

**Aplicación de un sistema de inferencia difusa  
basado en una red adaptable para el pronóstico del  
tipo de cambio libra esterlina - dolar americano**

por

**Ing. Alfredo Estrada Chávez**

**Tesis**

Presentada al Programa de Graduados en Programa de Graduados en Computación,

Información, Electrónica y Comunicaciones

del

Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Monterrey

como requisito parcial para obtener el grado académico de

**Maestro en Ciencias**

**con especialidad en Sistemas Inteligentes**

**Instituto Tecnológico y de Estudios Superiores de Monterrey**

**Campus Monterrey**

Monterrey, N.L. Julio del 2002

# Instituto Tecnológico y de Estudios Superiores de Monterrey

## Campus Monterrey

### Programa de Graduados en Programa de Graduados en Computación, Información, Electrónica y Comunicaciones

Los miembros del comité de tesis recomendamos que la presente tesis de Alfredo Estrada Chávez sea aceptada como requisito parcial para obtener el grado académico de Maestro en Ciencias, especialidad en:

#### Sistemas Inteligentes

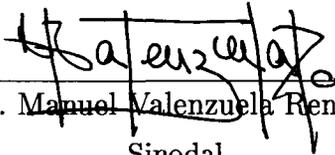
#### Comité de tesis:



---

Dr. Rogelio Soto Rodríguez

Asesor de la tesis



---

Dr. Manuel Valenzuela Rendón

Sinodal



---

Dr. Ramón Brena Pinero

Sinodal



---

Dr. David Garza Salazar

Programa de Graduados en Programa  
de Graduados en Computación,  
Información, Electrónica y  
Comunicaciones

Julio del 2002

A mis padres. En memoria de mis abuelos.

## Reconocimientos

Agradezco al Dr. Francisco J. Cantú Ortiz por las facilidades que me brindó para ingresar como asistente de excelencia en el Centro de Inteligencia Artificial y su confianza en los proyectos realizados para que de igual manera dispusiera del tiempo necesario para mis estudios.

Agradezco a mi asesor Dr. Rogelio Soto Rodríguez quien con su guía sutil encamino el desarrollo de la tesis conformando un ambiente de dedicación y disciplina.

Al Dr. Manuel Valenzuela Rendón le agradezco sus correcciones y sugerencias en la revisión del borrador y por su actitud proactiva en clase.

Al Dr. Ramón Brena Pinero le agradezco su amabilidad, cortesía y comentarios críticos.

Al coordinador de la maestría Dr. Hugo Terashima Marín por su actitud receptiva y vocación como profesor.

Enhorabuena.

ALFREDO ESTRADA CHÁVEZ

*Instituto Tecnológico y de Estudios Superiores de Monterrey  
Julio 2002*

# **Aplicación de un sistema de inferencia difusa basado en una red adaptable para el pronóstico del tipo de cambio libra esterlina - dolar americano**

Alfredo Estrada Chávez, M.C.  
Instituto Tecnológico y de Estudios Superiores de Monterrey, 2002

Asesor de la tesis: Dr. Rogelio Soto Rodríguez

El área de modelación o identificación difusa comprende la utilización de principios de un mecanismo de inferencia y reglas difusas para describir un sistema, como en este estudio de tesis, para fines de pronóstico en la cotización de divisas. Las reglas en un principio son definidas a partir de conocimiento implícito en patrones de datos sobre el dominio a modelar, para posteriormente aplicar un proceso de sintonización de las variables lingüísticas que las conforman a partir de vectores de entrenamiento entrada-salida estipulados; de esta forma se busca realizar un mapeo apropiado entre las entradas y salidas del sistema.

El proceso de sintonización es un proceso de aprendizaje en línea o en lote que en este estudio de tesis se aplica a arquitecturas de modelación que en forma genérica se conocen como sistema de inferencia difusa basado en una red adaptable, el cual comprende sistemas difusos del tipo Takagi-Sugeno-Kang (TSK).

Los sistemas difusos Takagi-Sugeno-Kang de orden cero y de primer orden pueden ser representados equivalentemente en forma de una red adaptable de manera tal que la sintonización de las reglas difusas en la base de conocimiento se transforma en un proceso de aprendizaje supervisado. En la red equivalente al sistema TSK de primer orden se utiliza un algoritmo de aprendizaje híbrido en que los parámetros de las premisas de las reglas son ajustados usando una estrategia de gradiente descendente y los del consecuente con un algoritmo de mínimos cuadrados recursivos. En la red del sistema TSK de orden cero se utiliza sólo el aprendizaje por gradiente descendente y los resultados de ambas arquitecturas son comparadas ante la variación en la configuración de las estructuras y sus parámetros para un mismo caso de estudio correspondiente al pronóstico del tipo de cambio libra esterlina - dolar americano.

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS  
SUPERIORES DE MONTERREY

CAMPUS MONTERREY

PROGRAMA DE GRADUADOS EN ELECTRONICA,  
COMPUTACION, INFORMACION Y COMUNICACIONES



APLICACION DE UN SISTEMA DE INFERENCIA  
DIFUSA BASADO EN UNA RED ADAPTABLE  
PARA EL PRONOSTICO DEL TIPO DE CAMBIO  
LIBRA ESTERLINA-DOLAR AMERICANO

TESIS

PRESENTADA COMO REQUISITO PARCIAL  
PARA OBTENER EL GRADO ACADEMICO DE:  
MAESTRO EN CIENCIAS  
CON ESPECIALIDAD EN SISTEMAS INTELIGENTES

ING. ALFREDO ESTRADA CHAVEZ

JULIO 2002

# INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

CAMPUS MONTERREY

PROGRAMA DE GRADUADOS EN ELECTRONICA,  
COMPUTACION, INFORMACION Y COMUNICACIONES



APLICACION DE UN SISTEMA DE INFERENCIA  
DIFUSA BASADO EN UNA RED ADAPTABLE  
PARA EL PRONOSTICO DEL TIPO DE CAMBIO  
LIBRA ESTERLINA-DOLAR AMERICANO

TESIS

PRESENTADA COMO REQUISITO PARCIAL  
PARA OBTENER EL GRADO ACADEMICO DE:  
MAESTRO EN CIENCIAS  
CON ESPECIALIDAD EN SISTEMAS INTELIGENTES

ING. ALFREDO ESTRADA CHAVEZ

JULIO 2002

# Índice General

|                                                                                              |           |
|----------------------------------------------------------------------------------------------|-----------|
| <b>Reconocimientos</b>                                                                       | <b>v</b>  |
| <b>Resumen</b>                                                                               | <b>vi</b> |
| <b>Índice de Figuras</b>                                                                     | <b>ix</b> |
| <b>Capítulo 1 Introducción</b>                                                               | <b>1</b>  |
| 1.1 Definición del Problema y Objetivos . . . . .                                            | 2         |
| 1.2 Hipótesis . . . . .                                                                      | 7         |
| <b>Capítulo 2 Fundamentos Teóricos de Sistemas Difusos</b>                                   | <b>9</b>  |
| 2.1 Introducción . . . . .                                                                   | 9         |
| 2.2 Conceptos sobre Lógica Difusa . . . . .                                                  | 9         |
| 2.2.1 Variables Lingüísticas y Conjuntos Difusos . . . . .                                   | 10        |
| 2.2.2 Proposiciones Difusas Compuestas y Operaciones en Conjuntos Difusos . . . . .          | 12        |
| 2.2.3 Reglas Difusas y Razonamiento Aproximado . . . . .                                     | 13        |
| 2.3 Sistemas Difusos . . . . .                                                               | 16        |
| 2.3.1 Base de Reglas Difusas . . . . .                                                       | 16        |
| 2.3.2 Máquina de Inferencia Difusa . . . . .                                                 | 18        |
| 2.3.3 Tipos de Sistemas Difusos . . . . .                                                    | 21        |
| <b>Capítulo 3 Algoritmos de Aprendizaje en Modelación Difusa</b>                             | <b>25</b> |
| 3.1 Introducción . . . . .                                                                   | 25        |
| 3.2 Aprendizaje por Gradiente Descendente . . . . .                                          | 25        |
| 3.2.1 Principios del Método de Gradiente Descendente . . . . .                               | 26        |
| 3.2.2 Algoritmo por Gradiente Descendente en la Sintonización de un Sistema Difuso . . . . . | 28        |
| 3.3 Aprendizaje por Estimación de Mínimos Cuadrados . . . . .                                | 30        |
| 3.3.1 Principios del Estimador por Mínimos Cuadrados . . . . .                               | 31        |
| 3.3.2 Estimación por Mínimos Cuadrados Recursivos . . . . .                                  | 34        |

|                   |                                                                                                         |           |
|-------------------|---------------------------------------------------------------------------------------------------------|-----------|
| 3.3.3             | Método de Mínimos Cuadrados Recursivos para Sistemas Variantes en el Tiempo . . . . .                   | 36        |
| <b>Capítulo 4</b> | <b>Caso de Estudio: Pronóstico del Tipo de Cambio Libra Esterlina - Dolar Americano</b>                 | <b>39</b> |
| 4.1               | Introducción . . . . .                                                                                  | 39        |
| 4.2               | Descripción del Caso de Estudio y de la Modelación Difusa Inicial . . .                                 | 40        |
| 4.3               | Resultados de la Identificación Difusa Sintonizada con un Algoritmo por Gradiente Descendente . . . . . | 44        |
| 4.4               | Resultados de la Identificación Difusa Sintonizada por un Algoritmo de Aprendizaje Híbrido . . . . .    | 45        |
| <b>Capítulo 5</b> | <b>Conclusiones</b>                                                                                     | <b>52</b> |
|                   | <b>Bibliografía</b>                                                                                     | <b>54</b> |
|                   | <b>Vita</b>                                                                                             | <b>55</b> |

