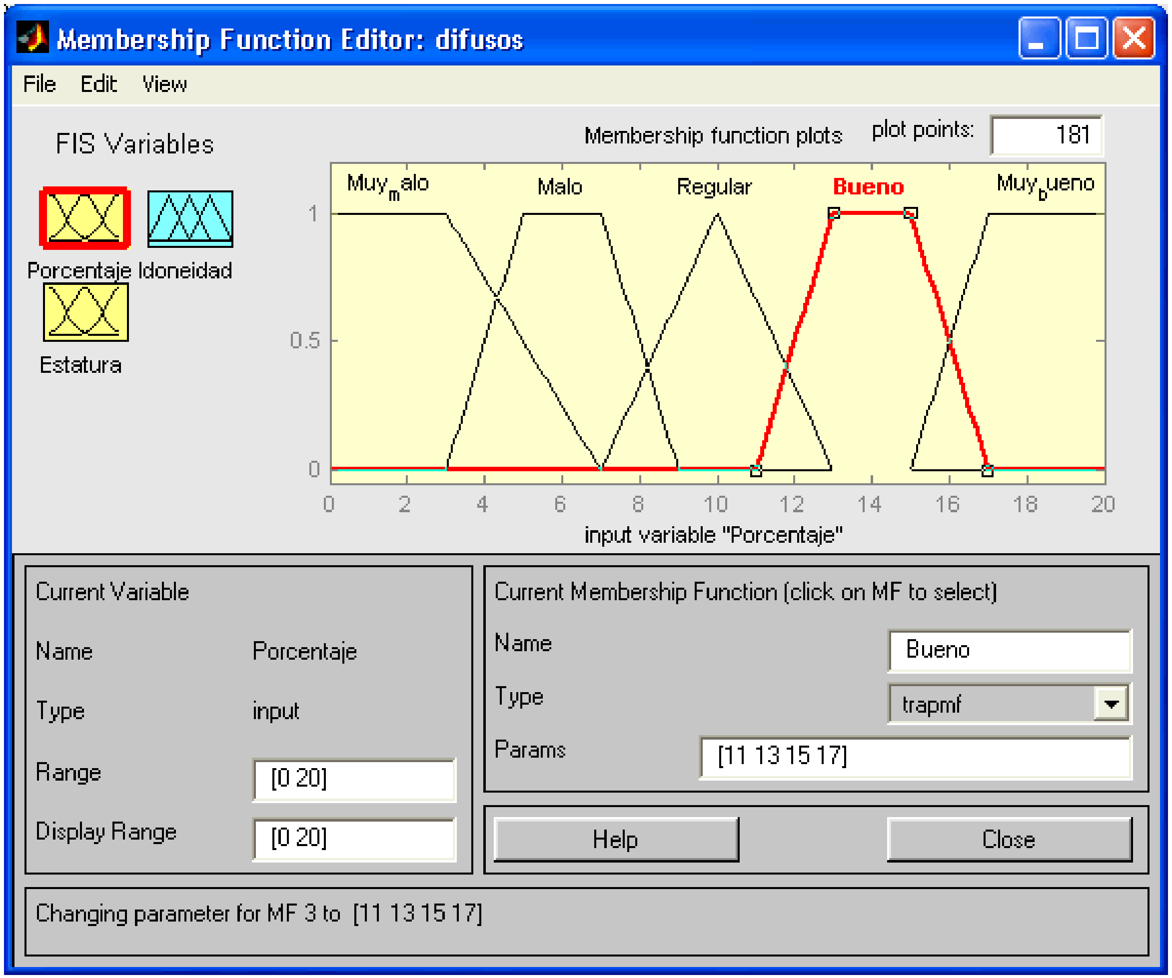
Variable “Porcentaje”



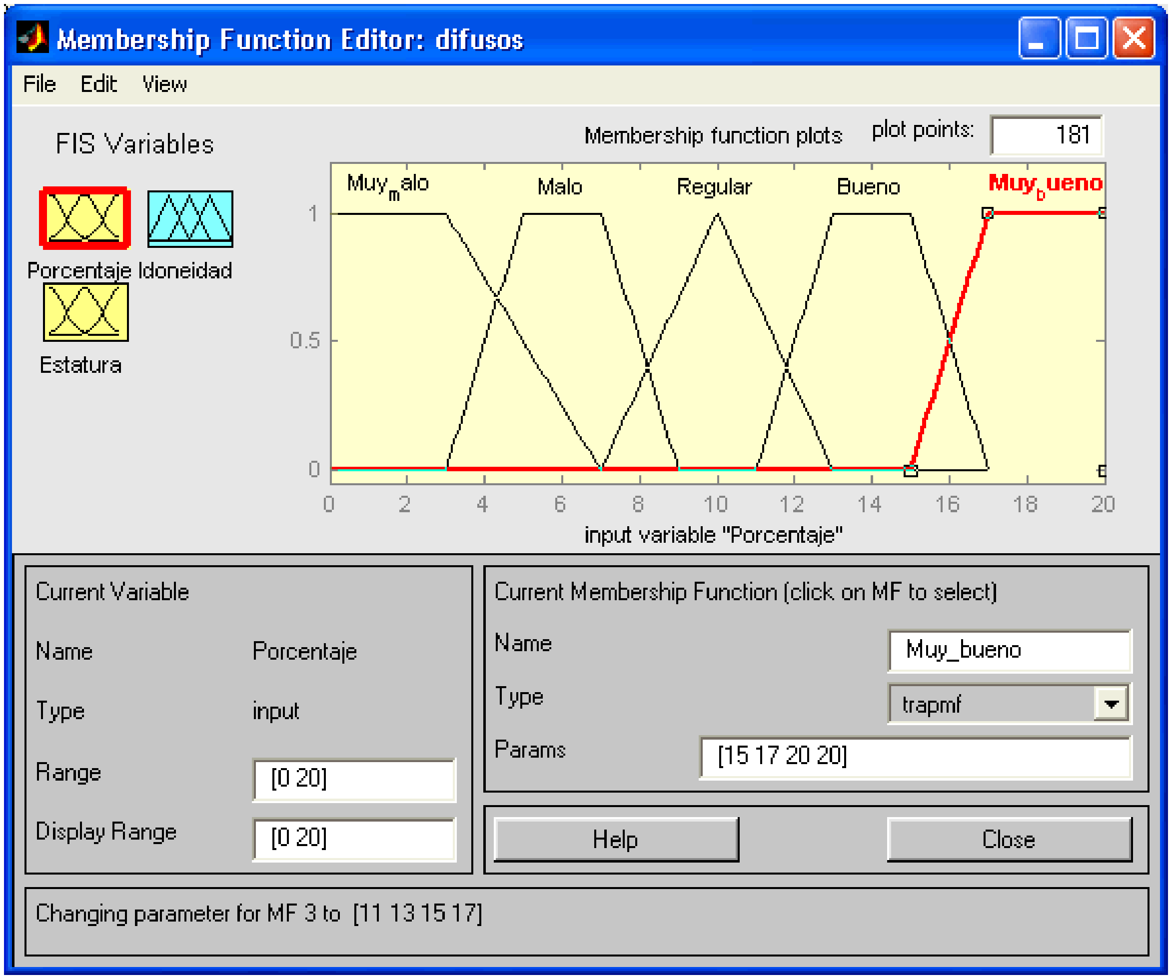
Variable “Porcentaje”



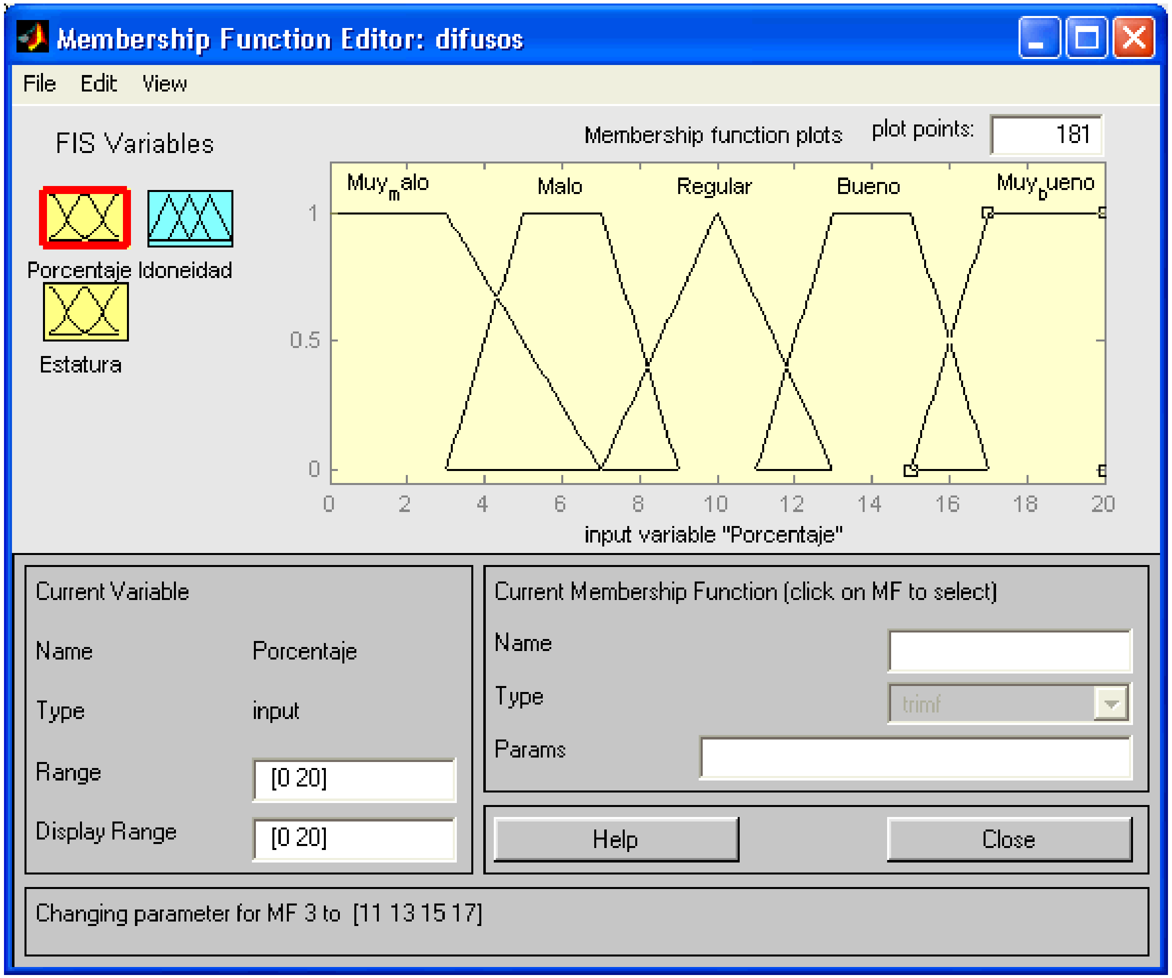
Variable “Porcentaje”



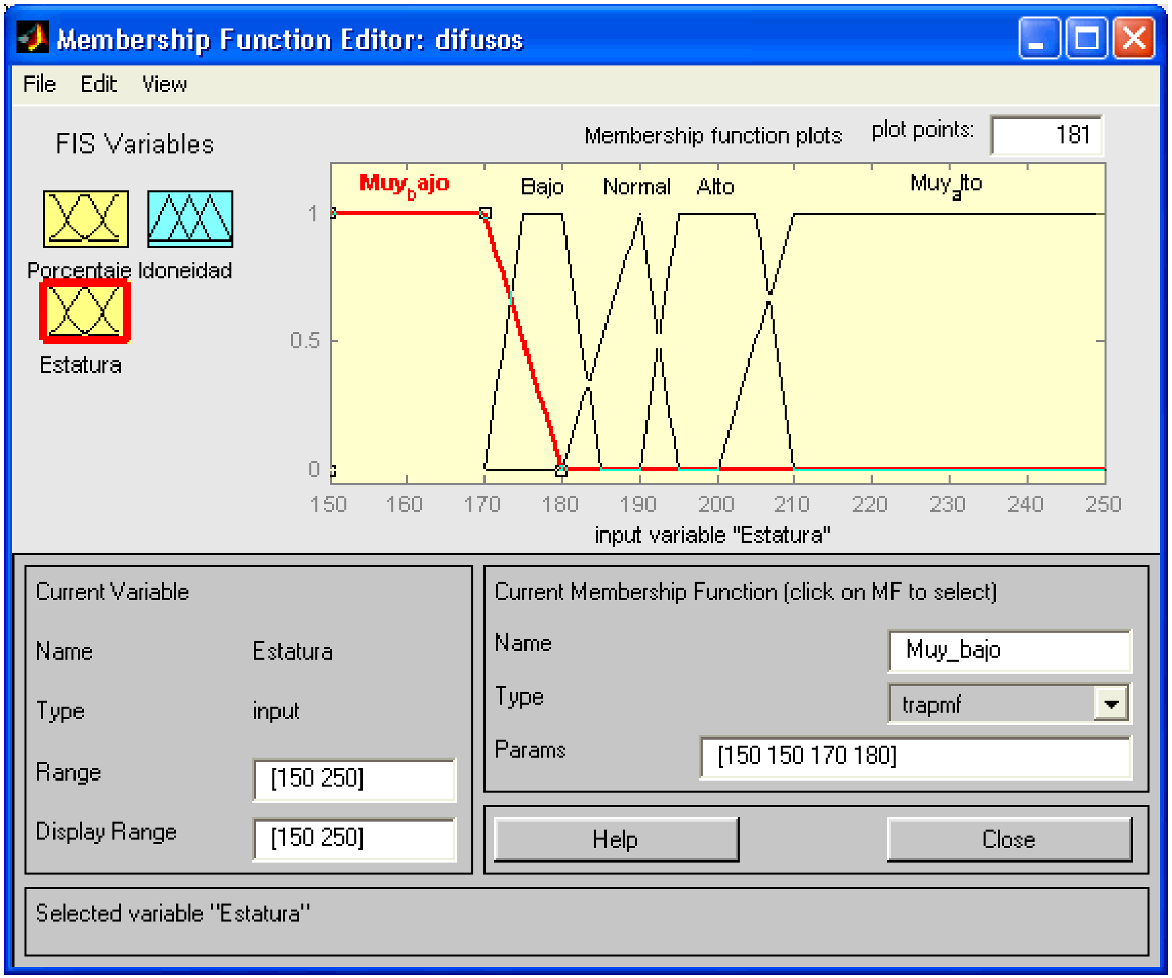
Variable “Porcentaje”



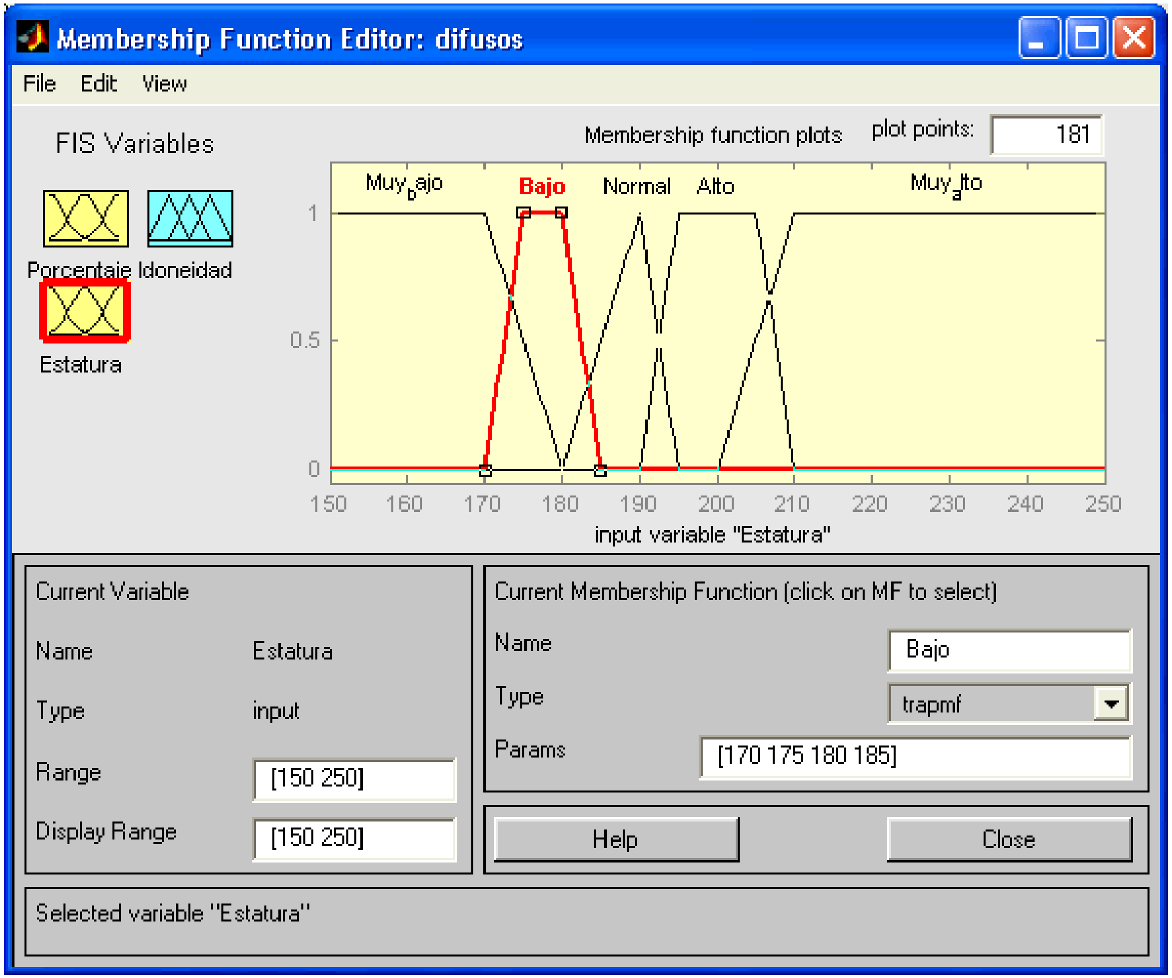
Variable "Porcentaje"



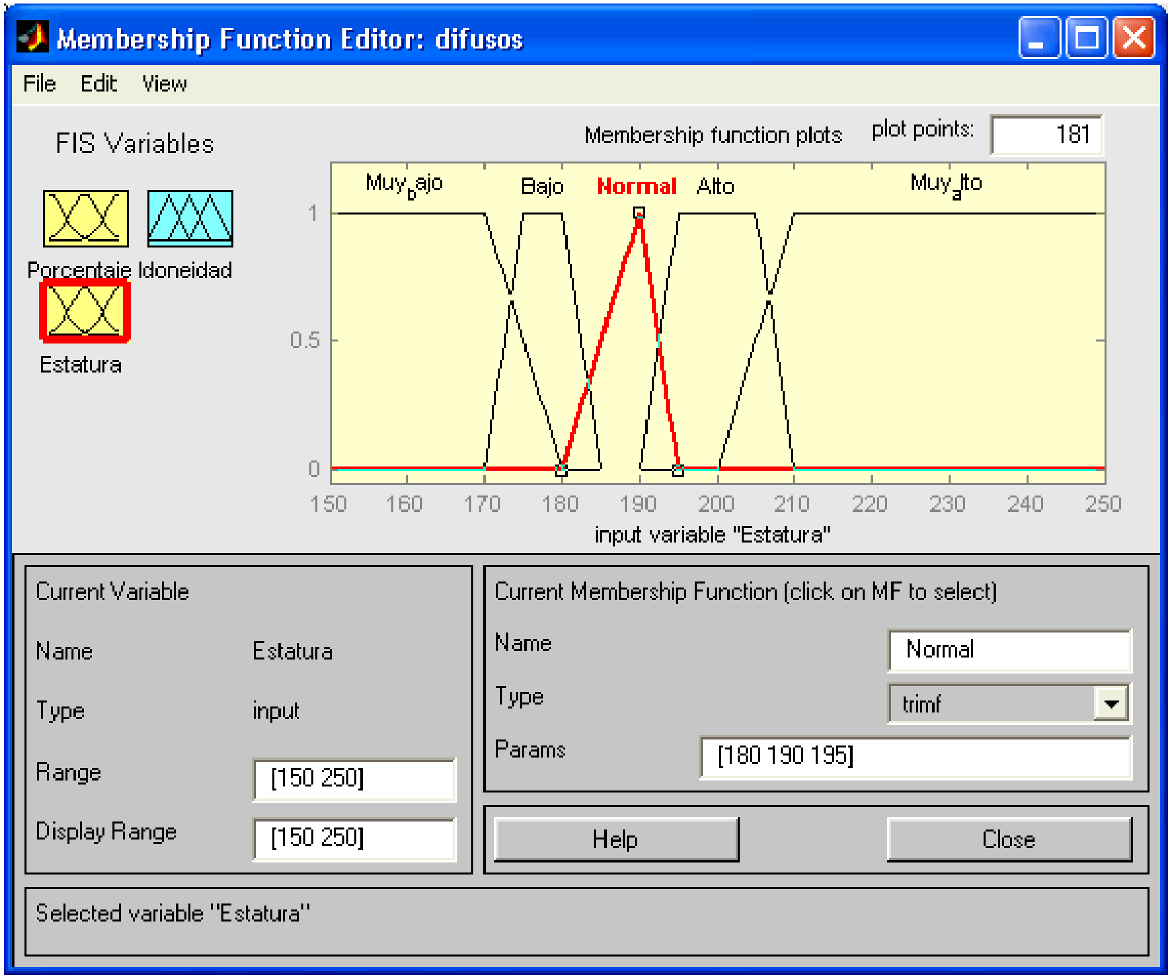
Variable “Estatura”



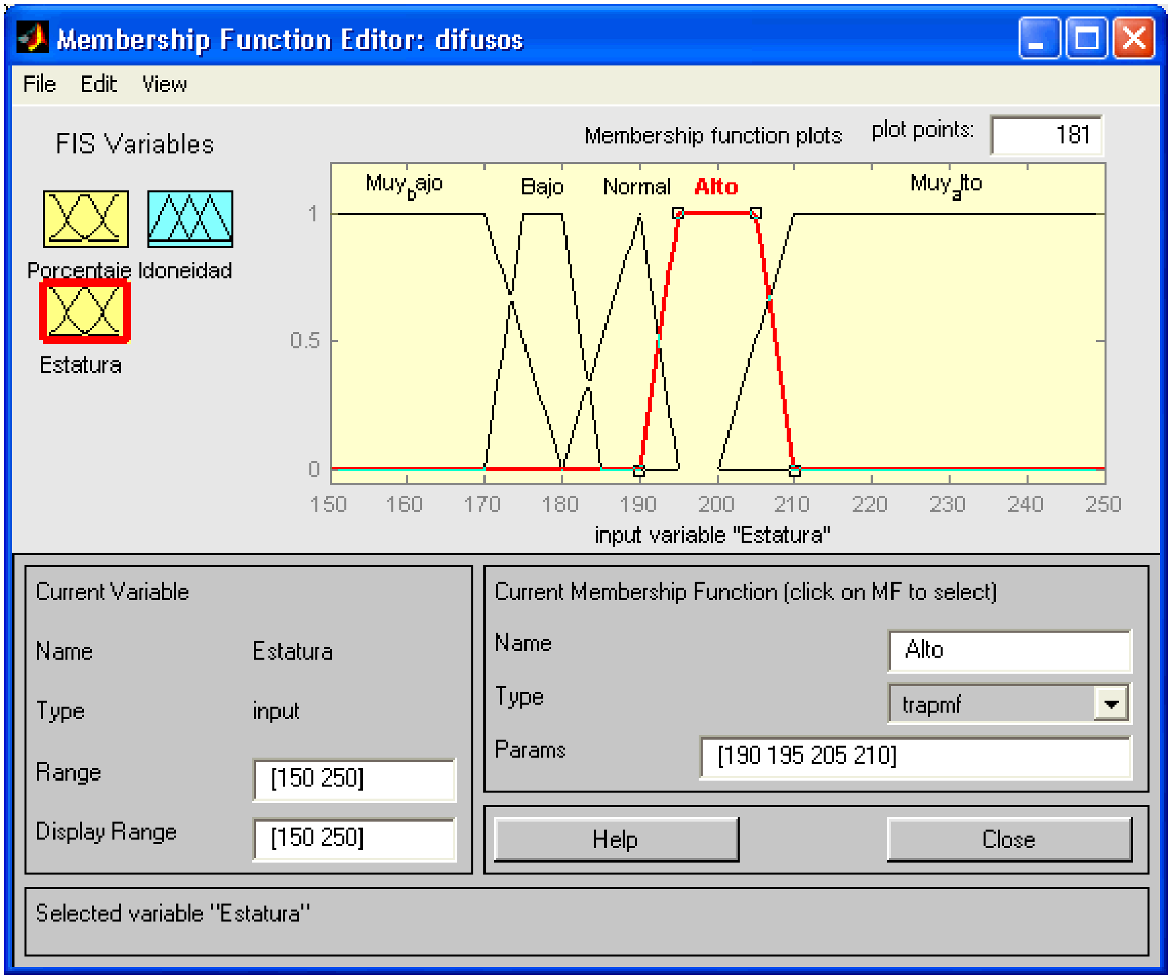
Variable "Estatura"



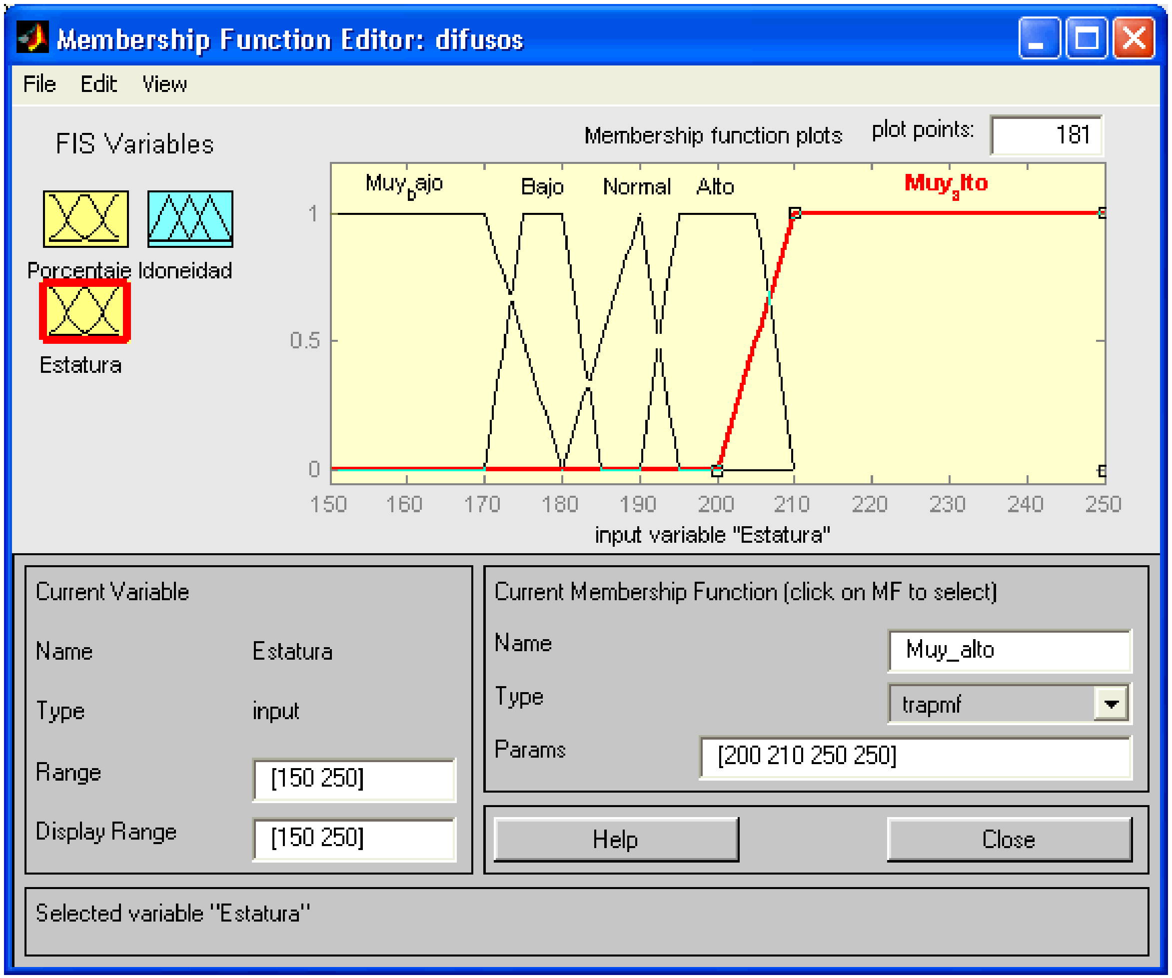
Variable "Estatura"



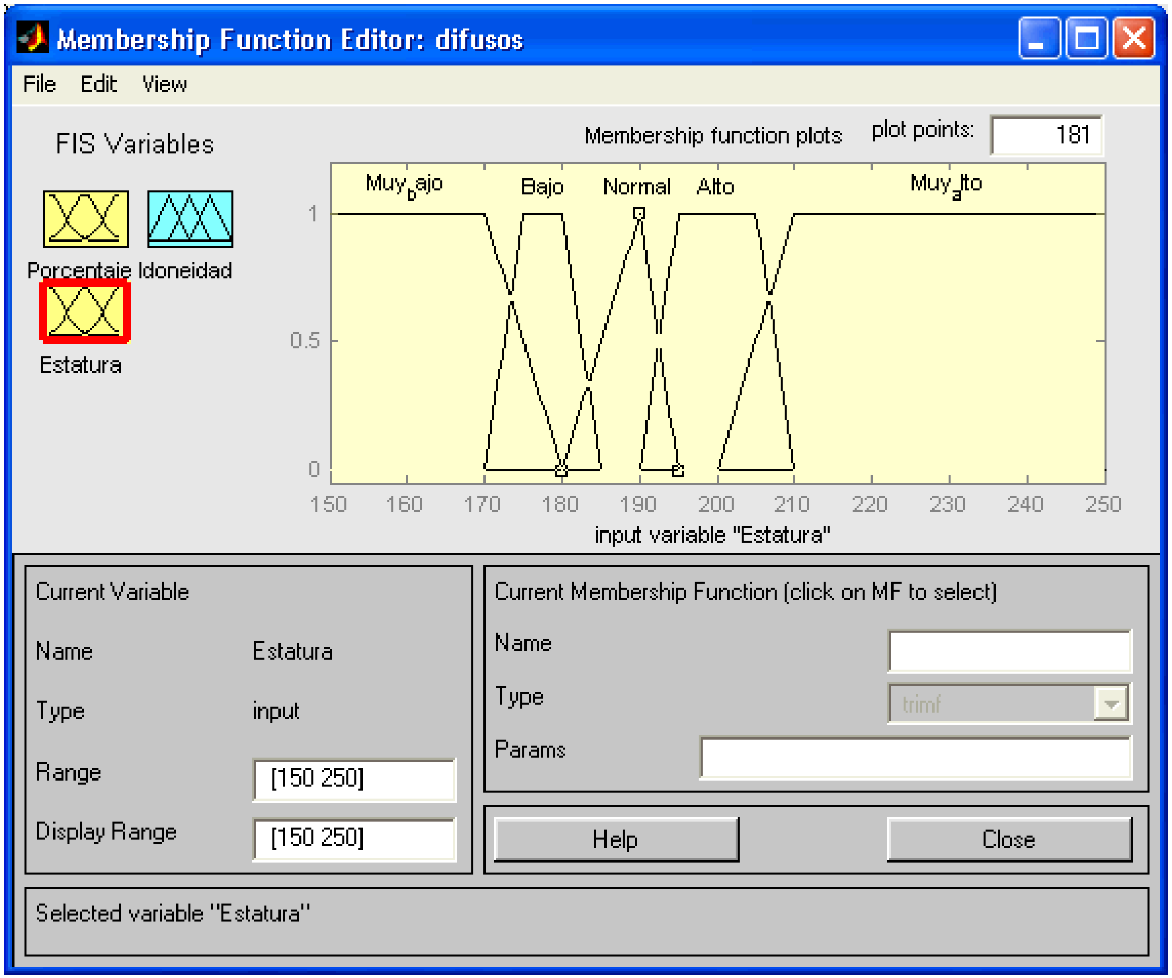
Variable "Estatura"



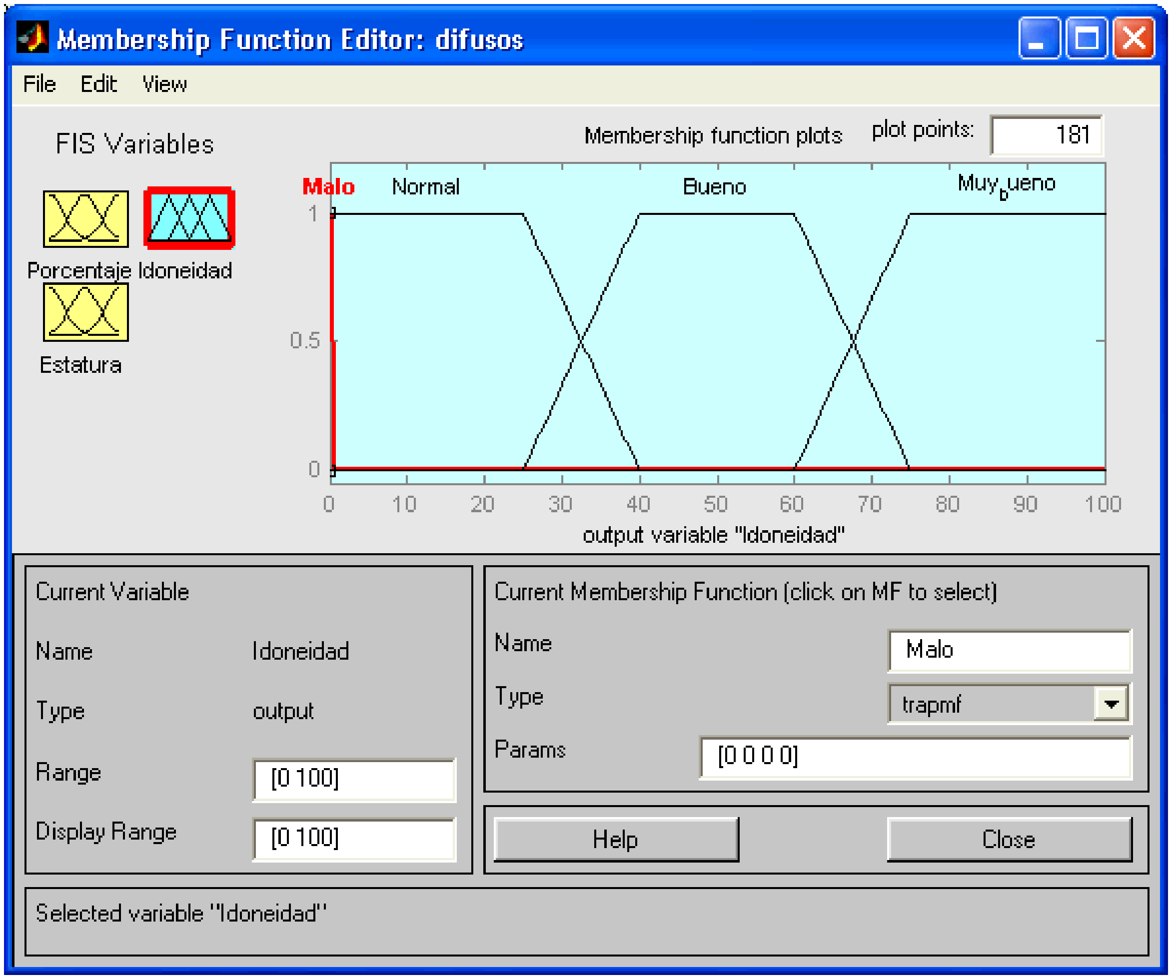
Variable "Estatura"



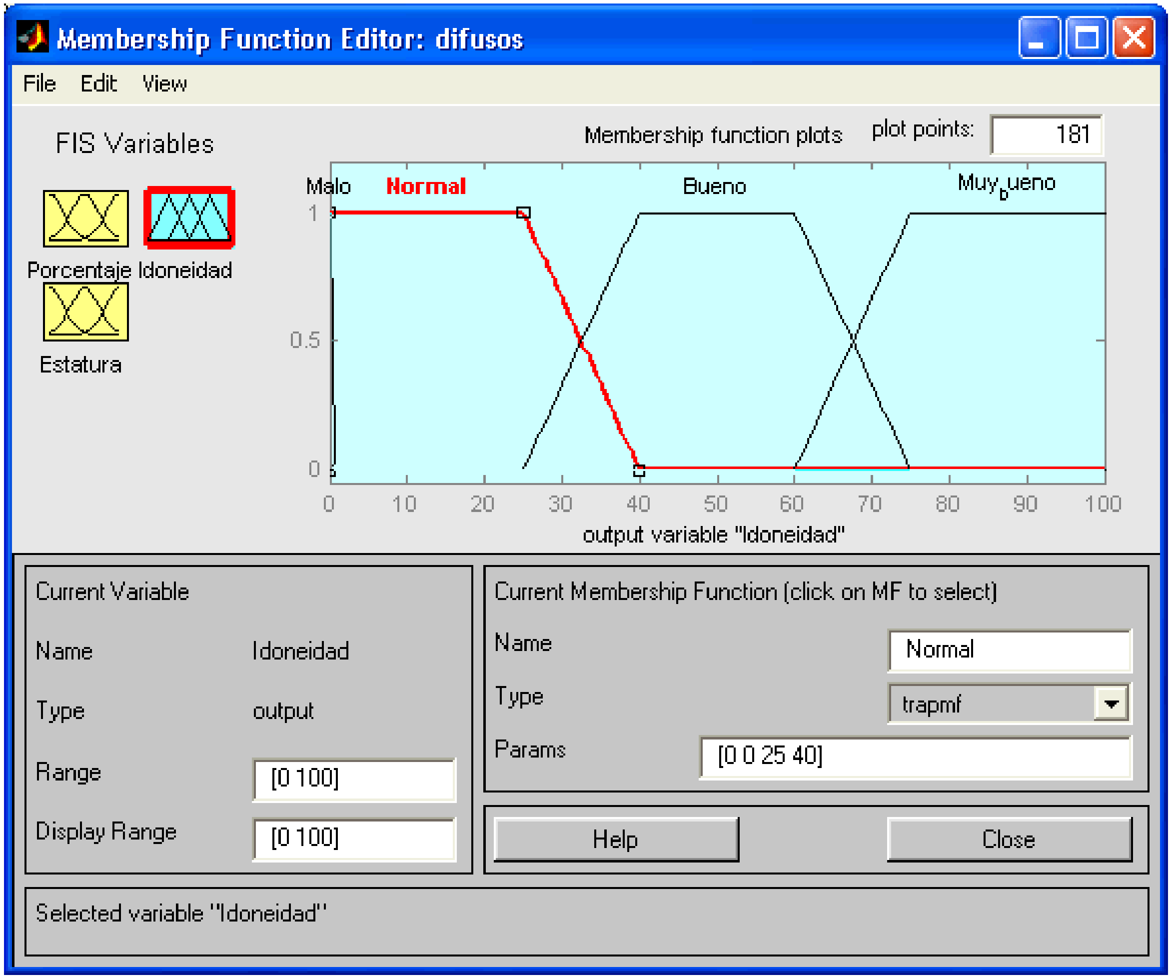
Variable "Estatura"



