

Sistemas Difusos

Lógica difusa y sistemas difusos aplicados al análisis inteligente de datos

Juan A. Botía Blaya y Mercedes Valdés Vela

juanbot@um.es, mvaldes@dif.um.es

Departamento de Ingeniería de la Información y las Comunicaciones

Universidad de Murcia

Conjuntos difusos

- Definición: los conjuntos difusos sirven para realizar una evaluación cualitativa de alguna cantidad física [Zad65].
- En los conjuntos difusos se establece un *grado de pertenencia*, de forma que un elemento pertenece a un conjunto difuso con cierto grado.
- Un *conjunto difuso* A en el dominio \mathcal{X} se define mediante un conjunto de pares ordenados:

$$A = \{(x, \mu_A(x)) \mid x \in \mathcal{X}\}$$

donde $\mu_A(x)$ es la *función de pertenencia* para el conjunto difuso A :

$$\mu_A : \mathcal{X} \rightarrow [0, 1]$$

- La función de pertenencia asigna a cada elemento $x \in \mathcal{X}$ un valor entre 0 y 1, dicho valor es el *grado de pertenencia* de x al conjunto A .
- \mathcal{X} es el *universo de discurso* (discreto o continuo)

Definiciones básicas

- El *soporte* de un conjunto difuso A es el conjunto de todos los puntos $x \in \mathcal{X}$ tales que su función de pertenencia es mayor que 0:

$$\text{soporte}(A) = \{x \in \mathcal{X} \mid \mu_A(x) > 0\}$$

- El *núcleo* de un conjunto difuso A es el conjunto de todos los puntos $x \in \mathcal{X}$ tales que su función de pertenencia es igual a 1:

$$\text{núcleo}(A) = \{x \in \mathcal{X} \mid \mu_A(x) = 1\}$$

- Un conjunto difuso A es *normal* si su núcleo es no vacío, es decir, si siempre podemos encontrar un punto $x \in \mathcal{X}$ tal que $\mu_A(x) = 1$.
- Se dice que A es un *conjunto difuso singleton* si su soporte es un solo punto $x \in \mathcal{X}$ con $\mu_A(x) = 1$.
- Un conjunto difuso A es *convexo* si y solo si para todo $x_1, x_2 \in \mathcal{X}$ y para todo $\lambda \in [0, 1]$:

$$\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \min\{\mu_A(x_1), \mu_A(x_2)\}$$

de forma alternativa, A es convexo si A_α es convexo, para todo $\alpha \in [0, 1]$.

Operaciones con cjtos. difusos

- **T -norma/Intersección difusa:** La intersección entre dos conjuntos A y B es otro conjunto difuso $C = A \cap B$ cuya función de pertenencia se calcula a partir de las funciones de pertenencia de A y B mediante una función $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ que realiza la agregación de dos grados de pertenencia de la siguiente forma:

$$\mu_C(x) = \mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x)) = \mu_A(x) \tilde{*} \mu_B(x)$$

donde $\tilde{*}$ es un operador binario para la función T .

- Satisfacen, al menos, los siguientes axiomas para todo $a, b, c \in [0, 1]$:

condición de límite: $T(0, 0) = 0, T(a, 1) = T(1, a) = a$

monótona: $T(a, b) \leq T(c, d)$ si $a \leq c$ y $b \leq d$

conmutativa: $T(a, b) = T(b, a)$

asociativa: $T(a, T(b, c)) = T(T(a, b), c)$

Operaciones con cjtos. difusos (II)

- Algunos de los operadores T -norma más frecuentes son:

mínimo: $T_{min}(a, b) = \min(a, b) = a \wedge b$ (Zadeh)

producto algebraico: $T_p(a, b) = ab$

producto acotado: $T_{pa}(a, b) = \max(0, a + b - 1) = 0 \vee (a + b - 1)$ (Luckasiewicz)

Producto drástico: $T_{pd}(a, b) = \{a, \text{ si } b = 1; b \text{ si } a = 1; 0 \text{ si } a, b < 1\}$

- T -conorma/Unión difusa: La unión de dos conjuntos difusos A y B es otro conjunto difuso $C = A \cup B$ cuya función de pertenencia se calcula a partir de las funciones de pertenencia de A y B mediante una función $S : [0, 1] \times [0, 1] \rightarrow [0, 1]$ que realiza la agregación de dos grados de pertenencia de la siguiente forma:

$$\mu_C(x) = \mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x)) = \mu_A(x) \tilde{+} \mu_B(x)$$

donde $\tilde{+}$ es un operador binario para la función S .

Operaciones con cjtos. difusos (II)

● satisfacen, al menos, los siguientes axiomas para todo $a, b, c \in [0, 1]$:

condición de límite: $S(1, 1) = 1, S(0, a) = S(a, 0) = a$

monótona: $S(a, b) \leq S(c, d)$ si $a \leq c$ y $b \leq d$

conmutativa: $S(a, b) = S(b, a)$

asociativa: $S(a, S(b, c)) = S(S(a, b), c)$

● Algunos de los operadores T -conormas más frecuentes son:

máximo: $S_{max}(a, b) = \text{máx}(a, b) = a \vee b$ (Zadeh)

suma algebraica: $S_s(a, b) = a + b - ab$

suma acotada: $S_{sa}(a, b) = \text{mín}(1, a + b) = 1 \wedge (a + b)$ (Luckasiewicz)

suma drástica : $S_{sd}(a, b) = \{a, \text{ si } b = 0; \text{ y si } a=0; 1 \text{ si } a, b > 0\}$

Funciones de pertenencia

- Función de pertenencia triangular: Esta función se define mediante tres parámetros $\{a, b, c\}$ de la siguiente forma:

$$\mu_{a,b,c}(x) = \begin{cases} 0, & \text{si } x \leq a \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b \\ \frac{c-x}{c-b}, & \text{si } b \leq x \leq c \\ 0, & \text{si } c \leq x \end{cases}$$

Una forma de expresión alternativa es la siguiente:

$$\mu_{a,b,c}(x) = \max \left(\min \left(\frac{x-a}{b-a}, \frac{c-x}{c-b} \right), 0 \right)$$

Los parámetros $\{a, b, c\}$ determinan las coordenadas de las tres esquinas del triángulo definido por la función.