## Índice de Figuras

|     |                                                                                                                            |    |
|-----|----------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Sistema de inferencia difusa . . . . .                                                                                     | 4  |
| 1.2 | (a)Sistema difuso del tipo TKS. (b)Red adaptable equivalente . . . . .                                                     | 5  |
| 2.1 | Función de membresía gaussiana . . . . .                                                                                   | 11 |
| 2.2 | Función de membresía de campana gaussiana . . . . .                                                                        | 11 |
| 2.3 | Condición de anchura para un sistema difuso con una variable en el antecedente de las reglas . . . . .                     | 17 |
| 2.4 | Ejemplo de funciones de membresia para un sistema difuso de dos entradas                                                   | 17 |
| 2.5 | Ejemplo de un sistema difuso del tipo TKS de primer orden. . . . .                                                         | 24 |
| 4.1 | Variación en las cotizaciones del tipo de cambio entre la libra esterlina y el dólar americano . . . . .                   | 40 |
| 4.2 | Error PEMA para los datos de evaluación, sistema TSK de orden cero .                                                       | 45 |
| 4.3 | Caso C del sistema TSK de orden cero utilizando entrenamiento en lote con un parámetro de aprendizaje de 0.01 . . . . .    | 47 |
| 4.4 | Caso A utilizando entrenamiento en lote con un parámetro de aprendizaje de 0.01 . . . . .                                  | 48 |
| 4.5 | Error PEMA para los datos de evaluación, sistema TSK de primer orden                                                       | 49 |
| 4.6 | Caso A del sistema TSK de primer orden utilizando entrenamiento en línea con un parámetro de aprendizaje de 0.01 . . . . . | 50 |
| 4.7 | Caso B del sistema TSK de primer orden utilizando entrenamiento en lote con un parámetro de aprendizaje de 0.001 . . . . . | 51 |

## Capítulo 1

### Introducción

El estudio de los problemas se da generalmente en términos de sistemas que son construidos como modelos conformados tanto de algunos aspectos de la realidad, como de objetos a procesar de una manera deseada por el hombre. El propósito de la construcción de modelos del primer tipo gira en torno al entendimiento de fenómenos reales, ya sean naturales o hechos por el hombre, para hacer predicciones adecuadas y para aprender a controlar los fenómenos de manera deseable; los modelos del segundo tipo son construidos con el propósito de identificar y prescribir operaciones por las cuales la concepción de un objeto pueda ser construida de una manera tal que una serie de objetivos deseables son satisfechos bajo ciertas restricciones [4].

En la construcción de un modelo se busca cumplir, entre otros aspectos, con los anteriores propósitos, al igual que minimizar una medida del error en la descripción de las relaciones e interacciones entre entradas-salidas del sistema que se identifica, esto a través de la incorporación y estructuración de una base de conocimiento que es sintonizada por medio de algoritmos de aprendizaje.

En esta tesis se estudia la identificación de sistemas para propósito de pronóstico utilizando enunciados condicionales y principios de mecanismos de inferencia difusos. Esta modelación difusa tiene sus características únicas que la hacen favorable para la solución de problemas de identificación de componentes no lineales.

Entre las principales ventajas de la modelación basada en enunciados condicionales difusos se encuentra la facilidad para la representación de conocimiento implícito en patrones de datos sobre un dominio en forma de reglas *Si-Entonces*, a través de las cuales se puede incluir información lingüística de expertos y combinarla con información numérica proveniente de sensores o de datos entrada-salida definidos [2] [3]; además se utilizan principios de un mecanismo de inferencia que captura aspectos del razonamiento humano y a través del cual se transforman reglas difusas en mapeos no lineales con los que se podría aproximar cualquier función arbitrariamente bien [6].

Otro aspecto a considerar es que los sistemas difusos pueden ser vistos desde dos perspectivas: por una parte, como sistemas basados en conocimiento que permiten modelar aspectos cualitativos del razonamiento humano a través de reglas condicionales con variables lingüísticas; por otro lado, como funciones que realizan mapeos o descripciones de sistemas complejos no lineales, donde tales funciones son representadas por fórmulas precisas y compactas que pueden ser ajustadas usando datos entrada-salida del sistema a modelar [7].

Un sistema difuso con las anteriores características puede tener una representación equivalente en forma de una red consistente de nodos y enlaces dirigidos que los conectan. Un esquema de este tipo es el propuesto por Jang [3], al cual se le agrega la funcionalidad de un sistema con aprendizaje supervisado, donde las variables lingüísticas y sus respectivas funciones de membresía son sintonizadas por un algoritmo basado en gradiente descendente, y los parámetros del consecuente de las reglas difusas son ajustados con una estrategia de mínimos cuadrados recursivos; es decir, se tiene un algoritmo de aprendizaje híbrido que se realiza en dos diferentes etapas secuenciales, buscando con esto darle a la red una capacidad de adaptación ante sistemas cuya dinámica varíe ligeramente en el tiempo, como en la identificación de componentes en línea. Esta arquitectura con aprendizaje híbrido y otra simplificada que sólo utiliza aprendizaje por gradiente descendente es el tópico de estudio de este trabajo de tesis, en conjunto con un caso de estudio relacionado a la cotización de divisas en que se evalúan algunos aspectos del desempeño de ambos esquemas, que en forma genérica se conocen como sistema de inferencia difusa basado en una red adaptable.

En las siguientes secciones se presenta la definición del problema en conjunto con una descripción de la arquitectura equivalente en red adaptable de un sistema difuso, así como los objetivos y alcances del estudio, para posteriormente plantear las hipótesis y hacer una revisión de los fundamentos y conceptos en que se basa este trabajo. Finalmente se presentan resultados experimentales de la modelación con arquitecturas base de un sistema de inferencia difusa con funcionalidad de red adaptable ante un mismo caso de estudio correspondiente al pronóstico del tipo de cambio libra esterlina - dolar americano.

## 1.1 Definición del Problema y Objetivos

En la naturaleza y en los procesos industriales es muy común encontrar sistemas con componentes o subsistemas no lineales. La identificación o modelación de tales sistemas es necesaria para la comprensión de la dinámica o de los movimientos de las variables que describen los estados del mismo a través del tiempo, esto con el propósito

de pronóstico o el diseño de controladores. En esta tesis se busca cumplir con los anteriores requerimientos de una modelación o identificación de sistemas utilizando sistemas difusos.

Las reglas difusas son del tipo *SI*  $x_1$  es  $A_1^1$  Y  $x_2$  es  $A_2^1$  *ENTONCES*  $y$  es  $B^1$  donde  $x_i$  para  $i = 1, 2$  son variables lingüísticas que pueden tomar algún valor  $A_i^1$ ,  $B^1$  lingüístico caracterizado por conjuntos difusos que son descritos a través de funciones de membresía  $\mu_{A_i^1}(x_i)$ ,  $\mu_{B^1}(y)$ .

Las funciones de membresía pueden tomar forma triangular, trapezoidal o de alguna forma de campana con máximo igual a 1 y mínimo igual a 0, como las obtenidas con las siguientes funciones:

$$\mu_{A_i^1}(x) = \frac{1}{1 + \left[ \left( \frac{x - \bar{x}_i}{\sigma_i} \right)^2 \right]^{b_i}} \quad (1.1)$$

ó

$$\mu_{A_i^1}(x) = \exp \left\{ - \left( \frac{x - \bar{x}_i}{\sigma_i} \right)^2 \right\} \quad (1.2)$$

donde  $\{\bar{x}_i, b_i, \sigma_i\}$  es un conjunto de parámetros para cada conjunto difuso, los cuales son almacenados en una base de datos, tal y como se muestra en la figura 1.1.

Lo que se busca en el área de identificación o modelación difusa es construir una base de reglas *Si-Entonces* con funciones de membresía apropiadas que describan la dinámica del sistema en el sentido de que haya un mapeo entre pares de estados entrada-salida estipulados, esto con el fin de que se cubran las posibles combinaciones de estados que describen en forma local el sistema bajo consideración. La relación de estados entrada-salida es obtenida siguiendo principios de un mecanismo de inferencia interpretable, como el que se muestra en la figura 1.1.

En un principio se aprovecha el conocimiento implícito en los pares de datos entrada-salida del proceso a modelar para definir la base de conocimiento, la cual está conformada por una serie de reglas difusas, que de ser completas, cubrirían todo el rango de posibles estados entrada-salida a través de variables lingüísticas cuyas funciones de membresía deberían estar igualmente espaciadas.

Después de fijarse el número de funciones de membresía para cada entrada, se busca que los parámetros de éstas estén definidos de manera que para cada componente del

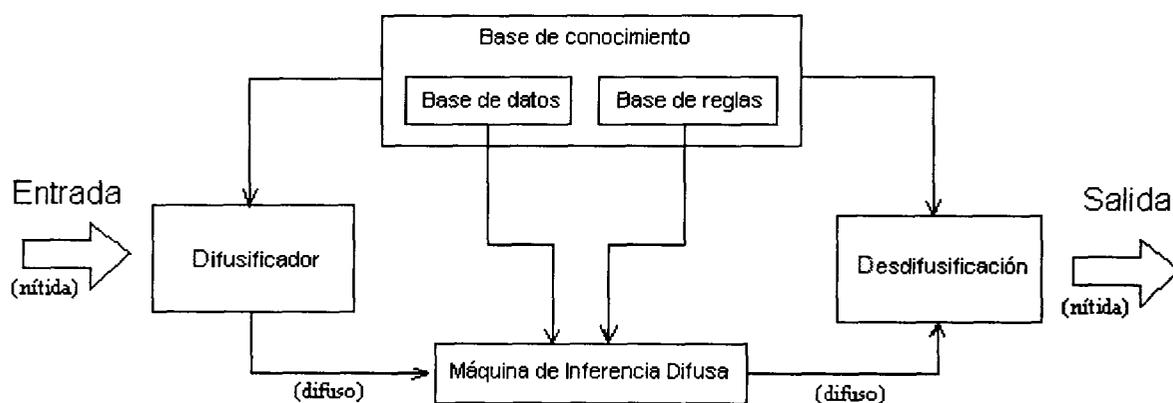


Figura 1.1: Sistema de inferencia difusa

vector  $x$  de entrada se encuentre un conjunto difuso  $A$  en cada variable del antecedente de una regla, tal que su grado de pertenencia sea mayor o igual a 0.5. Además se busca que la anchura entre dos funciones de membresía conlindantes sean semejantes e iguales a la longitud del intervalo entre valores máximos de tales conjuntos difusos adyacentes, de manera tal que al seguir principios de un sistema de inferencia difuso se obtengan salidas con transiciones suaves [1].