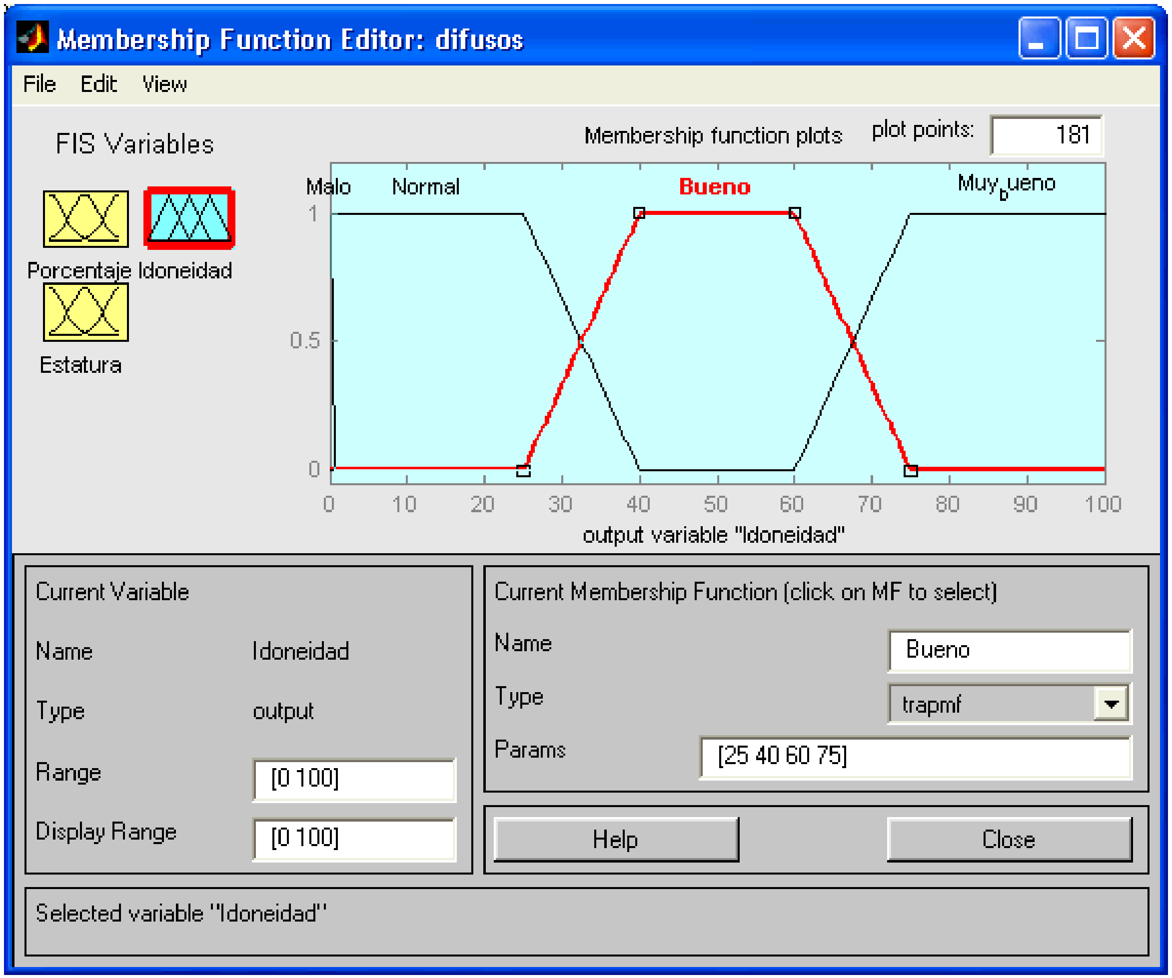
Variable “Idoneidad”



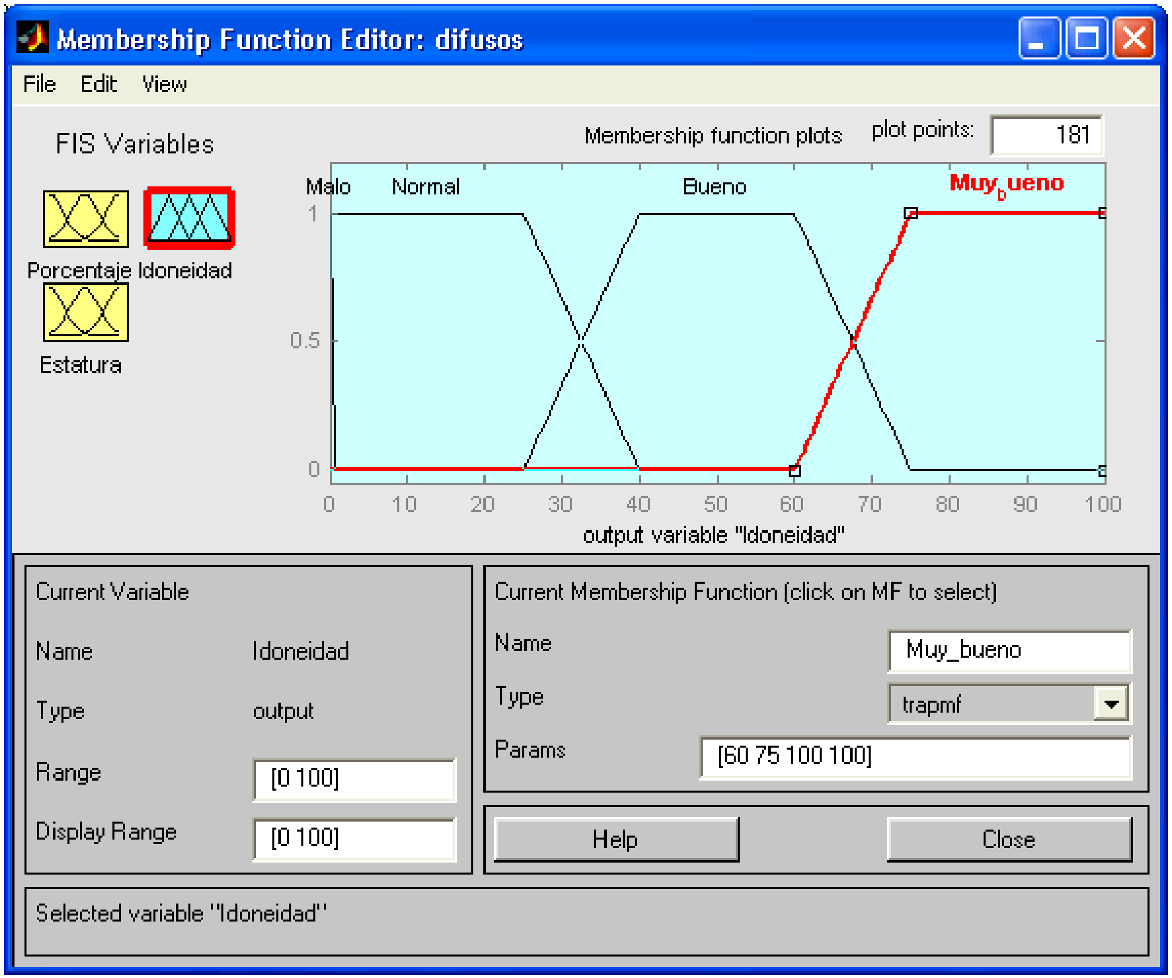
Variable "Idoneidad"



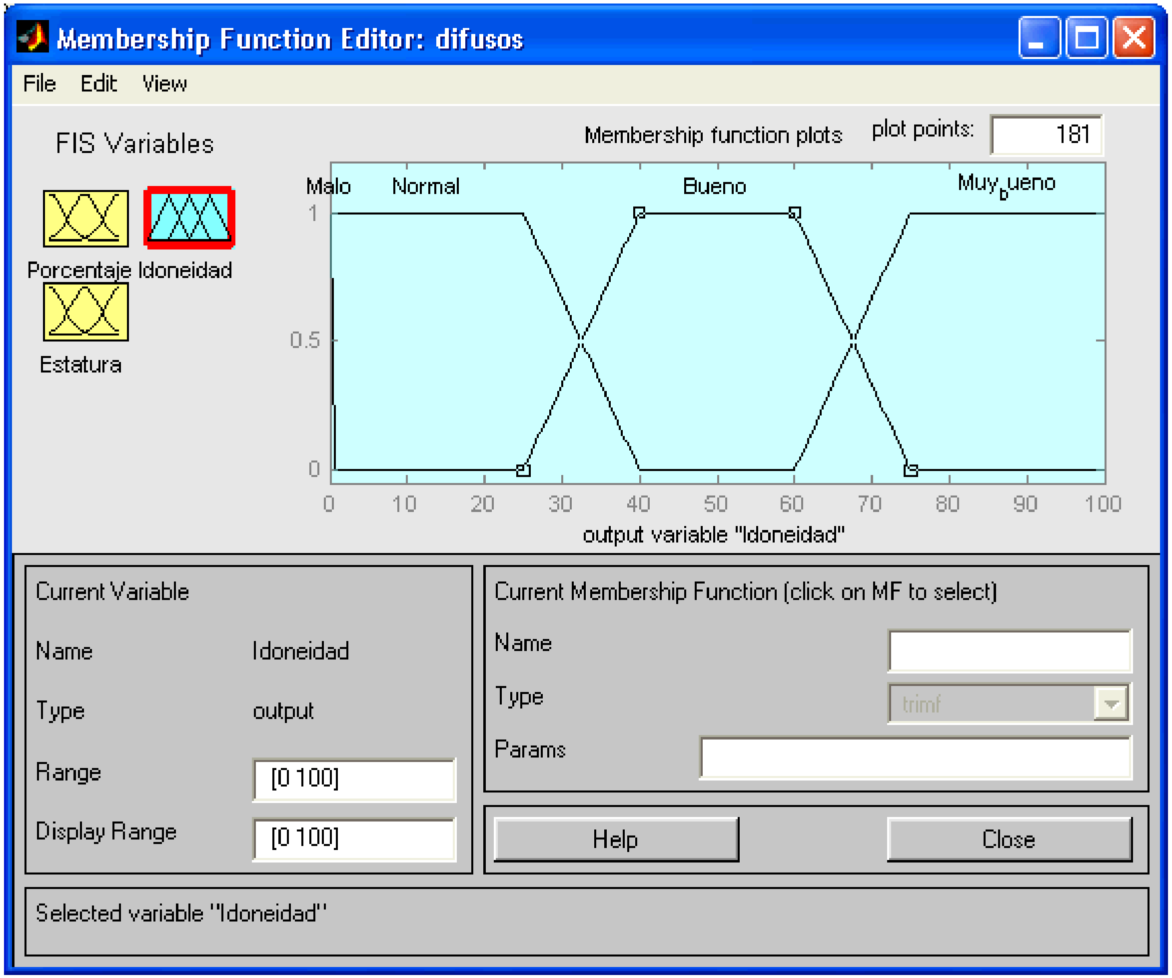
Variable “Idoneidad”



Variable "Idoneidad"



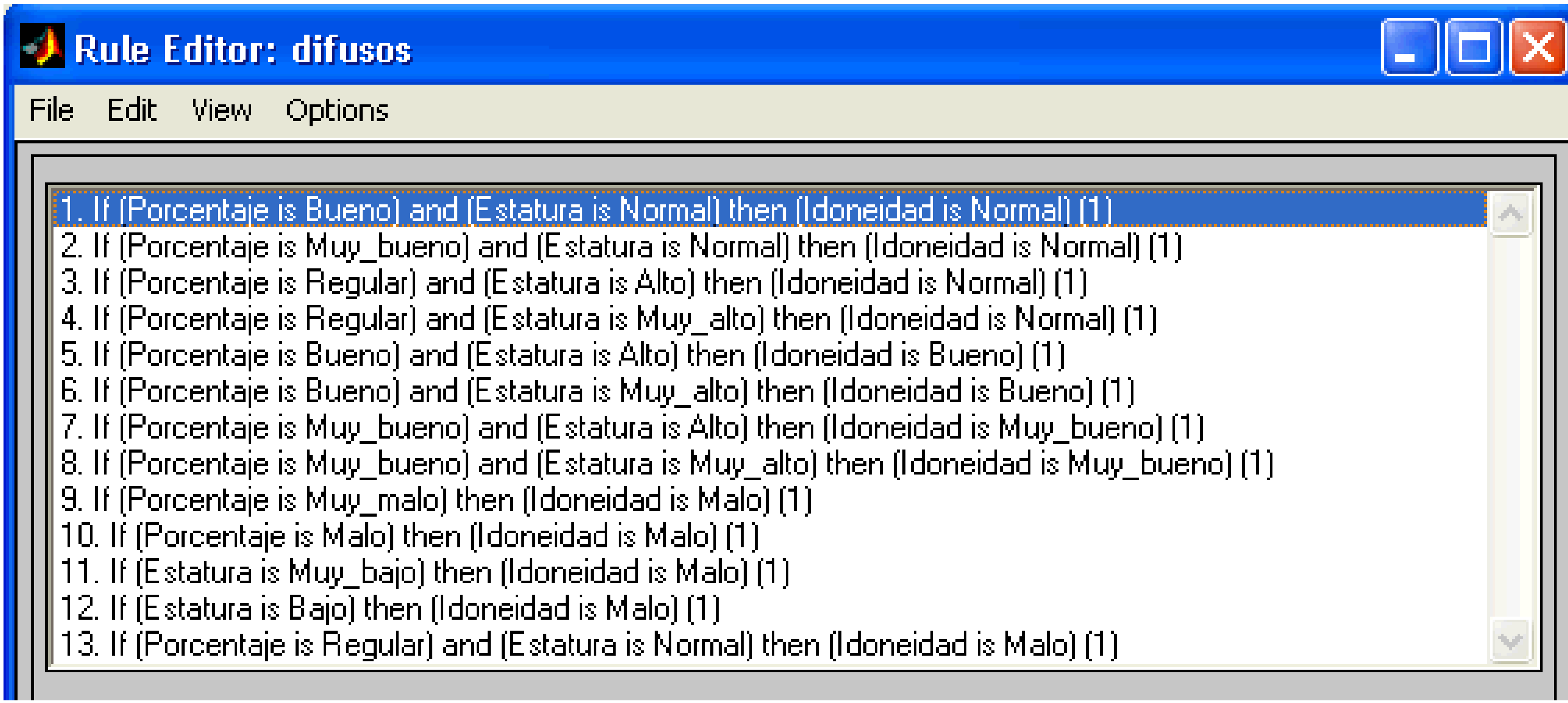
Variable "Idoneidad"



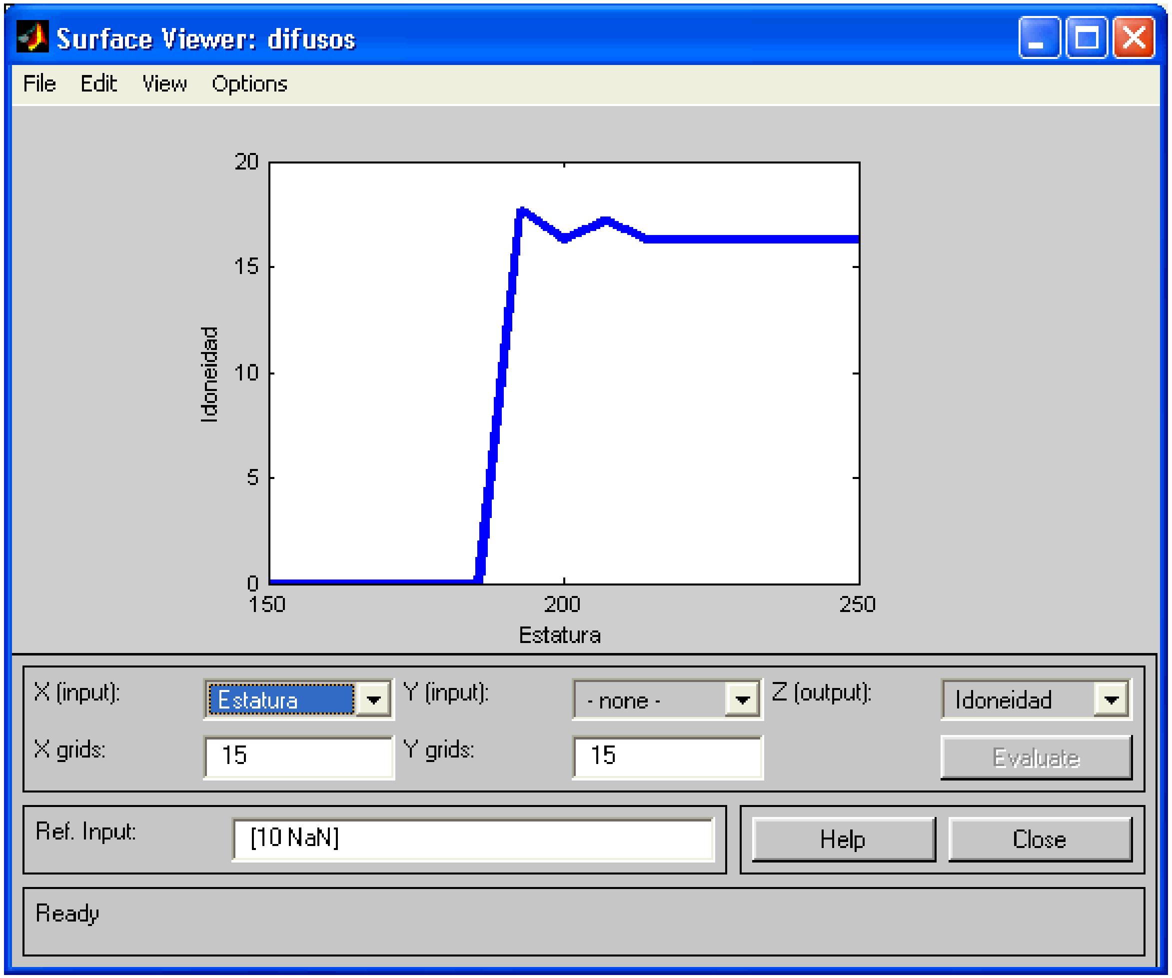
Reglas

% / Est.	Muy malo	Malo	Regular	Bueno	Muy bueno
Muy Bajo	Malo	Malo	Malo	Malo	Malo
Bajo	Malo	Malo	Malo	Malo	Malo
Normal	Malo	Malo	Malo	Normal	Normal
Alto	Malo	Malo	Normal	Bueno	Muy bueno
Muy Alto	Malo	Malo	Normal	Bueno	Muy bueno

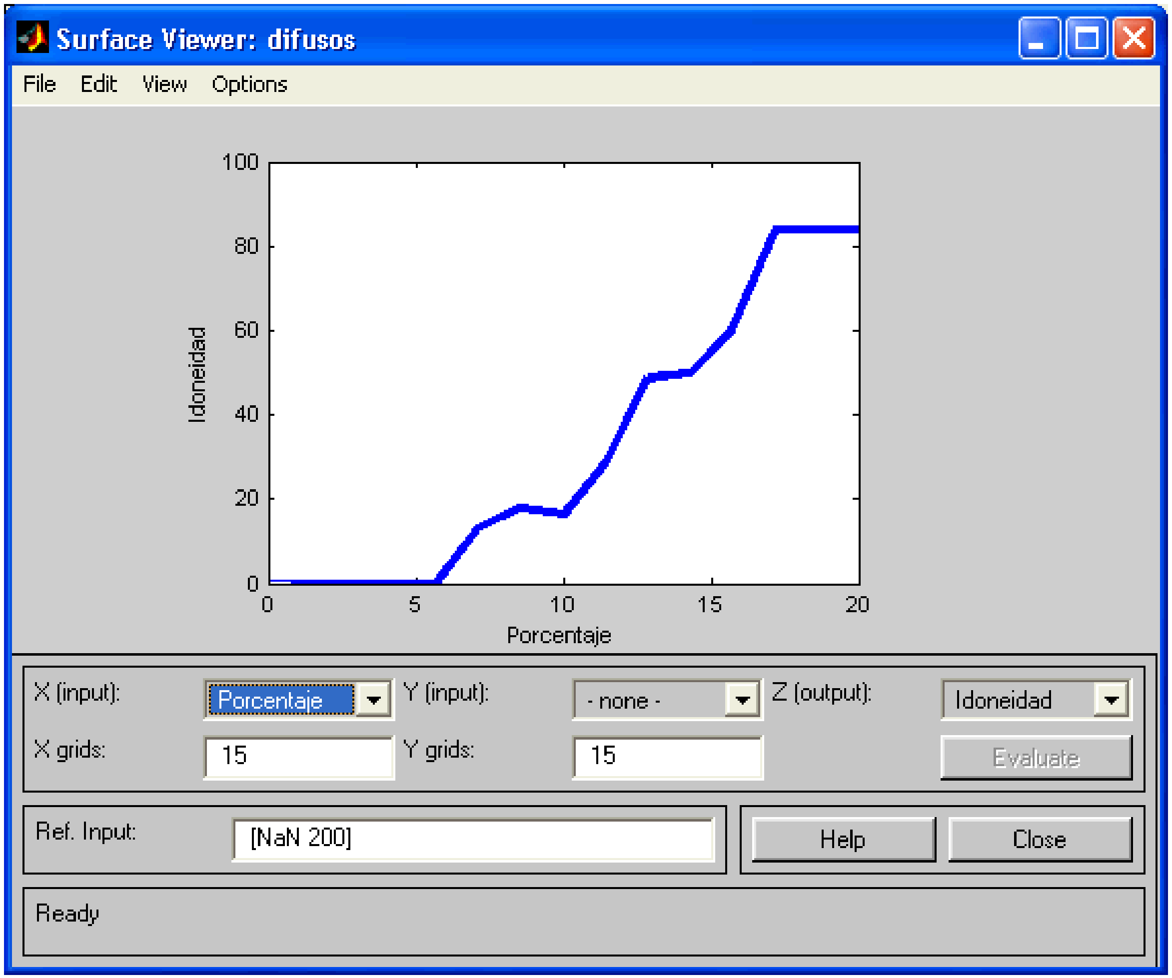
Reglas


Rule Editor: difusos
File Edit View Options
1. If (Porcentaje is Bueno) and (Estatura is Normal) then (Idoneidad is Normal) (1)
2. If (Porcentaje is Muy_bueno) and (Estatura is Normal) then (Idoneidad is Normal) (1)
3. If (Porcentaje is Regular) and (Estatura is Alto) then (Idoneidad is Normal) (1)
4. If (Porcentaje is Regular) and (Estatura is Muy_alto) then (Idoneidad is Normal) (1)
5. If (Porcentaje is Bueno) and (Estatura is Alto) then (Idoneidad is Bueno) (1)
6. If (Porcentaje is Bueno) and (Estatura is Muy_alto) then (Idoneidad is Bueno) (1)
7. If (Porcentaje is Muy_bueno) and (Estatura is Alto) then (Idoneidad is Muy_bueno) (1)
8. If (Porcentaje is Muy_bueno) and (Estatura is Muy_alto) then (Idoneidad is Muy_bueno) (1)
9. If (Porcentaje is Muy_malo) then (Idoneidad is Malo) (1)
10. If (Porcentaje is Malo) then (Idoneidad is Malo) (1)
11. If (Estatura is Muy_bajo) then (Idoneidad is Malo) (1)
12. If (Estatura is Bajo) then (Idoneidad is Malo) (1)
13. If (Porcentaje is Regular) and (Estatura is Normal) then (Idoneidad is Malo) (1)

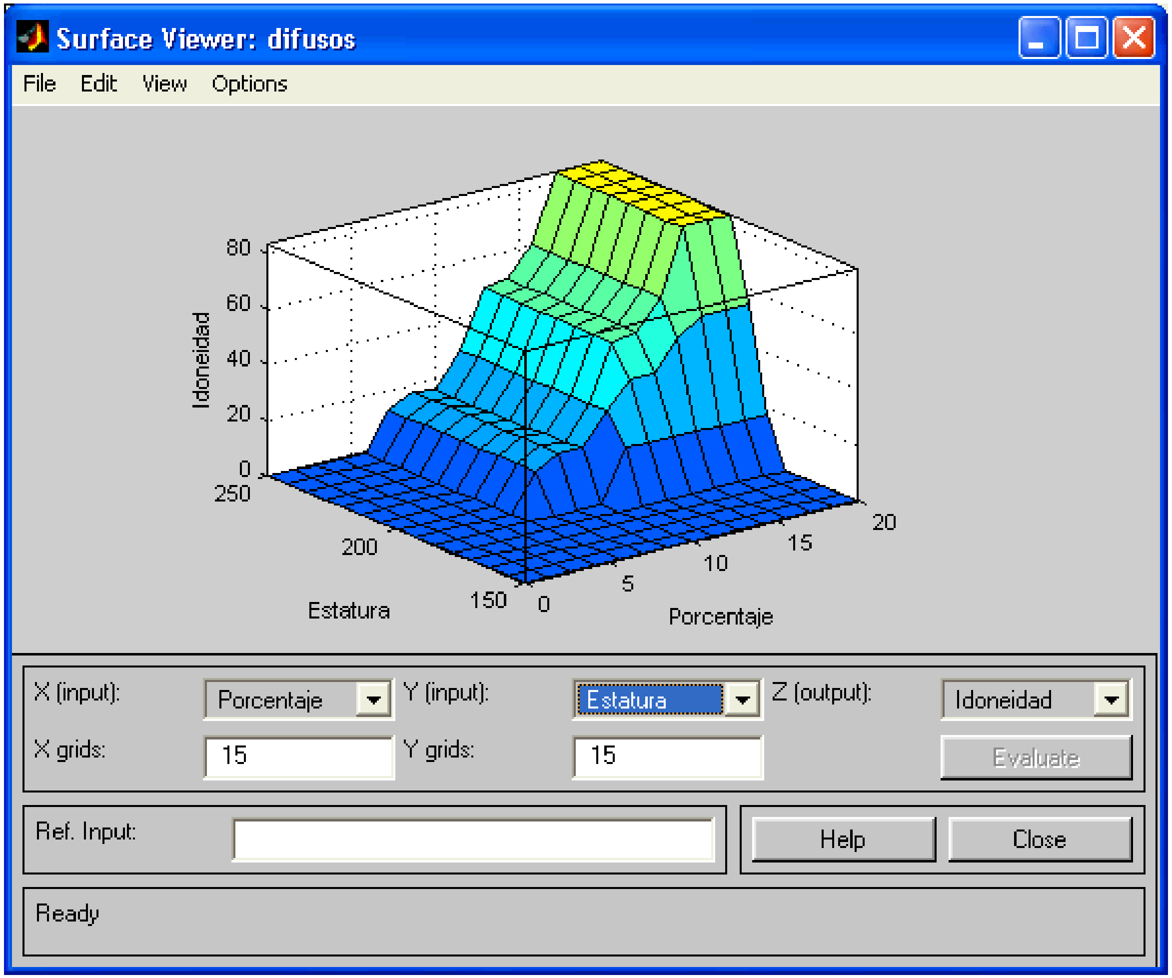
Reglas


Surface Viewer: difusos
File
Edit
View
Options
20
15
10
5
0
Idoneidad
150
200
250
Estatura
X (input):
Estatura
Y (input):
- none -
Z (output):
Idoneidad
X grids:
15
Y grids:
15
Evaluate
Ref. Input:
[10 NaN]
Help
Close
Ready

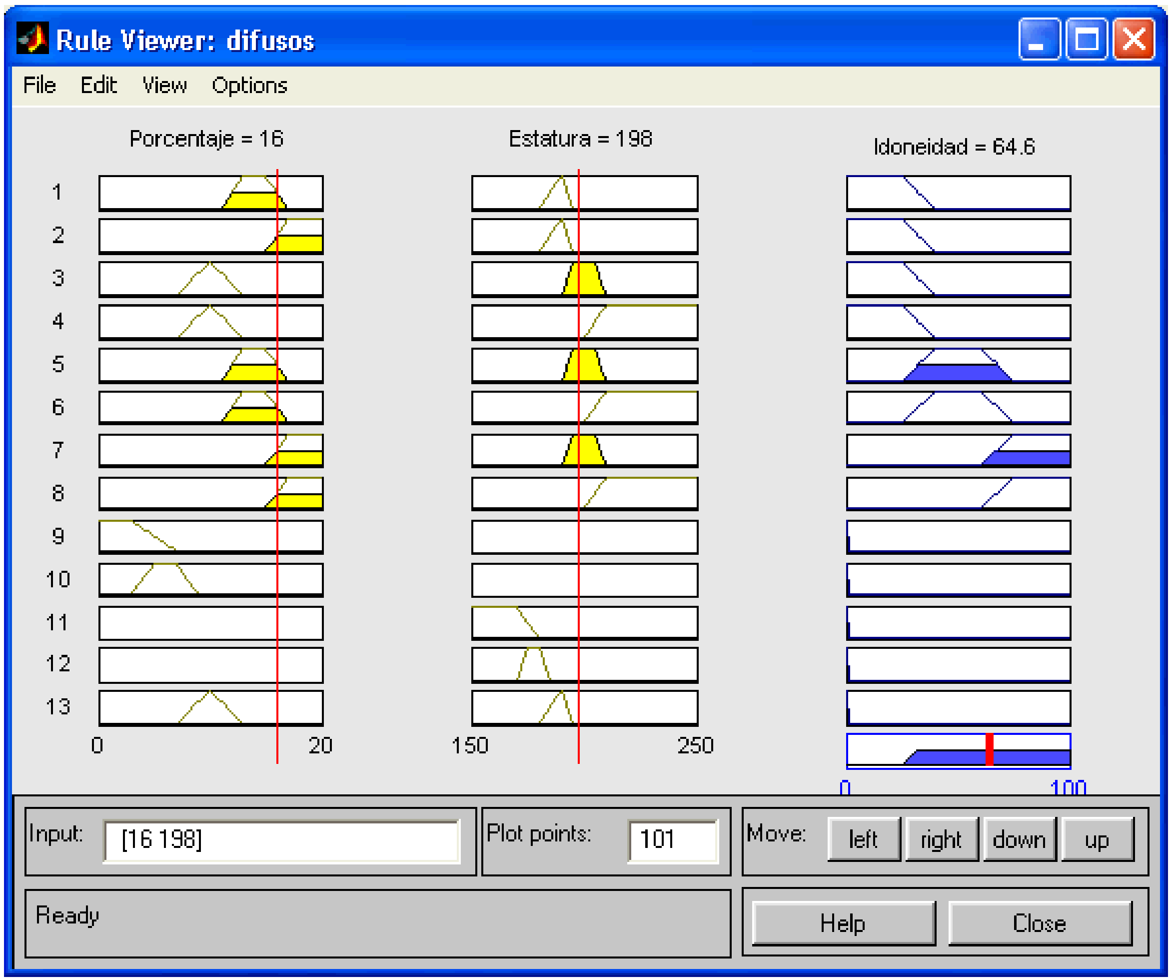
Reglas


Surface Viewer: difusos
File Edit View Options
100
80
60
40
20
0
Idoneidad
0
5
10
15
20
Porcentaje
X (input):
Porcentaje
Y (input):
- none -
Z (output):
Idoneidad
X grids:
15
Y grids:
15
Evaluate
Ref. Input:
[NaN 200]
Help
Close
Ready

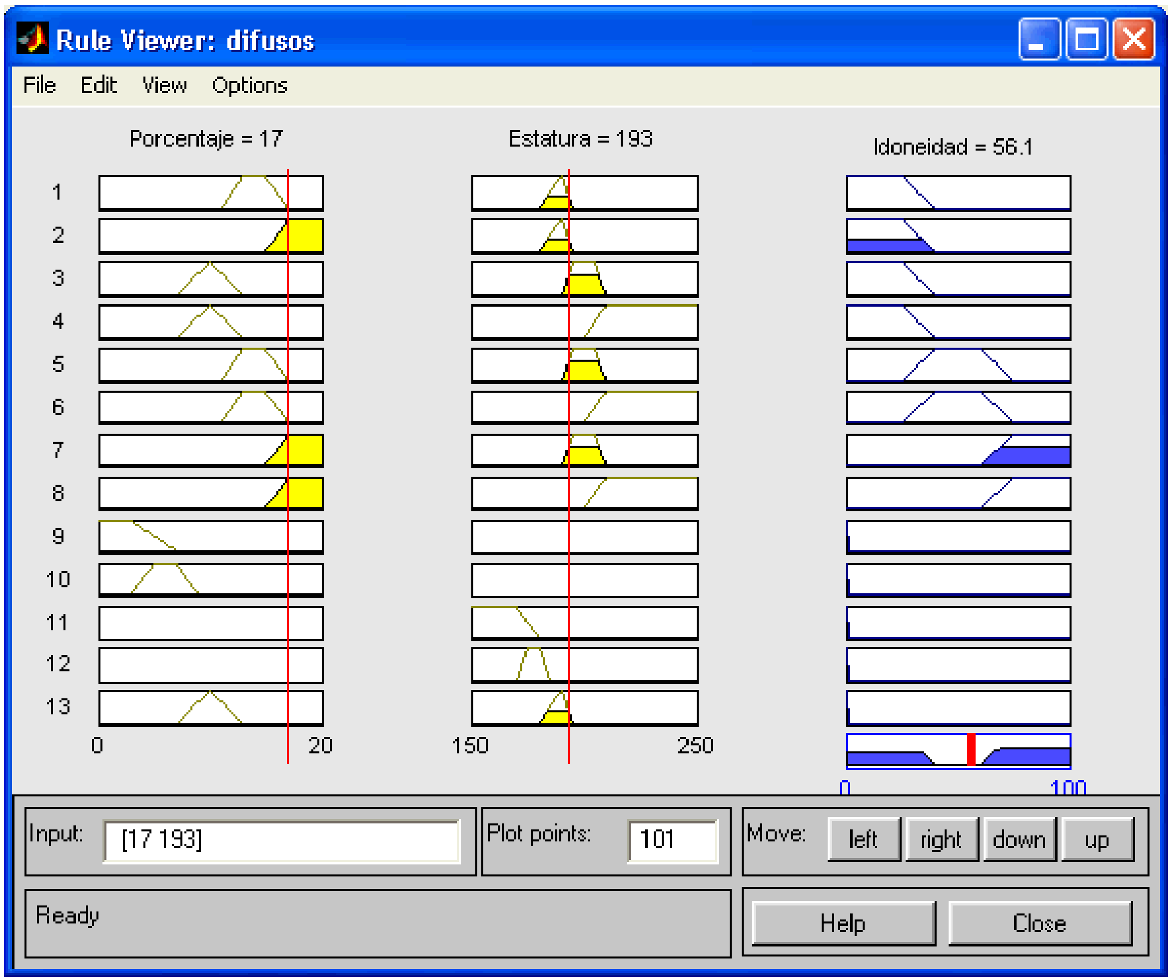
Reglas


Surface Viewer: difusos
File Edit View Options
Idoneidad
80
60
40
20
0
250
200
150
Estatura
0
5
10
15
20
Porcentaje
X (input):
Porcentaje
Y (input):
Estatura
Z (output):
Idoneidad
X grids:
15
Y grids:
15
Evaluate
Ref. Input:
Help
Close
Ready