Funciones de pertenencia (II)

- Función de pertenencia trapezoidal: se especifica mediante cuatro parámetros $\{a, b, c, d\}$ de la forma:

$$\mu_{a,b,c,d}(x) = \begin{cases} 0, & \text{si } x \leq a \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b \\ 1, & \text{si } b \leq x \leq c \\ \frac{d-x}{d-c}, & \text{si } c \leq x \leq d \\ 0, & \text{si } d \leq x \end{cases}$$

O expresado de forma más concisa:

$$\mu_{a,b,c,d}(x) = \max \left(\min \left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right), 0 \right)$$

Los parámetros $\{a, b, c, d\}$ determinan las coordenadas de las cuatro esquinas del trapecio definido por la función.

Funciones de pertenencia (III)

- Función de pertenencia *bell-shaped*: se especifica mediante tres parámetros $\{a, b, c\}$ de la forma:

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

donde el parámetro b es normalmente positivo (si fuera negativo, la forma de esta función sería la de una campana invertida)

- Función de pertenencia sigmoideal: se define mediante dos parámetros $\{a, c\}$ de la forma:

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

donde a controla la pendiente en el punto de cruce $x = c$. Dependiendo del signo del parámetro a , la función sigmoideal es abierta a derecha o izquierda.

Defuzzyficación

El proceso de defuzzificación permite asociar a un conjunto difuso un número no difuso. Esto se realiza para calcular el valor de salida de los modelos difusos. La defuzzificación puede realizarse de varias formas.

- Centro del área (COA): Asocia el centro del área formada por el número difuso. Este es uno de los métodos más utilizados. Matemáticamente se expresa de la siguiente forma:

$$y = \frac{\int_{\mathcal{Y}} \mu(y)ydy}{\int_{\mathcal{Y}} \mu(y)dy}$$

- Media del máximo (MOM): Realiza la media de los valores máximos del conjunto difuso. Matemáticamente se define de la siguiente forma:

$$y = \frac{\int_{\mathcal{Y}'} ydy}{\int_{\mathcal{Y}'} dy}$$

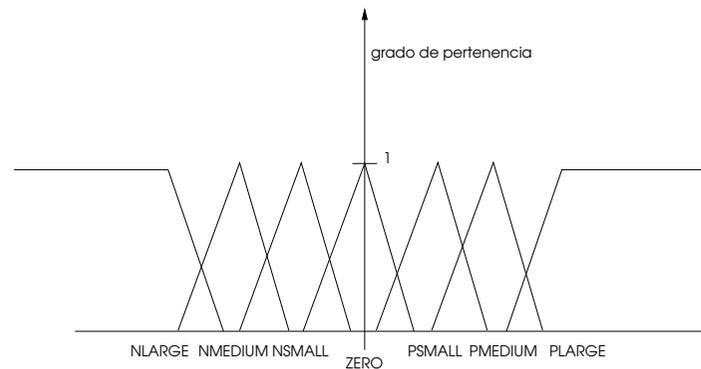
donde $\mathcal{Y}' = \{y \in \mathcal{Y} \mid \mu(y) = \mu^*\}$ y μ^* es el valor máximo de la función de pertenencia.

Modelado Difuso de Sistemas

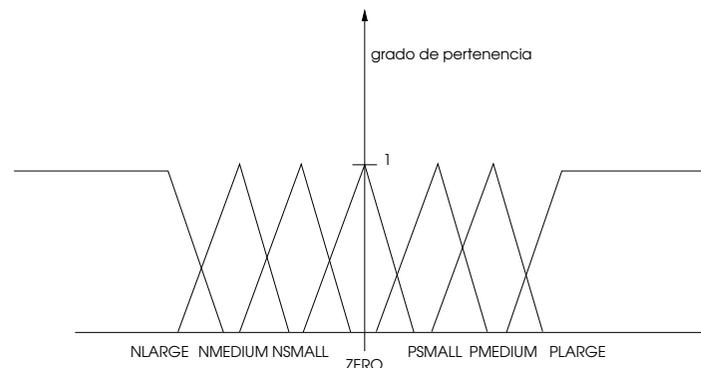
- El modelado difuso es un enfoque relativamente novedoso para la construcción de modelos de sistemas utilizando un lenguaje descriptivo basado en la lógica difusa con predicados difusos [SY93].
- Se describe el comportamiento de los sistemas de forma cualitativa, usando el lenguaje natural.
- Un modelo difuso es una descripción de un sistema con cantidades difusas.
- Las cantidades difusas se expresan como números difusos que pueden tener asociadas variables lingüísticas.

Modelado Difuso de Sistemas

- Ejemplo: el controlador de la carretilla para la maniobra de atraque
- Es una función $f(e_t)$ que proporciona salidas para el par (v_ϕ, v_l)
- Definimos el error de la trayectoria con siete conjuntos difusos triangulares

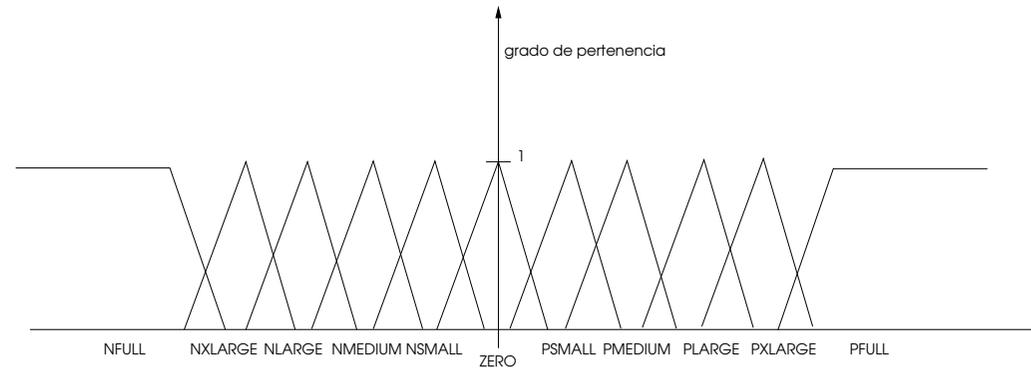


- También la velocidad angular



Modelado Difuso de Sistemas

- También la velocidad lineal con 11 conjuntos triangulares



- Y el conjunto de reglas queda

Antecedente	Consecuente	
Er. trayectoria (7fs)	V. angular (7fs)	V. lineal (11fs)
NLARGE	PLARGE	NZERO
NMEDIUM	PMEDIUM	NLARGE
NSMALL	PSMALL	NXLARGE
ZERO	ZERO	NFULL
PSMALL	NSMALL	NXLARGE
PMEDIUM	NMEDIUM	NLARGE
PLARGE	NLARGE	NZERO

Representación genérica de un modelo difuso

R_1 : SI x_{11} es A_{11} Y x_{12} es A_{12} Y ... Y x_{1p} es A_{1p} ENTONCES y es B_1

R_2 : SI x_{21} es A_{21} Y x_{22} es A_{22} Y ... Y x_{2p} es A_{2p} ENTONCES y es B_2

...