Al aumentar tanto el número de variables con que se describe un sistema como el número de reglas para modelarlo, se vuelve necesario un método efectivo para la sintonización de funciones de membresía, tal que se minimiza una medida del error en el mapeo no-lineal entre entradas-salidas del sistema. Es ante este requerimiento que se utiliza la arquitectura del llamado sistema de inferencia difusa basado en una red adaptable. Un esquema de este tipo es el propuesto por Jang [3], el cual transforma el sistema de inferencia difusa de la figura 1.1 en una estructura de red consistente de nodos y conexiones direccionales. Para el sistema difuso tipo Takagi-Sugeno-Kang (TSK) de primer orden con difusificador *singleton*, con principios de máquina de inferencia producto y funciones de membresía gaussiana, su representación equivalente en un red adaptable se muestra en la figura 1.2.

Al simplificar el sistema difuso del tipo Takagi-Sugeno-Kang de la figura 1.2 a uno de orden cero, es decir, con el consecuente descrito con un *singleton*  $\bar{y}^l$ , se obtiene una estructura con máquina de inferencia producto, difusificador *singleton*, desdifusificador de centro promedio y función de membresía gaussiana (1.2), la cual es descrita con la siguiente función

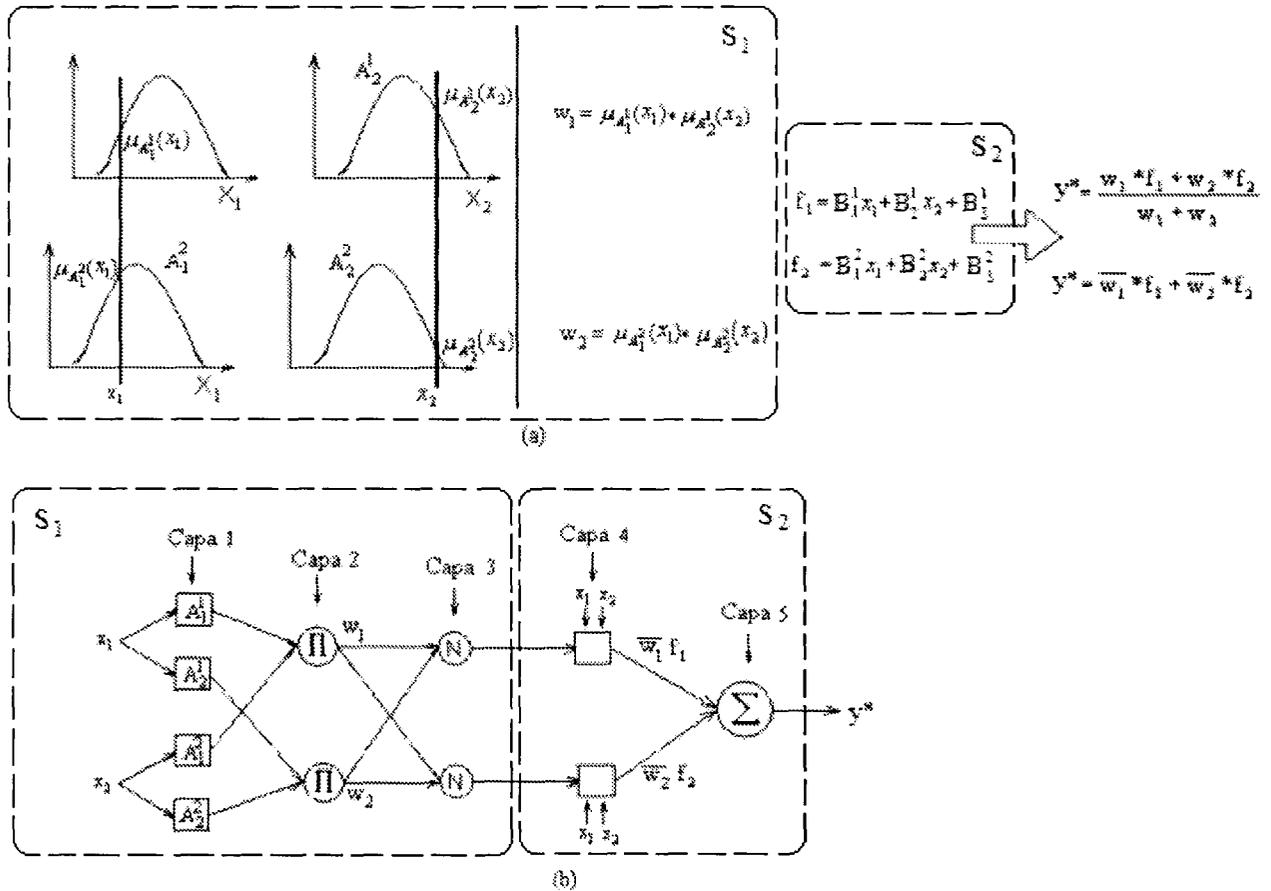


Figura 1.2: (a) Sistema difuso del tipo TKS. (b) Red adaptable equivalente

$$y^* = f(\mathbf{x}) = \frac{\sum_{l=1}^M \bar{y}^l \left[ \prod_{i=1}^n \exp \left( - \left( \frac{x_i - \bar{x}_i^l}{\sigma_i^l} \right)^2 \right) \right]}{\sum_{l=1}^M \left[ \prod_{i=1}^n \exp \left( - \left( \frac{x_i - \bar{x}_i^l}{\sigma_i^l} \right)^2 \right) \right]} \quad (1.3)$$

La anterior estructura simplificada resulta ser un sistema difuso con poder ilimitado de aproximación; es decir, existe una función de la forma (1.3) tal que sintonizando sus parámetros adecuadamente se puede ajustar cualquier función no lineal arbitrariamente bien [6].

Los parámetros de las funciones de membresía de las anteriores estructuras serían ajustados por un algoritmo de gradiente descendente tal que se minimiza una medida del error en la salida del modelo. El algoritmo de gradiente descendente, primeramente desarrollado por Werbos [8] en la década de los 70's, ha sido utilizado como regla básica

de aprendizaje en sistemas adaptables, pero muestra la tendencia de quedar atrapado en mínimos locales o converger lentamente; situación que se busca contrarrestar con un algoritmo de aprendizaje híbrido a ser aplicado a una arquitectura de sistema difuso TSK de primer orden como el del esquema de la figura 1.2 [3].

Los parámetros a ajustar en la red adaptable pueden ser descompuestos en dos grupos:  $S_1$  y  $S_2$ . El primer grupo  $S_1$  corresponde a los parámetros de las funciones de membresía que cubren el espacio de entradas y que caracterizan las variables lingüísticas de las premisas de las reglas difusas, y el segundo  $S_2$ , corresponde a los parámetros del consecuente del sistema difuso TSK, los cuales para cada regla  $l$  forman un función  $f_l$  en conjunto con las entradas  $\mathbf{x}$ . Para el esquema de la figura 1.2, los parámetros en  $S_2$  son  $B_1^1, B_2^1, B_3^1$  y  $B_1^2, B_2^2, B_3^2$ , por lo que se obtiene una salida  $y^*$  que es expresada como una combinación lineal de los parámetros del consecuente:

$$\begin{aligned}
 y^* &= \frac{w_1}{w_1+w_2}x_1B_1^1 + \frac{w_1}{w_1+w_2}x_2B_2^1 + \frac{w_1}{w_1+w_2}B_3^1 + \frac{w_2}{w_1+w_2}x_1B_1^2 + \frac{w_2}{w_1+w_2}x_2B_2^2 + \frac{w_2}{w_1+w_2}B_3^2 \\
 &= \bar{w}_1x_1B_1^1 + \bar{w}_1x_2B_2^1 + \bar{w}_1B_3^1 + \bar{w}_2x_1B_1^2 + \bar{w}_2x_2B_2^2 + \bar{w}_2B_3^2 \\
 &= \left[ \bar{w}_1x_1 \quad \bar{w}_1x_2 \quad \bar{w}_1 \right]^T \left[ B_1^1 \quad B_2^1 \quad B_3^1 \right] + \left[ \bar{w}_2x_1 \quad \bar{w}_2x_2 \quad \bar{w}_2 \right]^T \left[ B_1^2 \quad B_2^2 \quad B_3^2 \right] \\
 y^* &= \mathbf{F}_1(\mathbf{x}) \mathbf{B}^1 + \mathbf{F}_2(\mathbf{x}) \mathbf{B}^2
 \end{aligned} \tag{1.4}$$

El algoritmo híbrido de aprendizaje ajusta primero los parámetros de  $S_2$  por medio de un algoritmo de mínimos cuadrados recursivos al introducir  $k = 1, 2, \dots, P$  datos de entrenamiento al sistema y considerando valores fijos de  $S_1$ . Si el entrenamiento es en línea, para cada nuevo dato entrada-salida, se ajustan los parámetros de  $S_2$ , se obtiene la salida  $y^*$  del modelo y se compara con la estipulada en el dato, para así generar un error  $E_k$  que se retropropaga para sintonizar los elementos de  $S_1$  en las funciones de membresía por medio del algoritmo de gradiente descendente. Si el entrenamiento es fuera de línea o por lotes, se sigue un procedimiento análogo al anterior, sólo que los parámetros de  $S_1$  no son actualizados sino hasta que se obtiene la suma de los errores  $E$  para todo el conjunto de datos entrada-salida estipulados [3].