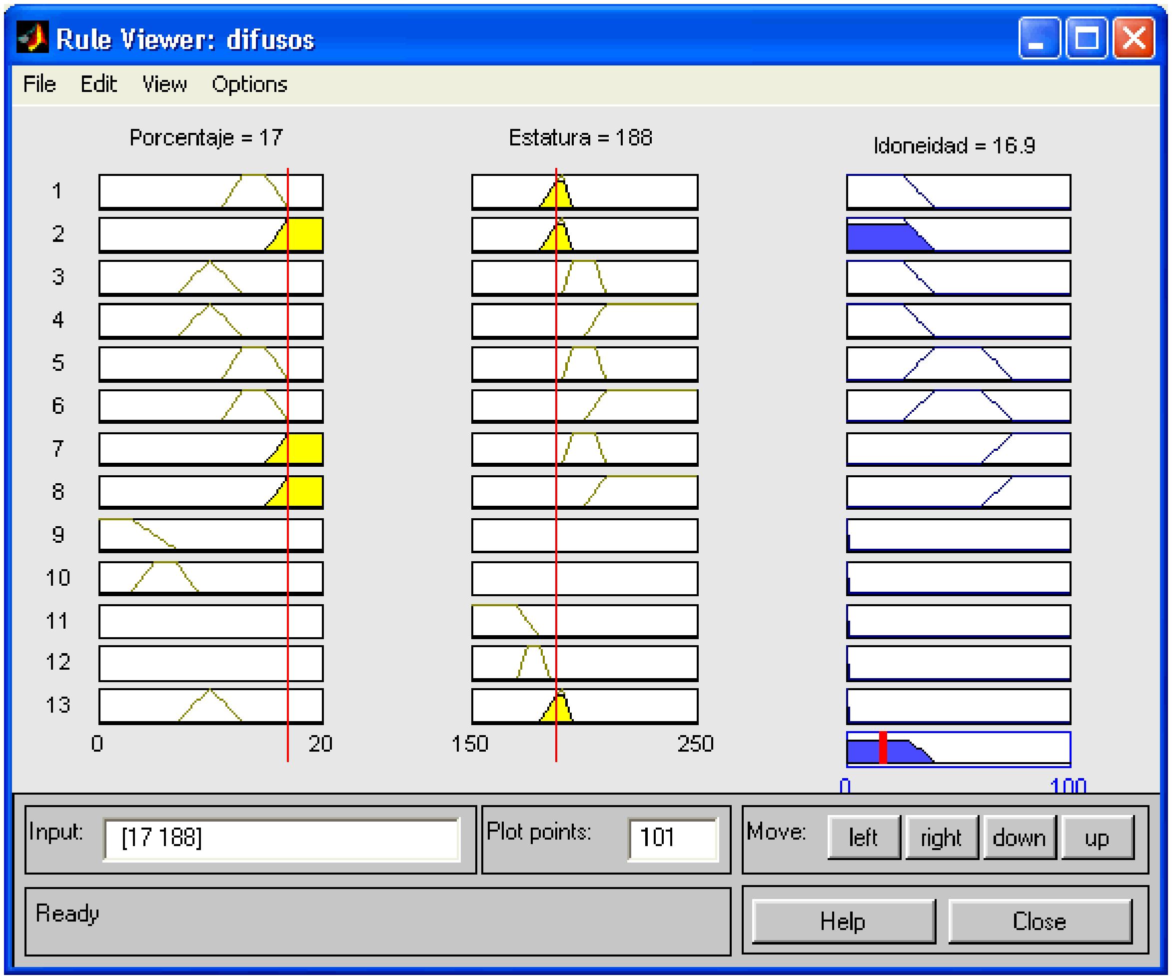
Ejemplos


Rule Viewer: difusos
File
Edit
View
Options
Porcentaje = 16
Estatura = 198
Idoneidad = 64.6
1
2
3
4
5
6
7
8
9
10
11
12
13
0
20
150
250
0
100
Input:
[16 198]
Plot points:
101
Move:
left
right
down
up
Ready
Help
Close

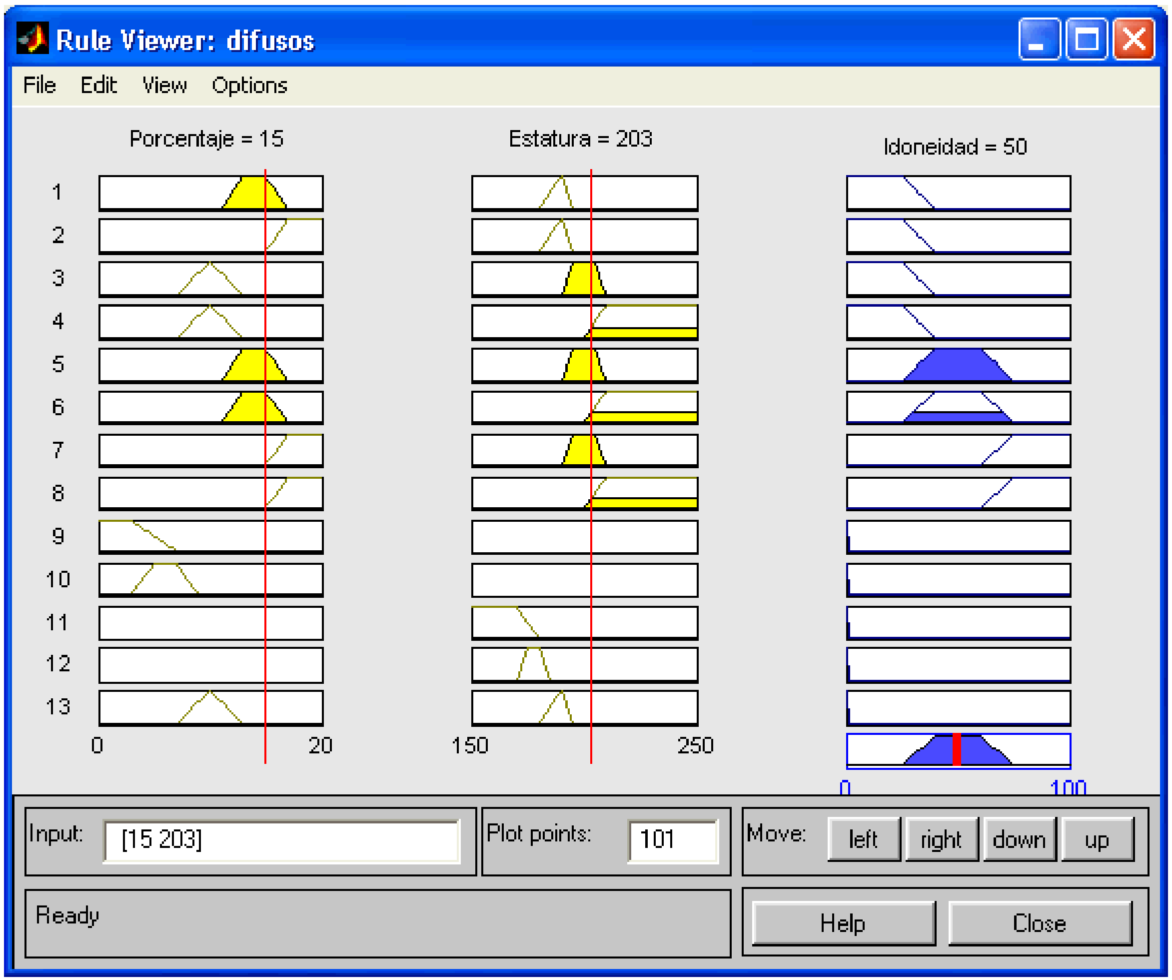
Ejemplos


Rule Viewer: difusos
File Edit View Options
Porcentaje = 17
Estatura = 193
Idoneidad = 56.1
1
2
3
4
5
6
7
8
9
10
11
12
13
0
20
150
250
0
100
Input:
[17 193]
Plot points:
101
Move:
left
right
down
up
Ready
Help
Close

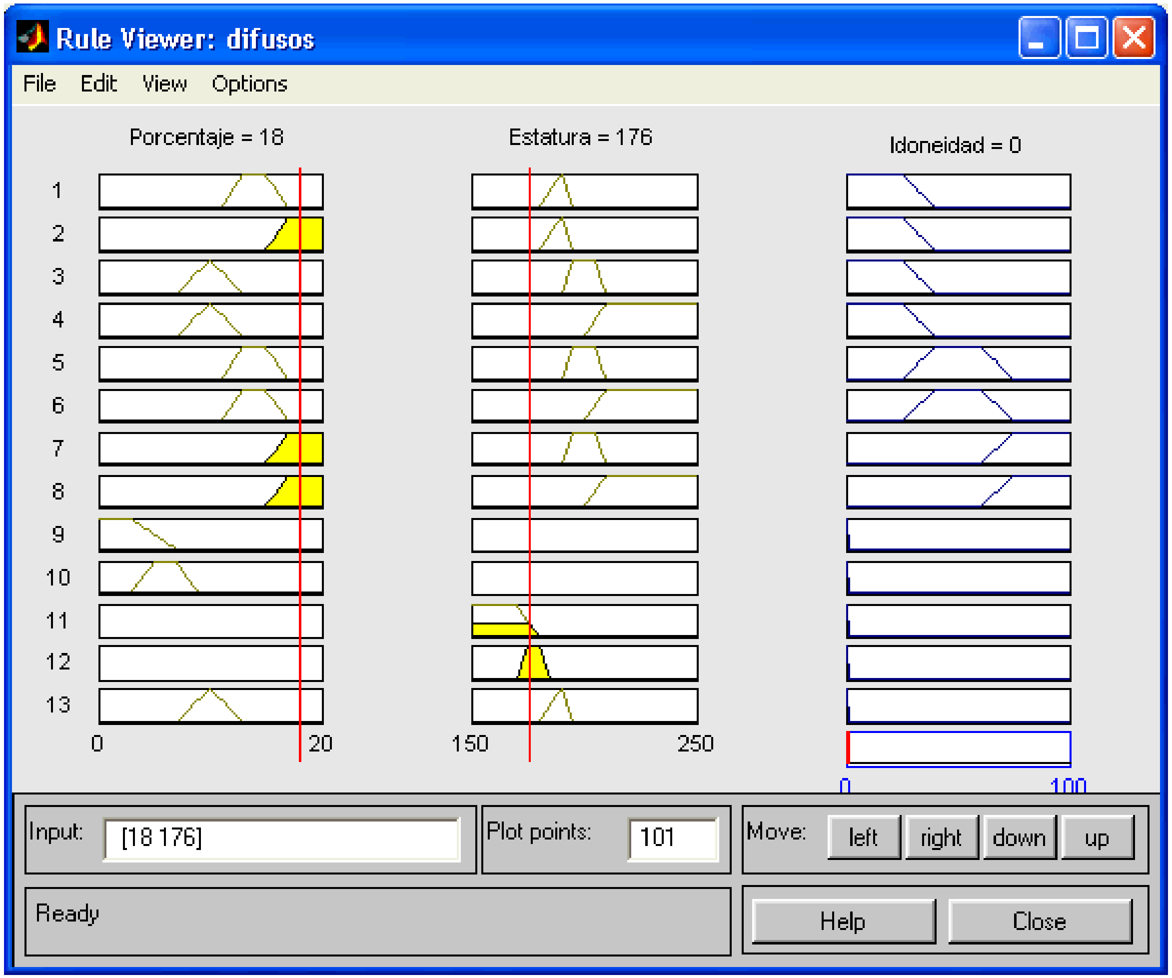
Ejemplos


Rule Viewer: difusos
File Edit View Options
Porcentaje = 17
Estatura = 188
Idoneidad = 16.9
1
2
3
4
5
6
7
8
9
10
11
12
13
0
20
150
250
0
100
Input:
[17 188]
Plot points:
101
Move:
left
right
down
up
Ready
Help
Close

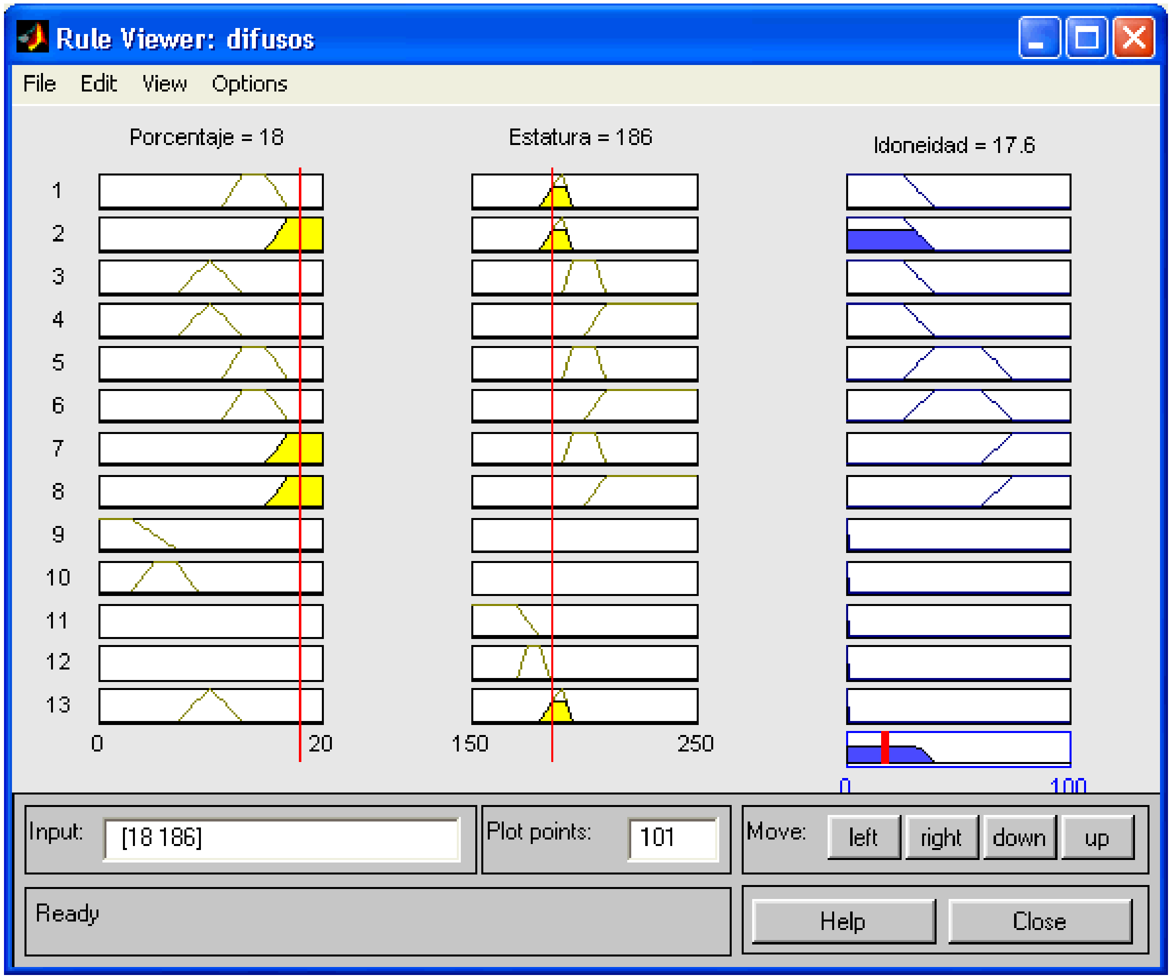
Ejemplos


Rule Viewer: difusos
File Edit View Options
Porcentaje = 15
Estatura = 203
Idoneidad = 50
1
2
3
4
5
6
7
8
9
10
11
12
13
0
20
150
250
0
100
Input:
[15 203]
Plot points:
101
Move:
left
right
down
up
Ready
Help
Close

Ejemplos


Rule Viewer: difusos
File Edit View Options
Porcentaje = 18
Estatura = 176
Idoneidad = 0
1
2
3
4
5
6
7
8
9
10
11
12
13
0
20
150
250
0
100
Input:
[18 176]
Plot points:
101
Move:
left
right
down
up
Ready
Help
Close

Ejemplos


Rule Viewer: difusos
File
Edit
View
Options
Porcentaje = 18
Estatura = 186
Idoneidad = 17.6
1
2
3
4
5
6
7
8
9
10
11
12
13
0
20
150
250
0
100
Input:
[18 186]
Plot points:
101
Move:
left
right
down
up
Ready
Help
Close

MÉTODOS FORMALES DE REPRESENTACIÓN DEL CONOCIMIENTO

- Nuevos modelos
  1. Lógicas no monótonas
  2. Razonamiento probabilístico
  3. Tratamiento de la incertidumbre
  4. Lógica difusa
  5. Modelos de credibilidad
- En lógica no monótona, la verdad de una afirmación puede estar basada en la falta de confianza en otra afirmación
- Los sistemas de primer orden son monótonos y crecientes, porque el número de declaraciones ciertas se incrementa estrictamente a lo largo de los procesos inferenciales

MÉTODOS FORMALES DE REPRESENTACIÓN DEL CONOCIMIENTO

- Ventajas de los sistemas monótonos
  - Cuando se deduce una nueva declaración no es preciso realizar ningún análisis de consistencia
  - Dada una declaración que acaba de ser demostrada, no es necesario recordar las declaraciones en las que se ha basado la demostración, ya que no van a desaparecer
- Sistemas no monótonos
  - En general, basan su estrategia en el hecho de que, normalmente, no disponemos de toda la información para resolver un problema
  - Se realizan suposiciones que pueden ser luego confirmadas o invalidadas, y dan lugar al llamado Razonamiento por Defecto

MÉTODOS FORMALES DE REPRESENTACIÓN DEL CONOCIMIENTO

- Razonamiento por defecto
  - Si X no es conocido, Entonces concluir Y
    - Pero en casi todos los sistemas formales gran parte del conocimiento está implícito
  - Si X no puede demostrarse, Entonces concluir Y
    - ¿Pero cómo podemos estar seguros de que X no puede demostrarse?
    - ¿Qué papel juegan las implicaciones temporales?
  - Si X no puede demostrarse en un determinado instante, Entonces concluir Y
    - Pero ahora la cuestión no es de carácter estrictamente lógico
    - Aparecen cuestiones relacionadas con la potencia computacional, la eficiencia de los procesos de búsqueda, etc
  - La cuestión es difícil de formalizar → Necesidad de definir esquemas no formales de representación del conocimiento

Otros Modelos Lógicos

Existen otras lógicas que permiten incorporar aspectos como el tiempo, la incertidumbre, que hechos dejen de ser ciertos…

- Lógicas modales
- Lógicas temporales
- Lógica difusa [fuzzy logic]
- Lógicas no monótonas
- …

Razonamiento Clásico

- Contenidos:
  - Aspectos generales del razonamiento
  - Interpretación diferencial
  - Elementos del razonamiento categórico
  - Procedimiento sistemático de razonamiento categórico
  - La corrección bayesiana
  - Necesidad de independencia en el esquema bayesiano
  - Otros problemas del esquema bayesiano
  - Redes Bayesianas

Razonamiento Clásico

- Punto de partida
  - Formalizar mecanismos y procesos inferenciales
  - Establecer y diseñar modelos de razonamiento
- Condicionantes
  - Características del dominio
  - Características del problema
- Dominios categóricos: (D1)
- Dominios estadísticos: (D2)
- Dominios cuasiestadísticos: (D3)
- Dominios difusos: (D4)
- Dominios reales: $\lambda 1(D1) + \lambda 2(D2) + \lambda 3(D3) + \lambda 4(D4)$

Métodos Categóricos

- Interpretación diferencial
  - En dominios de naturaleza puramente simbólica:

    ¿Cómo podemos definir un procedimiento encadenado y lógico para discriminar entre posibles soluciones candidatas, obtenidas a partir de datos y de verdades demostradas?

Métodos Categóricos

- Proceso:
  - Recopilación de información
  - Análisis de la importancia relativa de las manifestaciones
  - Análisis de posibles causas (establecimiento tentativo de relaciones causa-efecto)
  - Exclusión una a una de posibles interpretaciones
  - Fin del proceso…
    - No hay solución
    - Hay sólo una solución
    - Hay varias soluciones posibles

Métodos Categóricos

- Elementos del razonamiento categórico
  - Descripción del dominio de discurso
    - Manifestaciones posibles
    - Interpretaciones posibles
    - Relaciones causa-efecto
  - Ejemplo
    - Dominio: Predicciones meteorológicas
    - Manifestaciones: Color del crepúsculo,…
    - Interpretaciones o hipótesis: Posibilidad de lluvia,…
    - Relaciones causales: Si el crepúsculo… Y… Entonces

Métodos Categóricos

- Formalmente…
  - X = {manifestaciones} = {x1, x2,…, xn}
  - Y = {interpretaciones} = {y1, y2,…, ym}
  - En un caso concreto:
    - f = función booleana de X = f (x1,…, xn) tal que
      - f (xi) = 0 $\Leftrightarrow$ xi no es una manifestación de mi problema
      - f (xi) = 1 $\Leftrightarrow$ xi es una manifestación de mi problema
    - g = función booleana de Y = g (y1,…, ym) tal que
      - g (yj) = 0 $\Leftrightarrow$ yj no es una interpretación de mi problema
      - g (yj) = 1 $\Leftrightarrow$ yj es una interpretación de mi problema

Métodos Categóricos

- Las relaciones causales entre manifestaciones e interpretaciones se formalizan a través de la función de conocimiento E
  - E = E (X , Y)
- Problema lógico
  - Dadas unas manifestaciones caracterizadas por una función – f -, encontrar la función – g – que satisface
    - E : (f $\rightarrow$ g)
    - E : ($\neg$ g $\rightarrow$ $\neg$ f)
    - E = E (x1,…, xn, y1,…, ym)

Métodos Categóricos

- Ejemplo…
  - Sea un dominio D caracterizado por
    - Manifestaciones M = { m(1) , m(2) }
    - Interpretaciones I = { i(1) , i(2) }
  - En el que el conocimiento incluye las siguientes relaciones causales
    - Para que i(2) sea cierta, m(1) debe estar presente
    - Para que i(1) sea cierta, e i(2) sea falsa, m(2) debe estar presente
    - Para que i(2) sea cierta, e i(1) sea falsa, m(2) no debe estar presente
    - Si alguna manifestación está presente es porque se puede establecer alguna interpretación

Métodos Categóricos

- Formalización del ejemplo

  - R1: $i(2) \rightarrow m(1)$
  - R2: $i(1) \times \neg\, i(2) \rightarrow m(2)$
  - R3: $\neg\, i(1) \times i(2) \rightarrow \neg\, m(2)$
  - R4: $m(1) + m(2) \rightarrow i(1) + i(2)$
  - E = $\{ R1, R2, R3, R4 \}$

- En el dominio D aparece una situación en la que m(2) está presente, y m(1) está ausente… ¿Cuál es la interpretación lógica?

Métodos Categóricos

- $f[m(1), m(2)] = \neg m(1) \times m(2)$
- E: $(f \rightarrow g)$
- R1: $i(2) \rightarrow m(1) \equiv \neg i(2) \vee m(1)$
- R2: $i(1) \times \neg i(2) \rightarrow m(2) \equiv \neg i(1) \vee i(2) \vee m(2)$
- R3: $\neg i(1) \times i(2) \rightarrow \neg m(2) \equiv i(1) \vee \neg i(2) \vee \neg m(2)$
- R4: $m(1) + m(2) \rightarrow i(1) + i(2) \equiv [\neg m(1) \wedge \neg m(2)] \vee i(1) \vee i(2)$

- Sabemos que todas las declaraciones son ciertas
- Sabemos que $f = \neg m(1) \times m(2)$ también es cierta
- Para que R1 sea cierta, $[\neg i(2)]$ ha de ser cierta
- Para que R4 sea cierta, $[i(1)]$ ha de ser cierta
- Luego: $g = i(1) \times \neg i(2)$

Métodos Categóricos

- Si el dominio implicara tan sólo…
  - 30 manifestaciones
  - 600 interpretaciones
  - 125 relaciones causales…
  - … un tratamiento lógico convencional sería poco eficiente

↓

Necesitamos otra solución

Métodos Categóricos

- Procedimiento sistemático para el modelo categórico
  1. Identificación de M
  2. Identificación de I
  3. Construcción de E
  4. Construcción del conjunto completo de complejos de manifestaciones
  5. Construcción del conjunto completo de complejos de interpretaciones
  6. Construcción del conjunto completo de complejos manifestación-interpretación

Métodos Categóricos

- $M = \{ m(1), m(2), \ldots, m(n) \}$
- $I = \{ i(1), i(2), \ldots, i(t) \}$
- Número de complejos de manifestaciones = 2^n
- Número de complejos de interpretaciones = 2^t
- Número de complejos manifestación–interpretación = 2^{n+t}
- Carácter exhaustivo
- Elementos mutuamente excluyentes
- El conjunto de complejos manifestación – interpretación representa el total de situaciones idealmente posibles en el dominio… Pero no todas son posibles si tenemos en cuenta el conocimiento del sistema

Métodos Categóricos

- Recordemos que…
  - R1: $i(2) \rightarrow m(1)$
  - R2: $i(1) \times \neg\, i(2) \rightarrow m(2)$
  - R3: $\neg\, i(1) \times i(2) \rightarrow \neg\, m(2)$
  - R4: $m(1) + m(2) \rightarrow i(1) + i(2)$

  - En donde $M = \{ m(1), m(2) \}$, $I = \{ i(1), i(2) \}$

  - Si representamos la situación en término de complejos (sensible al criterio), podemos construir **M** e **I**, que contienen todas las combinaciones posibles de manifestaciones y de interpretaciones, respectivamente

Métodos Categóricos

m(1)	0	0	1	1
m(2)	0	1	0	1
	m1	m2	m3	m4

i(1)	0	0	1	1
i(2)	0	1	0	1
	i1	i2	i3	i4

M = { m1, m2, m3, m4 } **I** = { i1, i2, i3, i4 }

Métodos Categóricos

- Construcción del conjunto completo de complejos manifestación-interpretación
  - Base Lógica Expandida
    - BLE = **M** x **I** = { m1i1 , m1i2 , m1i3 , m1i4 ,
    m2i1 , m2i2 , m2i3 , m2i4 ,
    m3i1 , m3i2 , m3i3 , m3i4 ,
    m4i1 , m4i2 , m4i3 , m4i4 }
  - La solución a cualquier problema está en BLE, pero hay muchas combinaciones absurdas
  - El papel del conocimiento – E – es eliminarlas y pasar a una base lógica reducida E : (BLE $\rightarrow$ BLR)

Métodos Categóricos

	m1	m2	m3	m4	m1	m2	m3	m4	m1	m2	m3	m4	m1	m2	m3	m4
m (1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m (2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i (1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i (2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	i1	i1	i1	i2	i2	i2	i2	i3	i3	i3	i3	i4	i4	i4	i4

Métodos Categóricos

- R1: $i(2) \rightarrow m(1)$
  - Elimina de BLE: m1i2, m1i4, m2i2, m2i4

	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4
m (1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m (2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i (1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i (2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	i1	i1	i1	i2	i2	i2	i2	i3	i3	i3	i3	i4	i4	i4	i4

Métodos Categóricos

- R2: $i(1) \times \neg i(2) \rightarrow m(2)$
  - Elimina de BLE: m1i3, m3i3

	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4
m (1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m (2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i (1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i (2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	i1	i1	i1	i2	i2	i2	i2	i3	i3	i3	i3	i4	i4	i4	i4

Métodos Categóricos

- R3: $\neg$ i(1) x i(2) $\rightarrow$ $\neg$ m(2)
- Elimina de BLE: m2i2, m4i2

	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4
m (1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m (2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i (1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i (2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	i1	i1	i1	i2	i2	i2	i2	i3	i3	i3	i3	i4	i4	i4	i4

Métodos Categóricos

- R4: m(1) + m(2) → i(1) + i(2)
- Elimina de BLE: m2i1, m3i1, m4i1

	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4
m (1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m (2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i (1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i (2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	i1	i1	i1	i2	i2	i2	i2	i3	i3	i3	i3	i4	i4	i4	i4

Métodos Categóricos

- BLR = {m1i1, m3i2, m2i3, m4i3, m3i4, m4i4}

- IF: (1) El conocimiento es completo
- And: (2) El dominio está bien descrito
- Then: (1) La solución a cualquier problema está en BLR

Métodos Categóricos

En términos de complejos:

Problema planteado inicialmente

$f = \neg m(1) \times m(2) = m2$

¿Qué complejos de BLR contienen a m2?

m2i3

$\downarrow$

$g = i3 = i(1) \times \neg i(2)$

La interpretación i (1) es correcta, y además sabemos que i (2) es falsa

Métodos Categóricos

- Sea ahora la nueva regla:
- R5: $\neg$ i (1) x $\neg$ i (2) $\rightarrow$ m (1) + m (2)
- Elimina de BLE: m1i1

	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4
m (1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m (2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i (1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i (2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	i1	i1	i1	i2	i2	i2	i2	i3	i3	i3	i3	i4	i4	i4	i4

Métodos Categóricos

- Se abre debate…
  - ¿Qué ocurriría si aparece un problema caracterizado por

    $f' = \neg m(1) \times \neg m(2)$?

  - ¿Puede existir este problema?
  - En términos generales…
    - Las manifestaciones no son realmente esas
    - El conocimiento no es correcto
    - El dominio no está bien construido

Métodos Categóricos

- Eliminando ahora R5 y volviendo a la BLR inicial
- ¿Qué ocurriría si f" = m (1) x ¬ m (2) = m3?

	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4	m 1	m 2	m 3	m 4
m (1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m (2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i (1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i (2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	i1	i1	i1	i2	i2	i2	i2	i3	i3	i3	i3	i4	i4	i4	i4

Métodos Categóricos

- En este caso…
  - Los complejos de BLR son m3i2 y m3i4
  - g = i2 + i4 = [¬ i (1) x i (2)] + [i (1) x i (2)]
  - i (2) es cierta
  - De i (1) no sabemos nada
  - La incertidumbre aparece de forma espontánea en el modelo categórico… ¿por qué?

  - 7 manifestaciones y 24 interpretaciones obligan a discriminar entre 2,147,483,600 complejos manifestación-interpretación

Métodos Bayesianos

- Las interpretaciones categóricas son poco frecuentes en el mundo real
  - La existencia de una determinada causa no siempre conlleva la presencia de una manifestación
  - Ante una manifestación dada ¿podemos afirmar siempre y de forma categórica que existe una determinada causa?
- Probabilidad total y probabilidad condicional
  - Dado un universo y una lista de atributos ¿cuál es la probabilidad de que un determinado elemento de este universo presente tales atributos?

Métodos Bayesianos

- Sea un universo N
- Sea un conjunto de atributos f (x, y,…, t)
- N (f) = nº de elementos del universo que presenta los atributos f

$$P(f) = \frac{N(f)}{N}$$

Métodos Bayesianos

- Con nuestro lenguaje categórico…
  - Sean las interpretaciones i1,…,in que definen al complejo de interpretaciones Dj
  - Sea N (Dj) el número de elementos del universo cuyo "problema" es Dj
  - Sea N el número de elementos del universo
  - La probabilidad del complejo de interpretaciones Dj es…

$$P(Dj) = \frac{N(Dj)}{N}$$

Métodos Bayesianos

- La probabilidad condicional se parece a la total, pero puede ser definida como la probabilidad de las causas

- En la probabilidad condicional aparecen involucrados dos sucesos, en donde la ocurrencia del segundo depende de la ocurrencia del primero

Métodos Bayesianos

- Consideremos dos acciones posibles y dos resultados asociados a cada una de las acciones…
  - Acciones: A , B
  - Resultados: E , F
  - Se sabe que el resultado depende de la acción
  - Elegimos la acción por el método de la moneda: P (A) = 0.5 P (B) = 0.5

Métodos Bayesianos

- La acción A genera resultados E el 20% de las veces
- La acción B genera resultados E el 60% de las veces
- Ante un nuevo caso siempre podemos estimar la probabilidad a priori de un resultado E

  - Acción A $\quad P(A) = 0.5 \quad P(E/A) = 0.2$
  - Acción B $\quad P(B) = 0.5 \quad P(E/B) = 0.6$

  - $P(E) = P(E/A)\,P(A) + P(E/B)\,P(B) = 0.4$

Métodos Bayesianos

- Bayes introduce una forma de razonamiento a posteriori…
  - Dadas dos posibles acciones A y B, y ante un caso tratado con una de tales acciones ¿Cuál es la probabilidad de que la acción haya sido A, si la respuesta del sistema ha sido E?

$$P(A/E) = \frac{P(E/A)P(A)}{P(E)}$$

Métodos Bayesianos

- Obtención de la ecuación elemental del teorema de Bayes
  - Sea una población sobre parte de cuyos elementos ha sido efectuada una acción A de un número de posibles acciones
  - Todos los elementos de la población fueron tratados, con A o con cualquier otra acción, registrándose un cierto número de respuestas E

Métodos Bayesianos

- N = población
- Acciones posibles: A , $\neg$ A
- Resultados posibles: E , $\neg$ E
- n (A)
- n ($\neg$ A)
- n (E)
- n ($\neg$ E)
- n (A $\cap$ E) = n (E $\cap$ A)

Métodos Bayesianos

- Por definición de probabilidad condicional:

$$P(E/A)=\frac{n(A\cap E)}{n(A)}=\frac{n(A\cap E)/N}{n(A)/N}=\frac{P(A\cap E)}{P(A)}\rightarrow P(A\cap E)=P(A)P(E/A)$$

- Análogamente…

$$P(A/E)=\frac{n(A\cap E)}{n(E)}=\frac{n(A\cap E)/N}{n(E)/N}=\frac{P(A\cap E)}{P(E)}\rightarrow P(A\cap E)=P(E)P(A/E)$$

$$P(A)P(E/A)=P(E)P(A/E)$$

Métodos Bayesianos

- Obtención de una ecuación generalizada del teorema de Bayes
  - Sea una característica cualquiera X
  - Sea A el número de casos de una población estadística en los que X está presente
  - Sea ¬ A el número de casos de la misma población estadística en los que X está ausente
  - Sea P una prueba potencialmente resolutiva para investigar la característica X
  - Sea E el número de casos en los que la prueba da resultados positivos
  - Sea ¬ E el número de casos en los que la prueba da resultados negativos

Métodos Bayesianos

	A	¬A	TOTAL
E	a	b	a + b
¬E	c	d	c + d
TOTAL	a + c	b + d	N

- a = positivos reales
- b = falsos positivos
- c = falsos negativos
- d = negativos reales

Métodos Bayesianos

- Relaciones…
  - $P(E / A) = a / (a + c)$ sensibilidad
  - $P(\neg E / A) = c / (a + c)$
  - $P(A / E) = a / (a + b)$
  - $P(\neg A / E) = b / (a + b)$
  - $P(E / \neg A) = b / (b + d)$
  - $P(\neg E / \neg A) = d / (b + d)$ especificidad
  - $P(A / \neg E) = c / (c + d)$
  - $P(\neg A / \neg E) = d / (c + d)$
- Prevalencias
  - $P(E) = (a + b) / N$ $P(A) = (a + c) / N$
  - $P(\neg E) = (c + d) / N$ $P(\neg A) = (b + d) / N$

Métodos Bayesianos

- Tras conocer las probabilidades a priori ¿cómo se plantea la probabilidad condicional a posteriori de la característica X dado un resultado positivo de la prueba?

$$Bayes \rightarrow P(A/E) = \frac{P(E/A)P(A)}{P(E)} : P(E) = \frac{(a+b)}{N}$$

$$a = (a+c)P(E/A) : b = (b+d)P(E/\neg A)$$

$$(a+c) = N \times P(A) : (b+d) = N \times P(\neg A)$$

$$a = N \times P(A) \times P(E/A) : b = N \times P(\neg A) \times P(E/\neg A)$$

$$P(E) = P(A)P(E/A) + P(\neg A)P(E/\neg A)$$

$$P(A/E) = \frac{P(E/A)P(A)}{P(E/A)P(A) + P(E/\neg A)P(\neg A)}$$

Métodos Bayesianos

- La ecuación $$P(A/E) = \frac{P(E/A)P(A)}{P(E/A)P(A) + P(E/\neg A)P(\neg A)}$$

…es directamente generalizable. Así, si consideramos todos los Ai posibles obtenemos que:

$$P(A0/E) = \frac{P(E/A0)P(A0)}{\sum_i P(E/Ai)P(Ai)}$$

Que es una expresión general del teorema de Bayes

Métodos Bayesianos

- Volviendo al ejemplo categórico anterior…
  - BLR = {m1i1, m3i2, m2i3, m4i3, m3i4, m4i4}
  - Dado f = m(1) $\wedge \neg$ m(2) = m3 …
  - ¿Cuál es g? : g = i2 $\vee$ i4 = [$\neg$ i(1) $\wedge$ i(2)] $\vee$ [i(1) $\wedge$ i(2)]
- Según el planteamiento bayesiano, el problema se reduce ahora a evaluar…
  - P (i2 / m3)
  - P (i4 / m3)

Métodos Bayesianos

- De acuerdo con la expresión general del teorema de Bayes:

$$P(i2/m3)=\frac{P(m3/i2)P(i2)}{P(m3/i1)P(i1)+P(m3/i2)P(i2)+P(m3/i3)P(i3)+P(m3/i4)P(i4)}$$

$$P(i4/m3)=\frac{P(m3/i4)P(i4)}{P(m3/i1)P(i1)+P(m3/i2)P(i2)+P(m3/i3)P(i3)+P(m3/i4)P(i4)}$$

Métodos Bayesianos

- Así, un tratamiento exhaustivo del problema obliga a conocer:

P(i1)	P(i2)	P(i3)	P(i4)
P(m1/i1)	P(m1/i2)	P(m1/i3)	P(m1/i4)
P(m2/i1)	P(m2/i2)	P(m2/i3)	P(m2/i4)
P(m3/i1)	P(m3/i2)	P(m3/i3)	P(m3/i4)
P(m4/i1)	P(m4/i2)	P(m4/i3)	P(m4/i4)

- … y llegados a este punto toca abrir debate!!!

Métodos Bayesianos

- Necesidad de independencia en el modelo Bayesiano
  - Dados los datos de la tabla, evaluar P(A1/m1)
    - Utilizando el teorema de Bayes
    - Utilizando directamente los datos de la tabla

	Sólo A1	A1 y A2	Sólo A2	A3	Total
Sólo m1	B1	C1	D1	E1	N1
m1 y m2	B2	C2	D2	E2	N2
Sólo m2	B3	C3	D3	E3	N3
m3	B4	C4	D4	E4	N4
Total	B	C	D	E	N

Métodos Bayesianos

- P(A1) = (B+C)/N P(A2) = (C+D)/N P(A3) = E/N
- P(m1/A1) = (B1+B2+C1+C2)/(B+C)
- P(m1/A2) = (C1+C2+D1+D2)/(C+D)
- P(m1/A3) = (E1+E2)/E
- Utilizando el teorema de Bayes…
  - P(A1/m1) = P(m1/A1) P(A1)/ [P(m1/A1)P(A1)+P(m1/A2)P(A2)+P(m1/A3)P(A3)j]
  - P(m1/A1)P(A1) = (B1+B2+C1+C2)/N
  - P(m1/A2)P(A2) = (C1+C2+D1+D2)/N
  - P(m1/A3)P(A3) = (E1+E2)/N
  - P(A1/m1) = (B1+B2+C1+C2)/(N1+N2+C1+C2)
- Directamente de la tabla…
  - P(A1/m1) = (B1+B2+C1+C2)/(N1+N2)

Métodos Bayesianos

Otros problemas del esquema Bayesiano (1)

- Aparición secuencial de la información
  - Sea E1 el conjunto de toda la información disponible en un momento dado
  - Sea S1 un nuevo elemento de información
  - Llamaremos E al nuevo conjunto formado por E1 y S1
  - La nueva expresión del teorema de Bayes es…

$$P(Ii/E) = \frac{P(S1/Ii \wedge E1)P(Ii/E1)}{\sum_j P(S1/Ij \wedge E1)P(Ij/E1)}$$

Métodos Bayesianos

Otros problemas del esquema Bayesiano (2)

- Aplicación poco cuidadosa del modelo
  - Relación entre visión borrosa y glaucoma
  - P(visión borrosa) = 231/15315
  - P(glaucoma) = 320/15315
  - P(visión borrosa/glaucoma) = 150/320
  - Bayes $\rightarrow$ P(glaucoma/visión borrosa) = 0.67
  - ¿Le limpiamos las gafas al paciente?