R_r : SI x_{r1} es A_{r1} Y x_{r2} es A_{r2} Y ... Y x_{rp} es A_{rp} ENTONCES y es B_r

- p es el número de variables de entrada
- r es el número de reglas
- x_j con $1 \leq j \leq p$, es la j -ésima variable de entrada
- A_{ij} , con $1 \leq i \leq r$, es el conjunto difuso asociado a la variable de entrada j -ésima en la regla i -ésima
- y es la variable de salida
- B_i es el conjunto difuso asociado a la variable de salida en la i -ésima regla

Ventajas del modelado difuso

- Favorece la incorporación del conocimiento experto existente acerca del sistema que se está modelando
- Alta interpretabilidad
- Cuando existen datos entrada-salida sobre el sistema se pueden utilizar técnicas de identificación de sistemas clásicas para realizar el proceso de modelado
 - **Modelado Difuso Conducido por Datos**

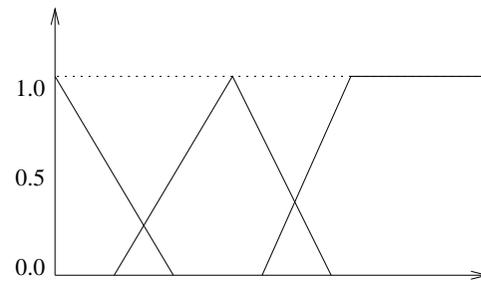
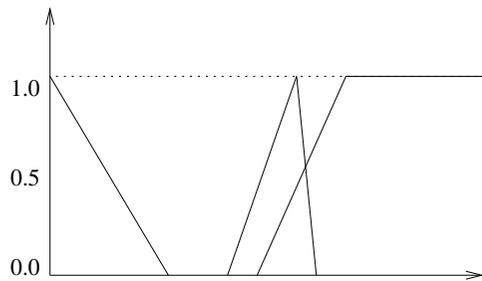
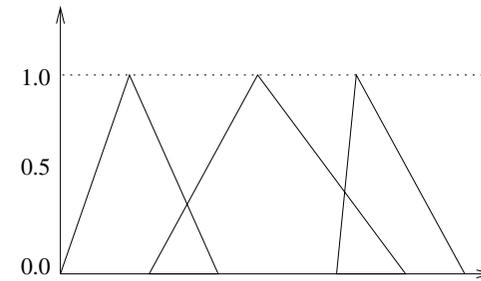
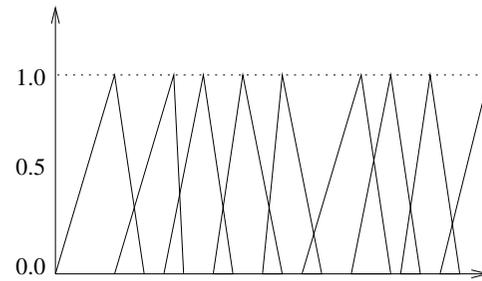
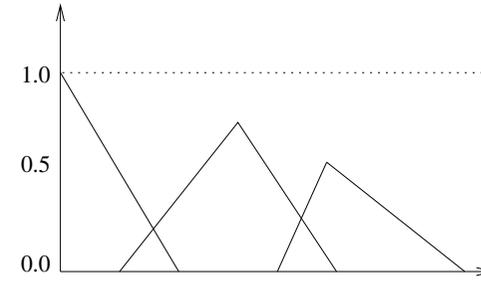
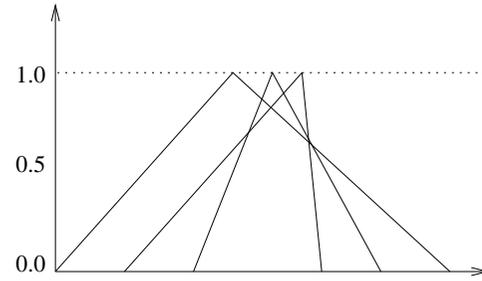
Características Deseables en un Modelo Difuso

- **Objetivo A: conseguir capacidad aproximativa**
 - los modelos difusos son aproximadores universales [Wan92, Kos92, Cas95]
- **Objetivo B: conseguir transparencia o interpretabilidad**
 - capacidad para incorporar conocimiento experto
 - explicar el funcionamiento de un sistema mediante reglas *SI-ENTONCES*
 - nos permiten entender la influencia de cada entrada en la salida

Características Deseables en un Modelo Difuso (II)