En principio las funciones de membresía capturan un significado lingüístico al describir un concepto en el dominio de cada variable lingüística que conforma el antecedente en reglas difusas, por lo que si se considera que tales conceptos implican términos cardinales u ordinales en el arreglo de valores lingüísticos, entonces al conformar una modelación difusa inicial del sistema en estudio y aplicar algoritmos de aprendizaje se podría dar una pérdida de interpretabilidad en los parámetros de las funciones de

membresía si hay un traslape o una intercalación de conjuntos difusos subsecuentes que guardan un orden jerárquico. Además se menciona en [3] que los sistemas con funciones de membresía de la forma (1.1) pierden su propiedad de aproximadores universales cuando en la sintonización el parámetro  $b_i$  toma valores negativos, por lo que para conservar tal propiedad se utilizarían exclusivamente funciones del tipo (1.2).

Considerando el uso de funciones de membresía gaussiana (1.2) en la conformación de una modelación difusa inicial, se evalúa la interpretabilidad después del entrenamiento en línea o en lote como el traslape o intercalación entre centros  $\bar{x}_i^l, \bar{x}_{i+1}^{l+1}$  de funciones de membresía en una variable lingüística  $x_i$  en el antecedente de reglas difusas  $l = 1 \dots M$ , por lo que uno de los objetivos sería evaluar tal concepto en la modelación correspondiente al caso de estudio.

Otro de los objetivos es comparar los resultados obtenidos de la red adaptable basada en un sistema difuso del tipo Takagi-Sugeno-Kang de primer orden y aprendizaje híbrido, con una estructura simplificada del tipo Takagi-Sugeno-Kang de orden cero y aprendizaje usando sólo el algoritmo por gradiente descendente.

Las estructuras de redes adaptables se compararían por medio de pruebas ante un mismo caso de estudio relacionado a la cotización de divisas al evaluar los resultados ante diferentes configuraciones de valores del parámetro de aprendizaje, número de funciones de membresía, número de reglas y aprendizaje por lote o en línea.

## 1.2 Hipótesis

Con la implementación de los sistemas difusos del tipo Takagi-Sugeno-Kang de primer orden y orden cero a los que se les agrega la funcionalidad de red adaptable con capacidad de aprendizaje híbrido y de sólo gradiente descendente respectivamente, se busca responder las siguientes hipótesis de investigación ante un mismo caso de estudio:

- Para ambos sistemas se obtienen mejores resultados al reducir los valores del parámetro de aprendizaje.
- Las variables lingüísticas de los sistemas difusos adaptables pierden su interpretabilidad inicial con el aprendizaje si no se cuenta con el suficiente número de conjuntos difusos que las definan, lo cual haría necesario también un mayor número de reglas.
- Si se aumenta el número de reglas se llega un punto óptimo en que el error es mínimo para datos de evaluación después del cual al incorporar más reglas

reduciendo el espacio de entrada que comprende cada una, aumenta el error del modelo. El modelo TSK de primer orden necesitaría un menor número de reglas para llegar a ese óptimo y por ende un menor conocimiento sobre el sistema.

- El error del sistema difuso en una red adaptable es menor cuando se conserva la interpretabilidad de los conjuntos difusos del antecedente de las reglas.

## Capítulo 3

# Algoritmos de Aprendizaje en Modelación Difusa

### 3.1 Introducción

El objetivo de este capítulo es presentar dos algoritmos de aprendizaje con que se sintoniza o refina la base de conocimiento de una modelación difusa inicial: algoritmo de aprendizaje por gradiente descendente y el algoritmo de estimación por estimación de mínimos cuadrados cuyos principios se toman de [2] para ser presentado en términos de un sistema difuso TSK de primer orden, tal y como se planteó en la definición del problema. Es decir, la presentación de ambos algoritmos de aprendizaje se basa principalmente en sistemas difusos que pueden ser representados en fórmulas compactas, las cuales al ser sintonizadas o ajustadas por vectores de datos entrada-salida, toman la funcionalidad de un sistema de inferencia difusa basado en una red adaptable.

La modelación difusa inicial corresponde a representar conocimiento cualitativo sobre el sistema a partir de pares de estados entrada-salida en forma de reglas *si-entonces* que son conjuntadas en una arquitectura de sistema difuso TSK; es decir, se parte del diseño de un sistema difuso equivalente a la arquitectura TSK de orden cero cuyas estructuras se sintonizan con un algoritmo de aprendizaje por gradiente descendente para posteriormente, retomando la modelación inicial, expandir las reglas difusas a una función lineal en término del vector de entrada al sistema de manera tal que los parámetros en el antecedente de las reglas se sintonizan con el algoritmo por gradiente descendente y las del consecuente con una estrategia por mínimos cuadrados recursivos, para conformar así el algoritmo híbrido de aprendizaje a aplicar en línea o en lote, tal y como se describe en la definición del problema, sección 1.1.

### 3.2 Aprendizaje por Gradiente Descendente

El método de gradiente descendente es un algoritmo básico de aprendizaje por medio del cual se busca ajustar los parámetros de la estructura de un modelo difuso

TSK de orden cero en proporción a la diferencia entre la salida generada por el mismo ante un patrón de entrada con la salida de un vector de entrenamiento con tal patrón de entrada. La media de la diferencia de las salidas al cuadrado es una medida del error  $E_k$  del modelo ante el vector de entrenamiento  $k$ , la cual es a su vez una función definida en el espacio de parámetros que se busca sintonizar.

### 3.2.1 Principios del Método de Gradiente Descendente

En el proceso de aprendizaje por gradiente descendente es necesario contar con vectores  $P$  de patrones entrada-salida definidos, de manera que  $k = 1, 2, \dots, P$ . El modelo a ajustar procesa un patrón de entrada por capas  $z$  secuenciales hasta obtener un patrón de salida en su última capa  $C$ ; es decir, cada capa  $z$  puede contener  $l$  funciones o nodos de procesamiento que producen una salida  $O_i^z$  a la siguiente función  $O_i^{z+1}$ , lo cual se expresa también como

$$O_i^z = O_i^z \left( O_1^{z-1}, \dots, O_{\#(z-1)}^{z-1}, \bar{x}_1^l, \bar{x}_2^l, \bar{x}_3^l, \dots \right) \quad (3.1)$$

donde la capa  $z - 1$  tiene  $\#(z - 1)$  nodos o funciones que son entradas para el nodo función  $O_i^z$ , el cual cuenta con los parámetros  $\bar{x}_1^l, \bar{x}_2^l, \bar{x}_3^l, \dots$  para producir una salida.

En la última capa del modelo se produce un patrón de salida que es comparado con aquel del patrón de entrada del vector de entrenamiento actual  $k$  para obtener el error  $E_k$  como

$$E_k = \sum_{m=1}^{\#(C)} \frac{1}{2} \left( T_{m,k} - O_{m,k}^C \right)^2 \quad (3.2)$$

donde  $T_{m,k}$  es el componente  $m$  del patrón de salida del actual vector de entrenamiento  $k$  y  $O_{m,k}^C$  es el componente  $m$  del vector de salida obtenido por el modelo al incorporar el patrón de entrada de tal vector de entrenamiento  $k$ . Una medida del error total es entonces

$$E = \sum_{k=1}^P E_k \quad (3.3)$$

Para minimizar el error  $E$  en el espacio de parámetros por el método de gradiente descendente, se calcula una tasa del error  $\partial E_k / \partial O$  para el vector de entrenamiento  $k$  y para cada componente  $m$  de la salida; lo cual a partir de (3.2) para el componente  $l$  quedaría expresado como

$$\frac{\partial E_k}{\partial O_{l,k}^C} = - (T_{l,k} - O_{l,k}^C) \quad (3.4)$$

De esta manera, la tasa de error se propaga a las capas anteriores ( $z = C - 1$ ) de funciones cuya salida toma parte en el cálculo de  $O_{l,k}^C$ . Aplicando la regla de la cadena, el mismo procedimiento se repite para funciones  $l$  de capas antecedentes  $z - 1$ , en la que la tasa de error para tal función  $\partial O_{l,k}^{z-1}$  es expresada como una combinación lineal de las tasas de error de las funciones en la capa actual  $z$

$$\frac{\partial E_k}{\partial O_{l,k}^{z-1}} = \sum_{m=1}^{\#(z)} \frac{\partial E_k}{\partial O_{m,k}^z} \frac{\partial O_{m,k}^z}{\partial O_{l,k}^{z-1}} \quad (3.5)$$

donde  $1 \leq z \leq L$ .

Para un parámetro  $\bar{x}_i^l$  en específico de una función de un nodo interno  $l$ , la anterior expresión se reescribiría como

$$\frac{\partial E_k}{\partial \bar{x}_i^l} = \sum_{O^* \in S} \frac{\partial E_k}{\partial O^*} \frac{\partial O^*}{\partial \bar{x}_i^l} \quad (3.6)$$

donde  $S$  es el conjunto de funciones cuyas salidas dependen de  $\bar{x}_i^l$ . Por lo tanto la derivada del error total con respecto a  $\bar{x}_i^l$  es

$$\frac{\partial E}{\partial \bar{x}_i^l} = \sum_{k=1}^P \frac{\partial E_k}{\partial \bar{x}_i^l} \quad (3.7)$$

Como resultado se tiene el gradiente de ajuste  $\Delta$  para un parámetro  $\bar{x}_i^l$  en específico de una función sería

$$\Delta \bar{x}_i^l = \alpha \frac{\partial E}{\partial \bar{x}_i^l} \quad (3.8)$$

donde  $\alpha$  es una constante de aprendizaje.