Métodos Bayesianos

Otros problemas del esquema Bayesiano (3)

- Consistencia matemática del modelo
  - Si P(Hipótesis/Evidencia) = x / $0 \leq x \leq 1$
  - Entonces P($\neg$ Hipótesis/Evidencia) = 1 – x
  - La misma evidencia apoya positivamente a una hipótesis y a la negación de esa misma hipótesis

¿?

Ejemplo Categórico-Bayesiano

- UN PROBLEMA PARA “HOUSE”
  - Una madre observa que su hija de 5 semanas tiene distress respiratorio. Esta situación se mantiene durante 5 días, y el proceso no parece remitir. La niña no presentaba este problema cuando nació. El examen médico muestra que la niña está bien nutrida aunque tiene hemangiomas en la parte trasera del cuello bajo la oreja izquierda. El examen del tórax por rayos X muestra una masa en el mediastino, que desplaza la tráquea. Acompañan a esta masa pequeños gránulos de calcio.

- A la vista de esta información, el clínico que atiende el caso debe discriminar entre las siguientes posibilidades para poder establecer un tratamiento adecuado:

  1: La paciente tiene un timo anormalmente grande. Es bien conocido que un timo anormalmente grande puede producir distress respiratorio

  2: La paciente tiene un gran hemangioma en el mediastino, que se manifiesta a través de los hemangiomas superficiales

  3: La paciente puede tener un quiste en el mediastino. La presencia de calcio junto a la masa detectada a través de rayos X sugiere esta posibilidad

Ejemplo Categórico-Bayesiano

- Además el clínico sabe que:
  - Si un paciente tiene un timo anormalmente grande, y además tiene o bien un hemangioma en el mediastino, o bien un quiste, entonces debe presentar distress respiratorio, y una masa en el mediastino
  - Si un paciente no tiene un hemangioma en el mediastino, entonces no debe presentar hemangiomas superficiales
  - Si un paciente no tiene un timo anormalmente grande, pero tiene un hemangioma en el mediastino, y además un quiste, entonces debe presentar una masa en el mediastino
- ¿A qué conclusión llegará House en este caso?
- ¿Cuál sería el diagnóstico si la prueba de r
  un resultado negativo?

Razonamiento Cuasi-Estadístico

- Factores de Certidumbre
  - Aspectos básicos del modelo
  - Combinación de evidencias
  - Propagación de incertidumbre
- Teoría Evidencial
  - Fundamentos teóricos
  - Combinación de evidencias
  - Credibilidad, Plausibilidad, Intervalo de Confianza
  - Casos particulares

Razonamiento Cuasi-Estadístico

- Bibliografía
  - Heckerman, Probabilistic interpretation for MYCIN´s certainty factors, Uncertainty in Artificial Intelligence, 1986
  - Shortliffe & Buchanan, A model of inexact reasoning in medicine, Mathematical Biosciences, vol.23, 1975
  - Shafer, A mathematical theory of evidence, Princeton University press, eds., 1976

Razonamiento Cuasi-Estadístico

- Evitar inconvenientes de modelos anteriores $\rightarrow$ Utilizar conocimiento heurístico
- Probabilidad condicional subjetiva
  - Medida numérica que relaciona dos sucesos, de forma que la ocurrencia de uno está condicionada por la ocurrencia del otro, pero en donde la relación no está avalada por amplios estudios estadísticos

Razonamiento Cuasi-Estadístico

- P (Ii / Sk) = x podría traducirse como:
  - “Si la manifestación Sk está presente, entonces hay una probabilidad subjetiva x de que la interpretación sea Ii
- La consistencia matemática del concepto de probabilidad exige que:
  - $\sum_i P(Ii / Sk) = 1$
- Cuando lo anterior no ocurre hay que normalizar:
  - $P(Ii / Sk) = x / \sum_i P(Ii / Sk)$
- La aparición de conceptos difíciles de definir como imprecisión, incertidumbre, falta de confianza, credibilidad, etc., exigen la puesta a punto de nuevos mecanismos para tratar el razonamiento

Razonamiento Cuasi-Estadístico

- Introduciendo conceptos…
  - P (Ii / Sk) = x puede interpretarse en términos de implicación

$$Sk \xrightarrow{x} Ii$$

  - “x” define la potencia evidencial de la implicación
  - Si x $\in$ [0 , 1) la implicación viene afectada de incertidumbre
  - La intensidad de la relación causal Ii / Sk se establece a través de la potencia evidencial x

Factores de Certidumbre

- El modelo de factores de certidumbre de Shortliffe y Buchanan
  - 1975. Naturaleza ad hoc. Sistema Mycin.
  - Dada una hipótesis, la potencia evidencial de una declaración se debe representar a través de dos medidas diferentes:
    - Medida de confianza creciente MB (h , e)
    - Medida de desconfianza creciente MD (h , e)
  - MB y MD son índices dinámicos que representan incrementos asociados a evidencias nuevas

Factores de Certidumbre

- Características fundamentales…
  - Si h es una hipótesis, y e es una evidencia, la misma evidencia e no puede, simultáneamente, incrementar la confianza en h y disminuir la confianza en h
  - MB (h , e) es el incremento de confianza en h dada la evidencia e
  - MD (h , e) es el incremento de la desconfianza en h dada la evidencia e

Factores de Certidumbre

- Formalismo…
  - Sea p(h) la confianza previa en h antes de e
  - Sea p(h/e) la confianza en h tras aparecer e
  - Sea 1 – p(h) la desconfianza en h antes de e
- Caso 1
  - Si p(h/e) > p(h) $\rightarrow$ la nueva evidencia produce un aumento de confianza en la hipótesis considerada

$$MB(h,e) > 0$$

$$MD(h,e) = 0$$

$$MB(h,e) = \frac{p(h/e) - p(h)}{1 - p(h)}$$

Factores de Certidumbre

- Caso 2
  - Si p(h/e) < p(h) $\rightarrow$ la nueva evidencia produce una disminución de la confianza inicialmente depositada en la hipótesis considerada

$$MB(h,e) = 0$$

$$MD(h,e) > 0$$

$$MD(h,e) = \frac{p(h) - p(h/e)}{p(h)}$$

Factores de Certidumbre

- Caso 3
  - Si p(h/e) = p(h) $\rightarrow$ la nueva evidencia es independiente de la hipótesis considerada, ya que ni aumenta la confianza ni aumenta la desconfianza

$$MB(h,e) = 0$$

$$MD(h,e) = 0$$

Factores de Certidumbre

- Valores límite:
  - p(h) es una probabilidad a priori en sentido clásico

$$MB(h,e)=1 \Leftrightarrow p(h)=1$$

$$MB(h,e)=\frac{\max[p(h/e),p(h)]-p(h)}{\max[0,1]-p(h)} \Leftrightarrow p(h)\neq 1$$

$$MD(h,e)=1 \Leftrightarrow p(h)=0$$

$$MD(h,e)=\frac{\min[p(h/e),p(h)]-p(h)}{\min[0,1]-p(h)} \Leftrightarrow p(h)\neq 0$$

Factores de Certidumbre

- Factor de certidumbre CF(h,e)
  - Combina las dos medidas anteriores según la expresión
    - $CF(h,e) = MB(h,e) – MD(h,e)$
  - Es de carácter formal, ya que una misma evidencia nunca puede incrementar, simultáneamente, la confianza y la desconfianza en la misma hipótesis
  - Sirve como un medio para facilitar la comparación entre potencias evidenciales de hipótesis alternativas (h1,…,hn) en relación con una misma evidencia e
- Rangos
  - $0 \leq MB(h,e) \leq 1$; $0 \leq MD(h,e) \leq 1$; $-1 \leq CF(h,e) \leq 1$

Factores de Certidumbre

- Comportamiento en casos extremos e hipótesis mutuamente excluyentes (1)
  - Si “h” es cierta $\rightarrow$ p(h/e) = 1

$$MB(h,e)=\frac{1-p(h)}{1-p(h)}=1$$

$$MD(h,e)=0$$

$$CF(h,e)=1$$

Factores de Certidumbre

- Comportamiento en casos extremos e hipótesis mutuamente excluyentes (2)
  - Si la negación de "h" es cierta $\rightarrow$ p($\neg$h/e) = 1

$$MD(h,e) = \frac{p(h) - 0}{p(h)} = 1$$

$$MB(h,e) = 0$$

$$CF(h,e) = -1$$

Factores de Certidumbre

- Comportamiento en casos extremos e hipótesis mutuamente excluyentes (3)
  - Los casos anteriores conducen a que…
    - $MB(\neg h,e) = 1 \leftrightarrow MD(h,e) = 1$
  - Si h1 y h2 son hipótesis mutuamente excluyentes, y sabemos que $MB(h1,e) = 1$, entonces podemos afirmar rotundamente que $MD(h2,e) = 1$

Factores de Certidumbre

- Evidencias independientes de la hipótesis
  - Sea h la hipótesis considerada, y sea e una evidencia. Si la evidencia es independiente de la hipótesis, entonces…

$$p(h/e) = p(h)$$

$$MB(h,e) = 0$$

$$MD(h,e) = 0$$

$$CF(h,e) = 0$$

Factores de Certidumbre

- Diferencias con las probabilidades condicionales
  - Los CFs de las hipótesis h y $\neg$h no son complementarios a la unidad, son opuestos entre sí
  - Si el apoyo que una evidencia presta a una hipótesis es bajo, no debería ser alto al apoyo a la negación de la hipótesis. Sobre todo en caso de información incompleta

Factores de Certidumbre

- Demostración de casos extremos…

$$p(h/e) > p(h)$$

$$CF(h,e) = MB(h,e) > 0$$

$$CF(\neg h,e) = -MD(h,e) < 0$$

$$MB(h,e) = \frac{p(h/e) - p(h)}{1 - p(h)} = \frac{1 - p(\neg h/e) - 1 + p(\neg h)}{p(\neg h)} =$$

$$= \frac{p(\neg h) - p(\neg h/e)}{p(\neg h)} = MD(\neg h,e) \rightarrow CF(h,e) = -CF(\neg h,e)$$

  - El mismo resultado se obtiene si p(h) > p(h/e)

Factores de Certidumbre

- ¿Cómo se manejan los CFs?
  - $0 < CF(h,e) \leq 1 \rightarrow$ La evidencia apoya a la hipótesis
  - $-1 \leq CF(h,e) < 0 \rightarrow$ La evidencia va en contra de la hipótesis
  - $CF(h,e) = 0$
    - La evidencia es independiente de la hipótesis
    - No se tiene información sobre la relación causal h/e

Factores de Certidumbre

- Combinación de evidencias S-B
  - Situación: Hay más de una evidencia relativa a la misma hipótesis
    - IF: e1 Then: h With: CF(h,e1)
    - IF: e2 Then: h With: CF(h,e2)
    - …
    - IF: en Then: h With: CF(h,en)
  - Problema:
    - Evaluar CF(h,E)
    - E = {e1, e2,…,en}

Factores de Certidumbre

- Primera aproximación (1)
  - Caso A: e1 y e2 contribuyen positivamente a la veracidad de la hipótesis

$$CF(h,e1) > 0$$

$$CF(h,e2) > 0$$

$$CF(h,e1 \wedge e2) = CF(h,e1) + CF(h,e2) - CF(h,e1) \times CF(h,e2)$$

Factores de Certidumbre

- Primera aproximación (2)
  - Caso B: e1 y e2 contribuyen negativamente a la veracidad de la hipótesis

$$CF(h,e1)<0$$

$$CF(h,e2)<0$$

$$CF(h,e1\wedge e2)=CF(h,e1)+CF(h,e2)+CF(h,e1)\times CF(h,e2)$$

Factores de Certidumbre

- Primera aproximación (3)
  - Caso C: e1 y e2 contribuyen una negativamente y la otra positivamente a la veracidad de la hipótesis

$$CF(h,e1)\times CF(h,e2)<0$$

$$CF(h,e1\wedge e2)=CF(h,e1)+CF(h,e2)$$

Factores de Certidumbre

- Ventajas de la primera aproximación
  - Nos previene de la hipotética situación de que ambas evidencias pudieran no ser completamente independientes
  - Las expresiones son generalizables directamente. En el caso de evidencias todas positivas obtenemos…

$$CF(h,E) = \sum_{i}^{n} CFi - \sum_{\substack{i \\ i<j}}^{n} CFi \times CFj + \sum_{\substack{i \\ i<j<k}}^{n} CFi \times CFj \times CFk - \ldots$$

Factores de Certidumbre

- Inconvenientes de la primera aproximación
  - Falta de asociatividad de la formulación
  - Sensibilidad de la formulación ante la aparición de evidencias contradictorias en estados avanzados del proceso de razonamiento

Factores de Certidumbre

- Ejemplo
  - 8 reglas apoyan a una misma hipótesis con CFi $\in$ [0.4 , 0.8]
  - $CF_{12345678} = 0.99$
  - Aparición tardía de una evidencia moderadamente contradictoria $CF_9 = -0.6$
  - $CF_{123456789} = 0.39$
  - $CF_{192345678} = 0.99$

Factores de Certidumbre

- Segunda aproximación

$$(a)\_CF(h,e1)>0:CF(h,e2)>0$$

$$CF(h,e1\wedge e2)=CF(h,e1)+CF(h,e2)-CF(h,e1)\times CF(h,e2)$$

$$(b)\_CF(h,e1)<0:CF(h,e2)<0$$

$$CF(h,e1\wedge e2)=CF(h,e1)+CF(h,e2)+CF(h,e1)\times CF(h,e2)$$

$$(c)\_CF(h,e1)\times CF(h,e2)<0$$

$$CF(h,e1\wedge e2)=\frac{CF(h,e1)+CF(h,e2)}{1-\min\{|CF(h,e1)|,|CF(h,e2)|\}}$$

Factores de Certidumbre

- Ventajas de la segunda aproximación
  - Las evidencias se pueden considerar en cualquier orden sin que el resultado final se vea afectado
  - Permite modelar los procesos de razonamiento sin tener que almacenar explícitamente los MBs y los MDs

Factores de Certidumbre

- Inconvenientes de la segunda aproximación
  - Si e1 implica lógicamente a e2, entonces CF(h,e1 y e2) debería ser igual a CF(h,e1)
  - De la aplicación de este modelo no se deduce el resultado anterior
  - Problema no resuelto de la combinación de evidencias S-B
  - Posibles soluciones
    - Estructurar muy bien las bases de conocimientos
    - Agrupar en una sola regla cláusulas con evidencias condicionalmente dependientes

Factores de Certidumbre

- Dos tipos de conocimiento inexacto
  - Imprecisión
    - Afecta a entidades, hechos y datos del dominio
  - Incertidumbre
    - Ligada a los procesos de Razonamiento
  - La imprecisión aparece porque…
    - La propia evidencia es imprecisa
    - La evidencia considerada es consecuencia de otra regla, y forma parte de un proceso de razonamiento que supone varias inferencias
  - Imprecisión e Incertidumbre pueden darse aislada o conjuntamente

Factores de Certidumbre

- Ejemplo:
  - Reglas
    - Cuando llueve mucho casi siempre me quedo en casa
    - Cuando me quedo en casa suelo leer
  - Representación del conocimiento
    - E1 = llover mucho
    - H1 = quedarse en casa
    - CF(H1,E1) = 0.95 (Indica incertidumbre en la $\rightarrow$)
    - E2 = quedarse en casa
    - H2 = leer
    - CF(H2,E2) = 0.75 (Indica incertidumbre en la $\rightarrow$)

Factores de Certidumbre

- Imprecisión:
  - En vez de “llover mucho” $\rightarrow$ “llueve bastante”
  - Llueve bastante = 0.75 Llueve mucho (arbitrario)
  - CF(Llueve_mucho , Llueve_bastante) = 0.75

$$e1 \xrightarrow{0.75} E1 \xrightarrow{0.95} H1 = E2 \xrightarrow{0.75} H2$$

Factores de Certidumbre

- Propagación de incertidumbre
  - Si E´ Entonces E con CF(E,E´) = x
  - Si E Entonces H con CF(H,E) = y

$$E' \xrightarrow{x} E \xrightarrow{y} H$$

  - Donde CF(H,E´) = CF(H,E) max {0 , CF(E , E´)}
  - Proceso de arrastre de la inexactitud y propagación

Factores de Certidumbre

- Inconveniente…
  - Si CF (E , E´) < 0 → CH (H , E´) = 0
- Alternativa de Heckerman
  - Dado que CF (¬ E , E´) = - CF (E , E´)
  - Cambiando el circuito inferencial…

$$E' \to \neg E \to H$$

  - CF (H , E´) = CF (H , ¬ E) max {0 , CF (¬ E , E´)}
  - Pero… ¿CF (H , ¬ E)?

Factores de Certidumbre

- Combinación lógica de evidencias
  - Las cláusulas de los antecedentes de las reglas suelen estar relacionadas a través de los correspondientes operadores lógicos ($\wedge$, $\vee$, $\neg$)
  - Conociendo CF (H1 , E) : CF (H2 , E) y sabiendo que E es toda la evidencia del antecedente…¿Cómo obtener los valores correspondientes a las expresiones siguientes?:
    - CF (H1 $\wedge$ H2 , E)
    - CF (H1 $\vee$ H2 , E)

Factores de Certidumbre

- Expresiones propuestas…
  - CF (H1 $\wedge$ H2 , E) = min {CF (H1 , E) , CF (H2 , E)}
  - CF (H1 $\vee$ H2 , E) = max {CF (H1 , E) , CF (H2 , E)}

$$(E2 \vee E3) \wedge E1 \wedge E4 \xrightarrow{CFr} H$$

- Ejemplo: Evaluación del antecedente y resultado final
- CF (H , E) =
  - = CFr x CF (antecedente , E) =
  - = CFr x CF {E1 $\wedge$ (E2 $\vee$ E3) $\wedge$ E4 , E} =
  - = CFr x min {CF(E1,E), max [CF(E2,E) , CF(E3,E)] ,CF(E4,E)}

Factores de Certidumbre

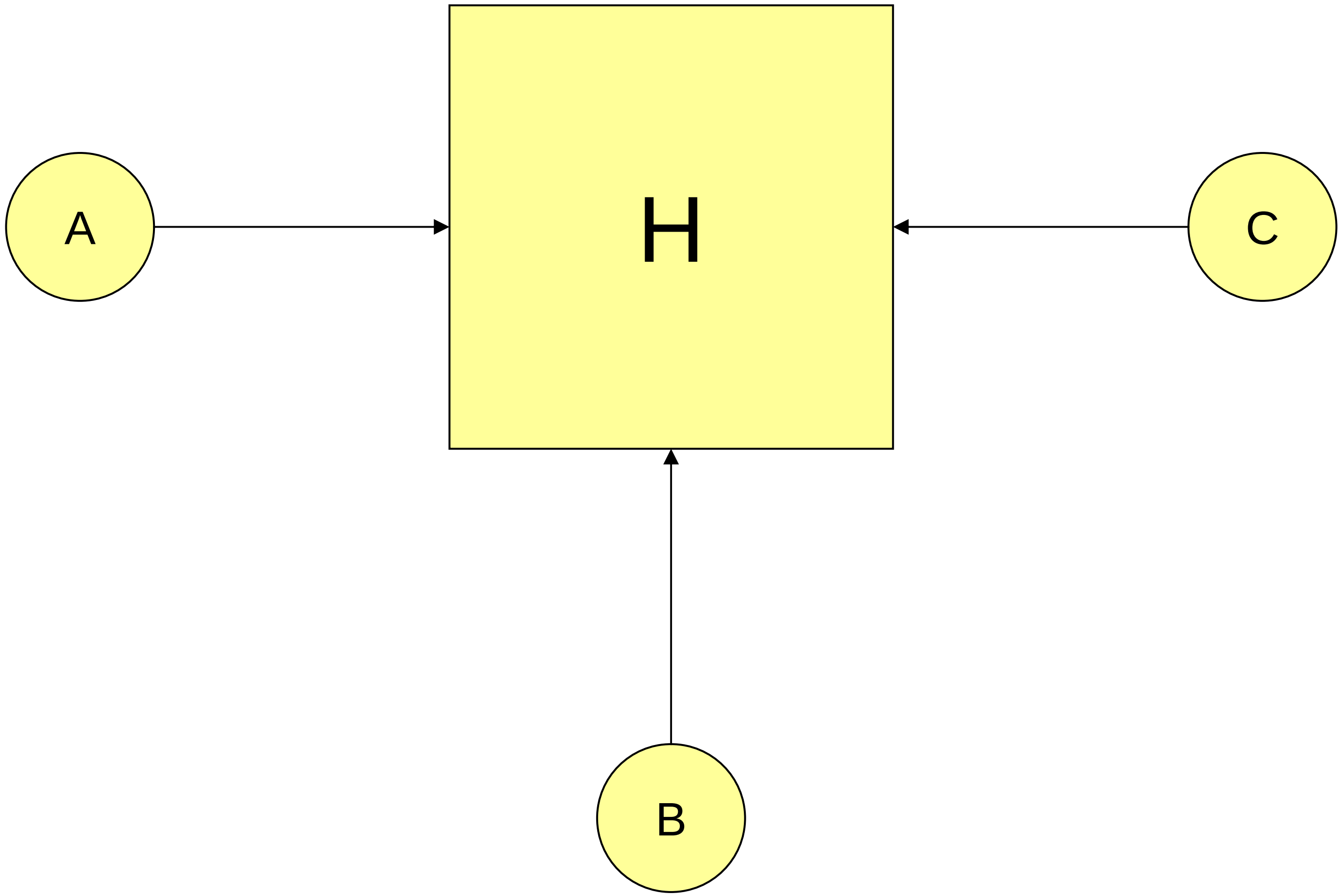
- Cuadro resumen
  - Combinación de evidencias sobre la misma hipótesis
    - + , + : CF(H , E) = CF1 + CF2 – CF1 x CF2
    - - , - : CF(H , E) = CF1 + CF2 + CF1 x CF2
    - + , - : CF(H , E) = [CF1 + CF2] / [1-min{|CF1|,|CF2|}]
  - Propagación de incertidumbre
    - $E' \rightarrow E \rightarrow H$ con CF(H,E) y CF(E,E´)
    - CF(H,E´) = CF(H,E) x max {0,CF(E,E´)}
  - Combinación lógica de evidencias
    - CF(H1 y H2 , E) = min {CF(H1,E) , CF(H2,E)}
    - CF(H1 o H2 , E) = max {CF(H1,E) , CF(H2,E)}

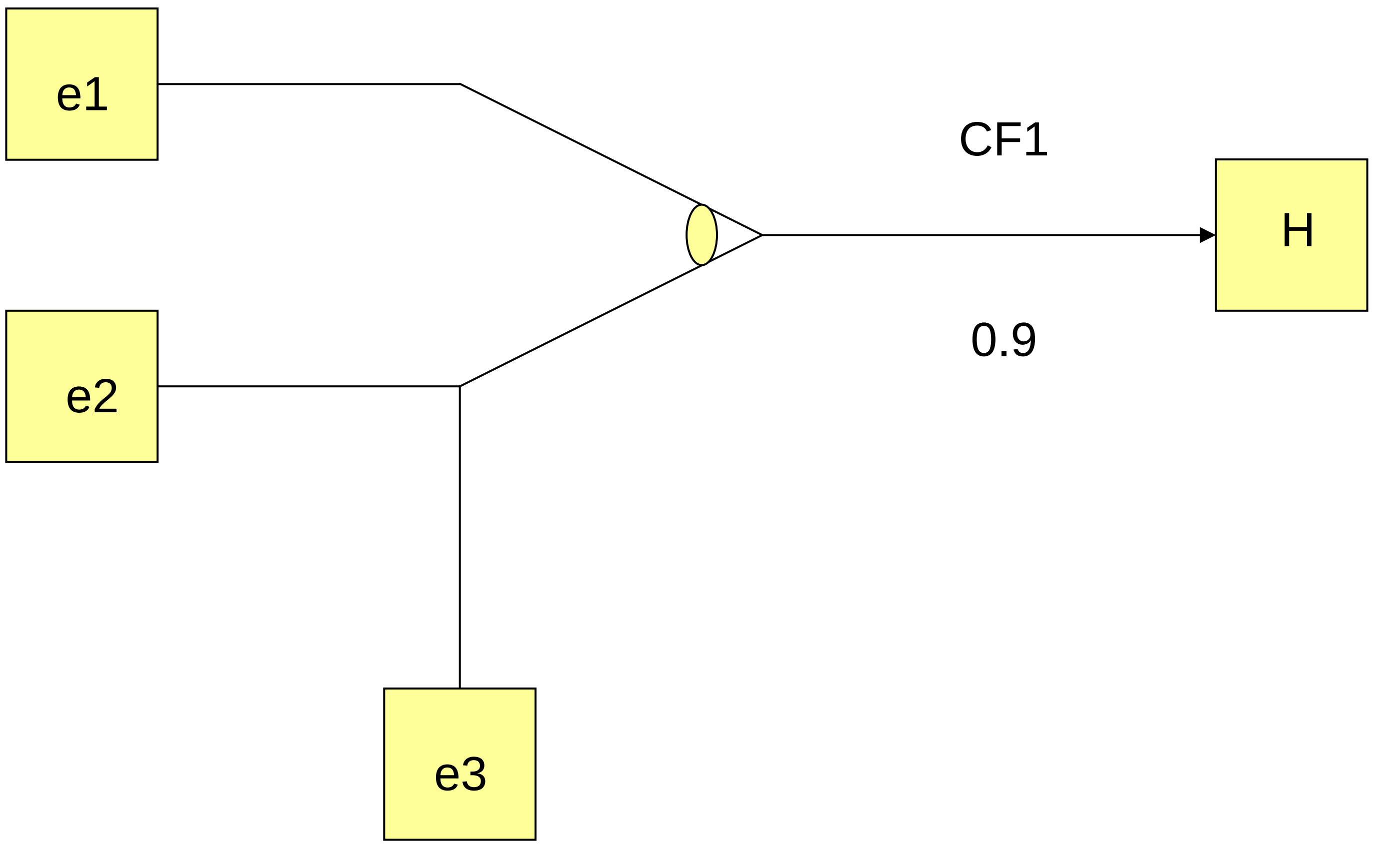
MODELOS CUASI-ESTADÍSTICOS

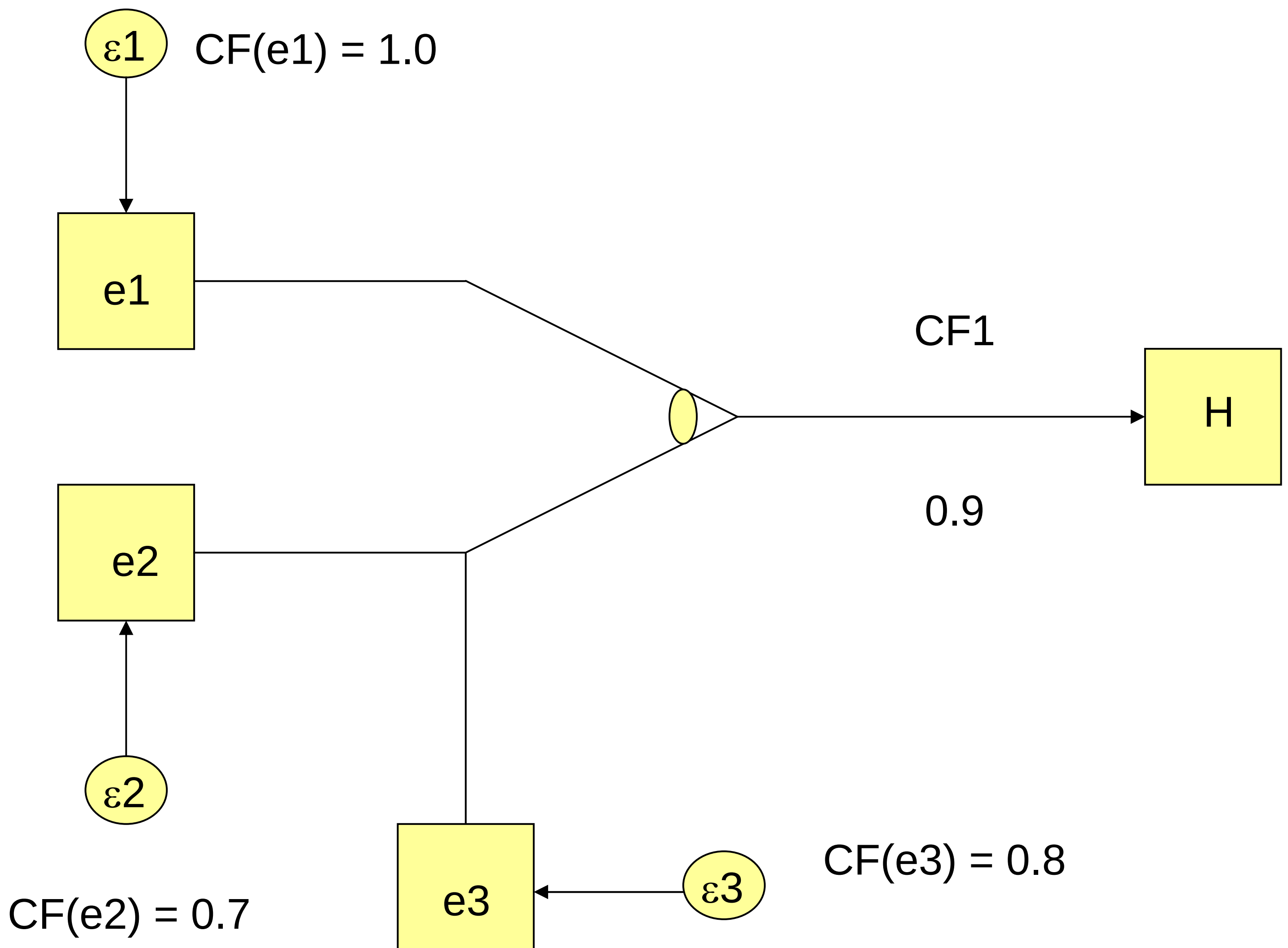
- Ejemplo
  - Durante un proceso regresivo y exhaustivo centrado en la hipótesis H, se activaron las reglas, y se obtuvieron las evidencias, que aparecen a continuación. Evaluar el factor de certidumbre final de la hipótesis H dada toda la evidencia disponible.

    - R1: e1 y (e2 o e3) $\rightarrow$ H CF(H/R1) = 0.9
    - R2: e4 $\rightarrow$ H CF(H/R2) = 0.8
    - R3: e5 y e6 $\rightarrow$ e4 CF(e4/R3) = 0.6
    - R4: e7 $\rightarrow$ H CF(H/R4) = 0.8
    - R5: e8 $\rightarrow$ e7 CF(e7/R5) = - 0.6
    - R6: e9 $\rightarrow$ e7 CF(e7/R6) = 0.8

    - CF(e1) = 1.0 CF(e2) = 0.7 CF(e3) = 0.8 CF(e5) = -1.0
    - CF(e6) = 0.8 CF(e8) = 0.6 CF(e9) = 0.7


A
H
C
B


e1
CF1
H
0.9
e2
e3


ε1
CF(e1) = 1.0
e1
CF1
H
0.9
e2
ε2
CF(e3) = 0.8
ε3
e3
CF(e2) = 0.7

- Rama A

$$Ea = \varepsilon 1, \varepsilon 2, \varepsilon 3$$

$$Antecedente\_A = e1 \wedge (e2 \vee e3)$$

$$CF(H, Ea) = CF(H / R1) \times \max\{0, CF(Antecedente\_A, Ea)\}$$

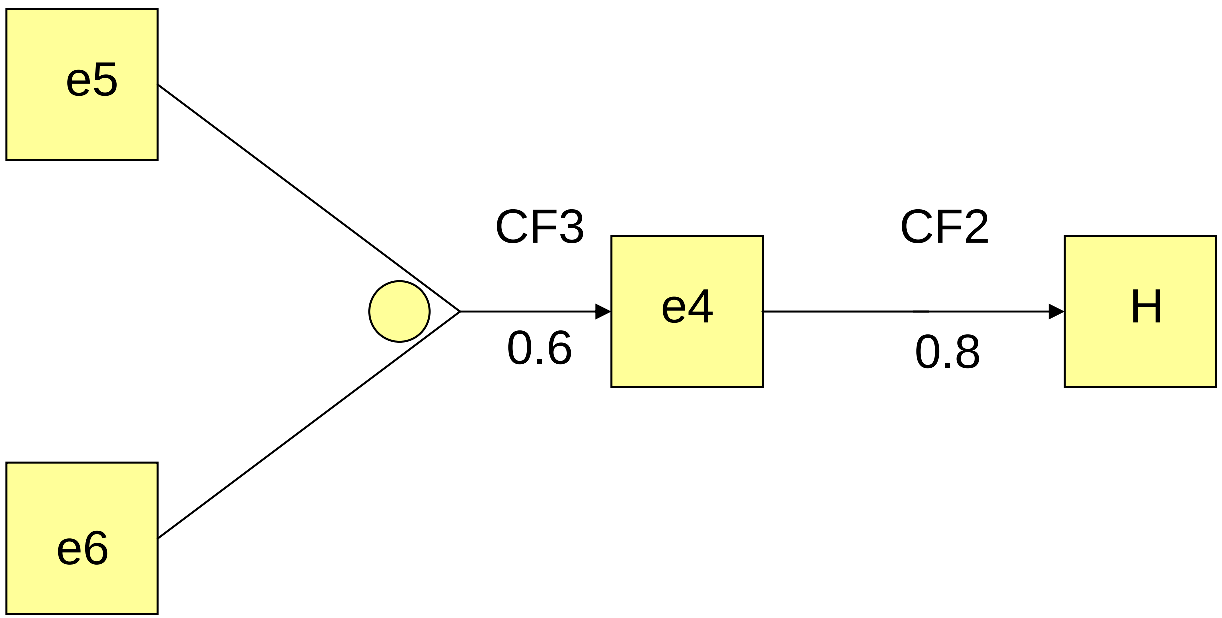
$$CF(Antecedente\_A, Ea) = CF\{e1 \wedge (e2 \vee e3), Ea\} =$$

$$= \min\{CF(e1, \varepsilon 1), \max[CF(e2, \varepsilon 2), CF(e3, \varepsilon 3)]\} =$$