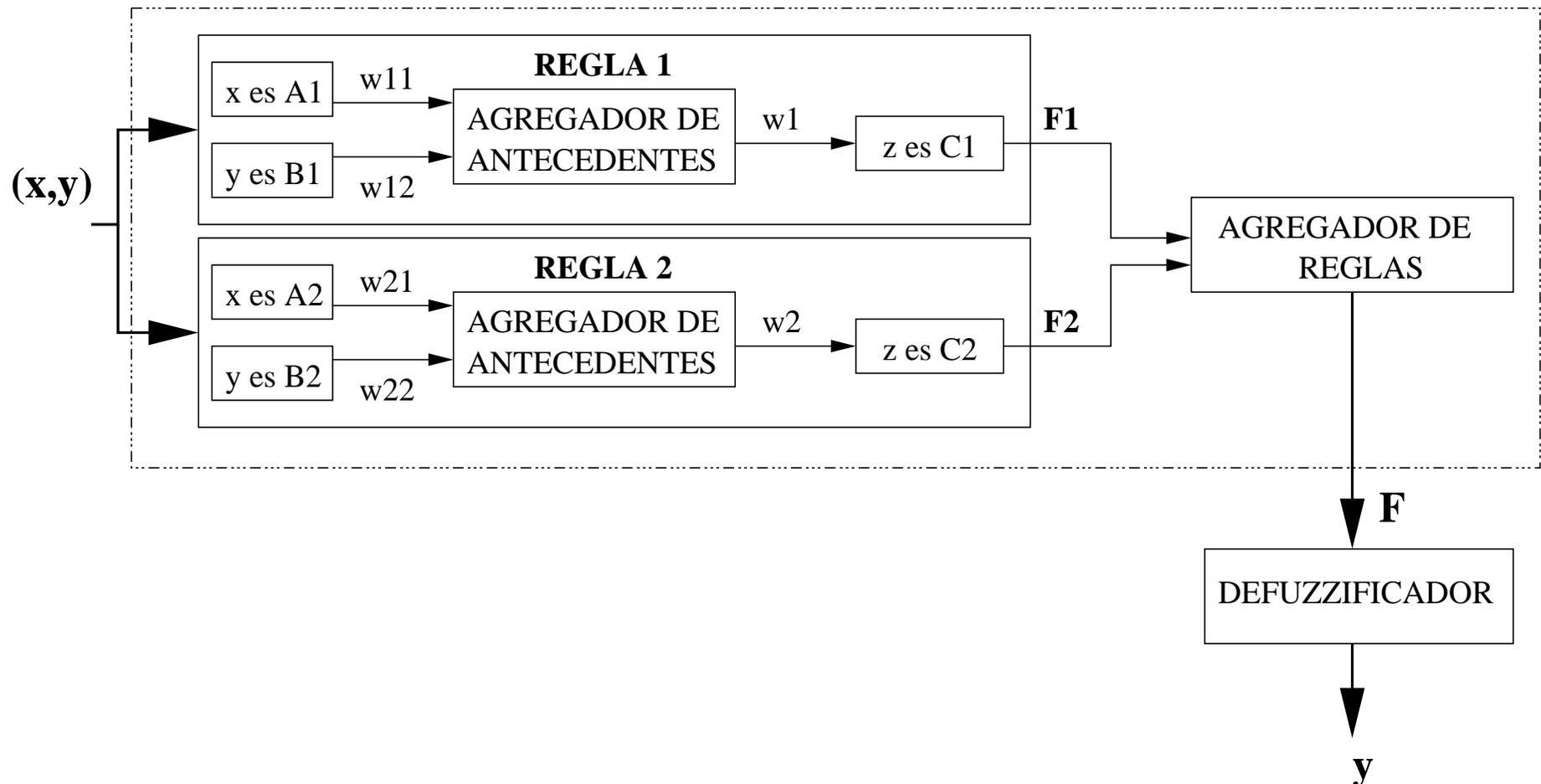
1. **Distinguibilidad:** término lingüístico con significado tanto más claro cuanto más se puedan diferenciar las funciones de pertenencia.
2. **Normalidad:** para cada función de pertenencia debe existir al menos un elemento en el universo con grado de pertenencia máximo (1)
3. **Número de funciones de pertenencia moderado:** número de entidades diferentes que pueden ser manejadas de forma eficiente por la memoria, entre 7 y 9
4. **Cubrimiento:** se debe cubrir todo el espacio de la variable representada

Características Deseables en un Modelo Difuso (III)



Elementos de un Modelo Difuso

Para un sistema con dos reglas



Interpretación de las conectivas difusas

- La entrada al sistema puede ser difusa o real (*crisp*)
- Para obtener la salida debida a la entrada
 1. determinar el **grado de cumplimiento** de cada una de las **proposiciones difusas atómicas** de los antecedentes
 2. calcular el grado de cumplimiento del antecedente de cada regla, agregando las proposiciones atómicas mediante la conectiva *Y*
 3. realizar las implicaciones *SI – ENTONCES*
 4. la agregación de las reglas debida a la conectiva *ADEMÁS*
 5. defusificar
- Para inferir una salida a partir de una entrada hay que proporcionar una interpretación de las conectivas difusas (**interpretación del mecanismo de inferencia**)

Clasificación de modelos difusos

- Dependiendo del número de variables de entrada y salida
 - **SISO** (*Single Inputs Single Output*) formados por una variable de entrada y una de salida
 - **MIMO** (*Multiple Inputs Multiple Outputs*): múltiples variables de entrada y múltiples variables de salida
 - **MISO** (*Multiple Inputs Single Output*) : modelos múltiples entradas y una sola variable de salida
- Variables involucradas:
 - **reglas completas**: en todas las reglas que conforman el modelo difuso, están involucradas las mismas variables de entrada
 - **reglas incompletas**: reglas formadas por subconjuntos (no necesariamente iguales) del conjunto total de variables (aquellas que no aparecen se les asume grado de pertenencia 1)
- según el Tipo de Consecuente
 - **Modelos con consecuente difuso**
 - **Modelos relacionales**
 - **Modelos TSK (Takagi-Sugeno-Kang**

Interpretación del método de razonamiento difuso

Modelos con consecuente difuso

- Conectiva Y (*x* es *A* Y *y* es *B*)

- $\min(\mu_A(c), \mu_B(y))$

- $\prod(\mu_A(c), \mu_B(y))$

- Implicación (Si *x* es *A* ENTONCES *y* es *B*)

- $\min(\mu_A(c), \mu_B(y))$

- $\prod(\mu_A(c), \mu_B(y))$

- Agregación de reglas (agregación equivale a disyunción)

R_1 : Si *x* es A_1 ENTONCES *z* es C_1

ADEMÁS

R_1 : Si *x* es A_2 ENTONCES *z* es C_2

...