Al utilizar una medida del error total  $\partial E$  en el cálculo de  $\Delta \bar{x}_i^l$ , se dice que el entrenamiento se realiza fuera de línea o por lotes; por otra parte, si se propaga el ajuste  $\Delta$  al presentar cada vector de entrenamiento por medio de  $\partial E_k$ , se dice que se tiene un entrenamiento en línea.

### 3.2.2 Algoritmo por Gradiente Descendente en la Sintonización de un Sistema Difuso

En la sintonización de sistemas difusos por medio del algoritmo de aprendizaje por gradiente descendente se tiene que especificar la arquitectura del modelo difuso para posteriormente definir los respectivos parámetros a ajustar por medio de vectores de patrones entrada-salida; es decir, en una primera etapa del diseño se especifica conocimiento cualitativo del sistema a modelar a través de reglas difusas para así definir una arquitectura con el difusificador, la máquina de inferencia y el desdifusificador que sean los más apropiados [7].

En este estudio, al genero de sistemas difusos al que se le aplica el algoritmo por gradiente descendente está conformado por una máquina de inferencia producto, difusificador singletón y desdifusificador de centro promedio. Este sistema difuso, definido por (2.28), se reescribe utilizando la función de membresía (2.4) en la descripción de los conjuntos difusos de las premisas:

$$y^* = f(\mathbf{x}) = \frac{\sum_{i=1}^M \bar{y}^l \left( \prod_{i=1}^n \exp \left( - \left( \frac{x_i - \bar{x}_i^l}{\sigma_i^l} \right)^2 \right) \right)}{\sum_{i=1}^M \left( \prod_{i=1}^n \exp \left( - \left( \frac{x_i - \bar{x}_i^l}{\sigma_i^l} \right)^2 \right) \right)} \quad (3.9)$$

donde  $M$  es el número de reglas que conforma la base de conocimiento la cual a su vez contiene una base de datos con los parámetros  $\bar{y}^l$ ,  $\bar{x}_i^l$  y  $\sigma_i^l$  a sintonizar.

Para aplicar el algoritmo de gradiente descendente al anterior sistema se cuenta con  $P$  vectores de patrones entrada-salida, lo cual se expresa como

$$[\mathbf{x}_k; y_k], \quad k = 1, 2, \dots, P \quad (3.10)$$

donde

$$\mathbf{x}_k \in U = x_1 \times \dots \times x_n \subset R^n \quad (3.11)$$

y

$$y_k \in V = \beta_y \subset R \quad (3.12)$$

Como el sistema difuso (3.9) cuenta sólo con un componente de salida  $f(\mathbf{x}_k)$ , la ecuación del error (3.2) dado un vector de entrenamiento de la forma (3.10) se escribe como

$$E_k = \frac{1}{2} [f(\mathbf{x}_k) - y_k]^2 \quad (3.13)$$

El algoritmo de gradiente descendente busca minimizar el error  $E_k$  modificando los parámetros de las funciones involucradas en el proceso de inferencia, esto es, ajustando la última capa de funciones de membresía del consecuente, para posteriormente sintonizar las funciones de membresía de los antecedentes de las reglas. En una descripción secuencial del proceso se tiene que un vector de entrada  $\mathbf{x}_k$  se incorpora al difusificador *singleton* y a una función gaussiana (2.4) para obtener en esta primera capa  $z^l = \prod_{i=1}^n \exp\left(-\left(\frac{x_i - \bar{x}_i^l}{\sigma_i^l}\right)^2\right)$ ; la salida de estas funciones se utilizan para en una segunda capa obtener, por medio de operaciones de unión y sumatorias ponderadas con pesos como en (2.24), los términos  $b = \sum_{l=1}^M z^l$  y  $a = \sum_{l=1}^M (\bar{y}^l z^l)$ ; por último en una tercera capa, la salida del sistema difuso es calculada como  $y^* = f(\mathbf{x}) = a/b$

El proceso de aprendizaje se realiza como un entrenamiento en línea  $E_k$  o en lote  $E$  por el que se sintonizan los parámetros  $\bar{y}^l$ ,  $\bar{x}_i^l$ ,  $\sigma_i^l$ , buscando con esto la estructura que mejor describa el sistema dinámico hasta un estado actual con que se cuente vectores de datos  $P$ . Para esto se aplica el método de gradiente descendente en un número de épocas  $q$  específicas; de esta forma, como la tasa del error (3.4) de la salida  $y^*$  depende de  $\bar{y}^l$  sólo a través de  $a = \sum_{l=1}^M (\bar{y}^l z^l)$ , se tiene que

$$\bar{y}^l(q+1) = \bar{y}^l(q) - \alpha \frac{\partial E_k}{\partial \bar{y}^l} \quad (3.14)$$

donde  $l = 1, 2, \dots, M$ ,  $q = 0, 1, 2, \dots$ , y  $\alpha$  es una constante de aprendizaje, tal y como se expresa en (3.8).

Aplicando la regla de la cadena (3.6) habiendo definido  $y^* = a/b$ ,  $a = \sum_{l=1}^M (\bar{y}^l z^l)$  y  $b = \sum_{l=1}^M z^l$  se obtiene

$$\frac{\partial E_k}{\partial \bar{y}^l} = (y^* - y_k) \frac{\partial y^*}{\partial a} \frac{\partial a}{\partial \bar{y}^l} = (y^* - y_k) \frac{1}{b} z^l \quad (3.15)$$

Por lo tanto, el método de entrenamiento para  $\bar{y}^l$  se reescribe como

$$\bar{y}^l(q+1) = \bar{y}^l(q) - \alpha \frac{(y^* - y_k)}{b} z^l \quad (3.16)$$

donde  $l = 1, 2, \dots, M$ ;  $q = 0, 1, 2, \dots$

Para calcular  $\bar{x}_i^l$  se sigue un procedimiento análogo al anterior por lo que se define

$$\bar{x}_i^l(q+1) = \bar{x}_i^l(q) - \alpha \frac{\partial E_k}{\partial \bar{x}_i^l} \Big|_q \quad (3.17)$$

donde  $l = 1, 2, \dots, M$ ;  $q = 0, 1, 2, \dots$ ;

Como la tasa del error (3.4) de la salida  $y^*$  depende de  $\bar{x}_i^l$  sólo a través de  $z^l$ , aplicando la regla de la cadena como en (3.6) se deriva que

$$\frac{\partial E_k}{\partial \bar{x}_i^l} = (y^* - y_k) \frac{\partial y^*}{\partial z^l} \frac{\partial z^l}{\partial \bar{x}_i^l} = (y^* - y_k) \frac{\bar{y}^l - y^*}{b} z^l \frac{2(x_{ki} - \bar{x}_i^l)}{\sigma_i^{l2}} \quad (3.18)$$

El método de entrenamiento para  $\bar{x}_i^l$  se obtiene al sustituir (3.18) en (3.17) resultando

$$\bar{x}_i^l(q+1) = \bar{x}_i^l(q) - \alpha \frac{(y^* - y_k)}{b} (\bar{y}^l(q) - y^*) z^l \frac{2(x_{ki} - \bar{x}_i^l(q))}{\sigma_i^{l2}(q)} \quad (3.19)$$

El mismo procedimiento se utiliza para obtener el algoritmo de entrenamiento para  $\sigma^l$  como

$$\sigma_i^l(q+1) = \sigma_i^l(q) - \alpha \frac{\partial E_k}{\partial \sigma_i^l} \Big|_q \quad (3.20)$$

$$\sigma_i^l(q+1) = \sigma_i^l(q) - \alpha \frac{(y^* - y_k)}{b} (\bar{y}^l(q) - y^*) z^l \frac{2(x_{ki} - \bar{x}_i^l(q))^2}{\sigma_i^{l3}(q)} \quad (3.21)$$

donde  $l = 1, 2, \dots, M$ ;  $q = 0, 1, 2, \dots$

### 3.3 Aprendizaje por Estimación de Mínimos Cuadrados

En un Sistema de Inferencia Difuso basado en una Red Adaptable que está conformado por un sistema difuso TSK de primer orden se cuenta con un vector  $\mathbf{B}^l = [B_1^l \dots B_n^l B_{n+1}^l]$  por regla difusa  $l = 1 \dots M$  y que en conjunto son el grupo  $S_2$  de parámetros a definir en un modelo.



de esta manera se tiene una matriz de diseño  $\mathbf{A}$  donde cada vector en función de un dato de entrenamiento se denota también como  $\left[ \mathbf{a}_i^T; y_i^* \right]$ , para conformar así la matriz de datos  $\left[ \mathbf{A}; \mathbf{y} \right]$ .