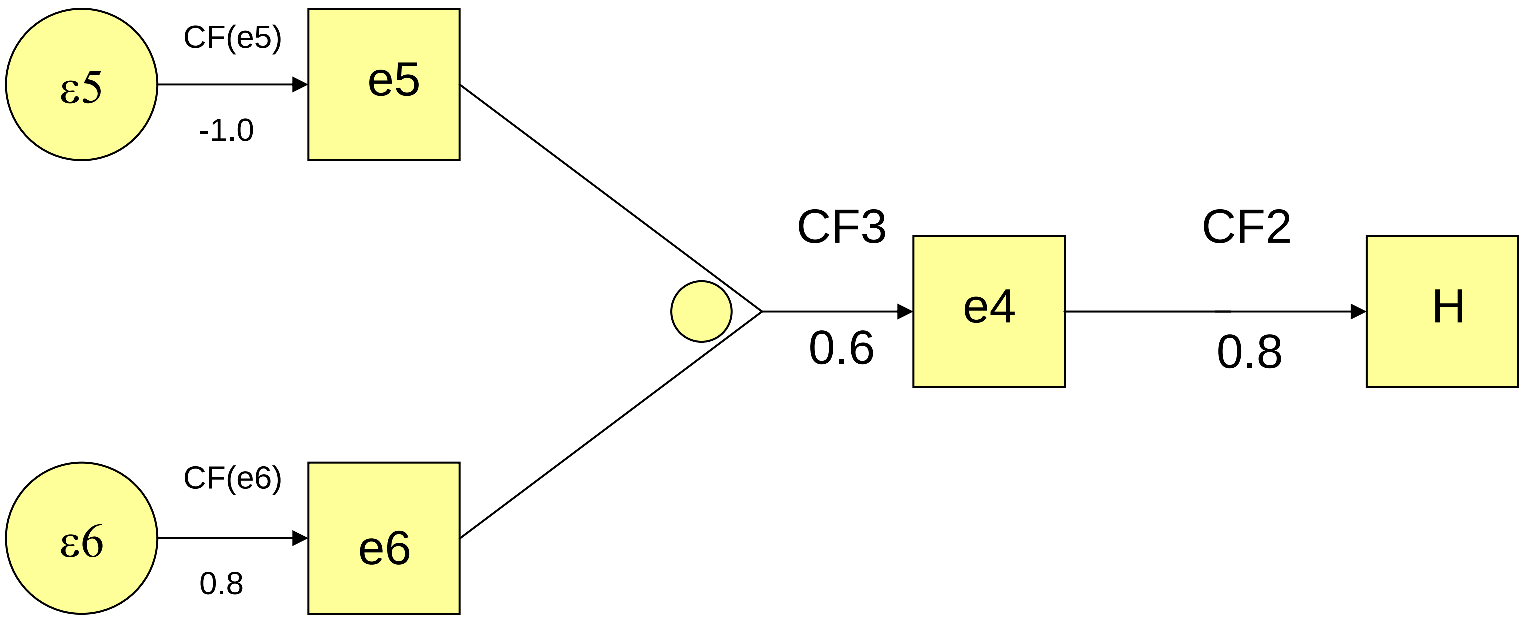
$$= \min\{1, \max(0.7, 0.8)\} = 0.8$$

$$CF(H, Ea) = 0.9 \times \max(0.0, 0.8) = 0.72$$

- Rama B



- Rama B



- Rama B

$$Eb = \varepsilon 5, \varepsilon 6$$

$$Antecedente\_B = e5 \wedge e6$$

$$CF(H, Eb) = CF(H / R2) \times \max\{0, CF(e4, Eb)\}$$

$$CF(e4, Eb) = CF(e4 / R3) \times \max\{0, CF(Antecedente\_B, Eb)\}$$

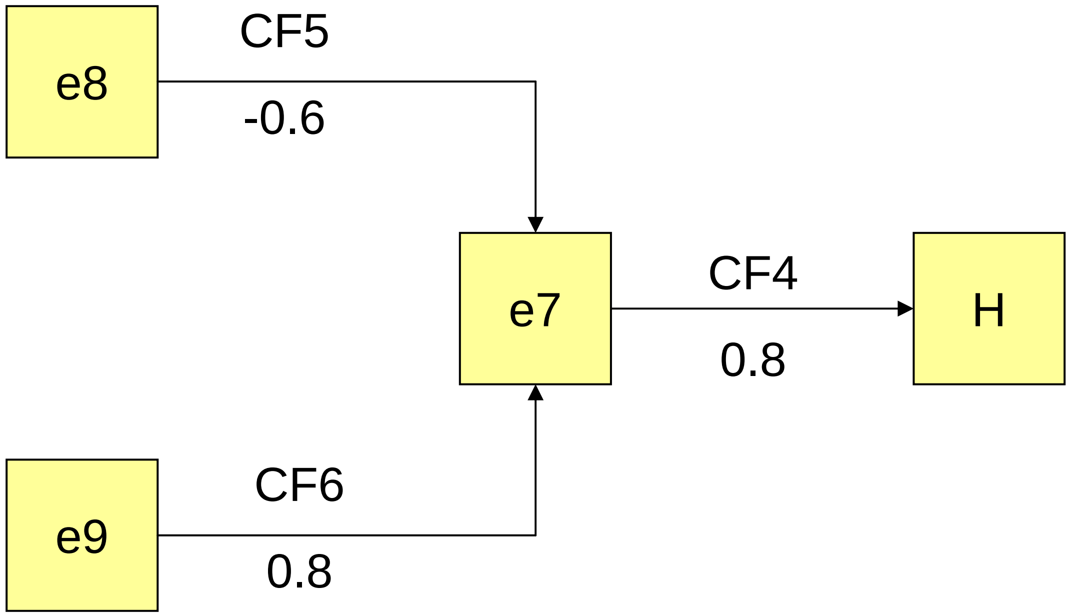
$$CF(Antecedente\_B, Eb) = CF\{e5 \wedge e6, Eb\} =$$

$$= \min\{CF(e5, \varepsilon 5), CF(e6, \varepsilon 6)\} = \min\{-1.0, 0.8)\} = -1.0$$

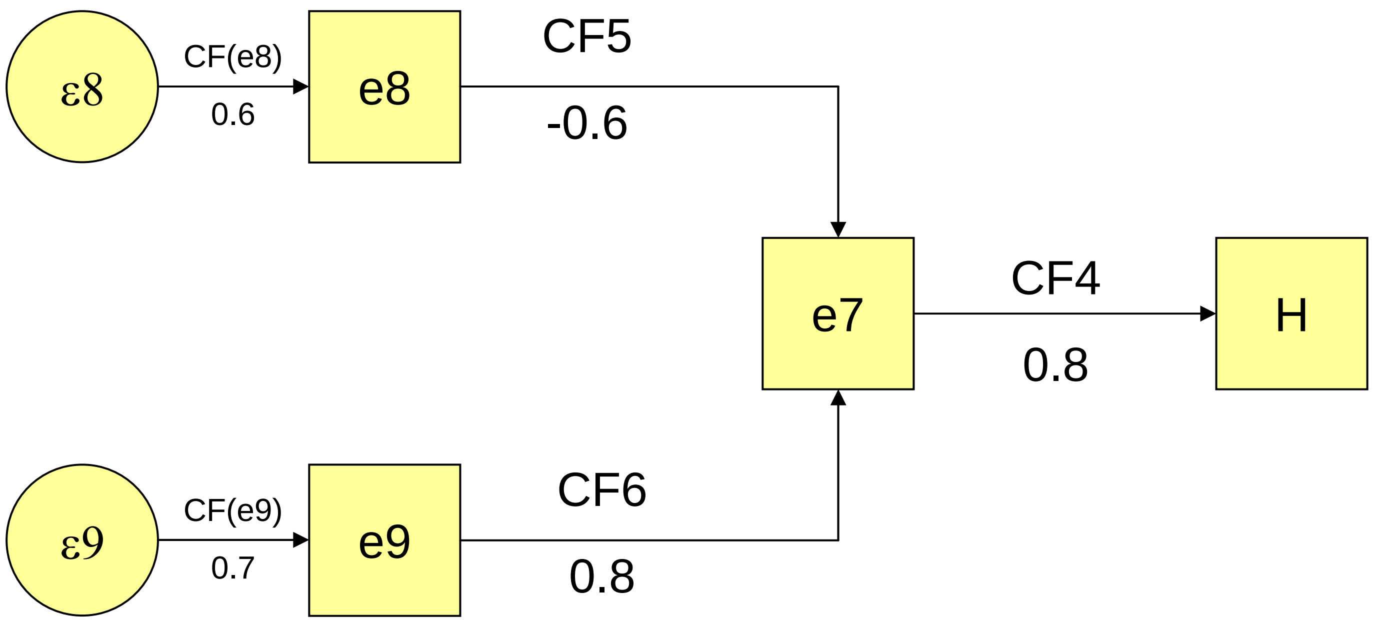
$$CF(e4, Eb) = CF(e4 / R3) \times \max(0.0, -1.0) = 0.0$$

$$CF(H, Eb) = CF(H / R2) \times \max(0.0, 0.0) = 0.0$$

- Rama C



- Rama C



- Rama C

$$Ec = \varepsilon 8, \varepsilon 9$$

$$Antecedente\_C = 2\_reglas$$

$$CF(H, Ec) = CF(H / R4) \times \max\{0, CF(e7, Ec)\}$$

$$CF(e7, Ec) = 2\_reglas \rightarrow Evidencias\_independientes\_sobre\_e7$$

$$Rama\_C1$$

$$CF(e7, \varepsilon 8) = CF(e7 / R5) \times \max\{0, CF(e8, \varepsilon 8)\} =$$

$$= -0.6 \times \max\{0.0, 0.6\} = -0.36$$

$$Rama\_C2$$

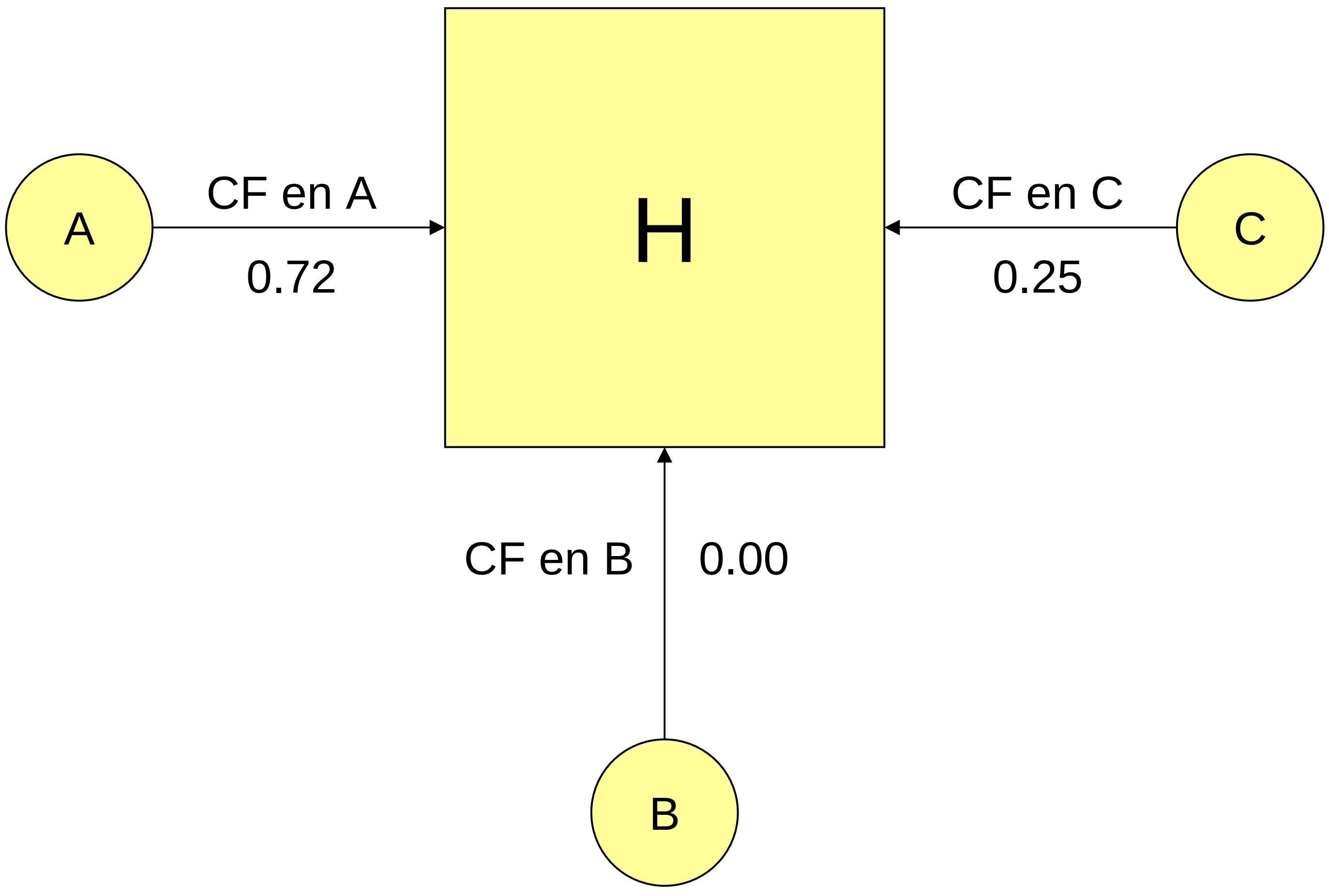
$$CF(e7, \varepsilon 9) = CF(e7 / R6) \times \max\{0, CF(e9, \varepsilon 9)\} =$$

$$= 0.8 \times \max\{0.0, 0.7\} = 0.56$$

- Rama C (continuación)

$$Luego: CF(e7, Ec) = \frac{0.56 - 0.36}{1 - \min\{|0.56|, |-0.36|\}} = \frac{0.20}{0.64}$$

$$CF(H, Ec) = CF(H / R4) \times \max\{0, CF(e7, Ec)\} = 0.8 \times \frac{0.20}{0.64} = 0.25$$

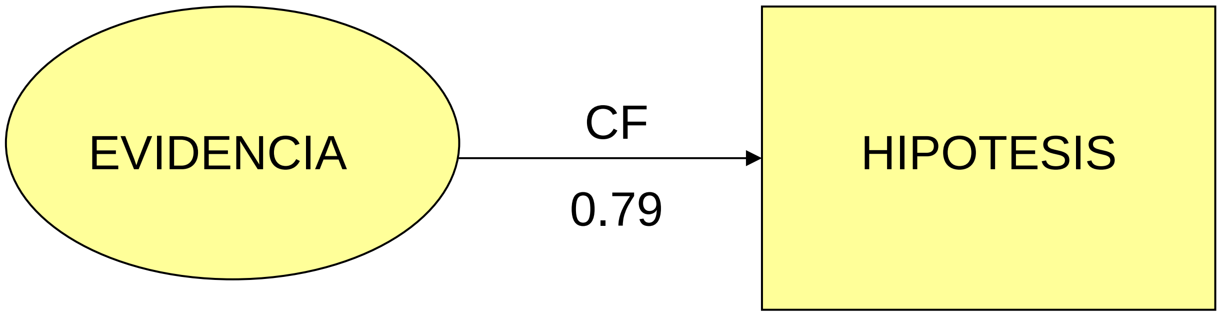

A
CF en A
0.72
H
CF en C
0.25
C
CF en B
0.00
B

$$CF(H, Ea) = 0.72$$

$$CF(H, Eb) = 0.00$$

$$CF(H, Ec) = 0.25$$

$$CF(H, E) = 0.72 + 0.25 - 0.72 \times 0.25 = 0.79$$


EVIDENCIA
CF
0.79
HIPOTESIS

Teoría Evidencial

- La Teoría Evidencial de Dempster y Shafer
  - Tiene una fuerte base teórica
  - Permite modelar de forma sencilla la incertidumbre asociada e evidencias e hipótesis
  - Permite considerar conjuntos de hipótesis sin que la confianza depositada en cada uno de ellos tenga que ser distribuida de ningún modo entre las hipótesis individuales
  - Permite reflejar de forma elegante la falta de conocimiento
  - Contiene a la teoría de la probabilidad como un caso particular
  - Contiene a algunas de las funciones combinatorias de evidencias del modelo de Shortliffe y Buchanan

Teoría Evidencial

- Marco de discernimiento (θ)
  - Conjunto finito de todas las hipótesis que se pueden establecer en el dominio del problema
  - Es un conjunto exhaustivo de hipótesis mutuamente excluyentes
  - El efecto de una evidencia sobre el conjunto global de hipótesis no viene determinado por la contribución de la confianza depositada en las hipótesis individuales

Teoría Evidencial

- Elementos de la teoría

$$\theta = \{h1, h2, \ldots, hn\}$$

$$A \subseteq \theta$$

$$\Gamma\theta = conjunto\_de\_subconjuntos\_de\_\theta$$

$$m = función\_básica\_de\_asignación\_de\_verosimilitud$$

$$e = evidencia$$

$$e : A \subseteq \theta \rightarrow m(A) = s / 0 \leq s \leq 1$$

$$m(\theta) = 1 - m(A) = 1 - s$$

$$m(B) = 0, \forall \subset \theta, B \neq \theta, B \neq A$$

Teoría Evidencial

- A es elemento focal si m(A) $\neq$ 0
- Condiciones para m:

$$\sum_{A \subset \theta} m(A) = 1$$

$$m(\varnothing) = 0$$

Teoría Evidencial

- Planteamiento probabilístico
  - $p(A) = s \rightarrow p(\neg A) = 1\text{-}s$

- Planteamiento evidencial
  - Si: $\theta = \{h1, h2, h3, h4\}$
  - Y: $A = \{h1, h2\}$
  - Con: $e : A \rightarrow m(A) = m(\{h1, h2\}) = s$
  - Entonces: $m(\theta) = m(\{h1, h2, h3, h4\}) = 1\text{-}s$
  - Y: $s \in [0 , 1]$

Teoría Evidencial

- Las evidencias…
  - No suelen aparecer solas
  - No tienen por qué referirse a los mismos elementos focales
    - $\theta = \{h1, h2, h3, h4\}$
    - e1: A1 = (h1, h2) / m1(h1, h2) = x
m1(h1, h2, h3, h4)= 1-x
    - e2: A2 = (h2, h3, h4) / m2(h2, h3, h4) = y
m2(h1, h2, h3, h4) = 1-y
    - e3: A3 = (h1) / m3(h1) = z
m3(h1, h2, h3, h4) = 1-z

Teoría Evidencial

- Combinación de evidencias…
  - Si e1: A1 / m1(A1) = x
  - Si e2: A2 / m2(A2) = y
  - Definimos m12(C) $= m1(A1) \otimes m2(A2) =$

    $= m1(A1) \times m2(A2)$
  - …Donde $C = A1 \cap A2$
  - Generalizando la expresión:

$$m12(C) = \sum_{C = Ai \cap Bj} m1(Ai) \times m2(Bj)$$

Teoría Evidencial

- La formulación coincide con la asignación de probabilidad a la intersección de dos sucesos independientes
- La teoría evidencial asume implícitamente independencia entre evidencias
- Se cumple la primera condición exigida a la función de asignación básica de verosimilitud

$$\sum_{C=Ai\cap Bj} m12(C) = 1$$

Teoría Evidencial

- Si la intersección de elementos focales es nula se introduce una peculiaridad en el modelo
  - e1: A1 / m1(A1) = x $0 < x \leq 1$
  - e2: A2 / m2(A2) = y $0 < y \leq 1$
  - $C = A1 \cap A2 = \varnothing$
  - $M12\ (C) = m12\ (\varnothing) = m1(A1) \times m2(A2) = xy \neq 0$
  - Se viola la segunda condición impuesta a la función de asignación básica de verosimilitud: $m12\ (\varnothing) = 0$

$$\downarrow$$

Normalizar el resultado

Teoría Evidencial

$$Grado\_de\_conflicto$$

$$K = \sum_{Ai \cap Bj = \varnothing} m1(Ai) \times m2(Bj)$$

$$Factor\_de\_normalizaci\acute{o}n$$

$$FN = \frac{1}{1 - K}$$

$$m12(C)_{normalizada} = \frac{\sum_{C = Ai \cap Bj} m1(Ai) \times m2(Bj)}{1 - K}$$

Teoría Evidencial

- Ejemplo
  - $\theta = \{h1, h2, h3, h4\}$
  - e1: $A1 = \{h1, h2\}$, $m1(A1) = 0.6$, $m1(\theta 1) = 0.4$
  - e2: $A2 = \{h2, h3, h4\}$, $m2(A2) = 0.7$, $m2(\theta 2) = 0.3$
  - e3: $A3 = \{h1\}$, $m3(A3) = 0.8$, $m3(\theta 3) = 0.2$
  - $\theta = \theta 1 = \theta 2 = \theta 3 = \{h1, h2, h3, h4\}$

Teoría Evidencial

- e1 y e2

	A1={h1,h2}	**θ1**
A2={h2,h3,h4}	**{h2}** **m12(h2)=0.42**	**{h2,h3,h4}** **m12{h2,h3,h4}=0.28**
θ2	**{h1,h2}** **m12{h1,h2}=0.18**	**θ12** **m12(θ12)=0.12**

Teoría Evidencial

- e1 y e2
  - Se cumplen las condiciones para la función de asignación básica de verosimilitud

$$\sum_{i=1}^{4} m12(Ci) = 1$$

$$m12(h2) + m12(h1, h2) + m12(h2, h3, h4) + m12(\theta 12) =$$

$$= 0.42 + 0.18 + 0.28 + 0.12 = 1.00$$

$$m12(\varnothing) = 0$$

Teoría Evidencial

- (e1 y e2) y e3

	{h2}	{h1,h2}	{h2,h3,h4}	θ12
{h1}	∅ m123(∅) 0.336	{h1} m123(h1) 0.144	∅ m123(∅) 0.224	{h1} m123(h1) 0.096
θ3	{h2} m123{h2} 0.084	{h1,h2} m123(h1,h2) 0.036	{h2,h3,h4} m123(h2,h3,h4) 0.056	θ123 m123(θ123) 0.024

Teoría Evidencial

- Grado de conflicto
  - K = 0.336 + 0.224 = 0.560
- Normalizando y agrupando términos
  - m123 (h1) = 0.545
  - m123 (h2) = 0.191
  - m123 (h1,h2) = 0.082
  - m123 (h2,h3,h4) = 0.127
  - m123 (h1,h2,h3,h4) = 0.055
  - $\sum = 1.000$

Teoría Evidencial

- Credibilidad
  - Indicador de la mínima confianza que podemos depositar en un elemento focal dado

$$Si\_\theta = marco\_de\_discernimiento$$

$$A \subseteq \theta = elemento\_focal$$

$$Cr(A) = \sum_{B \subseteq A} m(B)$$

Teoría Evidencial

- Plausibilidad
  - Indicador de la máxima confianza que podemos depositar en un elemento focal dado

$$Si\_\theta = marco\_de\_discernimiento$$

$$A \subseteq \theta = elemento\_focal$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

Teoría Evidencial

- Intervalo de Confianza
  - Segmento del espacio numérico [0 , 1] que tiene como valor mínimo el valor de la credibilidad del elemento focal, y como valor máximo el correspondiente valor de la plausibilidad
  - Representa la incertidumbre asociada al elemento focal
  - Puede utilizarse como índice dinámico de confianza

$$IC(A) = [Cr(A), Pl(A)]$$

Teoría Evidencial

- Casos:
  - $0 \leq Cr(A) \leq 1$ — $Cr(A) = 0$ y $Pl(A) = 1 \rightarrow$ No sé
  - $0 \leq Pl(A) \leq 1$ — $Cr(A) = 1$ y $Pl(A) = 1 \rightarrow$ Sí
  - $Cr(A) \leq Pl(A)$ — $Cr(A) = 0$ y $Pl(A) = 0 \rightarrow$ No
  - $Cr(A) \leq P(A) \leq Pl(A)$

Teoría Evidencial

- D/S versus S/B
  - Sean dos evidencias independientes que apoyan al mismo elemento focal $A \subseteq \theta$
    - e1 : A / m1(A) = s1 , m1(θ) = 1-s1
    - e2 : A / m2(A) = s2 , m2(θ) = 1-s2
    - m12(A) = s1s2 + s1(1-s2) + s2(1-s1) = s1 + s2 –s1s2
  - En este caso, considerando A como hipótesis en el modelo de Shortliffe y Buchanan estamos ante dos evidencias independientes que apoyan a la misma hipótesis
    - CF(H,E) = s1 + s2 –s1s2

EJEMPLO DEMSTER-SHAFER

Teoría Evidencial

CASO PRÁCTICO

Asesinato Evidencial

A Juan le habían roto el cuello mientras estaba sentado frente a la chimenea leyendo el periódico. Se lo habían roto desde atrás. “…lo agarraron por el mentón y por la frente, y lo desnucaron. Sobre las 23 horas más o menos” –dijo el forense- “hace falta una fuerza considerable para ello” –añadió.

La casa en la que se había cometido el crimen estaba muy aislada, y no había señal alguna de desplazamientos recientes de vehículos. Tampoco había pisadas marcadas en el barro. La noche anterior había llovido mucho. La policía concluyó que el asesino tenía que ser uno de los ocupantes de la mansión…pero ¿quién? Interrogaron a todos con el siguiente resultado:

Asesinato Evidencial

- María: Edad 27 años. 175 cm de estatura. Complexión fuerte. Camarera. Fue la que descubrió al cadáver y llamó a la policía después de alertar a los demás residentes de la mansión. En su testamento, Juan le deja una suma considerable de dinero. María era hija de la antigua camarera que había servido en la casa durante más de 40 años. Juan la quería mucho. María no tiene coartada, y aunque se muestra muy afectada por lo sucedido, fue vista por Alex, sobrino de Juan, entrando en la habitación del crimen sobre las 22:30 horas. María conocía el testamento de Juan.

Asesinato Evidencial

- Blas: Edad 65 años 170 cm de estatura. Enfermo desde hace años de artrosis. Mayordomo de la mansión. A punto de jubilarse. Últimamente discutía mucho con Juan, pero la cosa no pasaba nunca de ahí. Lo cierto es que Juan dejaba a Blas una importante herencia, como premio a los servicios prestados durante tanto tiempo. Tampoco Blas tiene coartada alguna. Dijo que aquella noche le dolían mucho los huesos, y que estaba descansando en su habitación. Blas conocía el testamento de Juan.

Asesinato Evidencial

- Alex: Edad 35 años. 185 cm de estatura. Deportista aficionado. Poco trabajador. Le gusta vivir la vida. Huérfano. Sobrino de Juan, con quien no se llevaba demasiado bien. Las relaciones sin embargo eran civilizadas, aunque Alex pasaba bastante de su tío desde que éste le dijo que le iba a rebajar la asignación mensual mientras no consiguiese un trabajo de verdad. A las 22:30 vio como María entraba en el salón de la chimenea, mientras él se dirigía hacia la biblioteca. A la hora del crimen estaba con Eva, su novia, en la biblioteca, manteniendo una videoconferencia con unos colegas igual de trabajadores que él. Este hecho fue comprobado por la policía. Precisamente para motivar laboralmente a Alex, Juan le había ocultado el contenido de su testamento que, no obstante, era muy generoso con su sobrino.

Asesinato Evidencial

- Eva: Edad 33 años. 172 cm de estatura. Complexión fuerte y aspecto atlético. Aparentemente aprecia mucho a Alex. Tiene una gran fortuna personal. Trabajadora infatigable, administra ella misma sus negocios. Siempre le decía a Alex que su tío tenía razón, que tenía que ponerse a trabajar. A la hora del crimen estaba con Alex, su novio, en la biblioteca, manteniendo una videoconferencia con unos colegas de Alex. Este hecho fue comprobado por la policía.

Asesinato Evidencial

- El inspector Castiñeiro ordenó toda la información del siguiente modo:
- - Sospechosos: María, Blas, Alex, Eva
- - Circunstancias: Fuerza Física, Codicia, Despecho, Coartada, Ubicación.

- El inspector Castiñeiro era hombre de gran experiencia y sumamente meticuloso, y sabía que no todas las circunstancias tienen el mismo peso a la hora de resolver un caso. Así que estableció una escala de pesos en las posibles pruebas:

Asesinato Evidencial

- Sin Evidencias: 0.00 (no afectan al caso en cuestión)
- Evidencias Circunstanciales: 0.25 (afectan colateralmente al caso)
- Evidencias Incriminatorias: 0.50 (tienen importancia para resolver el caso)
- Evidencias Inculpatorias: 0.75 (son muy importantes para resolver el caso)
- Evidencias Definitivas: 1.00 (son condición imprescindible para el crimen)

Asesinato Evidencial

- A continuación, el inspector Castiñeiro estudió su problema y decidió que, en base a su gran experiencia en la resolución de casos similares…:
- La Fuerza Física es una “Evidencia Definitiva”
- La Codicia es una “Evidencia Inculpatoria”
- El Despecho es una “Evidencia Incriminatoria”
- La Ausencia de Coartada es una “Evidencia Incriminatoria”
- La Ubicación de un Sospechoso es una “Evidencia Circunstancial”

Asesinato Evidencial

- Seguidamente aplicó su intuición al informe policial elaborado por el Oleiros Yard y estableció que:
- Fuerza Física (María, Alex, Eva) = 1.00
- Codicia (María, Blas) = 0.75
- Despecho (Alex) = 0.50
- Ausencia de Coartada (María, Blas) = 0.50
- Ubicación del Sospechoso (María) = 0.25
- ¿A qué conclusiones llegó el inspector Castiñeiro tras aplicar la teoría evidencial de Dempster y Shafer en la resolución de este interesante –aunque ficticio- problema policíaco?

Asesinato Evidencial

- $\Theta = \{M, A, E, B\}$
- Evidencias:
  - E1 = Fuerza
  - E2 = Codicia
  - E3 = Despecho
  - E4 = Coartada
  - E5 = Ubicación

Asesinato Evidencial

- E1
  - M1 {M, A, E} = 1.00
  - M1 {Θ1} = 0.00
- E2
  - M2 {M, B} = 0.75
  - M2 {Θ2} = 0.25
- E3
  - M3 {A} = 0.50
  - M3 {Θ3} = 0.50
- E4
  - M4 {M, B} = 0.50
  - M4 {Θ4} = 0.50
- E5
  - M5 {M} = 0.25
  - M5 {Θ5} = 0.75

Asesinato Evidencial

E2 Y E4	MB/0.75	Θ2/0.25
MB/0.50	MB/0.375	MB/0.125
Θ4/0.50	MB/0.375	Θ24/0.125

- M24 {M, B} = 0.875
- M24 {Θ24} = 0.125
- Cr2 {M} = Cr24 {M} = 0.00
- Pl2 {M} = Pl24 {M} = 1.00

Asesinato Evidencial

E24 Y E5	MB/0.875	Θ24/0.125
M/0.25	M/0.21875	M/0.03125
Θ5/0.75	MB/0.65625	Θ245/0.09375

- M245 {M} = 0.250
- M245 {M, B} = 0.656
- M245 {Θ245} = 0.094
- Cr245 {M} = 0.250
- Pl245 {M} = 1.000

Asesinato Evidencial

E245 Y E1	M/0.250	MB/0.656	Θ245/0.094
MAE/1.00	M/0.2500	M/0.6560	MAE/0.0940
Θ1/0.00	M/0.0000	MB/0.0000	Θ2451/0.0000

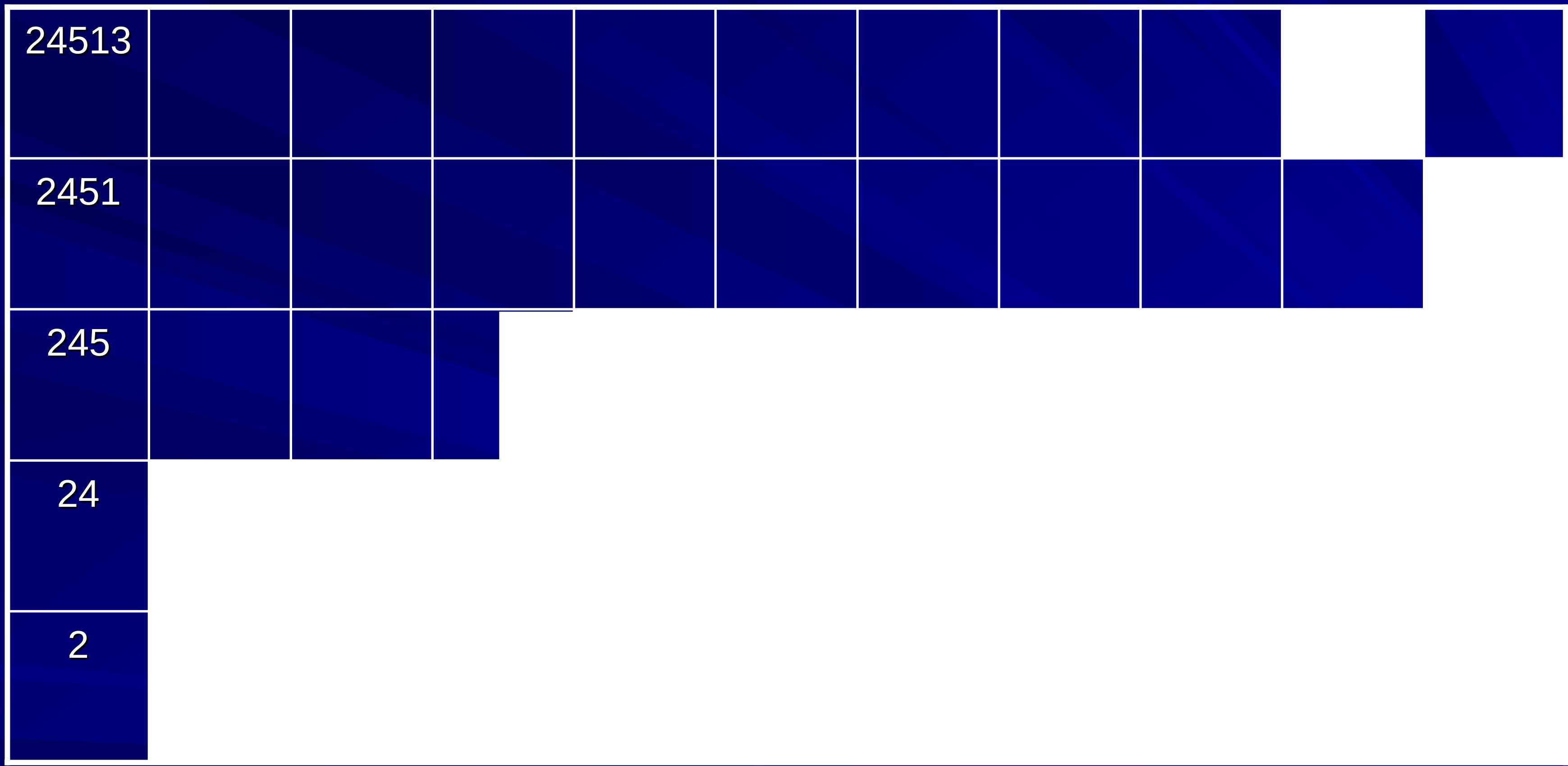
- M2451 {M} = 0.9060
- M2451 {M, A, E} = 0.0940
- Cr2451 {M} = 0.9060
- Pl2451 {M} = 1.0000

Asesinato Evidencial

E2451 Y E3	M/0.9060	MAE/0.0940
A/0.50	∅/0.45300	A/0.04700
Θ3/0.50	M/0.45300	MAE/0.04700

- K = 0.45300
- 1 – K = 0.54700
- M24513 {A} = 0.08592 → 0.086
- M24513 {M} = 0.82815 → 0.828
- M24513 {M, A, E} = 0.08592 → 0.086
- Cr24513 {M} = 0.828
- Pl24513 {M} = 0.914

Asesinato Evidencial (María)



UN CASO PRACTICO DE INGENIERIA DEL CONOCIMIENTO

VICENTE MORET BONILLO, PhD

SENIOR MEMBER, IEEE

	A0	A1	A2	A3
a1	0	0	1	1
a2	0	1	0	1

	A0	A1	A2	A3
a1	0	1	0	1
a2	0	0	1	1

	A0	A1	A2	A3	A4	A5	A6	A7
a1	0	0	0	0	1	1	1	1
a2	0	0	1	1	0	0	1	1
a3	0	1	0	1	0	1	0	1

	X0	X1	X2	X3
x1	0	0	1	1
x2	0	1	0	1

- S = {A0, A1, A2, A3}
- D = {X0, X1, X2, X3}

BLE = S × D = {A0X0, A0X1, A0X2, A0X3, A1X0, A1X1, A1X2, A1X3, A2X0, A2X1, A2X2, A2X3, A3X0, A3X1, A3X2, A3X3}

	A0	A1	A2	A3	A0	A1	A2	A3	A0	A1	A2	A3	A0	A1	A2	A3
a1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
a2	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
x1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
x2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	X0				X1				X2				X3			

- R1: Si el paciente presenta algún síntoma, entonces debe de haber algún diagnóstico asociado. (Nótese que esta regla no impide que haya problemas asintomáticos, tan solo establece que si hay síntomas es porque hay problemas)

- R2: Si el paciente tiene la patología x2, entonces tiene que presentar el síntoma a1.

- R3: Si el paciente tiene la patología x1, pero no tiene la patología x2, entonces tiene que presentar el síntoma a2.

- R4: Si el paciente no tiene la patología x1, pero tiene la patología x2, entonces no tiene que presentar el síntoma a2.

- R1: $(a1 = 1)$ OR $(a2 = 1) \rightarrow (x1 = 1)$ OR $(x2 = 1)$
- R2: $(x2 = 1) \rightarrow (a1 = 1)$
- R3: $(x1 = 1)$ AND $(x2 = 0) \rightarrow (a2 = 1)$
- R4: $(x1 = 0)$ AND $(x2 = 1) \rightarrow (a2 = 0)$

A1X0, A2X0, A3X0, A0X1, A1X1, A0X3, A1X3, A0X2, A2X2, A3X1

BLR = {A0X0, A1X2, A2X1, A2X3, A3X2, A3X3}

$$\frac{P(X1/A2)}{P(X3/A2)}$$

- $A0 = [0, 0]$
- $A1 = [0, 1]$
- $A2 = [1, 0]$
- $A3 = [1, 1]$

- $X0 = [0, 0]$
- $X1 = [0, 1]$
- $X2 = [1, 0]$
- $X3 = [1, 1]$

- ANTECEDENTE $\rightarrow$ CONSECUENTE $\equiv$ NOT ANTECEDENTE OR CONSECUENTE

- R1: $(a1 = 1)$ OR $(a2 = 1) \rightarrow (x1 = 1)$ OR $(x2 = 1) \equiv (a1$ OR $a2) \rightarrow (x1$ OR $x2)$
- R2: $(x2 = 1) \rightarrow (a1 = 1) \equiv x2 \rightarrow a1$
- R3: $(x1 = 1)$ AND $(x2 = 0) \rightarrow (a2 = 1) \equiv (x1$ AND NOT $x2) \rightarrow a2$
- R4: $(x1 = 0)$ AND $(x2 = 1) \rightarrow (a2 = 0) \equiv$ (NOT $x1$ AND $x2) \rightarrow$ NOT $a2$

- R1: [NOT a1 AND NOT a2] OR x1 OR x2
- R2: a1 OR NOT x2
- R3: a2 OR NOT x1 OR x2
- R4: NOT a2 OR x1 OR NOT x2

- R1 es compatible con los vectores: [0 0, – –] , [– –, 1 –] , [– –, – 1]
- R2 es compatible con los vectores: [1 –, – –] , [– –, – 0]
- R3 es compatible con los vectores: [– 1, – –] , [– –, 0 –] , [– –, – 1]
- R4 es compatible con los vectores: [– 0, – –] , [– –, 1 –] , [– –, – 0]

VECTORES	R1	R2	R3	R4	CONSISTENCIA
[0 0, 0 0]	SI	SI	SI	SI	SI
[0 1, 0 0]	NO	SI	SI	SI	NO
[1 0, 0 0]	NO	SI	SI	SI	NO
[1 1, 0 0]	NO	SI	SI	SI	NO
[0 0, 0 1]	SI	NO	SI	SI	NO
[0 1, 0 1]	SI	NO	SI	NO	NO
[1 0, 0 1]	SI	SI	SI	SI	SI
[1 1, 0 1]	SI	SI	SI	NO	NO
[0 0, 1 0]	SI	SI	NO	SI	NO
[0 1, 1 0]	SI	SI	SI	SI	SI
[1 0, 1 0]	SI	SI	NO	SI	NO
[1 1, 1 0]	SI	SI	SI	SI	SI
[0 0, 1 1]	SI	NO	SI	SI	NO
[0 1, 1 1]	SI	NO	SI	SI	NO
[1 0, 1 1]	SI	SI	SI	SI	SI
[1 1, 1 1]	SI	SI	SI	SI	SI

- $[0\ 0,\ 0\ 0] \equiv A0X0$
- $[1\ 0,\ 0\ 1] \equiv A2X1$
- $[0\ 1,\ 1\ 0] \equiv A1X2$
- $[1\ 1,\ 1\ 0] \equiv A3X2$
- $[1\ 0,\ 1\ 1] \equiv A2X3$
- $[1\ 1,\ 1\ 1] \equiv A3X3$

- R1:	IF	$S = [0\ 0]$	THEN	$D = [0\ 0]$		
- R2:	IF	$S = [0\ 1]$	THEN	$D = [1\ 0]$		
- R3:	IF	$S = [1\ 0]$	THEN	$D = [0\ 1]$	OR	$D = [1\ 1]$
- R4:	IF	$S = [1\ 1]$	THEN	$D = [1\ 0]$	OR	$D = [1\ 1]$

- CASO 1: El paciente no presenta el síntoma 1, pero presenta el síntoma 2
    - Vector de Síntomas Asociado: S = [0 1]
    - Regla Activada y Ejecutada: R2
    - Vector de Diagnósticos Asociado: D = [1 0]
    - Conclusión: El paciente presenta la patología 1 y podemos descartar la patología 2

- CASO 2: El paciente presenta el síntoma 1, pero no presenta el síntoma 2
    - Vector de Síntomas Asociado: S = [1 0]
    - Regla Activada y Ejecutada: R3
    - Vectores Diagnósticos Asociados: D = [0 1] D = [1 1]
    - Conclusión: Podemos establecer con total certeza la patología 2, pero la evidencia disponible es contradictoria con respecto a la patología 1, por lo que no podemos ni confirmarla ni descartarla.