R_r : Si *x* es A_r ENTONCES *z* es C_r

- $\mu_{R_{set}}(x, z) = \max(\mu_{R_{c1}}(x, z), \dots, \mu_{R_{cr}}(x, z))$

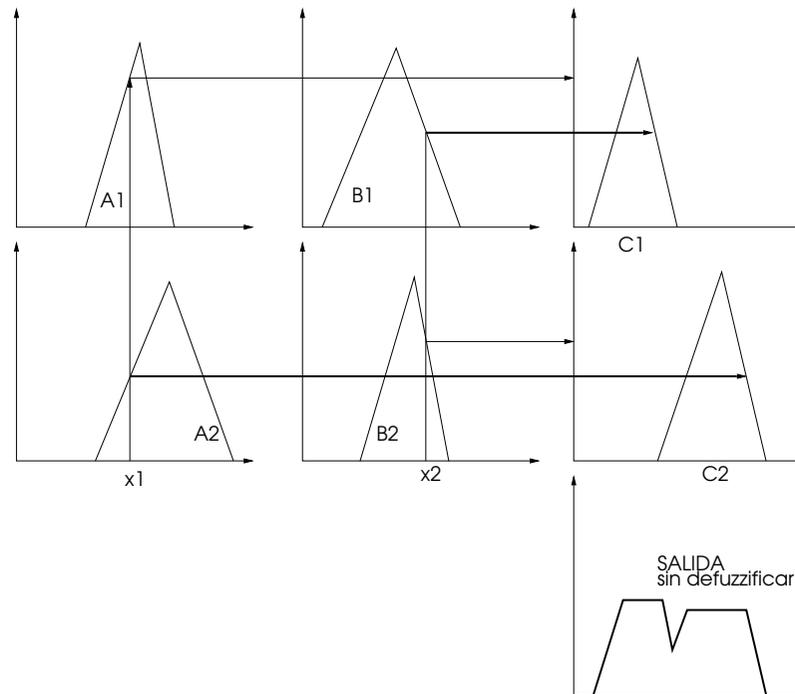
- $\mu_{R_{set}}(x, z) = \sum_{algebraica}(\mu_{R_{c1}}(x, z), \dots, \mu_{R_{cr}}(x, z))$

Ejemplo

● Para el par de reglas

R_1 : SI x es A_1 ENTONCES z es C_1

R_2 : SI x es A_2 ENTONCES z es C_2



Interpretación del método de razonamiento difuso

Modelos TSK: los consecuentes de las reglas son funciones polinomiales de las variables de entrada

Si x_1 es A_{11} Y ... Y x_p es A_{1p} ENTONCES $y_1 = a_{11}(x_1, \dots, x_p)$ Y ... Y $y_h = a_{1h}(x_1, \dots, x_p)$
Si x_1 es A_{21} Y ... Y x_p es A_{2p} ENTONCES $y_1 = a_{21}(x_1, \dots, x_p)$ Y ... Y $y_h = a_{2h}(x_1, \dots, x_p)$
⋮
Si x_1 es A_{r1} Y ... Y x_p es A_{rp} ENTONCES $y_1 = a_{r1}(x_1, \dots, x_p)$ Y ... Y $y_h = a_{rh}(x_1, \dots, x_p)$



Geoméricamente, las reglas del modelo TSK corresponden a una aproximación de la función original mediante una combinación de funciones lineales.



Las funciones lineales pueden remplazarse por otras no lineales



Interpretación

1. Supongamos una única variable a la salida
2. hay que calcular el grado de disparo de cada regla, τ_i , con $i = 1, \dots, r$

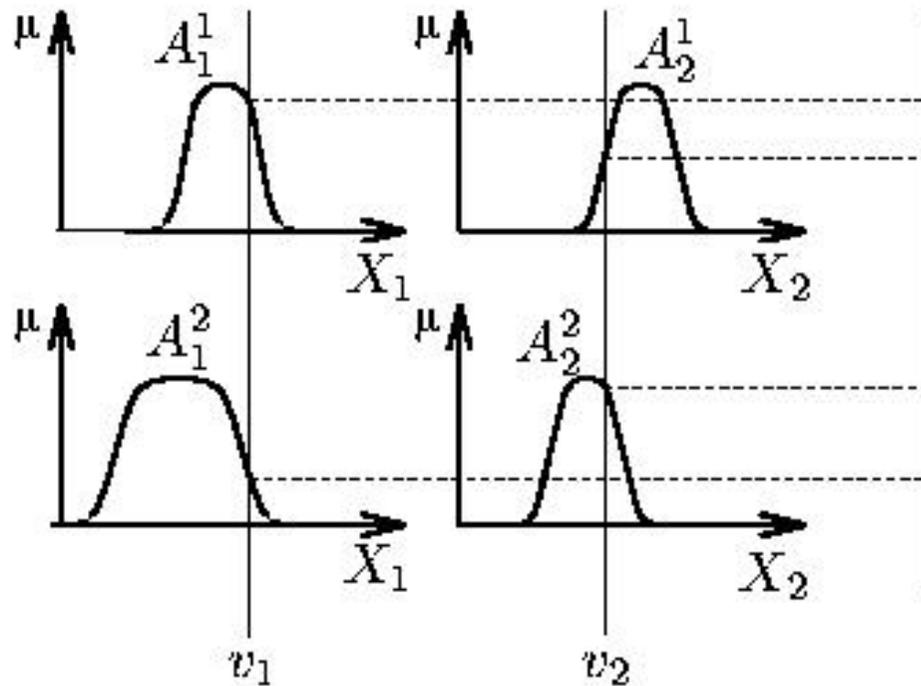
$$\tau_i = T(A_{i1}(x_1), \dots, A_{ip}(x_p)),$$

3. La salida y se obtiene de la siguiente forma:

$$y = \frac{\sum_{i=1}^r \tau_i y_i}{\sum_{i=1}^r \tau_i}$$

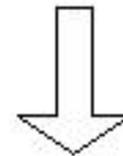
siendo y_i el consecuente de la regla i -ésima

Ejemplo



$$\tau^1 \quad y^1 = p_0^1 + p_1^1 v_1 + p_2^1 v_2$$

$$\tau^2 \quad y^2 = p_0^2 + p_1^2 v_1 + p_2^2 v_2$$



$$y^* = \frac{\tau^1 y^1 + \tau^2 y^2}{\tau^1 + \tau^2}$$

Modelado difuso guiado por datos

- En análisis de datos, buscamos una función objetivo f , que aproximamos mediante \hat{f} , obtenida a partir de datos de aprendizaje
- En el modelado difuso tratamos de encontrar un sistema de inferencia difusa para representar \hat{f} y aproximar f
- Debemos entonces determinar
 - El tipo de sistema de inferencia (TSK, etc)
 - Interpretación de las conectivas
 - El número de conjuntos difusos para cada variable de entrada y salida
 - El número de reglas
 - Los conjtos. difusos de cada regla
 - Otros