Para una matriz cuadrada con  $K = M$  vectores iniciales de entrenamiento, el vector de parámetros desconocidos se obtiene como

$$\mathbf{B} = \mathbf{A}^{-1} \mathbf{y} \quad (3.25)$$

En la identificación de sistemas es común tomar en cuenta un mayor número de vectores de entrenamiento que de parámetros ( $K > M$ ), por lo cual es necesario ponderar los vectores a través de la estimación de parámetros  $\mathbf{B}_i$  en que se incluye un parámetro de ruido  $e$  - siendo en un inicio un factor unitario- que refleja la diferencia de patrones que considera el modelo para la descripción del sistema. De esta manera, el ajuste de los parámetros de identificación con  $K > M$  patrones de datos se realiza minimizando la suma cuadrada del error definida como

$$E(\mathbf{B}) = \sum_{i=1}^K (y_i^* - \mathbf{a}_i^T \mathbf{B})^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{A}\mathbf{B})^T (\mathbf{y} - \mathbf{A}\mathbf{B}) \quad (3.26)$$

Para la obtención de la estimación de parámetros  $\hat{\mathbf{B}} = \mathbf{B}$  que minimiza la suma cuadrada del error, se introducen las siguientes principios o lemmas en el cálculo de matrices:

### **Lemma 3.1 Propiedad de la transpuesta de matrices**

Teniendo  $\mathbf{A}$  y  $\mathbf{B}$  como matrices compatibles, entonces

$$(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T \mathbf{A}^T \quad (3.27)$$

donde  $T$  denota la transpuesta de una matriz.

### **Lemma 3.2 Propiedad de inversa de matrices**

Teniendo  $\mathbf{A}$  y  $\mathbf{B}$  como matrices compatibles y no singulares, entonces

$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1} \quad (3.28)$$

donde el superscript  $-1$  denota la inversa de una matriz.

### Lemma 3.3 Función cuadrática óptima

Teniendo cualquier función cuadrática en  $f(\mathbf{x})$  que en notación matricial se escriba como

$$f(\mathbf{x}) = \mathbf{xMx} + 2\mathbf{b}^T\mathbf{x} + c \quad (3.29)$$

donde  $\mathbf{M}$  es una matriz simétrica definida positiva, entonces  $f(\mathbf{x})$  alcanza su mínimo en  $\mathbf{x} = -\mathbf{M}^{-1}\mathbf{b}$ .

La ecuación de la suma cuadrada del error (3.26) es una función cuadrática que resulta al notar que  $\mathbf{B}^T\mathbf{A}^T\mathbf{y} = \mathbf{y}^T\mathbf{AB}$ , de manera que se obtiene la siguiente expansión

$$E(\mathbf{B}) = (\mathbf{y}^T - \mathbf{B}^T\mathbf{A}^T)(\mathbf{y} - \mathbf{AB}) = \mathbf{B}^T\mathbf{A}^T\mathbf{AB} - 2\mathbf{y}^T\mathbf{AB} + \mathbf{y}^T\mathbf{y} \quad (3.30)$$

Al ser una función cuadrática  $E(\mathbf{B}) = f(\mathbf{x})$  se le aplica el Lemma 3.3 con  $\mathbf{M} = \mathbf{A}^T\mathbf{A}$ ,  $\mathbf{b}^T = -\mathbf{y}^T\mathbf{A}$  y  $c = \mathbf{y}^T\mathbf{y}$ , obteniéndose la estimación de mínimos cuadrados como

$$\hat{\mathbf{B}} = (\mathbf{A}^T\mathbf{A})^{-1} \mathbf{A}^T\mathbf{y} \quad (3.31)$$

En la anterior derivación se considera que cada uno de los vectores iniciales tiene la misma ponderación en la estimación de parámetros del modelo, aunque para sistemas variantes en el tiempo, un nuevo patrón entrada-salida puede tener un mayor peso en la descripción paramétrica del sistema. Es por esto que la ecuación de la suma de cuadrados se puede generalizar agregándole una matriz de pesos simétrica y definida positiva

$$E_W(\mathbf{B}) = (\mathbf{y} - \mathbf{AB})^T \mathbf{W} (\mathbf{y} - \mathbf{AB}) \quad (3.32)$$

donde  $\mathbf{W}$  es una matriz diagonal

$$\mathbf{W} = \begin{bmatrix} \lambda^{K-1} & 0 & \dots & 0 \\ 0 & \lambda^{K-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \quad (3.33)$$

para  $0 < \lambda \leq 1$ .

La minimización de esta ecuación se obtiene de manera análoga a (3.31), resultando en

$$\hat{\mathbf{B}}_W = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y} \quad (3.34)$$

de manera que si  $\mathbf{W}$  es una matriz identidad, ambas ecuaciones son equivalentes.

### 3.3.2 Estimación por Mínimos Cuadrados Recursivos

Los parámetros iniciales  $\hat{\mathbf{B}}$  se obtienen resolviendo el sistema de ecuaciones de la forma (1.4) a partir de  $K \geq M$  patrones de vectores iniciales que representen valores genéricos en la activación de cada regla. Subsecuentes vectores de entrenamiento  $K+1$  hasta  $P$  podrían actualizar en línea el valor de los parámetros  $\hat{\mathbf{B}}$  sin volver a conformar la matriz  $\mathbf{A}$  y realizar de nuevo la estimación de parámetros con la solución (3.31). Más bien se utiliza el algoritmo de mínimos cuadrados recursivos, el cual se basa en una operación análoga obtenida con la expresión:

$$\hat{\mathbf{B}}_{K+1} = \left( \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \quad (3.35)$$

En la derivación del algoritmo se introducen las matrices auxiliares  $\mathbf{C}_K$  y  $\mathbf{C}_{K+1}$  de  $M \times M$  como

$$\mathbf{C}_K = (\mathbf{A}^T \mathbf{A})^{-1} \quad (3.36)$$

$$\begin{aligned} \mathbf{C}_{K+1} &= \left( \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} \\ &= \left( \begin{bmatrix} \mathbf{A}^T & \mathbf{a} \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} \\ &= (\mathbf{A}^T \mathbf{A} + \mathbf{a} \mathbf{a}^T)^{-1} \end{aligned} \quad (3.37)$$

Además considerando que

$$\mathbf{C}_{K+1} = (\mathbf{C}_K^{-1} + \mathbf{a} \mathbf{a}^T)^{-1} \quad (3.38)$$

se obtiene una relación entre las dos matrices como

$$\mathbf{C}_K^{-1} = \mathbf{C}_{K+1}^{-1} - \mathbf{a} \mathbf{a}^T \quad (3.39)$$

De esta manera las ecuaciones (3.31) y (3.35) se reescriben respectivamente como

$$\hat{\mathbf{B}}_K = \mathbf{C}_K \mathbf{A}^T \mathbf{y} \quad (3.40)$$

$$\hat{\mathbf{B}}_{K+1} = \mathbf{C}_{K+1} (\mathbf{A}^T \mathbf{y} + \mathbf{a} y^*) \quad (3.41)$$

La integración de estas ecuaciones se realiza eliminando el término  $\mathbf{A}^T \mathbf{y}$  de la segunda (3.41) al observar que  $\mathbf{A}^T \mathbf{y} = \mathbf{C}_K^{-1} \hat{\mathbf{B}}_K$  obteniendo

$$\hat{\mathbf{B}}_{K+1} = \mathbf{C}_{K+1} (\mathbf{C}_K^{-1} \hat{\mathbf{B}}_K + \mathbf{a} y^*) \quad (3.42)$$

que dada la relación (3.39) entre matrices, esta ecuación se expresa en términos de  $\mathbf{C}_{K+1}$  como

$$\begin{aligned} \hat{\mathbf{B}}_{K+1} &= \mathbf{C}_{K+1} \left[ (\mathbf{C}_{K+1}^{-1} - \mathbf{a} \mathbf{a}^T) \hat{\mathbf{B}}_K + \mathbf{a} y^* \right] \\ \hat{\mathbf{B}}_{K+1} &= \hat{\mathbf{B}}_K + \mathbf{C}_{K+1} \mathbf{a} (y^* - \mathbf{a}^T \hat{\mathbf{B}}_K) \end{aligned} \quad (3.43)$$

En esta ecuación se nota que la actualización de parámetros se obtiene calculando un error  $(y^* - \mathbf{a}^T \hat{\mathbf{B}}_K)$  a partir del nuevo vector de datos y el actual vector de parámetros; tal error es multiplicado a su vez por un vector de ganancia en la adaptación  $\mathbf{C}_{K+1} \mathbf{a}$ .

Como se aprecia en la ecuación (3.38), el término  $\mathbf{C}_{K+1}$  involucra la inversión de una matriz  $M \times M$  la cual sería substituida por una fórmula incremental al aplicar el siguiente lemma.

### Lemma 3.4 Fórmula de inversión de matrices

Teniendo  $\mathbf{R}$  y  $\mathbf{I} + \mathbf{T} \mathbf{R}^{-1} \mathbf{S}$  como matrices no singulares cuadradas, entonces

$$(\mathbf{R} + \mathbf{S} \mathbf{T})^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{S} (\mathbf{I} + \mathbf{T} \mathbf{R}^{-1} \mathbf{S})^{-1} \mathbf{S} \mathbf{R}^{-1} \quad (3.44)$$

Aplicando el Lemma 3.4 a la expresión (3.38) con  $\mathbf{R} = \mathbf{C}_K^{-1}$ ,  $\mathbf{S} = \mathbf{a}$  y  $\mathbf{T} = \mathbf{a}^T$  se obtiene la siguiente fórmula incremental para  $\mathbf{C}_{K+1}$  como:

$$\mathbf{C}_{K+1} = \mathbf{C}_K - \mathbf{C}_K \mathbf{a} \left( \mathbf{I} + \mathbf{a}^T \mathbf{C}_K \mathbf{a} \right)^{-1} \mathbf{a}^T \mathbf{C}_K \quad (3.45)$$

$$\mathbf{C}_{K+1} = \mathbf{C}_K - \frac{\mathbf{C}_K \mathbf{a} \mathbf{a}^T \mathbf{C}_K}{1 + \mathbf{a}^T \mathbf{C}_K \mathbf{a}}$$

En resumen, las fórmulas para obtener la actualización de los parámetros  $\hat{\mathbf{B}}$  por el algoritmo de mínimos cuadrados recursivos se reescriben como:

$$\mathbf{C}_{K+1} = \mathbf{C}_K - \frac{\mathbf{C}_K \mathbf{a}_{K+1} \mathbf{a}_{K+1}^T \mathbf{C}_K}{1 + \mathbf{a}_{K+1}^T \mathbf{C}_K \mathbf{a}_{K+1}} \quad (3.46)$$

$$\hat{\mathbf{B}}_{K+1} = \hat{\mathbf{B}}_K + \mathbf{C}_{K+1} \mathbf{a}_{K+1} \left( y_{K+1}^* - \mathbf{a}_{K+1}^T \hat{\mathbf{B}}_K \right) \quad (3.47)$$

donde  $\left[ \mathbf{a}_{K+1}^T; y_{K+1}^* \right]$  está en términos de un nuevo vector de entrenamiento agregado a los  $K$  iniciales con que se forma la matriz  $\left[ \mathbf{A}; \mathbf{y} \right]$  para la obtención de los parámetros iniciales  $\mathbf{B}_K$ ; de esta manera las fórmulas (3.46) y (3.47) se utilizarían en la actualización de  $\mathbf{B}_{K+1}$  para los vectores restantes  $K + 1$  hasta  $P$  disponibles en el entrenamiento de la red adaptable.