- Sea Pos(AiXj) la posibilidad de una asociación AiXj
    - $Pos(AiXj) = 0 \leftrightarrow AiXj \notin BLR$
    - $Pos(AiXj) = 1 \leftrightarrow AiXj \in BLR$

Por otra parte:

- Sea P(AiXj) la probabilidad de una asociación AiXj
    - $P(AiXj) = 0 \leftrightarrow AiXj \notin BLR$
    - $0 < P(AiXj) \leq 1 \leftrightarrow AiXj \in BLR$

$$A_i \xrightarrow{P(X_j/A_i)} X_j$$

Las probabilidades condicionales se construyen a partir de las probabilidades conjuntas y de las prevalencias (o probabilidades totales) del primer suceso. Así:

$$P(X_j/A_i) = \frac{P(A_i X_j)}{P(A_i)}$$

$$A_1 \xrightarrow{P(X_2/A_1)=\frac{P(A_1\,X_2)}{p(A_1)}} X_2$$

$$A_2 \xrightarrow{P(X_1/A_2)} X_1 \qquad P(X_1/A_2) = \frac{P(A_2X_1)}{P(A_2)}$$

$$A_2 \xrightarrow{P(X_3/A_2)} X_3 \qquad P(X_3/A_2) = \frac{P(A_2X_3)}{P(A_2)}$$

$$\frac{P(X_1\,/A_2)}{P(X_3\,/\,A_2)} = \frac{P(A_2X_1)}{P(A_2X_3)}$$

S ↓ / D →		X0 [00]	X1 [01]	X2 [10]	X3 [11]	TOTAL
A0	[00]	512				512
A1	[01]			75		75
A2	[10]		23		37	60
A3	[11]			5	60	65
TOTAL		512	23	80	97	712

S↓ / D→		X0	X1	X2	X3	Prevalencias
		[00]	[01]	[10]	[11]	
A0	[00]	0.719				0.719
A1	[01]			0.105		0.105
A2	[10]		0.032		0.052	0.084
A3	[11]			0.007	0.084	0.091
Prevalencias		0.719	0.032	0.112	0.136	0.999

$$P(X_0/A_0) = \frac{P(A_0X_0)}{P(A_0)} = \frac{0.719}{0.719} = 1.000$$

$$P(X_1/A_2) = \frac{P(A_2X_1)}{P(A_2)} = \frac{0.032}{0.084} = 0.381$$

$$P(X_2/A_1) = \frac{P(A_1X_2)}{P(A_1)} = \frac{0.105}{0.105} = 1.000$$

$$P(X_2/A_3) = \frac{P(A_3X_2)}{P(A_3)} = \frac{0.007}{0.091} = 0.077$$

$$P(X_3/A_2) = \frac{P(A_2X_3)}{P(A_2)} = \frac{0.052}{0.084} = 0.619$$

$$P(X_3/A_3) = \frac{P(A_3X_3)}{P(A_3)} = \frac{0.084}{0.091} = 0.923$$

$$A_0 \xrightarrow{1.000} X_0 \Big\} : \sum = 1.000$$

$$A_1 \xrightarrow{1.000} X_2 \Big\} : \sum = 1.000$$

$$\left.\begin{array}{l} A_2 \xrightarrow{0.381} X_1 \\ A_2 \xrightarrow{0.619} X_3 \end{array}\right\} : \sum = 1.000$$

$$\left.\begin{array}{l} A_3 \xrightarrow{0.077} X_2 \\ A_3 \xrightarrow{0.923} X_3 \end{array}\right\} : \sum = 1.000$$

- R1: IF $S = [00]$ THEN $D = [00]$ WITH $P(D/S) = 1.000$
- R2: IF $S = [01]$ THEN $D = [10]$ WITH $P(D/S) = 1.000$
- R3: IF $S = [10]$ THEN $D = [01]$ WITH $P(D/S) = 0.381$
- R4: IF $S = [10]$ THEN $D = [11]$ WITH $P(D/S) = 0.619$
- R5: IF $S = [11]$ THEN $D = [10]$ WITH $P(D/S) = 0.077$
- R6: IF $S = [11]$ THEN $D = [11]$ WITH $P(D/S) = 0.923$

- CASO 1:
    - Síntomas: S = [01]
    - Reglas Activadas y Ejecutadas: R2
    - Diagnósticos Asociados: D = [10]
    - Conclusión: Los síntomas confirman la patología 1 y descartan la patología 2 sin ninguna incertidumbre.

- CASO 2:
    - Síntomas: S = [10]
    - Reglas Activadas y Ejecutadas: R3 y R4
    - Diagnósticos Asociados: D = [01] (P=0.381) y D = [11] (P=0.619)
    - Conclusión: Los síntomas confirman la patología 2 y la presencia de la patología 1 es 1.6 veces más probable que su ausencia.

$$P(A_0/X_0) = \frac{P(A_0X_0)}{P(X_0)} = \frac{0.719}{0.719} = 1.000$$

$$P(A_1/X_2) = \frac{P(A_1X_2)}{P(X_2)} = \frac{0.105}{0.112} = 0.938$$

$$P(A_2/X_1) = \frac{P(A_2X_1)}{P(X_1)} = \frac{0.032}{0.032} = 1.000$$

$$P(A_2/X_3) = \frac{P(A_2X_3)}{P(X_3)} = \frac{0.052}{0.136} = 0.382$$

$$P(A_3/X_2) = \frac{P(A_3X_2)}{P(X_2)} = \frac{0.007}{0.112} = 0.062$$

$$P(A_3/X_3) = \frac{P(A_3X_3)}{P(X_3)} = \frac{0.084}{0.136} = 0.618$$

$$\left. X_0 \xrightarrow{1.000} A_0 \right\} : \sum = 1.000$$

$$\left. X_1 \xrightarrow{1.000} A_2 \right\} : \sum = 1.000$$

$$\left. \begin{array}{l} X_2 \xrightarrow{0.938} A_1 \\ X_2 \xrightarrow{0.062} A_3 \end{array} \right\} : \sum = 1.000$$

$$\left. \begin{array}{l} X_3 \xrightarrow{0.382} A_2 \\ X_3 \xrightarrow{0.618} A_3 \end{array} \right\} : \sum = 1.000$$

- R1: IF $D = [00]$ THEN $S = [00]$ WITH $P(S/D) = 1.000$
- R2: IF $D = [01]$ THEN $S = [10]$ WITH $P(S/D) = 1.000$
- R3: IF $D = [10]$ THEN $S = [01]$ WITH $P(S/D) = 0.938$
- R4: IF $D = [10]$ THEN $S = [11]$ WITH $P(S/D) = 0.062$
- R5: IF $D = [11]$ THEN $S = [10]$ WITH $P(S/D) = 0.382$
- R6: IF $D = [11]$ THEN $S = [11]$ WITH $P(S/D) = 0.618$

S↓ / D→		X0	X1	X2	X3	TOTAL
		[00]	[01]	[10]	[11]	
A0	[00]	512		15		527
A1	[01]			75		75
A2	[10]		23		37	60
A3	[11]			5	60	65
TOTAL		512	23	95	97	727

S↓ / D→		X0	X1	X2	X3	Prevalencias
		[00]	[01]	[10]	[11]	
A0	[00]	0.704		0.021		0.725
A1	[01]			0.103		0.103
A2	[10]		0.032		0.051	0.083
A3	[11]			0.007	0.083	0.090
Prevalencias		0.704	0.032	0.131	0.134	1.001

1. $A0 \rightarrow X0$: $P(X0 / A0) = P(A0X0) / P(A0) = 0.968$
2. $A0 \rightarrow X2$: $P(X2 / A0) = P(A0X2) / P(A0) = 0.029$
3. $X0 \rightarrow A0$: $P(A0 / X0) = P(A0X0) / P(X0) = 1.000$
4. $X2 \rightarrow A0$: $P(A0 / X2) = P(A0X2) / P(X2) = 0.160$

1. Disponer de un número estadísticamente significativo de casos de uso.
2. Identificar síntomas individuales y diagnósticos individuales.
3. Generar los vectores de las correspondientes asociaciones de síntomas y de diagnósticos, utilizando el criterio binario de numeración.
4. Construir con los casos de uso la tabla de contingencia correspondiente.
5. Calcular las prevalencias y las probabilidades conjuntas.
6. Calcular las probabilidades condicionales de las relaciones causales. Este paso se puede realizar desde los síntomas hacia los diagnósticos, desde los diagnósticos hacia los síntomas, o en ambos sentidos.
7. Verificar la consistencia de los cálculos.
8. Construir las reglas correspondientes.

$A0 = [0000]$	$A1 = [0001]$	$A2 = [0010]$	$A3 = [0011]$
$A4 = [0100]$	$A5 = [0101]$	$A6 = [0110]$	$A7 = [0111]$
$A8 = [1000]$	$A9 = [1001]$	$A10 = [1010]$	$A11 = [1011]$
$A12 = [1100]$	$A13 = [1101]$	$A14 = [1110]$	$A15 = [1111]$

X0	= [00000]	X1	= [00001]	X2	= [00010]	X3	= [00011]
X4	= [00100]	X5	= [00101]	X6	= [00110]	X7	= [00111]
X8	= [01000]	X9	= [01001]	X10	= [01010]	X11	= [01011]
X12	= [01100]	X13	= [01101]	X14	= [01110]	X15	= [01111]
X16	= [10000]	X17	= [10001]	X18	= [10010]	X19	= [10011]
X20	= [10100]	X21	= [10101]	X22	= [10110]	X23	= [10111]
X24	= [11000]	X25	= [11001]	X26	= [11010]	X27	= [11011]
X28	= [11100]	X29	= [11101]	X30	= [11110]	X31	= [11111]

- IF $S = [0101]$ THEN $D = [10100]$ WITH $P(D/S) = a$ $(0 < a \leq 1)$

  - A5 $\rightarrow$ $[0101]_{\text{Base 2}}$ $=$ $[5]_{\text{Base 10}}$
  - X20 $\rightarrow$ $[10100]_{\text{Base 2}}$ $=$ $[20]_{\text{Base 10}}$

- IF $S = [5]$ THEN $D = [20]$ WITH $P(D/S) = a$ $(0 < a \leq 1)$

- IF $S = [20251]$ THEN $D = [4578]$ WITH $P(D/S) = a$ $(0 < a \leq 1)$

POSICION	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01
VALOR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	0	0	1	1	0	1	1

POSICION	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01
VALOR	0	0	1	0	0	0	1	1	1	1	0	0	0	1	0

La interpretación, de acuerdo con lo establecido en la regla anterior, es que –en este caso concreto- el paciente presenta los síntomas [1, 2, 4, 5, 9, 10, 11, 12, 15] a los que pueden asociarse las patologías [2, 6, 7, 8, 9, 13] con una probabilidad determinada – $P(D/S) = a$, $(0 < a \leq 1)$-.

1. $A \xrightarrow{P(B/A)=0} B$
2. $A \xrightarrow{P(B/A)=1} B$
3. $A \xrightarrow{P(B/A)=x \in (0,1)} B$

1. Tener distinta relevancia (es decir, ser unos más importantes que otros)
2. Tener distinto grado de severidad (es decir, desviarse unos más que otros de los valores de referencia considerados como normales)
3. Tener distinta relevancia y distinto grado de severidad

			SEVERIDAD				
			NULA	LEVE	MODERADO	SEVERO	MUY SEVERO
			0,00	0,25	0,50	0,75	1,00
RELEVANCIA	NULA	0,00	0,0000	0,0000	0,0000	0,0000	0,0000
	POCO	0,25	0,0000	0,0625	0,1250	0,1875	0,2500
	SIGNIFICATIVO	0,50	0,0000	0,1250	0,2500	0,3750	0,5000
	MUCHO	0,75	0,0000	0,1875	0,3750	0,5625	0,7500
	CRITICO	1,00	0,0000	0,2500	0,5000	0,7500	1,0000

$$R1: S = [00] \xrightarrow{1.000} D = [00]$$

$$R2: S = [01] \xrightarrow{1.000} D = [10]$$

$$R3: S = [10] \xrightarrow{0.381} D = [01]$$

$$R4: S = [10] \xrightarrow{0.619} D = [11]$$

$$R5: S = [11] \xrightarrow{0.077} D = [10]$$

$$R6: S = [11] \xrightarrow{0.923} D = [11]$$

$$E(S) \xrightarrow{0.7500} S = [11] \xrightarrow{0.077} D = [10]$$

$$E(S) \xrightarrow{0.7500} S = [11] \xrightarrow{0.923} D = [11]$$

- CASO 1:
  - $I\,(D = [10]) = 0.7500 \times 0.077 \;= 0.05775$

- CASO 2:
  - $I\,(D = [11]) = 0.7500 \times 0.923 = 0.69225$

$$S\,[11] \rightarrow \frac{P(D = [11])}{P(D = [10])} = \frac{0.923}{0.077} = \frac{0.69225}{0.05775} \approx 12$$

- Modelo Categórico:
  - $S\ [11] \rightarrow D\ [10]$: Es posible: Posibilidad = 1
  - $S\ [11] \rightarrow D\ [11]$: Es posible: Posibilidad = 1

- Modelo Probabilístico:
  - $S\ [11] \rightarrow D\ [10]$: Es posible: Probabilidad = 0.077
  - $S\ [11] \rightarrow D\ [11]$: Es posible: Probabilidad = 0.923

- Modelo Impreciso:
  - $E\ (S = [11]) = 0.75 \rightarrow D\ [10]$: Es posible: Confianza = 0.05775
  - $E\ (S = [11]) = 0.75 \rightarrow D\ [11]$: Es posible: Confianza = 0.69225

$$(a_1 = 1)\ AND\ (a_2 = 1) \xrightarrow{0.077} D = [01]$$

$$(a_1 = 1)\ AND\ (a_2 = 1) \xrightarrow{0.923} D = [11]$$

- CASO 1:

$$\{RC\ [(a_1 = 1)] = 0.3750\}\ AND\ \{RC\ [(a_2 = 1)] = \ 0.0625 \xrightarrow{0.077} D = \ [01]$$

- CASO 2:

$$\{RC\ [(a_1 = 1)] = 0.3750\}\ AND\ \{RC\ [(a_2 = 1)] = \ 0.0625 \xrightarrow{0.923} D = \ [11]$$

  - CASO 1:
    - I (D = [01]) = min {0.3750 , 0.0625} × 0.077 = 0.0625 × 0.077 = 0.0048

  - CASO 2:
    - I (D = [11]) = min {0.3750 , 0.0625} × 0.923 = 0.0625 × 0.923 = 0.0577

$$\frac{D\ [11]}{D\ [01]} = \frac{0.923}{0.077} = \frac{0.69225}{0.05775} = \frac{0.0577}{0.0048} \approx 12$$

- Un paciente llega a la consulta quejándose de los síntomas a1 y a2.
- Ambos síntomas, conjuntamente, están relacionados con las asociaciones diagnósticas [01] y [11]
- También sabemos que, en ausencia de otras evidencias, la asociación diagnóstica [11] es más frecuente que la asociación diagnóstica [01], por lo que el clínico decide tratar D = [11]
- En cualquier caso ambos síntomas, a1 y a2, son críticos para obtener el diagnóstico correcto.
- A continuación el clínico valora los síntomas, que resultan ser ambos muy severos, para prescribir un tratamiento adecuado. Por lo tanto, de acuerdo con la regla 6 de este ejemplo:

$$\{RC\ [(a_1 = 1)] = 1.0000\}\ AND\ \{RC\ [(a_2 = 1)] = 1.0000 \xrightarrow{0.923} D = [11]$$

- $I\ (D = [11]) = \min\{1.0000\ , 1.0000\} \times 0.923 = 1.0000 \times 0.923 = 0.923$
- El Indice de Severidad del diagnóstico D = [11] está en torno al 90%

- A la vista de estos resultados, el clínico prescribe un tratamiento, y emplaza al paciente a una segunda visita, en la cual constata que ambos síntomas han evolucionado positivamente de muy severos a severos. Por lo tanto:

$$\{RC\ [(a_1 = 1)] = 0.7500\}\ AND\ \{RC\ [(a_2 = 1)] = 0.7500 \xrightarrow{0.923} D = [11]$$

- I (D = [11]) = min {0.7500 , 0.7500} × 0.923 = 0.7500 × 0.923 = 0.692
- El Índice de Severidad del diagnóstico D = [11] está en torno al 70%

- Seguidamente, el clínico realiza ajustes en el tratamiento, y emplaza al paciente a una tercera visita, en la cual no observa mejorías en el primer síntoma, pero sí en el segundo, que pasa de severo a moderado. Por lo tanto:

$$\{RC\ [(a_1 = 1)] = 0.7500\}\ AND\ \{RC\ [(a_2 = 1)] = \ 0.5000 \xrightarrow{0.923} D = \ [11]$$

- I (D = [11]) = min {0.7500 , 0.5000} × 0.923 = 0.5000 × 0.923 = 0.462
- El Índice de Severidad del diagnóstico D = [11] está en torno al 45%

1. Apneas Obstructivas	=	x1
2. Apneas Centrales	=	x2
3. Apneas Mixtas	=	x3
4. Hipopneas Obstructivas	=	x4
5. Hipopneas Centrales	=	x5
6. Hipopneas Mixtas	=	x6

1. El flujo respiratorio en vías aéreas = a1
2. Los movimientos respiratorios abdominales = a2
3. Los movimientos respiratorios torácicos = a3
4. La saturación de oxígeno en sangre arterial = a4

S↓ / D→		X0	X1	X2	X4	X8	X16	X32	TOTAL
		000000	000001	000010	000100	001000	010000	100000	
A0	**0000**	14288							**14288**
A3	**0011**		4	1	61				**66**
A5	**0101**		3	1	85				**89**
A7	**0111**		14	3	347				**364**
A9	**1001**		4		66	2	2	12	**86**
A11	**1011**		4	1	100	1		11	**117**
A13	**1101**		8	2	164	1	1	22	**198**
A14	**1110**		12	2	267	1	3	13	**298**
A15	**1111**		66	60	1499	149	136	454	**2364**
TOTAL		**14288**	**115**	**70**	**2589**	**154**	**142**	**512**	**17870**

S↓ / D→		X0	X1	X2	X4	X8	X16	X32	Prevalencias
		000000	000001	000010	000100	001000	010000	100000	
A0	0000	0.7995523							0.7995523
A3	0011		0.0002238	0.0000560	0.0034135				0.0036933
A5	0101		0.0001679	0.0000560	0.0047566				0.0049804
A7	0111		0.0007834	0.0001679	0.0194180				0.0203693
A9	1001		0.0002238		0.0036933	0.0001119	0.0001119	0.0006715	0.0048125
A11	1011		0.0002238	0.0000560	0.0055960	0.0000560		0.0006156	0.0065473
A13	1101		0.0004477	0.0001119	0.0091774	0.0000560	0.0000560	0.0012311	0.0110800
A14	1110		0.0006715	0.0001119	0.0149412	0.0000560	0.0001679	0.0007275	0.0166760
A15	1111		0.0036933	0.0033576	0.0838836	0.0083380	0.0076105	0.0254057	0.1322888
Prevalencias		0.7995523	0.0064354	0.0039172	0.1448797	0.0086178	0.0079463	0.0286514	1.0000000

S↓ / D→		X0	X1	X2	X4	X8	X16	X32
		000000	000001	000010	000100	001000	010000	100000
A0	**0000**	1.0000000						
A3	**0011**		0.0606067	0.0151517	0.9242526			
A5	**0101**		0.0337080	0.0112360	0.9550589			
A7	**0111**		0.0384616	0.0082418	0.9532983			
A9	**1001**		0.0465120		0.7674474	0.0232560	0.0232560	0.1395359
A11	**1011**		0.0341880	0.0085470	0.8546990	0.0085470		0.0940169
A13	**1101**		0.0404041	0.0101010	0.8282845	0.0050505	0.0050505	0.1111113
A14	**1110**		0.0402684	0.0067114	0.8959728	0.0033557	0.0100671	0.0436241
A15	**1111**		0.0279188	0.0253807	0.6340945	0.0630287	0.0575296	0.1920473

P (X0/A0)	= 1.000	P (X1/A3)	= 0.061	P(X1/A5)	= 0.034
P (X1/A7)	= 0.038	P (X1/A9)	= 0.047	P(X1/A11)	= 0.034
P (X1/A13)	= 0.040	P (X1/A14)	= 0.040	P(X1/A15)	= 0.028
P (X2/A3)	= 0.015	P (X2/A5)	= 0.011	P(X2/A7)	= 0.008
P (X2/A11)	= 0.009	P (X2/A13)	= 0.010	P(X2/A14)	= 0.007
P (X2/A15)	= 0.025	P (X4/A3)	= 0.924	P(X4/A5)	= 0.955
P (X4/A7)	= 0.953	P (X4/A9)	= 0.767	P(X4/A11)	= 0.855
P (X4/A13)	= 0.828	P (X4/A14)	= 0.896	P(X4/A15)	= 0.634

P (X8/A9)	= 0.023	P (X8/A11)	= 0.009	P(X8/A13)	= 0.005
P (X8/A14)	= 0.003	P (X8/A15)	= 0.063	P(X16/A9)	= 0.023
P (X16/A13)	= 0.005	P (X16/A14)	= 0.010	P(X16/A15)	= 0.058
P (X32/A9)	= 0.140	P (X32/A11)	= 0.094	P(X32/A13)	= 0.111
P (X32/A14)	= 0.044	P (X32/A15)	= 0.192		

1. $S = [0000] \quad \rightarrow \quad D = [000000] \quad PROBABILIDAD = 1.000$
2. $S = [0011] \quad \rightarrow \quad D = [000001] \quad PROBABILIDAD = 0.061$
3. $S = [0011] \quad \rightarrow \quad D = [000010] \quad PROBABILIDAD = 0.015$
4. $S = [0011] \quad \rightarrow \quad D = [000100] \quad PROBABILIDAD = 0.924$
5. $S = [0101] \quad \rightarrow \quad D = [000001] \quad PROBABILIDAD = 0.034$
6. $S = [0101] \quad \rightarrow \quad D = [000010] \quad PROBABILIDAD = 0.011$
7. $S = [0101] \quad \rightarrow \quad D = [000100] \quad PROBABILIDAD = 0.955$
8. $S = [0111] \quad \rightarrow \quad D = [000001] \quad PROBABILIDAD = 0.038$
9. $S = [0111] \quad \rightarrow \quad D = [000010] \quad PROBABILIDAD = 0.008$
10. $S = [0111] \quad \rightarrow \quad D = [000100] \quad PROBABILIDAD = 0.953$
11. $S = [1001] \quad \rightarrow \quad D = [000001] \quad PROBABILIDAD = 0.047$
12. $S = [1001] \quad \rightarrow \quad D = [000100] \quad PROBABILIDAD = 0.767$
13. $S = [1001] \quad \rightarrow \quad D = [001000] \quad PROBABILIDAD = 0.023$
14. $S = [1001] \quad \rightarrow \quad D = [010000] \quad PROBABILIDAD = 0.023$
15. $S = [1001] \quad \rightarrow \quad D = [100000] \quad PROBABILIDAD = 0.140$
16. $S = [1011] \quad \rightarrow \quad D = [000001] \quad PROBABILIDAD = 0.034$
17. $S = [1011] \quad \rightarrow \quad D = [000010] \quad PROBABILIDAD = 0.009$
18. $S = [1011] \quad \rightarrow \quad D = [000100] \quad PROBABILIDAD = 0.855$
19. $S = [1011] \quad \rightarrow \quad D = [001000] \quad PROBABILIDAD = 0.009$

20. $S = [1011] \rightarrow D = [100000]$ PROBABILIDAD $= 0.094$
21. $S = [1101] \rightarrow D = [000001]$ PROBABILIDAD $= 0.040$
22. $S = [1101] \rightarrow D = [000010]$ PROBABILIDAD $= 0.010$
23. $S = [1101] \rightarrow D = [000100]$ PROBABILIDAD $= 0.828$
24. $S = [1101] \rightarrow D = [001000]$ PROBABILIDAD $= 0.005$
25. $S = [1101] \rightarrow D = [010000]$ PROBABILIDAD $= 0.005$
26. $S = [1101] \rightarrow D = [100000]$ PROBABILIDAD $= 0.111$
27. $S = [1110] \rightarrow D = [000001]$ PROBABILIDAD $= 0.040$
28. $S = [1110] \rightarrow D = [000010]$ PROBABILIDAD $= 0.007$
29. $S = [1110] \rightarrow D = [000100]$ PROBABILIDAD $= 0.896$
30. $S = [1110] \rightarrow D = [001000]$ PROBABILIDAD $= 0.003$
31. $S = [1110] \rightarrow D = [010000]$ PROBABILIDAD $= 0.010$
32. $S = [1110] \rightarrow D = [100000]$ PROBABILIDAD $= 0.044$
33. $S = [1111] \rightarrow D = [000001]$ PROBABILIDAD $= 0.028$
34. $S = [1111] \rightarrow D = [000010]$ PROBABILIDAD $= 0.025$
35. $S = [1111] \rightarrow D = [000100]$ PROBABILIDAD $= 0.634$
36. $S = [1111] \rightarrow D = [001000]$ PROBABILIDAD $= 0.063$
37. $S = [1111] \rightarrow D = [010000]$ PROBABILIDAD $= 0.058$
38. $S = [1111] \rightarrow D = [100000]$ PROBABILIDAD $= 0.192$

$$ICN(D_j) = 100 \times Probabilidad\,(D_j / S)$$

1.			ICN (Dj)	=	0	→	Dj Descartado
2.	0	<	ICN (Dj)	≤	25	→	Dj es Muy Poco Probable
3.	25	<	ICN (Dj)	≤	50	→	Dj es Poco Probable
4.	50	<	ICN (Dj)	<	75	→	Dj es Probable
5.	75	≤	ICN (Dj)	<	100	→	Dj es Muy Probable
6.			ICN (Dj)	=	100	→	Dj Confirmado

R2:	IF	1.	No hay disminución del flujo respiratorio
	AND	2.	No hay esfuerzo respiratorio abdominal anormal
	AND	3.	Hay esfuerzo respiratorio torácico anormal
	AND	4.	Hay desaturación significativa en sangre arterial
	THEN	1.	Evento Posible = Hipopnea
	AND	2.	Tipo de Evento = Mixto
	AND	3.	Plausibilidad = Muy Poco Probable
	AND	4.	ICN= 6.1

1. De todos los episodios apneicos confirmados (Apnea o Hipopnea), la distribución resultó ser la siguiente:
    a. Porcentaje de Hipopneas = 77 %
    b. Porcentaje de Apneas = 23 %
2. De todos los episodios apneicos confirmados como Apneas o como Hipopneas, la distribución, para los diferentes tipos de eventos, resultó ser la siguiente:
    a. Porcentaje de Eventos Mixtos = 7 %
    b. Porcentaje de Eventos Centrales = 6 %
    c. Porcentaje de Eventos Obstructivos = 87 %

	EXPERTO HUMANO CONSIDERADO COMO GOLD STANDARD				
MODELO	Evento Hipopnea	Evento Apnea	Evento Obstructivo	Evento Central	Evento Mixto
	95 %	86 %	94 %	80 %	65 %

S	Hipopneas	Apneas	Tipo de Evento	Capacidad Predictiva
[0011]	98 %	100 %	91 %	95 %
[0101]	100 %	100 %	87 %	93 %
[0111]	100 %	100 %	87 %	93 %
[1001]	98 %	95 %	89 %	94 %
[1011]	95 %	83 %	84 %	89 %
[1101]	79 %	52 %	90 %	84 %
[1110]	96 %	86 %	84 %	90 %
[1111]	91 %	86 %	92 %	92 %

Redes Bayesianas

- Aprendizaje
  - cuál es la mejor hipótesis (más probable) dados los dato?
- Red Bayesiana (RB)
  - Red de creencia
- Clasificación de una RB
  - Aprendizaje Estructural.
    - Obtener la estructura de la red
  - Aprendizaje Paramétrico
    - Dada la estructura, obtener las probabilidades asociadas

Redes Bayesianas

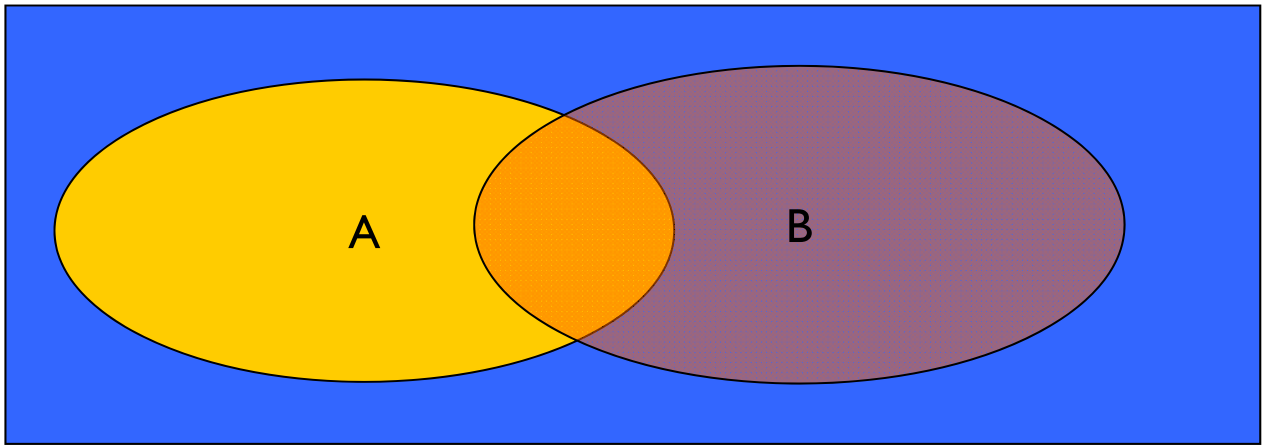
- P(A, B)

  Probabilidad de ambas A y B.

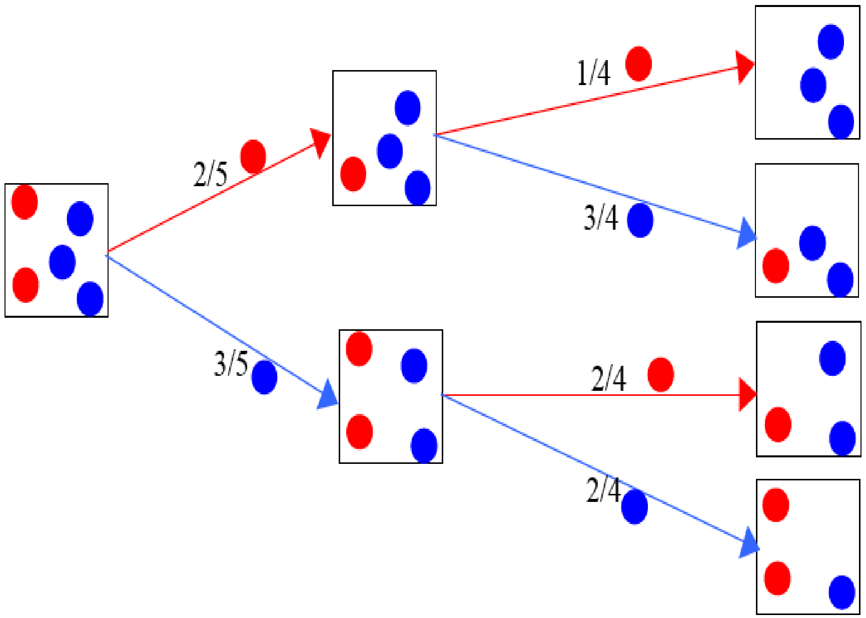
- P(A|B)

  Probabilidad condicional de A dado B.

$$P(A,B)<P(A|B)$$



Redes Bayesianas

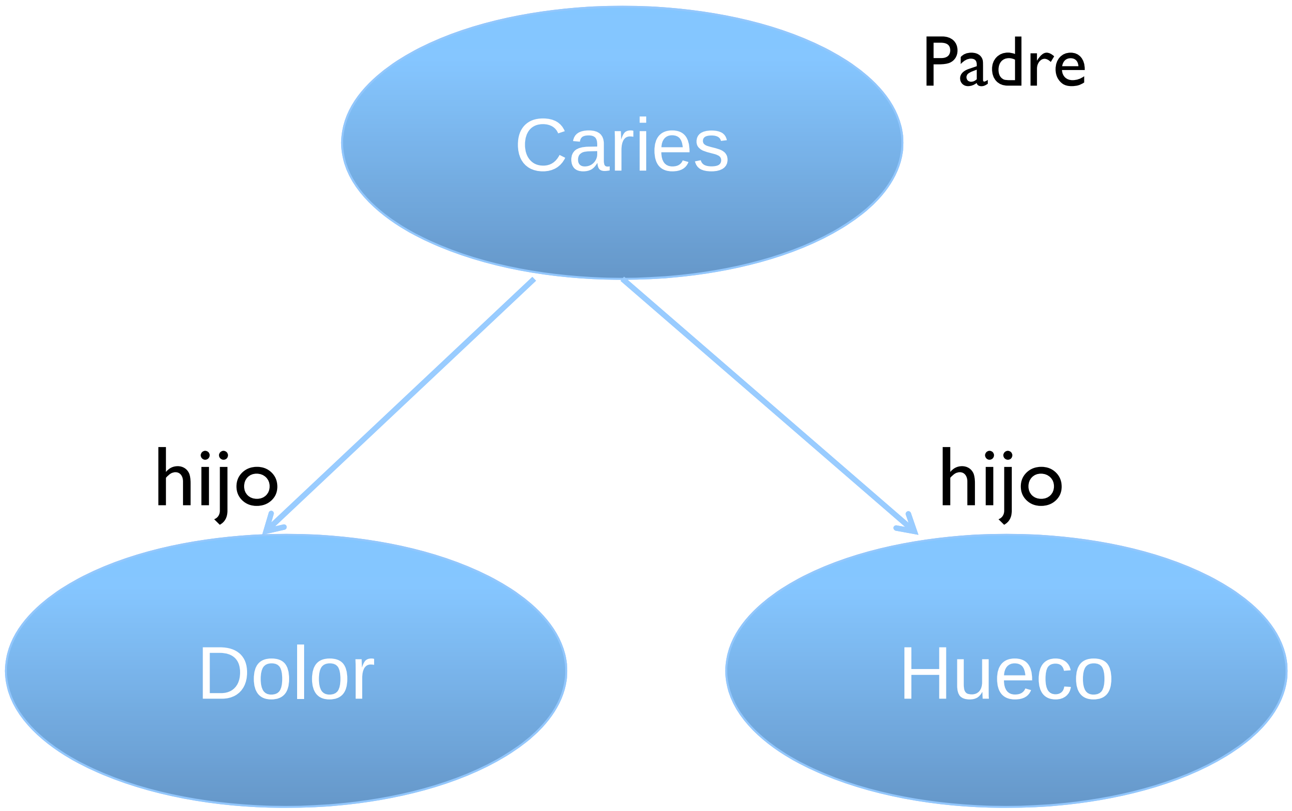


- Condicional.
  - Probabilidad de que la 2ª canica sea roja dado que la 1ª es azul
  - P(R2|R1)
- Conjunta
  - Probabilidad de que la 1ª sea roja y la 2ª azul
  - P(R1 y R2)

Redes Bayesianas

- Una red bayesiana es un grafo dirigido acíclico que consta de:
  - Un conjunto de nodos, uno por cada variable aleatoria del “mundo”
  - Un conjunto de arcos dirigidos que conectan los nodos
  - Cada nodo tiene una probabilidad condicional asociada
  - Cada arco *X* a *Y* indica una influencia directa de *X* sobre *Y*

Redes Bayesianas



RAZONAMIENTO PROBABILÍSTICO

- Elementos Básicos del Modelo Bayesiano
- Aplicación Elemental de la Regla de Bayes
- El Modelo Bayesiano en Sistemas Expertos
- Ventajas e Inconvenientes del Modelo Bayesiano
- Introducción a la Teoría de Grafos
- Fundamentos de las Redes de Creencia
- Inferencia en Redes de Creencia
- Aprendizaje en Redes de Creencia

RAZONAMIENTO PROBABILÍSTICO

- **Elementos Básicos del Modelo Bayesiano**
  - Sea *A* un evento.
  - La colección de todos los posibles eventos elementales Ω se conoce como espacio de eventos.
  - La probabilidad de un evento *A* se denota como *p*(*A*)
  - Toda función de probabilidad *p* debe satisfacer tres axiomas:

(1) La probabilidad de cualquier evento elemental A es no negativa, es decir, $\forall A \in \Omega$: $p(A) \geq 0$.