Modelado Neuro-Difuso

- los sistemas de inferencia difusos proporcionan un mecanismo intuitivo y de alto nivel para representar el conocimiento
- las redes neuronales [Cyb88, Bis95], poseen un alto grado de adaptabilidad y capacidad de aprendizaje y generalización
- las herramientas que se nutren de estas dos áreas son un mecanismo eficiente a la hora de modelar sistemas reales: las **redes adaptativas neuro-difusas** [CW96, Jan93, JSM97]
- El enfoque más extendido es el de usar backpropagation en sistemas difusos
 1. el controlador difuso se transforma en una red neuronal
 2. la red se entrena mediante *backpropagation*
 3. han de utilizarse operadores diferenciables en el FRM y funciones de pertenencia diferenciables para los conjuntos difusos

Modelado Neuro-Difuso con Redes Neuronales con Funciones de Base Radial (RBF)

Arquitectura

- capa oculta formada funciones de base radial R_i , típicamente R_i es gaussiana:

$$R_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right)$$

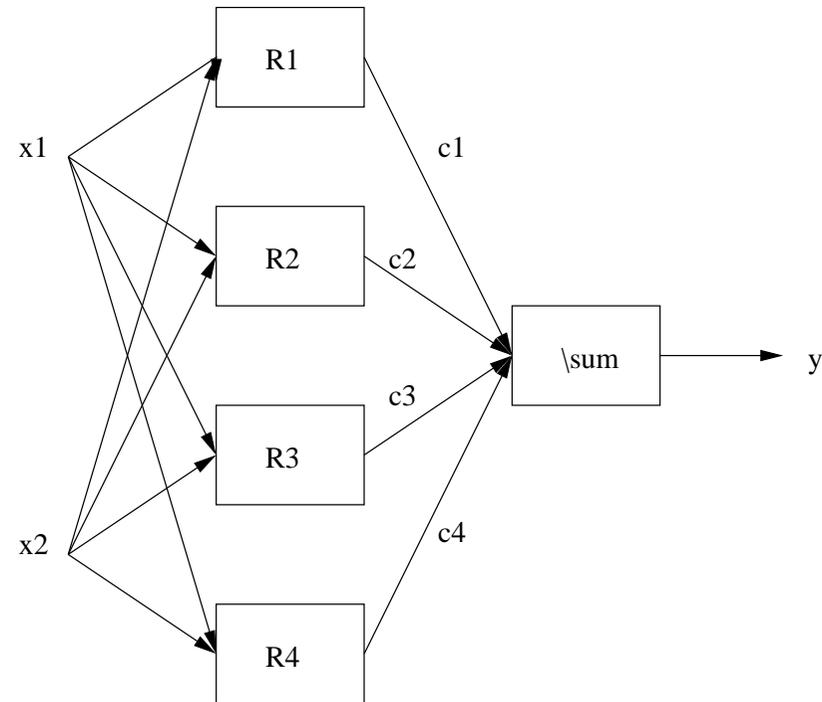
- La capa de salida tiene un solo nodo
 - suma ponderada de las salidas de cada neuronal de la capa oculta:

$$y = \sum_{i=1}^c w_i R_i(x)$$

- media ponderada de los valores de salida asociados con cada neurona:

$$y = \frac{\sum_{i=1}^c w_i R_i(x)}{\sum_{i=1}^c R_i(x)}$$

Modelado Neuro-Difuso con Redes de Funciones de Base Radial (RBFN)



Equivalencia con TSK

- Si asignamos una función lineal a la salida de cada neurona:

$$w_i = f_i(x) = a_i^T x + b_i$$

donde a_i y b_i son los parámetros asociados a la salida de la neurona i -ésima

- la salida obtenida por la red RBFN es:

$$y = \frac{\sum_{i=1}^c R_i(x) f_i(x)}{\sum_{i=1}^M R_i(x)}$$

siendo idéntica a la producida por un sistema de inferencia difuso TSK de primer orden si

- El número de neuronas es igual al número de reglas
- RBFN y FIS tienen el mismo método de agregación (suma ponderada o media ponderada) para calcular la salida total.
- Las funciones de base radial son de pertenencia **multidimensional** para el antecedente de cada regla del sistema
- La salida de cada regla del FIS y de la red RBFN debe tener la misma función de respuesta:
 - en el caso de una función lineal de primer orden, $f_i(x) = a_i^T x + b_i$, resulta equivalente a un sistema de inferencia TSK de primer orden,
 - en el caso de una constante, $f_i(x) = b_i$, entonces resulta equivalente a un sistema de inferencia TSK de orden cero.

Ajuste de parámetros para una red RBF-TSK

- El algoritmo de entrenamiento actualiza de forma incremental los parámetros (tanto de los antecedentes como de los consecuentes) en base a los datos entrada-salida que se le presentan
- La función objetivo en este caso es el error cuadrático medio.
- El grado de disparo de una regla dada equivale al grado de disparo de una neurona y viene dado mediante la composición suma-producto de la forma:

$$R_i(\vec{x}) = \exp \left[- \sum_{j=1}^p \frac{|x_j - c_{ij}|^2}{2\sigma_{ij}^2} \right] = \prod_{j=1}^p \exp \left[- \frac{|x_j - c_{ij}|^2}{2\sigma_{ij}^2} \right] = \prod_{j=1}^p \mu_{A_{ij}}$$

donde: $c_i = [c_{i1}, \dots, c_{ip}]^T$ es el vector de centros de la regla i -ésima
 $\sigma_i = [\sigma_{i1}, \dots, \sigma_{ip}]^T$ es el vector de varianzas de la regla i -ésima A_{ij} , $j = 1, \dots, p$
son los conjuntos difusos del antecedente de la regla i -ésima, definido cada uno por una función de pertenencia gaussiana con parámetros (c_{ij}, σ_{ij}) .