### 3.3.3 Método de Mínimos Cuadrados Recursivos para Sistemas Variantes en el Tiempo

Durante el proceso de identificación por mínimos cuadrados recursivos se realiza un actualización de parámetros que involucra la obtención de  $\mathbf{C}_K = \left( \mathbf{A}^T \mathbf{A} \right)^{-1}$  para el grupo completo de datos de entrenamiento de  $k$  hasta  $P$ . Este proceso de refinamiento del modelo se lleva a cabo a través de la ganancia en la adaptación  $\mathbf{C}_{K+1} \mathbf{a}_{K+1}$  en la ecuación (3.47), la cual decrece con el número de datos  $k$  por el efecto en la minimización del error (3.26) para sistemas invariantes en el tiempo:

$$\lim_{k \rightarrow \infty} P_k = \lim_{k \rightarrow \infty} \frac{1}{k} \left( \frac{1}{k} \mathbf{A}^T \mathbf{A} \right)^{-1} = 0 \quad (3.48)$$

Este resultado es consecuencia de que a todos los vectores  $k$  que conforman la matriz  $\mathbf{A}$  se les asigne un mismo peso en la estimación de parámetros, lo cual sería válido para sistemas invariantes en el tiempo en que un decremento en la adaptación significaría que se está llegando al óptimo.

Por otra parte, en los sistemas en que la dinámica va cambiando, los nuevos vectores tendrían una mayor importancia en la estimación de los parámetros del modelo por lo que se incorpora un factor de atenuación  $\lambda$  al proceso recursivo de minimización del error, el cual le da un menor énfasis a los vectores anteriores al darle un mayor peso al más reciente:

$$E(\mathbf{B}) = \sum_{k=K}^P \lambda^{P-k} (y_k - \mathbf{a}_k^T \mathbf{B})^2 = (\mathbf{y} - \mathbf{A}\mathbf{B})^T \mathbf{W} (\mathbf{y} - \mathbf{A}\mathbf{B}) \quad (3.49)$$

Esta ecuación, la cual es equivalente a (3.32), es la que describe la suma de cuadrados generalizada cuya minimización resultó en

$$\hat{\mathbf{B}}_K = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y} \quad (3.50)$$

la cual es la base del método de mínimos cuadrados recursivos para sistemas variantes en el tiempo.

Siguiendo un procedimiento análogo al de la sección anterior, se plantea la incorporación de un nuevo vector en la identificación paramétrica

$$\hat{\mathbf{B}}_{K+1} = \left( \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \lambda \mathbf{W} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \lambda \mathbf{W} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \quad (3.51)$$

$$\hat{\mathbf{B}}_{K+1} = (\lambda \mathbf{A}^T \mathbf{W} \mathbf{A} + \mathbf{a} \mathbf{a}^T)^{-1} (\lambda \mathbf{A}^T \mathbf{W} \mathbf{y} + \mathbf{a} y^*)$$

De esta manera las matrices auxiliares  $\mathbf{C}_K$  y  $\mathbf{C}_{K+1}$  quedan definidas como

$$\mathbf{C}_K = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \quad (3.52)$$

$$\mathbf{C}_{K+1} = \left( \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \lambda \mathbf{W} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} = (\lambda \mathbf{A}^T \mathbf{W} \mathbf{A} + \mathbf{a} \mathbf{a}^T)^{-1} \quad (3.53)$$

cuya relación es

$$\mathbf{C}_{K+1} = (\lambda \mathbf{C}_K^{-1} + \mathbf{a} \mathbf{a}^T)^{-1} \quad (3.54)$$

$$\lambda \mathbf{C}_K^{-1} = \mathbf{C}_{K+1}^{-1} - \mathbf{a}\mathbf{a}^T \quad (3.55)$$

Reescribiendo (3.50) y (3.51) en término de las matrices auxiliares

$$\hat{\mathbf{B}}_K = \mathbf{C}_K \mathbf{A}^T \mathbf{W} \mathbf{y} \quad (3.56)$$

$$\hat{\mathbf{B}}_{K+1} = \mathbf{C}_{K+1} (\lambda \mathbf{A}^T \mathbf{W} \mathbf{y} + \mathbf{a} y^*) \quad (3.57)$$

Estas ecuaciones se conjuntan a través del término  $\mathbf{A}^T \mathbf{W} \mathbf{y}$

$$\hat{\mathbf{B}}_{K+1} = \mathbf{C}_{K+1} (\lambda \mathbf{C}_K^{-1} \hat{\mathbf{B}}_K + \mathbf{a} y^*) \quad (3.58)$$

Esta última ecuación se reescribe en términos de  $\mathbf{C}_{K+1}$  incorporando (3.55) y manipulando algebraicamente

$$\hat{\mathbf{B}}_{K+1} = \hat{\mathbf{B}}_K + \mathbf{C}_{K+1} \mathbf{a} (y^* - \mathbf{a}^T \hat{\mathbf{B}}_K) \quad (3.59)$$

Para evitar la inversa en el cálculo de la matriz  $\mathbf{C}_{K+1}$  se aplica de nuevo el Lemma 3.4 a la ecuación (3.54) pero ahora con  $\mathbf{R} = \lambda \mathbf{C}^{-1}$ ,  $\mathbf{S} = \mathbf{a}$ ,  $\mathbf{I} = \mathbf{W}$  y  $\mathbf{T} = \mathbf{a}^T$ , lo que completa el algoritmo de mínimos cuadrados recursivos

$$\mathbf{C}_{K+1} = \frac{1}{\lambda} \mathbf{C}_K - \frac{1}{\lambda} \mathbf{C}_K \mathbf{a} (\mathbf{W} + \mathbf{a}^T \frac{\mathbf{C}_K}{\lambda} \mathbf{a})^{-1} \mathbf{a}^T \frac{\mathbf{C}_K}{\lambda} \quad (3.60)$$

$$\mathbf{C}_{K+1} = \frac{1}{\lambda} \left( \mathbf{C}_K - \frac{\mathbf{C}_K \mathbf{a} \mathbf{a}^T \mathbf{C}_K}{\lambda + \mathbf{a}^T \mathbf{C}_K \mathbf{a}} \right)$$

En resumen, para sistemas variantes en el tiempo, las fórmulas a utilizar en la identificación paramétrica por mínimos cuadrados recursivos se expresan como

$$\mathbf{C}_{K+1} = \frac{1}{\lambda} \left( \mathbf{C}_K - \frac{\mathbf{C}_K \mathbf{a}_{K+1} \mathbf{a}_{K+1}^T \mathbf{C}_K}{\lambda + \mathbf{a}_{K+1}^T \mathbf{C}_K \mathbf{a}_{K+1}} \right) \quad (3.61)$$

$$\hat{\mathbf{B}}_{K+1} = \hat{\mathbf{B}}_K + \mathbf{C}_{K+1} \mathbf{a}_{K+1} (y_{K+1}^* - \mathbf{a}_{K+1}^T \hat{\mathbf{B}}_K) \quad (3.62)$$

## Capítulo 4

# Caso de Estudio: Pronóstico del Tipo de Cambio Libra Esterlina - Dolar Americano

### 4.1 Introducción

En este capítulo se presenta un problema de modelación para propósito de pronóstico en el contexto de un sistema de inferencia difusa basado en una red adaptable. El sistema en específico en este caso de estudio corresponde a la cotización del tipo de cambio entre la libra esterlina y el dolar americano, el cual se modela a partir de la configuración en series de tiempo de los datos estipulados.

La configuración de los datos en forma de series de tiempo permite la conformación de vectores entrada-salida que sirven para la estructuración de reglas difusas. Tales enunciados condicionales describiendo los estados del sistema son una base de conocimiento que en conjunto con un sistema de inferencia definen una arquitectura específica de modelación difusa a la que al aplicarle algoritmos de aprendizaje toman la funcionalidad de una red adaptable, tal y como se ha venido estudiando.

La forma en que se plantea este caso de estudio busca ir de acorde a la definición del problema y a las hipótesis de investigación por lo que se utilizan arquitecturas de sistemas TSK de orden cero y primer orden en la configuración inicial de conocimiento implícito en los patrones de datos, para que a partir de esta identificación difusa inicial se aplique el algoritmo de aprendizaje correspondiente a la estructura que se sintoniza. Por último se presentan resultados de la aplicación de ambas arquitecturas de sistema de inferencia difusa basado en una red adaptable ante diferentes configuraciones de valores del parámetro de aprendizaje, número de funciones de membresía, número de reglas y aprendizaje por lote o en línea.