(2) La probabilidad del espacio de eventos es uno, es decir, $p(\Omega) = 1$.

(3) Si k eventos $A_1, A_2, \ldots, A_k$ son mutuamente excluyentes (es decir, no pueden ocurrir simultáneamente), entonces la probabilidad de que al menos uno de esos eventos ocurra es la suma de las probabilidades individuales $p(A_1 \cup A_2 \cup \ldots \cup A_k) = \sum_{i=1}^{k} p(A_i)$

RAZONAMIENTO PROBABILÍSTICO

- A partir de estos axiomas podemos deducir otros. Así, de los axiomas 1 y 2 podemos obtener el siguiente resultado:
  - $\forall A \in \Omega : 0 \leq p(A) \leq 1$
- El complementario de *A* (representado como $\neg A$) contiene la colección de todos los eventos elementales en Ω excepto *A*
- Como *A* y $\neg A$ son mutuamente excluyentes y $A \cup \neg A = \Omega$
  - $p(A) + p(\neg A) = p(A \cup \neg A) = p(\Omega) = 1$
- Si suponemos que $B \in \Omega$ es otro evento, la probabilidad de que A ocurra sabiendo que *B* ocurre, representado por $p(A/B)$ se conoce como la probabilidad condicional de *A* dado *B*

$$p(A/B) = \frac{p(A \cap B)}{p(B)}$$

RAZONAMIENTO PROBABILÍSTICO

- A partir de esta ecuación es sencillo obtener la regla de Bayes, base de los modelos bayesianos

$$p(A/B) = \frac{p(B/A)\,p(A)}{p(B)}$$

- La ecuación básica de Bayes puede generalizarse

$$p(A/B) = \frac{p(B/A)\,p(A)}{p(B/A)\,p(A) + p(B/\neg A)\,p(\neg A)}$$

RAZONAMIENTO PROBABILÍSTICO

$$p(A_0 / B) = \frac{p(B / A_0)\, p(A_0)}{\sum_i p(B / A_i)\, p(A_i)}$$

$$p(E / S) = \frac{p(S / E)\, p(E)}{p(S)}$$

¿No podría obtenerse también de antemano la probabilidad *p*(*E*/*S*)? La respuesta a estas preguntas es que el conocimiento obtenido por diagnóstico es más débil que el conocimiento causal.

RAZONAMIENTO PROBABILÍSTICO

- Bayes nos permite obtener la probabilidad de una hipótesis dada, en base a las probabilidades condicionales de observar una evidencia dada una hipótesis, $p(E/H)$, y a las probabilidades a priori de todas las hipótesis, $p(H)$.

- Múltiples evidencias:

$$p(H_0 / E_1, E_2, \ldots, E_n) = \frac{p(E_1, E_2, \ldots, E_n / H_0)\, p(H_0)}{\sum_i p(E_1, E_2, \ldots, E_n / H_i)\, p(H_i)}$$

RAZONAMIENTO PROBABILÍSTICO

- El problema con esta nueva versión de la regla de Bayes es la explosión combinatoria
- Así para un sencillo problema en el que 10 evidencias apuntan a 3 posibles hipótesis nos encontramos con que el número de probabilidades condicionales que tenemos que almacenar es de $3 \times 210 = 3 \times 1024 = 3072$, un número claramente elevado para un problema claramente sencillo.
- La solución al problema de la explosión combinatoria consiste en suponer que las evidencias son condicionalmente independientes.

Razonamiento Probabilístico

Dos evidencias *E*1 y *E*2 son condicionalmente independientes cuando su probabilidad conjunta dada una determinada hipótesis *H* equivale al producto de las probabilidades condicionales de cada evento dado *H*, o lo que es lo mismo $p(E1, E2/H) = p(E1/H)\, p(E2/H)$.

$$p(H_0 / E_1, E_2, \ldots, E_n) = \frac{p(E_1 / H_0)\; p(E_2 / H_0) \cdots p(E_n / H_0)\; p(H_0)}{\sum_i p(E_1 / H_i)\; p(E_2 / H_i) \cdots p(E_n / H_i)\; p(H_i)}$$

Suponiendo independencia condicional, el número de probabilidades condicionales a almacenar pasa a ser (nº H) $\times$ (nº E).

Para el ejemplo con 10 evidencias y 3 hipótesis el resultado es 30. Número claramente inferior al resultado de 3072 obtenido suponiendo dependencia entre la hipótesis.

Razonamiento Probabilístico

- Supongamos que tenemos un conjunto exhaustivo formado por tres hipótesis mutuamente excluyentes (*H*1, *H*2 y *H*3) y dos evidencias independientes (*E*1 y *E*2)

	$i=1$	$i=2$	$i=3$
$p(H_i)$	0.5	0.3	0.2
$p(E_1/H_i)$	0.4	0.8	0.3
$p(E_2/H_i)$	0.7	0.9	0.0

$$p(H_1/E_1) = \frac{(0.4 \times 0.5)}{(0.4 \times 0.5) + (0.8 \times 0.3) + (0.3 \times 0.2)} = 0.40$$

$$p(H_2/E_1) = \frac{(0.8 \times 0.3)}{(0.4 \times 0.5) + (0.8 \times 0.3) + (0.3 \times 0.2)} = 0.48$$

$$p(H_3/E_1) = \frac{(0.3 \times 0.2)}{(0.4 \times 0.5) + (0.8 \times 0.3) + (0.3 \times 0.2)} = 0.12$$

Razonamiento Probabilístico

$$p(H_1 / E_1E_2) = \frac{(0.4 \times 0.7 \times 0.5)}{(0.4 \times 0.7 \times 0.5) + (0.8 \times 0.9 \times 0.3) + (0.3 \times 0.0 \times 0.2)} = 0.39$$

$$p(H_2 / E_1E_2) = \frac{(0.8 \times 0.9 \times 0.3)}{(0.4 \times 0.7 \times 0.5) + (0.8 \times 0.9 \times 0.3) + (0.3 \times 0.0 \times 0.2)} = 0.61$$

$$p(H_1 / E_1E_2) = \frac{(0.3 \times 0.0 \times 0.2)}{(0.4 \times 0.7 \times 0.5) + (0.8 \times 0.9 \times 0.3) + (0.3 \times 0.0 \times 0.2)} = 0.00$$

Razonamiento Probabilístico

La principal ventaja de los métodos bayesianos reside en que están fuertemente fundados en la teoría de la probabilidad, sin embargo su principal dificultad estriba en la gran cantidad de probabilidades que es necesario obtener para construir una base de conocimientos.

Una aproximación para resolver estos problemas son las redes de creencia (belief networks). Una red de creencia es un tipo especial de diagrama de influencia en el que los nodos representan variables aleatorias. Pearl (1988) demostró que el uso de redes de creencia permite construir bases de conocimiento probabilisticas consistentes, sin imponer innecesarias asunciones de independencia condicional. Estas redes también aseguran que la evidencia a favor de una hipótesis no será construida por soporte parcial de su negación, y que explicaciones consistentes pueden ser obtenidas mediante el rastreo de las creencias hasta los puntos iniciales de la red.

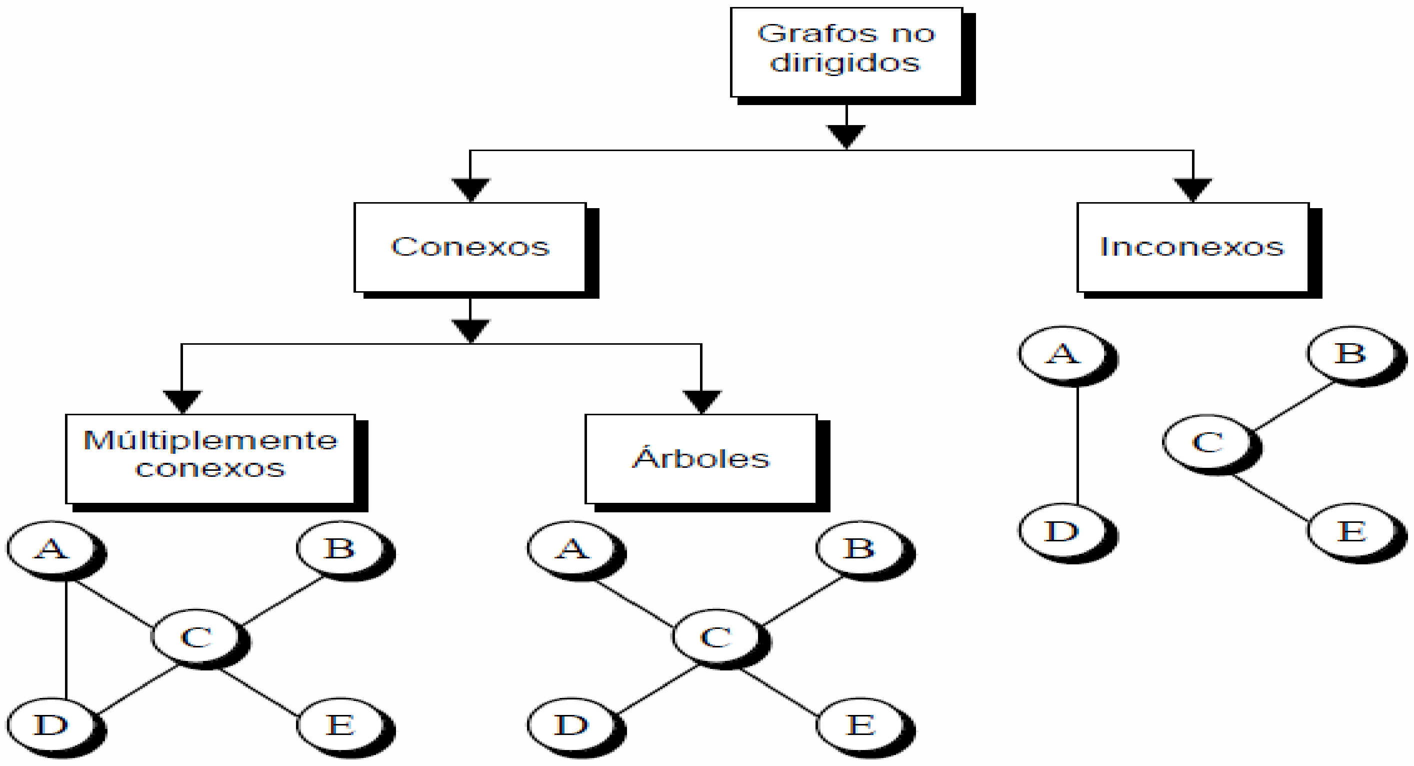
Razonamiento Probabilístico

Antes de entrar en detalle con los modelos de redes de creencia vamos a realizar una breve introducción a la teoría de grafos.

Un grafo (o red) G se compone de un par de conjuntos $G = (X, L)$ donde $X = \{X_1, X_2, ..., X_n\}$ es un conjunto finito de elementos (nodos) y L es un conjunto de aristas, es decir, un subconjunto de pares ordenados de elementos distintos de X. Así, si una arista une a los nodos X_i y X_j se denotará mediante L_{ij}.

Se dice que la arista que une los nodos X_i y X_j es dirigida cuando $L_{ij} \in L$ y $L_{ji} \notin L$, y se denota como $X_i \rightarrow X_j$. Por otro lado la arista entre los nodos X_i y X_j es no dirigida cuando $L_{ij} \in L$ y $L_{ji} \in L$ y se representa como $X_i — X_j$. Un grafo en el cual todas las aristas son dirigidas se denomina grafo dirigido, y un grafo en el que todas sus aristas son no dirigidas se denomina grafo no dirigido.

Razonamiento Probabilístico



Tipos de grafos no dirigidos

Razonamiento Probabilístico

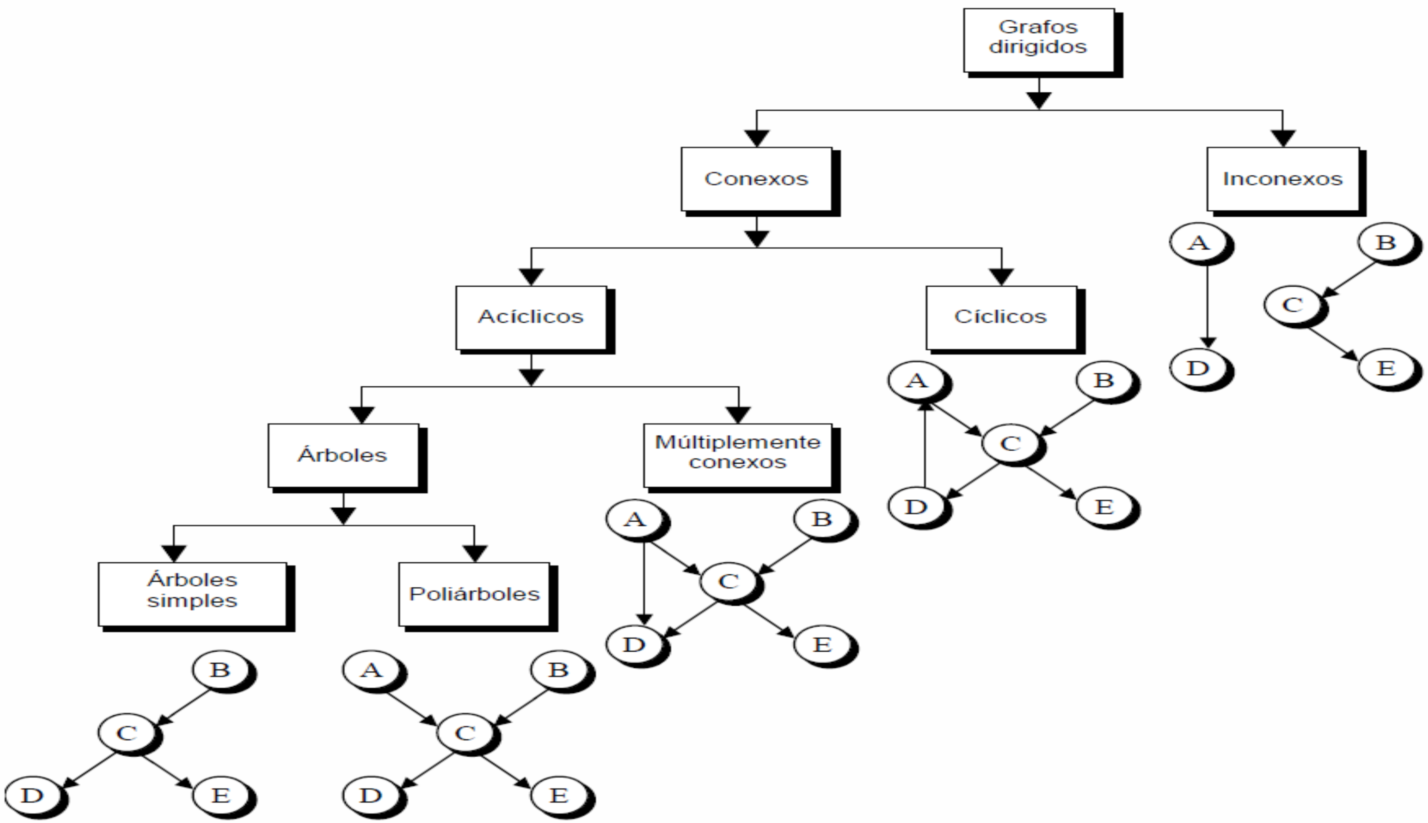


Figura 1.2 Tipos de grafos dirigidos

Razonamiento Probabilístico

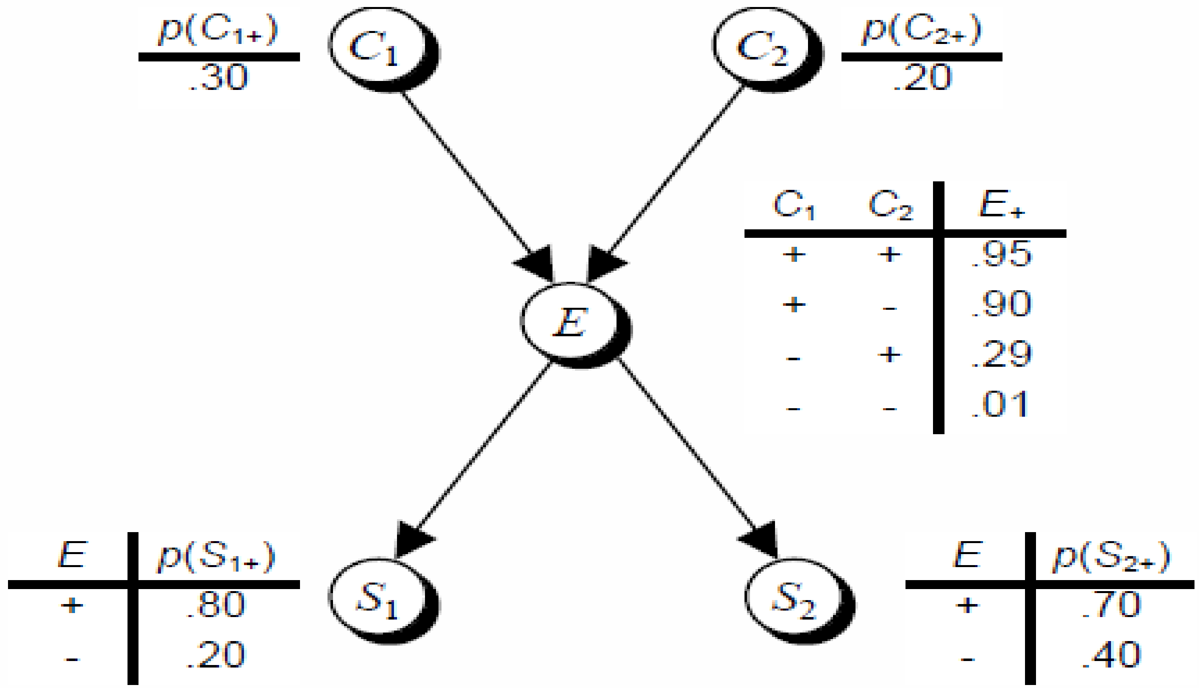
Redes de creencia como grafos dirigidos acíclicos

(1) Nodos.- que representan proposiciones o variables aleatorias y que representaremos como $\{x_1, x_2, \ldots, x_n\}$.

(2) Arcos dirigidos.- que representan la existencia de influencias causales directas entre las proposiciones que une. El significado implícito de una flecha que vaya del nodo xi al nodo x_j es el de que x_i ejerce una influencia directa sobre x_j (se suele decir que xi es un nodo padre de x_j).

(3) Tablas de probabilidad condicional.- que miden la potencia de las influencias mediante el uso de probabilidades condicionadas $p(x_i/\text{Padres}(x_j))$. Por cada nodo hay una tabla de probabilidad condicional que sirve para cuantificar los efectos de los padres sobre dicho nodo.

Razonamiento Probabilístico



Ejemplo de red bayesiana

Razonamiento Probabilístico

- Cada nodo de la red representan un conjunto de proposiciones exhaustivas y mutuamente excluyentes
- *C*1 representa a las proposiciones *C*1+ y *C*1-, que representan respectivamente “*C*1 = True” y “*C*1 = False”
- Cada nodo tiene una distribución de probabilidad por cada una de las combinaciones de sus nodos condicionantes
- *E* está condicionada por *C*1 y *C*2 por ello hay cuatro posibles distribuciones de probabilidad de *E*, que corresponden con las cuatro posibles combinaciones de *C*1 y *C*2.
- Se representa sólo la probabilidad de *E*+, ya que la probabilidad de *E*- puede obtenerse como 1 – *p*(*E*+)
- Todo esto siempre y cuando las proposiciones sean binarias, es decir, que sólo tomen dos posibles valores

Razonamiento Probabilístico

- También es posible la existencia de proposiciones multivaluadas y proposiciones que toman valores continuos.
- Las variables que toman valores continuos pueden tener funciones de densidad de probabilidad como, por ejemplo, la curva o campana de Gauss.
- De esta forma en vez de tener una tabla de probabilidades en el nodo tendríamos los parámetros (media y varianza) que caracterizarían a la curva gaussiana.
- De todas formas lo más corriente suele ser discretizar los valores continuos mediante intervalos, por ejemplo la edad podría tomar los valores “<18”, “18 – 65” y “>65”.

Razonamiento Probabilístico

- La ausencia de arcos refleja aserciones de independencia condicional.
- Así si no hay ningún arco entre $C1$ y $C2$ quiere decir que la probabilidad de $C2$ no depende de $C1$.
- De la misma forma $S1$ no depende directamente de $C1$, sino que están conectados a través del nodo E.
- En este caso se dice que $S1$ es independiente de $C1$ dado E

Razonamiento Probabilístico

- Separación dependiente de la dirección (d-separación)
  - Sean *X*, *Y*, y *Z* tres subconjuntos disjuntos de nodos en un grafo dirigido acíclico *D*; entonces se dice que *Z* d-separa *X* e *Y* si y sólo si, a lo largo de todo el camino no dirigido entre cualquier nodo de *X* y cualquier nodo de *Y*, existe un nodo intermedio *A* tal que, o bien:
    - *A* no es un nodo de aristas convergentes en el camino y *A* está en *Z*, o bien
    - *A* es un nodo de aristas convergentes en el camino y ni *A* ni sus descendientes está en *Z*.

Razonamiento Probabilístico

- Cuando *Z* d-separa *X* e *Y*, se escribe *I* (*X*, *Y* / *Z*) para indicar que *X* e *Y* son condicionalmente independientes dado *Z*

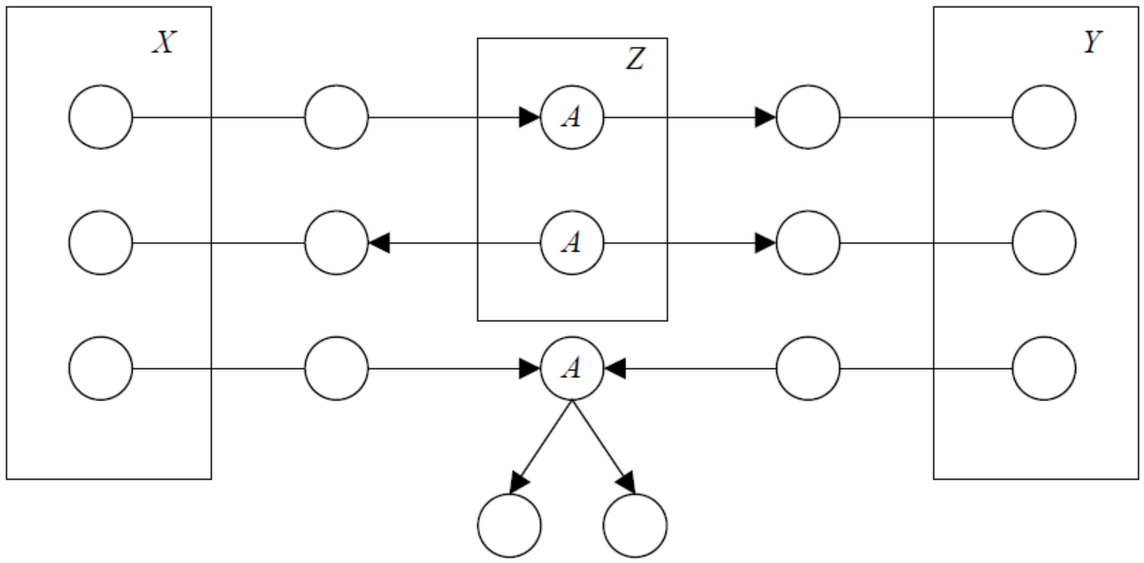
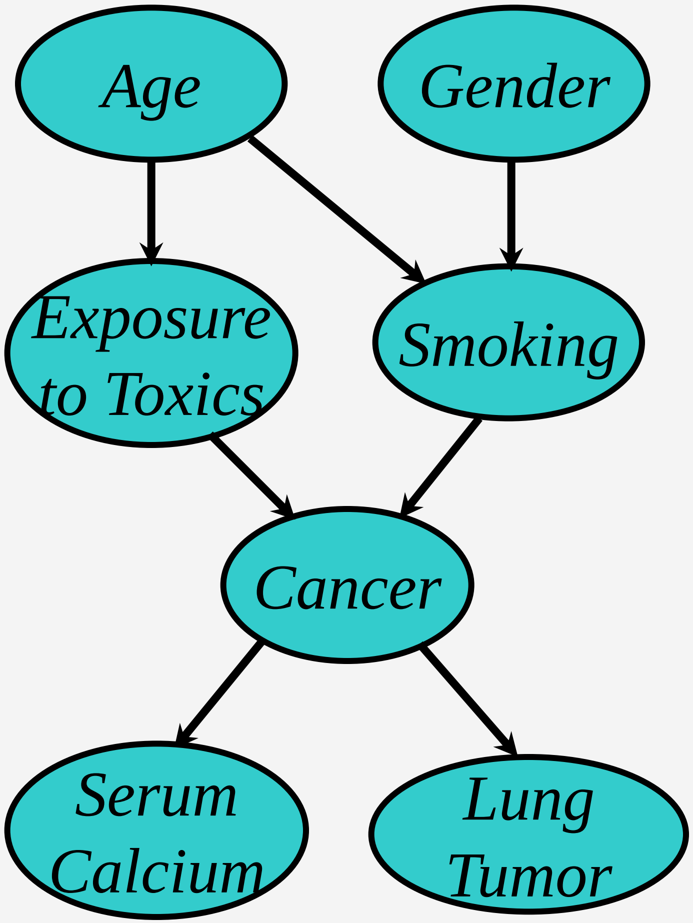
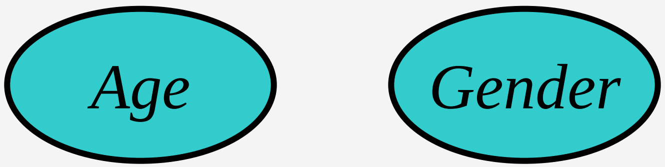


Figura 1.4 Criterio de d-separación, los dos primeros casos corresponden a la primera definición de d-separación y el tercero a la segunda definición.

A Bayesian Network



Independence



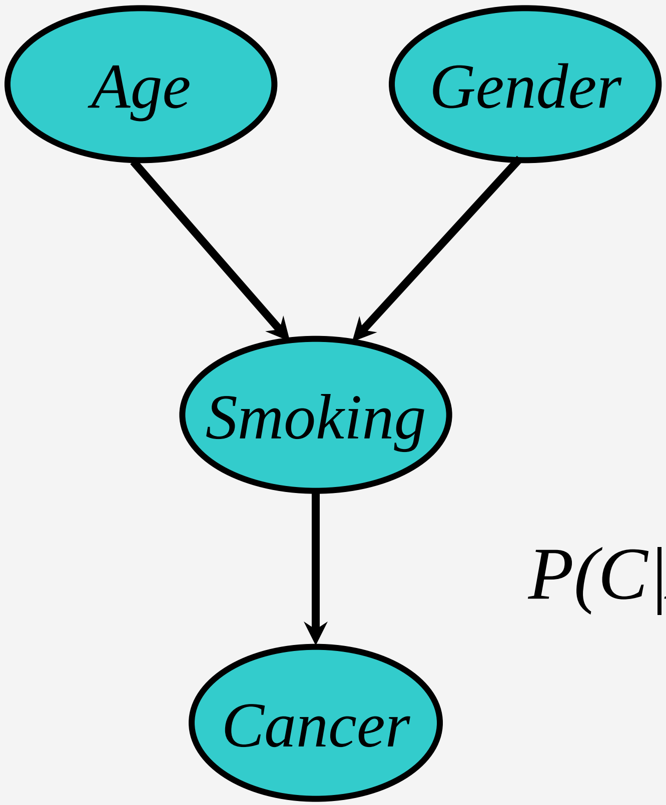
Age and *Gender* are independent.

$$P(A,G) = P(G)P(A)$$

$$P(A|G) = P(A) \quad A \perp G$$
$$P(G|A) = P(G) \quad G \perp A$$

$$P(A,G) = P(G|A)\,P(A) = P(G)P(A)$$
$$P(A,G) = P(A|G)\,P(G) = P(A)P(G)$$

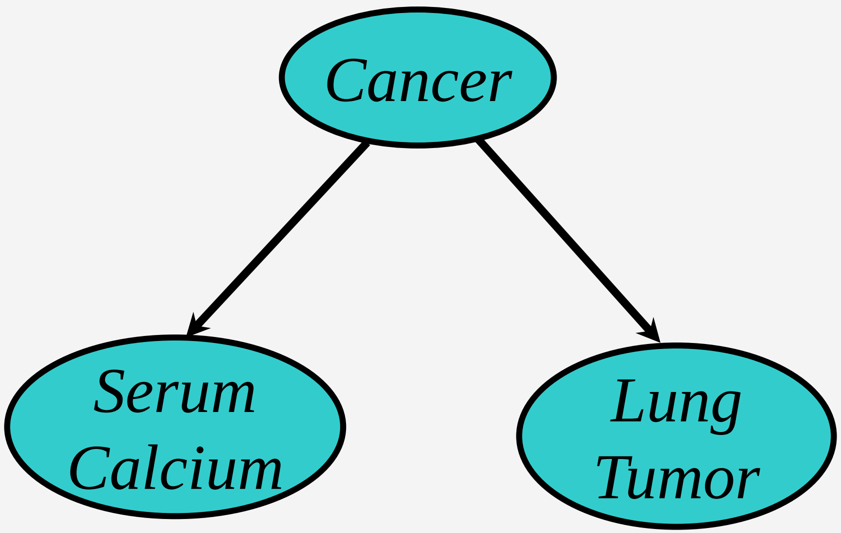
Conditional Independence



Cancer is independent of *Age* and *Gender* given *Smoking*.

$$P(C|A,G,S) = P(C|S) \quad C \perp A,G \mid S$$

More Conditional Independence: Naïve Bayes

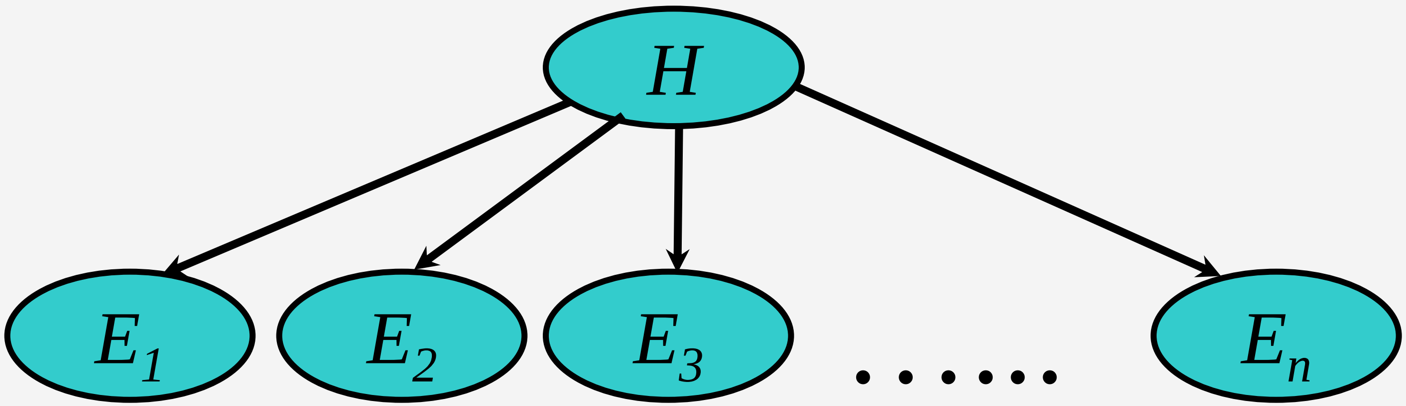


Serum Calcium and *Lung Tumor* are dependent

Serum Calcium is independent of *Lung Tumor*, given *Cancer*

$$P(L|SC,C) = P(L|C)$$

Naïve Bayes in general

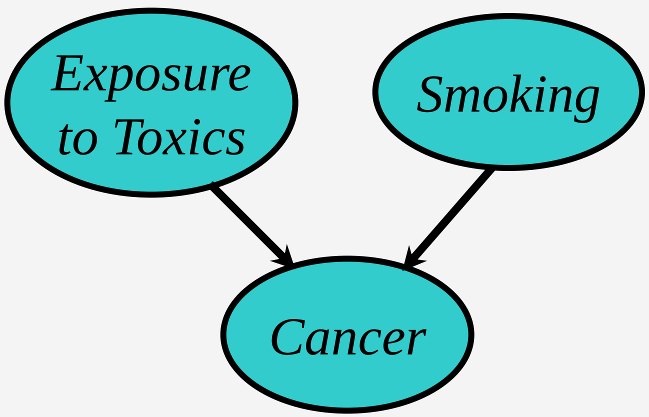


2n + 1 parameters:

$$P(h)$$

$$P(e_i \mid h), P(e_i \mid \bar{h}), \quad i = 1, \ldots, n$$

More Conditional Independence: Explaining Away



Exposure to Toxics and *Smoking* are independent

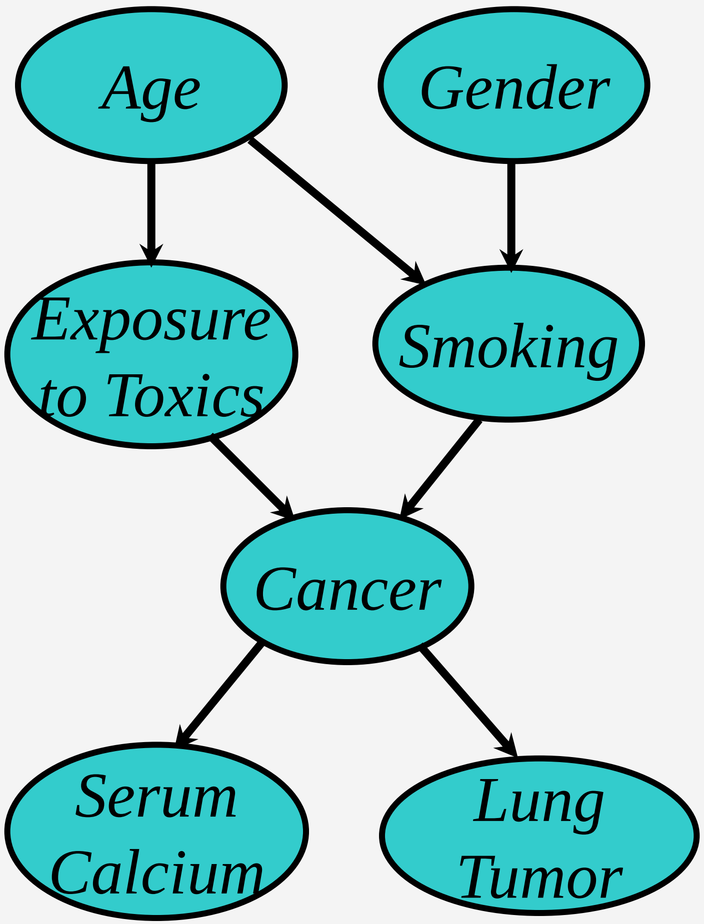
$$E \perp S$$

Exposure to Toxics is **dependent** on *Smoking*, given *Cancer*

$$P(E = heavy \mid C = malignant) >$$

$$P(E = heavy \mid C = malignant, S{=}heavy)$$

Put it all together



$$P(A, G, E, S, C, L, SC) =$$

$$P(A) \cdot P(G) \cdot$$

$$P(E \mid A) \cdot P(S \mid A, G) \cdot$$

$$P(C \mid E, S) \cdot$$

$$P(SC \mid C) \cdot P(L \mid C)$$

Razonamiento Probabilístico

- Distribución de probabilidad conjunta
  - Dada una serie de variables aleatorias $\{x1, x2, ..., xn\}$ ordenadas arbitrariamente, podemos calcular la probabilidad conjunta de las variables, $p(x1, x2, ..., xn)$, como:
    - $p(x1, x2, ..., xn) = p(xn / xn\text{-}1, ..., x1) \ldots p(x3 / x2, x1)\, p(x2 / x1)\, p(x1)$
  - Esta ecuación puede simplificarse si tenemos en cuenta que el conjunto de nodos padre de un determinado nodo *xi*, representado como *Padres*(*xi*) “protege” a *xi* de la influencia de sus otros predecesores.
  - Es decir, si tenemos un conjunto de nodos $\{y1, ..., yn\}$ predecesores de *xi*, pero que no pertenecen a *Padres*(*xi*), entonces:
    - $p(xi / Padres(xi), y1, ..., yn) = p(xi / Padres(xi))$

Razonamiento Probabilístico

- De esta forma podemos expresar la distribución de probabilidad conjunta de la red de creencia como:

$$p(x_1,\ldots,x_n) = \prod_{i=1}^{n} p(x_i \,/\, Padres(x_i))$$

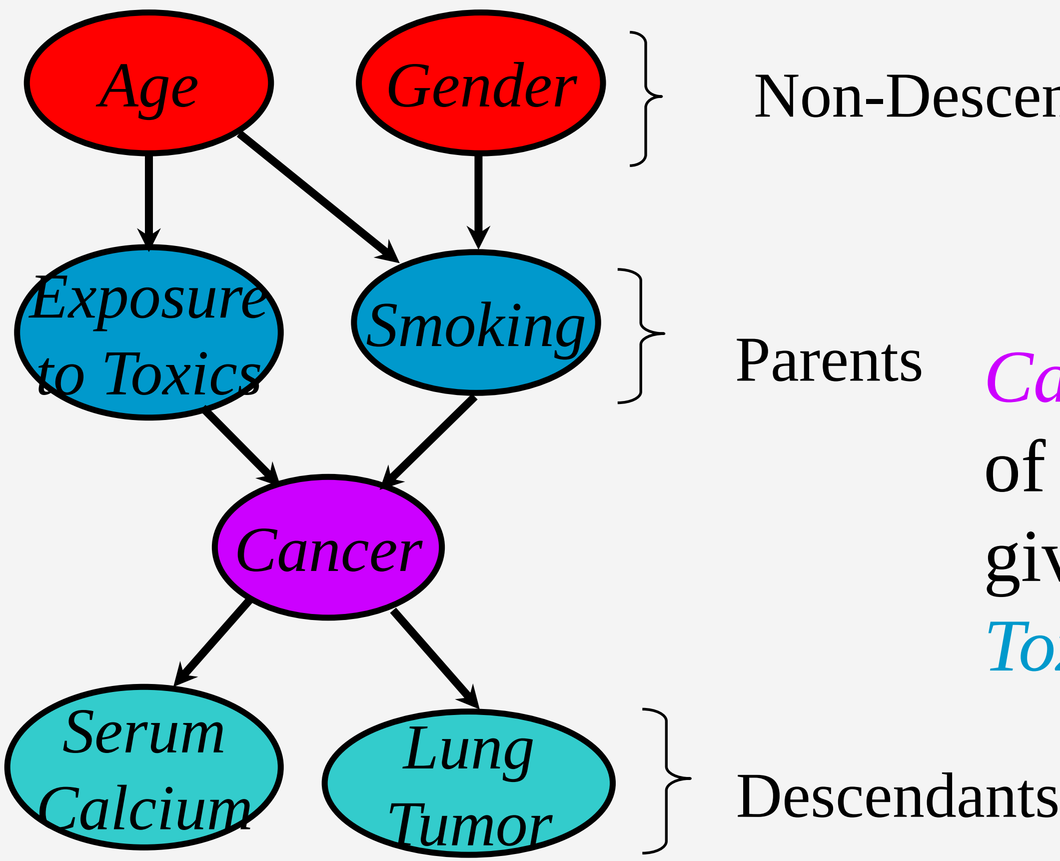
General Product (Chain) Rule for Bayesian Networks

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \boldsymbol{Pa_i})$$

$$\boldsymbol{Pa_i} = parents(X_i)$$

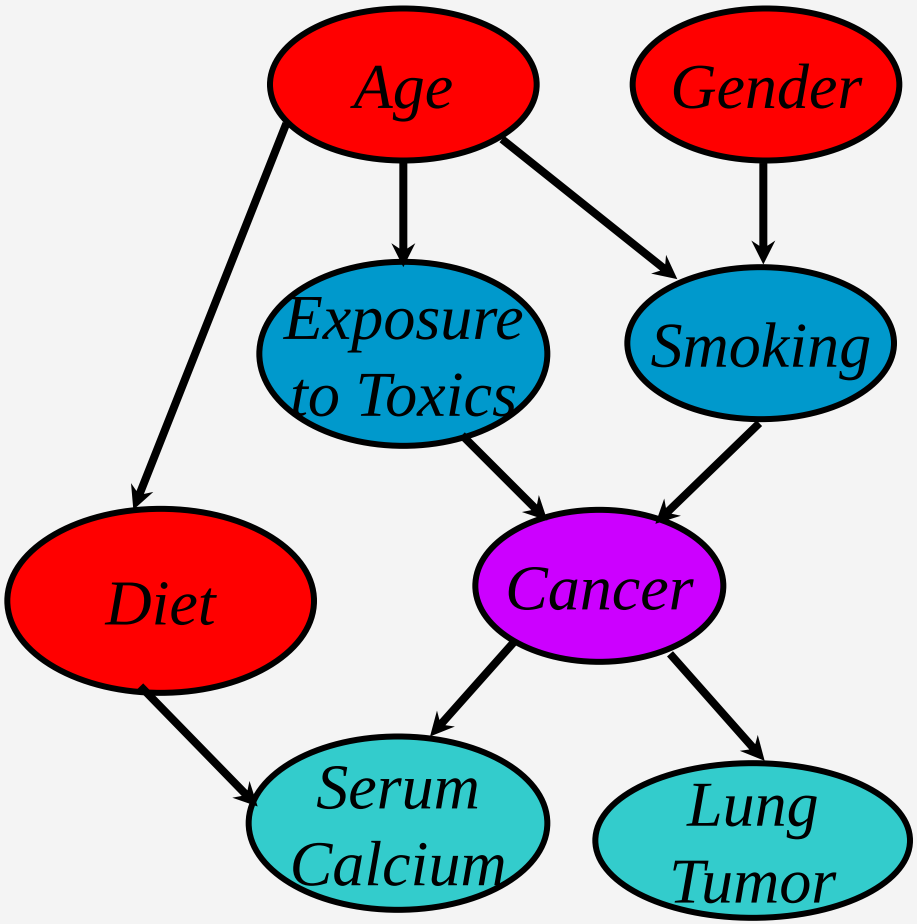
Conditional Independence

A variable (node) is conditionally independent of its non-descendants given its parents.