Ajuste de parámetros para una red RBF-TSK (II)

- La salida total del modelo se calcula sumando las contribuciones individuales de cada regla utilizando la media ponderada
- En el caso de una función lineal de primer orden:

$$f_i(x) = \theta_{i1}x_1 + \dots + \theta_{ip}x_n + \theta_{i(p+1)}$$

donde $\theta_i = [\theta_{i1}, \dots, \theta_{i(p+1)}]^T = [a_i^T; b_i]^T$ es el vector de parámetros del consecuente

- Para el patrón de entrenamiento k -ésimo, el error E_k es:

$$E_k = \frac{1}{2} (y_k - t_k)^2$$

donde y_k es la salida real del modelo, y t_k es la salida deseada para el vector de entrada k -ésimo.

Ajuste de parámetros para una red RBF-TSK (III)

- La regla para realizar la actualización es la siguiente:

$$v_i^{new} = v_i^{old} + \eta \Delta v_i = v_i^{old} - \eta \frac{\partial E_k}{\partial v_i}$$

donde: $i = 1, \dots, r$, $v_i = [c_i^T; \sigma_i^T; \theta_i^T]^T$ son los parámetros asociados a la regla i -ésima y η es el ratio de aprendizaje.

- Los gradientes negativos de E_k con respecto a cada parámetro se calculan de la siguiente forma:

$$\Delta c_{ji} = -\frac{\partial E_k}{\partial c_{ji}} = (t_k - y_k) \frac{f_i(x) - y_k}{z} R_i(x) \frac{x_j - c_{ji}}{\sigma_{ji}^2}$$

$$\Delta \sigma_{ji} = -\frac{\partial E_k}{\partial \sigma_{ji}} = (t_k - y_k) \frac{f_i(x) - y_k}{z} R_i(x) \frac{(x_j - c_{ji})^2}{\sigma_{ji}^3}$$

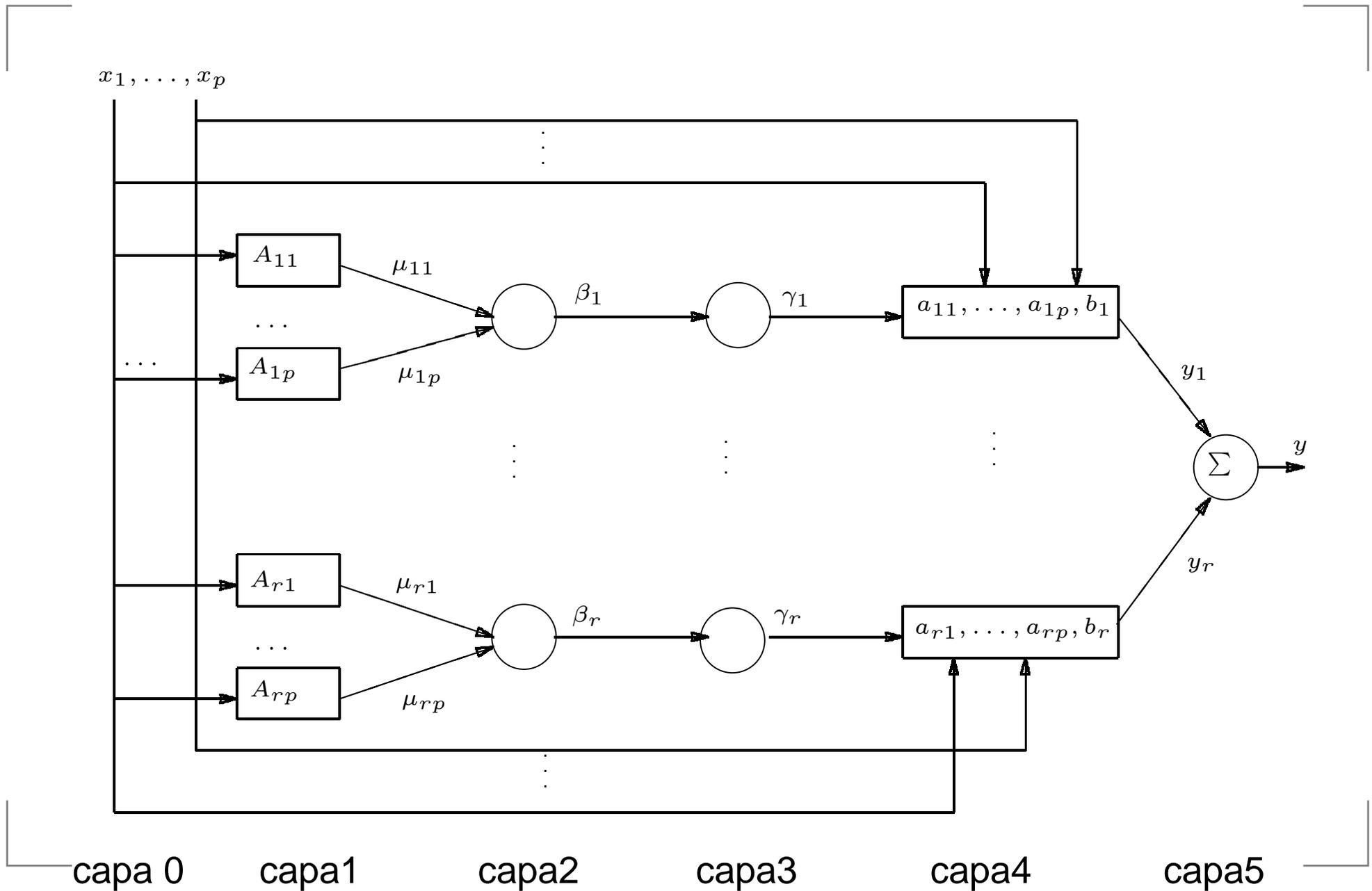
$$\Delta \theta_{ij} = -\frac{\partial E_k}{\partial \theta_{ij}} = (t_k - y_k) \frac{1}{z} R_i(x) x_j$$

donde $i = 1, \dots, r$, $j = 1, \dots, p$, y $z = \sum_{i=1}^c R_i(x)$.