Cancer is independent of *Age* and *Gender* given *Exposure to Toxics* and *Smoking*.

Another non-descendant



Cancer is independent of *Diet* given *Exposure to Toxics* and *Smoking*.

Razonamiento Probabilístico

- Inferencia en Redes de Creencia
  - La forma más simple de inferencia que puede aparecer en una red de creencia es aquella que se refiere al cálculo de las probabilidades $p(Q=q0)$ para una determinada variable de consulta Q y un determinado valor $q0$.
  - Sin embargo la frase “inferencia en redes de creencia” generalmente se refiere al cálculo de las probabilidades $p(Q=q0 / E=e0)$, en donde Q es la variable de consulta, E es una lista de evidencias observadas y $e0$ son sus correspondientes valores.

Razonamiento Probabilístico

- **Inferencias por diagnóstico**.
  - Que van de las evidencias a las causas.
  - También se conoce como razonamiento abductivo ya que intenta encontrar las causas que mejor explican los efectos observados.

Razonamiento Probabilístico

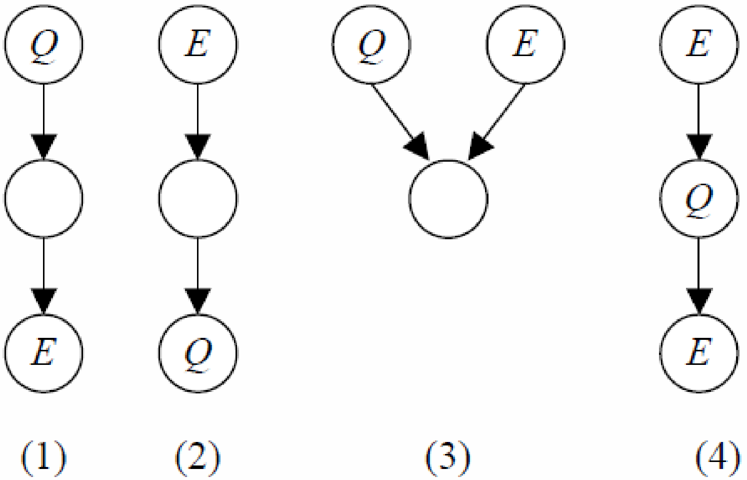
- **Inferencias causales.**
  - Que van de las causas a los efectos, $p(S1/E)$.
  - También se conoce como razonamiento deductivo o razonamiento predictivo cuando se refiere a efectos que se van a manifestar en el futuro

Razonamiento Probabilístico

- **Inferencias intercausales**.
  - Cuando las inferencias se realizan entre las causas de un efecto común.
  - Así hemos visto antes como cambiaba la probabilidad $p(C1/E)$ cuando se añade una nueva evidencia que también apunta a E, $p(S1/E, C2)$, es decir, sabiendo que se tiene sida y hemofilia la probabilidad de drogadicción disminuye.
  - Es lo que se conoce como “explaining away” o disculparse dando explicaciones.

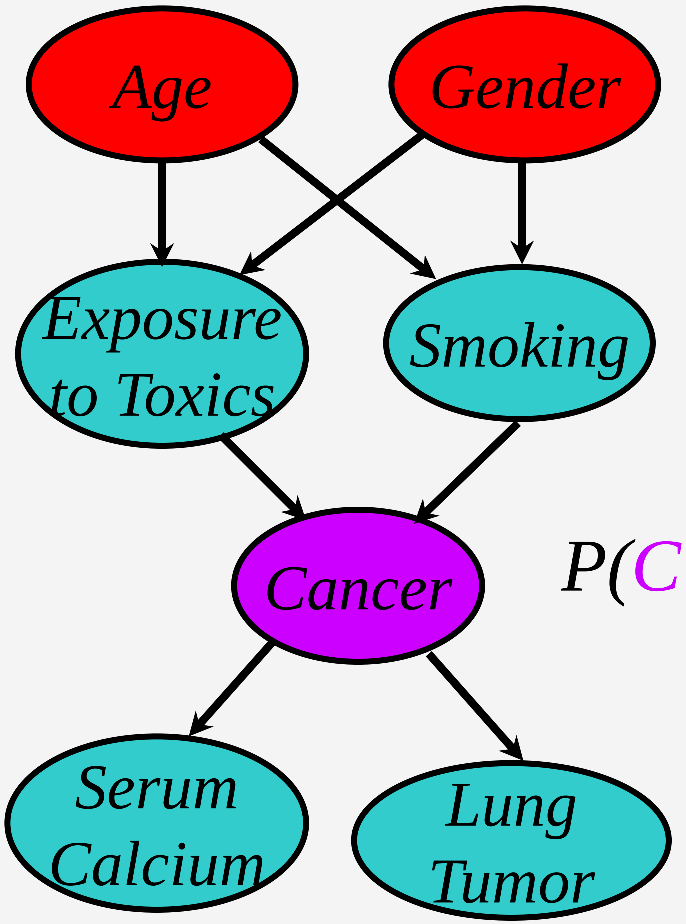
Razonamiento Probabilístico

- **Inferencias mixtas.**
  - Son una combinación de una o varias de las inferencias anteriores. $p(E/S1, C1)$.



Tipos de inferencia (1) diagnóstico, (2) causal, (3) intercausal y (4) mixta.

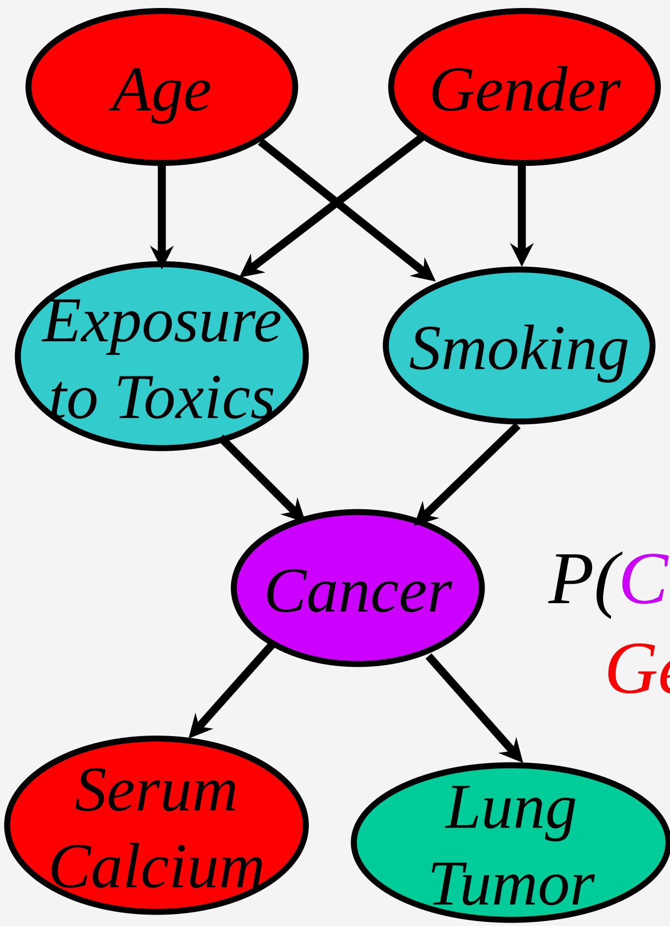
Predictive Inference



How likely are elderly males to get malignant cancer?

$P(C{=}malignant \mid Age{>}60, Gender{=}\ male)$

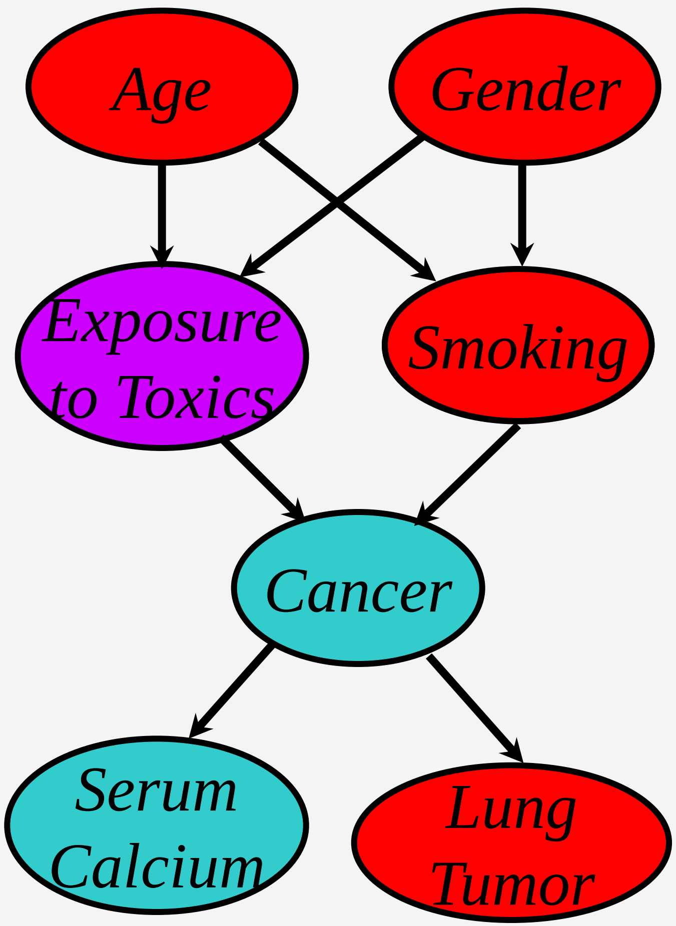
Combined



How likely is an elderly male patient with high Serum Calcium to have malignant cancer?

P(C=malignant | Age>60, Gender= male, Serum Calcium = high)

Explaining away



- If we see a lung tumor, the probability of heavy smoking and of exposure to toxics both go up.
- If we then observe heavy smoking, the probability of exposure to toxics goes back down.

Inference in Belief Networks

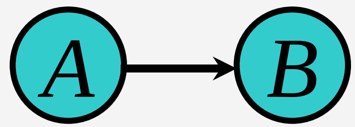
- Find $P(Q=q|\boldsymbol{E}= \boldsymbol{e})$
  - Q the query variable
  - $\boldsymbol{E}$ set of evidence variables

$$P(q \mid \boldsymbol{e}) = \frac{P(q, \boldsymbol{e})}{P(\boldsymbol{e})}$$

$X_1, \ldots, X_n$ are network variables except Q, $\boldsymbol{E}$

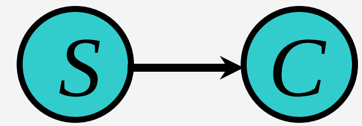
$$P(q, \boldsymbol{e}) = \sum_{x_1, \ldots, x_n} P(q, e, x_1, \ldots, x_n)$$

Basic Inference



$$P(b) = \text{?}$$

Product Rule



- $P(C,S) = P(C|S)\,P(S)$

$S\Downarrow$ $C\Rightarrow$	*none*	*benign*	*malignant*
no	0.768	0.024	0.008
light	0.132	0.012	0.006
heavy	0.035	0.010	0.005

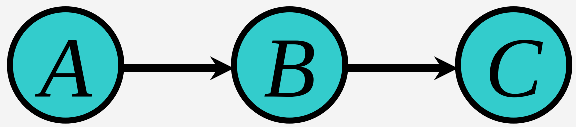
Marginalization

S⇓ *C*⇒	*none*	*benign*	*malig*	total
no	0.768	0.024	0.008	.80
light	0.132	0.012	0.006	.15
heavy	0.035	0.010	0.005	.05
total	0.935	0.046	0.019	

P(Smoke)

P(Cancer)

Basic Inference



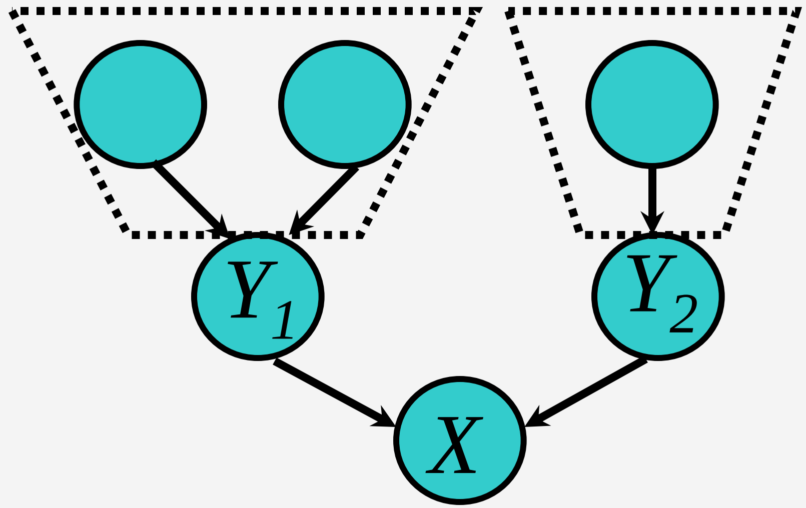
$$\underbrace{P(b)} = \sum_a P(a, b) = \sum_a P(b \mid a)\, P(a)$$

$$P(c) = \sum_b P(c \mid b)\, \overbrace{P(b)}$$

$$P(c) = \sum_{b,a} P(a, b, c) = \sum_{b,a} P(c \mid b)\, P(b \mid a)\, P(a)$$

$$= \sum_b P(c \mid b) \underbrace{\sum_a P(b \mid a)\, P(a)}_{P(b)}$$

Inference in trees



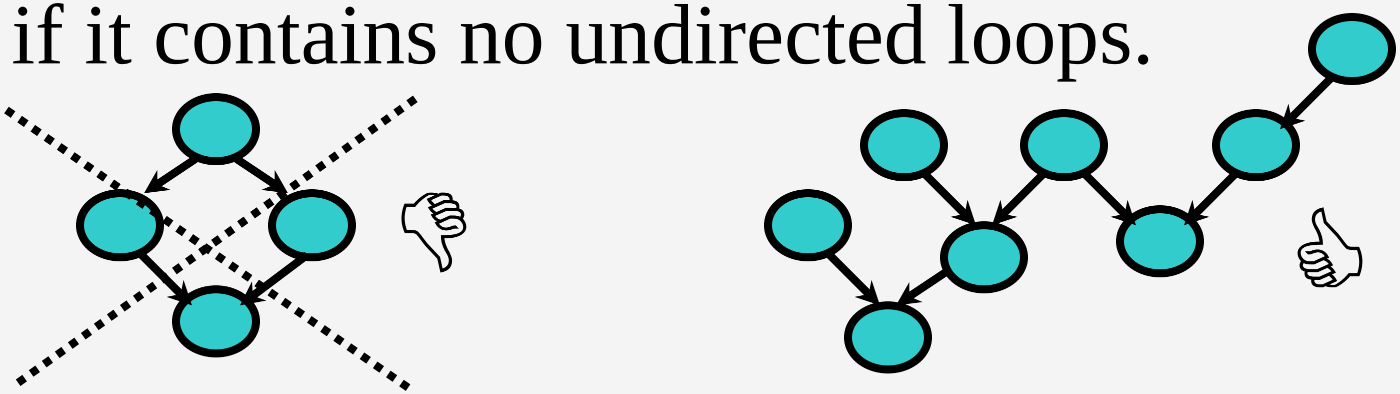
$$P(x) = \sum_{y_1, y_2} P(x \mid y_1, y_2)\, P(y_1, y_2)$$

because of independence of Y_1, Y_2:

$$= \sum_{y_1, y_2} P(x \mid y_1, y_2)\, P(y_1)\, P(y_2)$$

Polytrees

- A network is *singly connected* (a *polytree*) if it contains no undirected loops.



Theorem: Inference in a singly connected network can be done in linear time.

Main idea: in variable elimination, need only maintain distributions over single nodes.

Razonamiento Probabilístico

- La distribución de probabilidad conjunta puede utilizarse para computar cualquier probabilidad de interés de la red.
- Así, por ejemplo, supongamos que estamos interesados en conocer la probabilidad a priori de padecer la enfermedad *E*, es decir, estamos interesados en conocer la probabilidad $p(E+)$.
- Esta probabilidad puede ser obtenida de la distribución de probabilidad conjunta de la siguiente forma:

$$p(E_+) = \sum_{C_{1i}, C_{2i}, S_{1i}, S_{2i}} p(C_{1i}, C_{2i}, E_+, S_{1i}, S_{2i})$$

Razonamiento Probabilístico

- Si lo que queremos es conocer la probabilidad de padecer la enfermedad *E* sabiendo que tenemos los síntomas *S*1 y *S*2, es decir, estamos interesados en conocer la probabilidad *p*(*E*+/*S*1+, *S*2+).
- La propia definición de probabilidad condicional nos permite decir que *p*(*E*+/*S*1+, *S*2+) = *p*(*E*+, *S*1+, *S*2+) / *p*(*S*1+, *S*2+) y estas últimas probabilidades pueden ser obtenidas también obtenidas de la distribución de probabilidad conjunta:

$$p(E_+ / S_{1+}, S_{2+}) = \frac{p(E_+, S_{1+}, S_{2+})}{p(S_{1+}, S_{2+})} = \frac{\sum_{C_{1i}, C_{2i}} p(C_{1i}, C_{2i}, E_+, S_{1+}, S_{2+})}{\sum_{C_{1i}, C_{2i}, E_i} p(C_{1i}, C_{2i}, E_i, S_{1+}, S_{2+})}$$

Razonamiento Probabilístico

- Una forma más eficiente de calcular estas probabilidades es utilizar la estructura de independencia contenida en la propia red de creencia

$$p(E_+) = \sum p(S_{1i}/E_+)p(S_{2i}/E_+)p(E_+/C_{1i},C_{2i})p(C_{1i})p(C_{2i})$$

$$p(E_+) = \sum_{S_{1i},S_{2i}} p(S_{1i}/E_+)p(S_{2i}/E_+) \sum_{C_{1i},C_{2i}} p(E_+/C_{1i},C_{2i})p(C_{1i})p(C_{2i})$$

$$p(E_+) = \sum_{S_{1i}} \left[p(S_{1i}/E_+) \sum_{S_{2i}} p(S_{2i}/E_+) \right] \sum_{C_{1i}} \left[p(C_{1i}) \sum_{C_{2i}} \left[p(E_+/C_{1i},C_{2i})p(C_{2i}) \right] \right]$$

Razonamiento Probabilístico

Para el caso de esta red en concreto es fácil comprobar que $\sum_{S_{1i},S_{2i}} p(S_{1i} / E_+) p(S_{2i} / E_+) = 1$ con lo que el cálculo de $p(E_+)$ se reduce a

$$p(E_+) = \sum_{C_{1i}} \left[p(C_{1i}) \sum_{C_{2i}} \left[p(E_+ / C_{1i}, C_{2i}) p(C_{2i}) \right] \right] =$$

$$= p(C_{1+}) \left[p(E_+ / C_{1+}, C_{2+}) p(C_{2+}) + p(E_+ / C_{1+}, C_{2-}) p(C_{2-}) \right] +$$
$$+ p(C_{1-}) \left[p(E_+ / C_{1-}, C_{2+}) p(C_{2+}) + p(E_+ / C_{1-}, C_{2-}) p(C_{2-}) \right] =$$

$$= 0.30 \times [0.95 \times 0.20 + 0.90 \times 0.80] + 0.70 \times [0.29 \times 0.20 + 0.01 \times 0.80] =$$
$$= 0.30 \times 0.91 + 0.70 \times 0.066 = 0.319$$

Razonamiento Probabilístico

Actuando de la misma forma podemos calcular las probabilidades a priori de S_{1+} y S_{2+}

$$P(S_{1+}) = \sum_{E_i} p(S_{1+} / E_i) p(E_i) = p(S_{1+} / E_+) p(E_+) + p(S_{1+} / E_-) p(E_-) =$$

$$= 0.80 \times 0.319 + 0.20 \times 0.681 = 0.392$$

$$P(S_{21+}) = \sum_{E_i} p(S_{2+} / E_i) p(E_i) = p(S_{21+} / E_+) p(E_+) + p(S_{2+} / E_-) p(E_-) =$$

$$= 0.70 \times 0.319 + 0.4 \times 0.681 = 0.496$$

Razonamiento Probabilístico

De este modo obtenemos las probabilidades a priori de cada nodo. Sin embargo lo que suele ser más interesante en las redes de creencia es el cálculo de las probabilidades a posteriori, es decir, cual es la probabilidad de tener la enfermedad E, si se tiene un determinado síntoma S_1 o S_2. Esto puede hacerse aplicando la regla de Bayes entre los nodos de la red de la siguiente forma:

$$p(E_+ / S_{1+}) = \frac{p(S_{1+} / E_+) p(E_+)}{p(S_{1+})} = \frac{0.8 \times 0.319}{0.392} = 0.652$$

$$p(E_+ / S_{2+}) = \frac{p(S_{2+} / E_+) p(E_+)}{p(S_{2+})} = \frac{0.7 \times 0.319}{0.496} = 0.451$$

Si los dos síntomas aparecen simultáneamente el cálculo debería ser

$$p(E_+ / S_{1+}, S_{2+}) = \frac{p(S_{1+}, S_{2+} / E_+) p(E_+)}{p(S_{1+}, S_{2+})}$$

Razonamiento Probabilístico

El valor de $p(S_{1+}, S_{2+} / E_+)$ no puede extraerse directamente de las tablas de probabilidades condicionales, sin embargo podemos aplicar la separación condicional para obtener una expresión más sencilla

$$p(S_{1+}, S_{2+} / E_+) = p(S_{1+} / E_+) p(S_{2+} / E_+)$$

con lo que llegamos a

$$p(E_+ / S_{1+}, S_{2+}) = \frac{p(S_{1+} / E_+) p(S_{2+} / E_+) p(E_+)}{p(S_{1+}, S_{2+})} = \alpha \times p(S_{1+} / E_+) p(S_{2+} / E_+) p(E_+)$$

Razonamiento Probabilístico

En este caso se ha representado $1/p(S_{1+}, S_{2+})$ como α para indicar que no se necesita calcular este valor a partir de los datos de la red, sino que puede obtener por normalización. Así tenemos que

$$p(E_+ / S_{1+}, S_{2+}) = \alpha \times p(S_{1+} / E_+) p(S_{2+} / E_+) p(E_+) = \alpha \times 0.8 \times 0.7 \times 0.319 = \alpha \times 0.179$$

$$p(E_- / S_{1+}, S_{2+}) = \alpha \times p(S_{1+} / E_-) p(S_{2+} / E_-) p(E_-) = \alpha \times 0.2 \times 0.4 \times 0.681 = \alpha \times 0.054$$

Como las probabilidades deben sumar la unidad obtenemos que

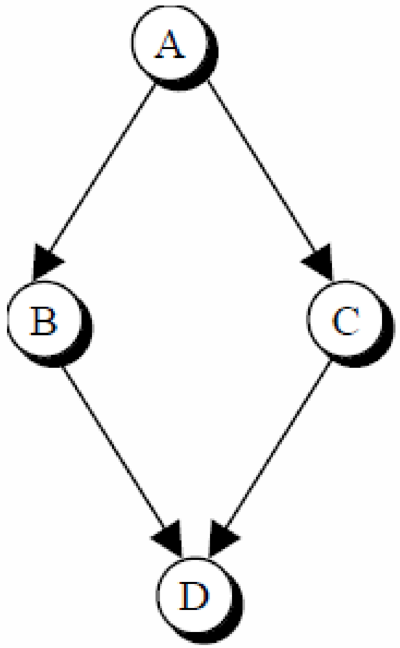
$$\alpha \times 0.179 + \alpha \times 0.054 = 1 \quad \Rightarrow \quad \alpha \times 0.233 = 1 \quad \Rightarrow \quad \alpha = \frac{1}{0.233} = 4.288$$

por lo tanto

$$p(E_+ / S_{1+}, S_{2+}) = \alpha \times 0.179 = 0.766$$
$$p(E_- / S_{1+}, S_{2+}) = \alpha \times 0.054 = 0.234$$

Razonamiento Probabilístico

Kim y Pearl (1983) desarrollaron un algoritmo eficiente para la inferencia en redes con topología de poliárbol. La principal característica de este algoritmo es que su complejidad es lineal en el tamaño de la red (es decir en el número de nodos y aristas que la componen), a diferencia del método de fuerza bruta que requiere un número exponencial de operaciones para realizar la propagación.



Razonamiento Probabilístico

- ***Métodos aproximados de inferencia***
  - Los métodos aproximados de inferencia son aquellos métodos que calculan las probabilidades condicionales de los nodos de forma aproximada, y se utilizan en aquellos casos en los que los algoritmos exactos no son aplicables, o son computacionalmente costosos.
  - La idea básica de los métodos aproximados consiste en generar una muestra de posibles escenarios de la red (en donde cada escenario define una red completamente ejemplificada) y luego utilizar la muestra generada para calcular valores aproximados de las probabilidades de ciertos sucesos dada una determinada evidencia.
  - Los métodos de propagación aproximada pueden clasificarse en dos tipos:
    - *métodos de simulación estocástica*, que generan la muestra usando mecanismos aleatorios
    - *métodos de búsqueda determinista*, que generan la muestra de forma sistemática.

Razonamiento Probabilístico

- Aprendizaje en Redes de Creencia
  - El aprendizaje paramétrico se basa fundamentalmente en encontrar los valores de los parámetros de las distribuciones de probabilidad condicional que maximizan la verosimilitud de los datos suministrados.
  - Estos datos se suelen denominar datos de entrenamiento y se dividen en una serie de casos que se suponen son independientes entre si.

Razonamiento Probabilístico

- Consideremos la estimación de la tabla de probabilidades condicionales del nodo *E* de la red del ejemplo.
- A partir de los datos de entrenamiento podemos contar el número de casos en los que el valor de *E* es True, para cada una de las combinaciones de valores de *C*1 y *C*2, y que representamos como *N*(*E*+, *C*1+, *C*2+), *N*(*E*+, *C*1+, *C*2-), *N*(*E*+, *C*1-, *C*2+) y *N*(*E*+, *C*1-, *C*2-).
- Calculados estos valores podemos dar una estimación de la probabilidad *p*(*E*+ / *C*1+, *C*2+) de la siguiente forma:

$$p(E_+ / C_{1+}, C_{2+}) \approx \frac{N(E_+, C_{1+}, C_{2+})}{N(C_{1+}, C_{2+})}$$

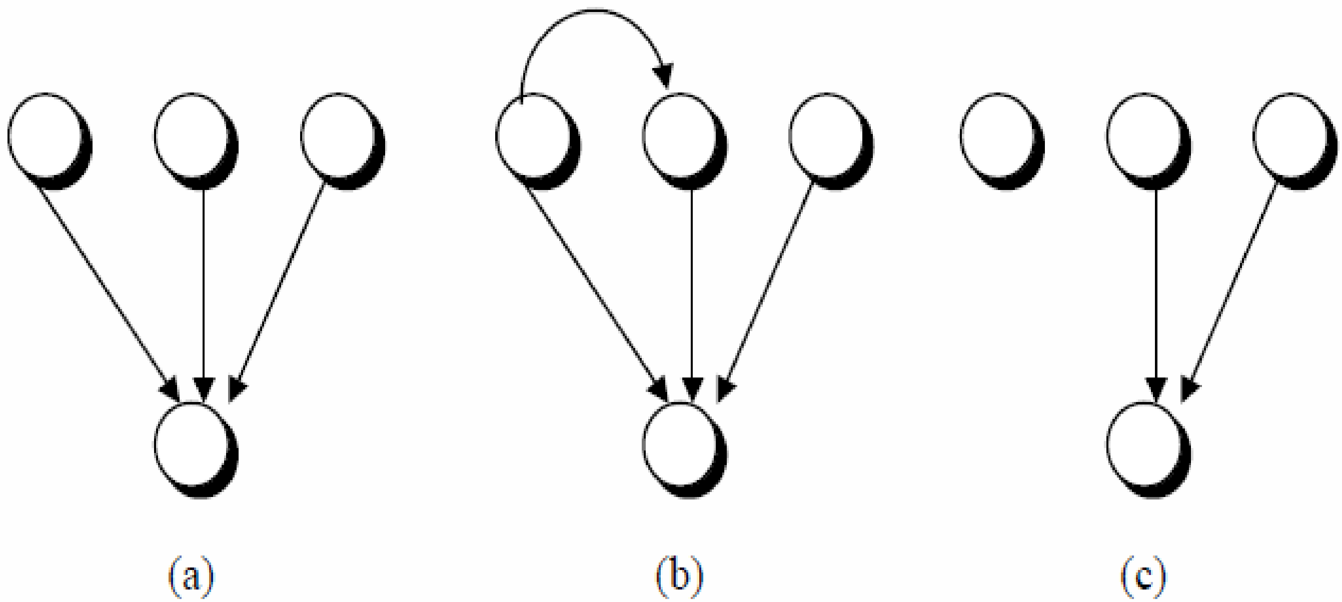
- Actuando de la misma forma para el resto de valores de *C*1+ y *C*2+ podemos obtener la tabla de probabilidades condicionales del nodo *E*.

Razonamiento Probabilístico

- ***Aprendizaje estructural***

El proceso de aprendizaje estructural aparece como paso previo al aprendizaje paramétrico y consiste en decidir que nodos están unidos por un arco y cual es la dirección de dicho arco. En este proceso se pueden cometer dos tipos de errores: añadir un arco de más, con lo que estaremos aumentando el número de parámetros a estimar y estaremos realizando asunciones equívocas acerca de causalidad y la estructura del dominio y; omitir un arco, cuyo efecto no podrá ser compensado por la estimación de los demás parámetros además de perder causalidad y afectar a la estructura del dominio

Razonamiento Probabilístico



Aprendizaje estructural: (a) red a estimar, (b) error por adición de arcos, (c) error por eliminación de arcos.

Razonamiento Probabilístico

Un aspecto que deben tener en cuenta los algoritmos de aprendizaje estructural es que existen redes que son equivalentes. Decimos que dos redes son equivalentes cuando representan las mismas declaraciones de independencia. Por ejemplo supongamos las redes mostradas en la Figura 1.8.

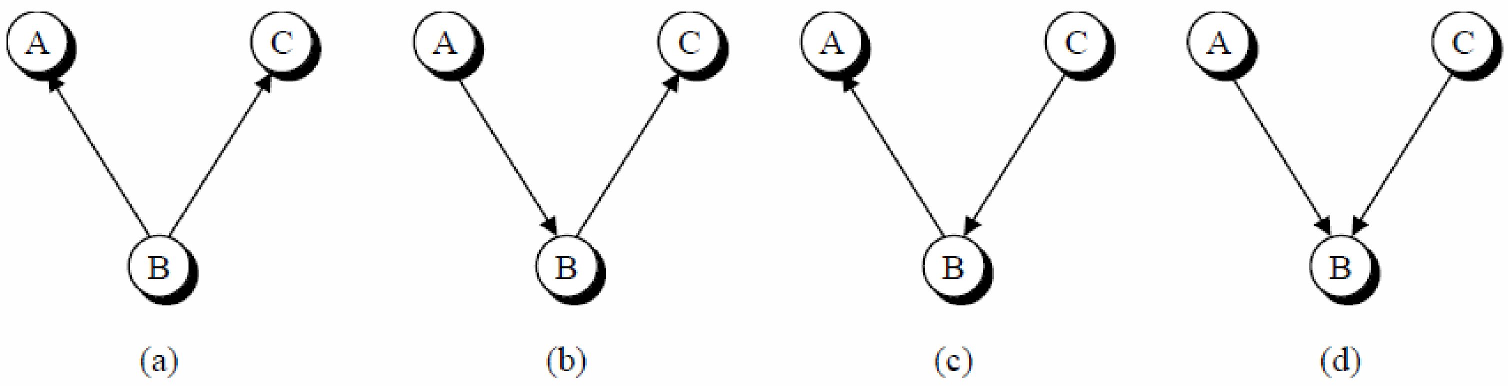


Figura 1.8 Distintos tipos de redes con tres nodos y dos arcos.

Razonamiento Probabilístico

Estas tres primeras redes (a), (b) y (c) son equivalentes porque representan la misma estructura de independencia. Para probar esto podemos ver que la función de probabilidad conjunta de estas redes es:

$$p_a(A,B,C) = p(B)p(A/B)p(C/B)$$

$$p_b(A,B,C) = p(A)p(B/A)p(C/B)$$

$$p_c(A,B,C) = p(C)p(B/C)p(A/B)$$

Razonamiento Probabilístico

Realizando sencillas operaciones sobre las probabilidades conjuntas $p_b\,(A, B, C)$ y $p_c\,(A, B, C)$ aplicando la regla de Bayes podemos ver como resultan ser equivalentes a $p_a\,(A, B, C)$.

$$p_b(A,B,C) = p(A)p(B/A)p(C/B) \xrightarrow{Bayes} = p(A)\frac{p(A/B)p(B)}{p(A)}p(C/B)$$

$$\xrightarrow{simplificando} = p(B)p(A/B)p(C/B) = p_a(A,B,C)$$

$$p_c(A,B,C) = p(C)p(B/C)p(A/B) \xrightarrow{Bayes} = p(C)\frac{p(C/B)p(B)}{p(C)}p(A/B)$$

$$\xrightarrow{simplificando} = p(B)p(A/B)p(C/B) = p_a(A,B,C)$$

Razonamiento Probabilístico

Sin embargo la red de la Figura 1.8 (d) no es equivalente a las demás ya que su distribución de probabilidad conjunta es

$$p_d(A,B,C) = p(A)p(C)p(B/A,C)$$