Modelado Neuro-Difuso con ANFIS

- La arquitectura ANFIS (*Adaptive Neuro-based Fuzzy Inference System*) [Jan93, JSM97] es funcionalmente equivalente a un sistema TSK.
- La desarrollada a continuación se corresponde con el modelo TSK de consecuente lineal.
- Los nodos situados en la misma capa realizan funciones similares.
- Los nodos representados con rectángulos son adaptativos, mientras que los nodos representados por círculos no lo son

ANFIS, Arquitectura



Funcionalidad en las capas

- Capa 0: corresponde a las entradas x_1, \dots, x_p .
- Capa 1: los nodos de esta capa son adaptativos, y cada uno calcula el grado de pertenencia μ_{ij} de la entrada j -ésima al conjunto difuso A_{ij} . Su función de pertenencia puede definirse de varias maneras, teniendo en cuenta que ha de ser **diferenciable**. Por ejemplo, podría ser una función *bell-shaped*:

$$A_{ij} = \frac{1}{1 + \left| \frac{x_j - c_j^i}{a_j^i} \right|^{2b_j^i}} \quad i = 1, \dots, c, \quad j = 1, \dots, p,$$

donde x_j es la entrada y $\{a_j^i, b_j^i, c_j^i\}$ es el conjunto de parámetros. O por ejemplo, una gaussiana:

$$A_{ij} = e^{-\left(\frac{x_j - c_j^i}{a_j^i}\right)^2} \quad i = 1, \dots, c, \quad j = 1, \dots, p,$$

donde x_j es la entrada y $\{a_j^i, c_j^i\}$ es el conjunto de parámetros.

Funcionalidad en las capas (II)

- Capa 2: los nodos de esta capa son no adaptativos y se define la salida del nodo i -ésimo como el producto de sus entradas:

$$\beta_i = \prod_{j=1}^p \mu_{ij} \quad i = 1, \dots, r,$$

donde cada salida β_i corresponde al grado de disparo de la i -ésima regla (usando la T -norma del producto).

- Capa 3: los nodos de esta capa son no adaptativos. El nodo i -ésimo obtiene el grado de disparo normalizado de la regla i -ésima

$$\gamma_i = \frac{\beta_i}{\sum_{k=1}^r \beta_k}$$

Funcionalidad en las capas (III)

- Capa 4: cada nodo de esta capa es adaptativo y los parámetros del nodo i -ésimo son $a_{i1}, \dots, a_{ip}, b_i$. Su salida se corresponde con la salida parcial de la i -ésima regla,

$$y_i = \gamma_i (a_{i1}x_1 \dots + a_{ip}x_p + b_i)$$

- Capa 5: el nodo de esta capa es no adaptativo y su salida se define como la suma de las salidas parciales y_i :

$$y = \sum_{i=1}^r y_i$$

Ajuste de parámetros en ANFIS

- Basado en una combinación de mínimos cuadrados y backpropagation
- Los nodos de la capa 1 y cuatro se corresponden, respectivamente, con
 - los parámetros de los conjuntos difusos de los antecedentes de un sistema TSK
 - los coeficientes de las funciones lineales de sus consecuentes

Mínimos cuadrados para los parámetros en los consecuentes

- Si los parámetros de los antecedentes quedan fijos, la salida y puede ser reescrita entonces de la siguiente manera

$$(1) \quad \begin{aligned} y &= \gamma_1(a_{11}x_1 + \dots + a_{1p}x_p + b_1) \\ &+ \dots + \\ &\gamma_r(a_{r1}x_1 + \dots + a_{rp}x_p + b_r) \end{aligned}$$

y esta ecuación puede reescribirse como:

$$(2) \quad \begin{aligned} f &= \gamma_1x_1a_{11} + \dots + \gamma_1x_pa_{1p} + \gamma_1b_1 \\ &+ \dots + \\ &\gamma_rx_1a_{r1} + \dots + \gamma_rx_pa_{rp} + \gamma_rb_r \end{aligned}$$

- lineal en los parámetros de los consecuentes $a_{i1}, \dots, a_{ip}, b_i$ con $i = 1, \dots, p$.
- Se puede aplicar mínimos cuadrados

Método global

- combinar el método de *backpropagation* con el método de cuadrados mínimos para modificar los parámetros de la red
- cada época está compuesta por una pasada hacia adelante y una pasada hacia atrás.
 - Hacia adelante para cada vector de entrada, se evalúa la red hasta la capa 4, y los parámetros de los consecuentes son identificados mediante el método de cuadrados mínimos
 - se calculan los errores para cada par del conjunto de entrenamiento y, en la pasada hacia atrás se propagan las señales del error y los parámetros de las premisas son modificados por el mecanismo clásico de *backpropagation*

Referencias

- [Bis95] Christopher M. Bishop, *Neural networks for pattern recognition*, Clarendon Press, Oxford, 1995.
- [Cas95] J. L. Castro, *Fuzzy logic controllers are universal approximators*, IEEE Transactions on System, Man and Cybernetics **25** (1995), no. 4, 629–635.
- [CW96] K. B. Cho and B. H. Wang, *Radial basis function based adaptive fuzzy systems and their application to system identification and prediction*, Fuzzy Sets and Systems **83** (1996), 325–339.
- [Cyb88] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Mathematics of Control, Signals, and Systems **2** (1998), no. 4, 303–314.
- [Jan93] J.S. Jang, *Anfis: Adaptive-network-based fuzzy inference systems*, IEEE Trans. on Systems, Man, and Cybernetics **23** (1993), no. 03, 665–685.
- [JSM97] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neuro-fuzzy and soft computing*, Matlab Curriculum, Prentice Hall, 1997.
- [Kos92] B. Kosko, *Fuzzy systems as universal approximators*, Proceedings of the IEEE International Conference on Fuzzy Systems, 1992, pp. 1153–1161.
- [SY93] M. Sugeno and T. Yasukawa, *A fuzzy logic based approach to qualitative modeling*, IEEE Transactions on Fuzzy Systems **1** (1993), no. 1, 7–31.
- [Wan92] L. X. Wang, *Fuzzy systems are universal approximators*, Proceedings of the International Conference on Fuzzy Systems (San Diego. USA), 1992, pp. 1163–1170.
- [Zad65] Lofti A. Zadeh, *Fuzzy sets*, Information and Control **8** (1965), 338–353.