## 4.2 Descripción del Caso de Estudio y de la Modelación Difusa Inicial

El universo de discurso  $U$  del presente caso de estudio corresponde a la cotización del tipo de cambio entre la libra esterlina y el dólar americano por semana, en específico del jueves, desde comienzos de 1990 hasta finales de junio del 2000. Estos valores correspondientes a 548 semanas se presentan en la figura 4.1.

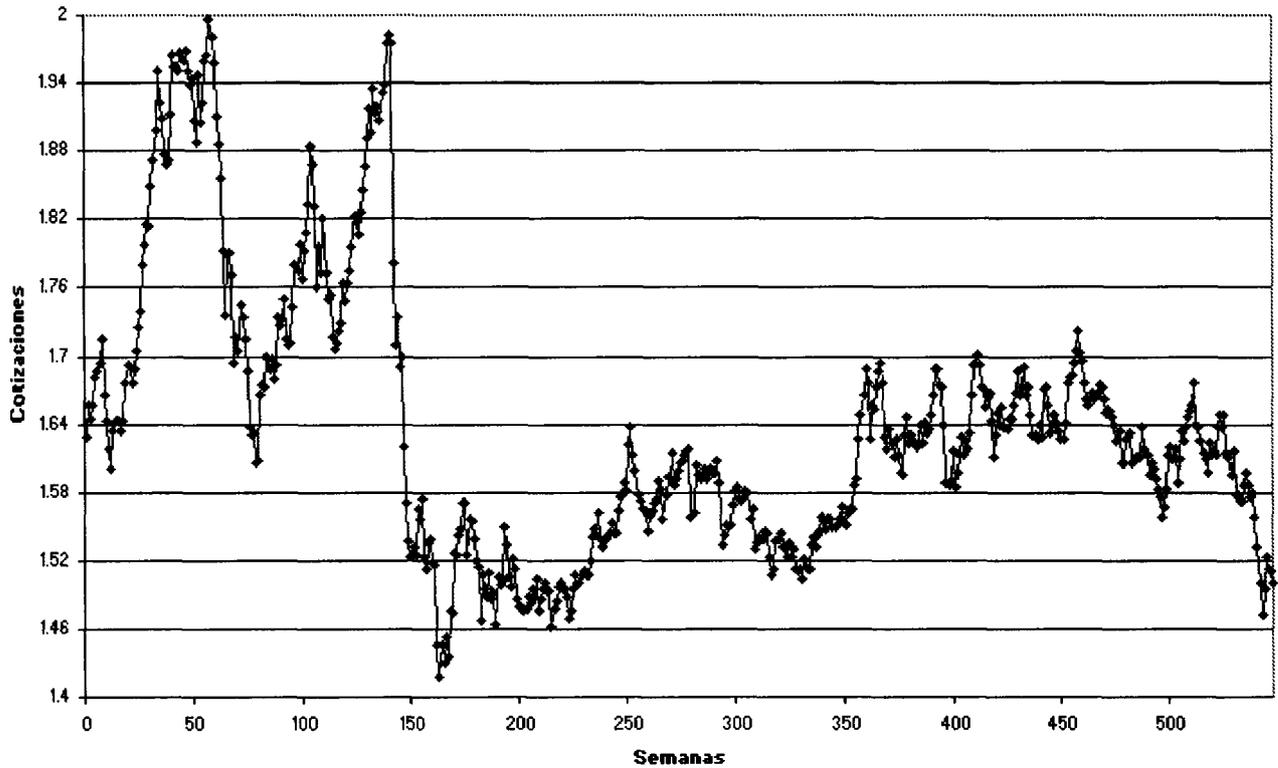


Figura 4.1: Variación en las cotizaciones del tipo de cambio entre la libra esterlina y el dólar americano

Para atrapar la dinámica en la variación del tipo de cambio se realiza una modelación por series de tiempo a partir del siguiente esquema

$$[X(k - n + 1), X(k - n + 2), \dots, X(k); X(k + 1)] \quad (4.1)$$

donde  $k$  es un contador de semanas y  $n$  es el número de semanas anteriores que se toman en cuenta para atrapar la dinámica del estado  $k + 1$ .

Un esquema de modelación semejante por series de tiempo se presenta en [5] donde

se estudia el efecto de las ventanas de tiempo para pronósticos en horizontes cortos utilizando una arquitectura de redes adaptables diferente a este estudio. En [5] se asevera que la mayoría de los estudios en la literatura adoptan la práctica de arbitrariamente dividir los datos disponibles en un grupo de entrenamiento (de muestra) para la construcción del modelo y un grupo de evaluación (fuera de muestra) para la validación del modelo. En este estudio también se dividen los datos históricos - presentados en la figura 4.1 - en dos secciones, la primera comprendida de 1990 hasta finales del primer trimestre del 2000 se utiliza para la conformación de la modelación inicial y la posterior sintonización por algoritmos de aprendizaje; la segunda sección de datos, corresponde a 13 semanas, se emplea en la evaluación del modelo para comparación de arquitecturas de modelación difusa, tal y como se planteó en la definición del problema y en las hipótesis de investigación de las secciones 1.1 y 1.2 respectivamente.

A partir de la configuración de vectores  $\left[ \mathbf{x}_k : y_k \right]$  de patrones entrada-salida, donde  $\mathbf{x}_k = \left[ X(k-n+1) \ X(k-n+2) \ \cdots \ X(k) \right]^T$  y  $y_k = X(k+1)$  para  $k = 1 \dots P$ , se procede con la modelación por sistemas difusos para lo cual, tal y como resume Wang [7], se debe empezar por obtener una colección de reglas *Si-Entonces* difusas a partir de expertos humanos o basadas en conocimiento del dominio. El siguiente paso es combinar estas reglas en un sólo sistema. Diferentes sistemas difusos utilizan diferentes principios para esta combinación.

El planteamiento del dominio para la modelación difusa se definió por el universo de discurso  $U$  y se delimita a  $P = 535$  datos de entrenamiento, por lo que la base de conocimiento representa conocimiento cualitativo sobre la variación en las cotizaciones configuradas en forma de series de tiempo a partir del esquema (4.1). Siguiendo tal esquema, en este caso se definen  $n = 2$  semanas y la actual en la ventana de series de tiempo como antecedente de las reglas difusas y se obtiene una cotización de la siguiente semana como consecuente, para completar así el mes de datos; es decir, se construye la base de reglas al examinar los patrones de datos entrada-salida como descripciones vagas, lo que representaría enunciados como

SI la cotización 2 semanas anteriores está alrededor de 1.65 Y

la cotización 1 semanas anteriores está alrededor de 1.66 Y

la cotización de la semana actual está alrededor de 1.677 ENTONCES

la cotización de la próxima semana está cerca a 1.68 dólares/libra

Reglas lingüísticas como la anterior expresan conocimiento implícito sobre el sistema y donde la palabra alrededor debe ser precisada en términos matemáticos por

medio de funciones de membresía que definen y especifican conocimiento humano.

Como parte del diseño de un sistema difuso se debe fijar el número de funciones de membresía y buscar que los parámetros de éstas estén definidos de manera que para cada componente del vector  $x$  se encuentre un conjunto difuso  $A_i$  en cada variable  $x_i$  del antecedente de una regla, tal que su grado de pertenencia sea mayor o igual a 0.5. Por otra parte, las reglas deben tener significativamente diferente activación para que en el momento en que se integren a una arquitectura de sistema difuso TSK de primer orden, los parámetros del consecuente se puedan determinar por la estimación de mínimos cuadrados (3.31). Entonces, dado el planteamiento del problema y buscando cumplir las anteriores condiciones, se utilizan funciones de membresía gaussiana (2.4) de las mismas características de amplitud  $\sigma_i$  por variable  $x_i$  pero con diferente distribución y desfazamiento de centros entre variables  $x_i, x_{i+1}$  subsecuentes.

Como conocimiento experto adicional, se considera que las semanas más anteriores tienen una menor influencia en la cotización de la siguiente semana, lo que se representa con la asignación de un menor número de funciones de membresía, con una variación de rango de valores que cubren. Al variar la amplitud  $\sigma_i$ , varía el número de conjuntos difusos por variable  $x_i$  y el número de reglas difusas necesarias para la descripción de la dinámica del sistema, la cual es una descripción por modelación difusa en el sentido que haya un mapeo entre pares de estados entrada-salida para combinaciones de estados definidos en el dominio. Para explorar los anteriores planteamientos - con las implicaciones dadas en las hipótesis de investigación de la sección 1.2 - se definen tres estructuras de modelación difusa inicial, las cuales se enumeran en la tabla 4.1.

Tabla 4.1: Funciones de membresía por variable para los sistemas difusos de este caso de estudio

|        | $x_3$ | $x_2$ | $x_1$ | $Y$ | Número de reglas |
|--------|-------|-------|-------|-----|------------------|
| Caso A | 6     | 6     | 7     | 7   | 9                |
| Caso B | 6     | 7     | 12    | 12  | 12               |
| Caso C | 7     | 8     | 8     | 8   | 14               |

Para definir las reglas  $l$  en cada uno de los tres casos de modelación difusa se calculó el promedio de la variación en la cotización de una semana a otra, siendo de 0.016 dolares a cada lado, por lo que en la selección de conjuntos difusos entre variables  $x_i^l$  para configurar una regla, debe haber un desfazamiento de centros de por lo menos 0.032 dolares, buscando con esto que cada regla tenga significativamente diferente activación, lo que sería necesario para que en un sistema difuso TSK de primer orden, los parámetros del consecuente se puedan determinar por (3.31). Por otra parte

se debe tener un número suficiente de reglas para la descripción adecuada del sistema; a este respecto cabe destacar que como se utilizan funciones de membresía gaussiana - ver figura 2.1 - todos los elementos en vectores de datos tienen soporte de un conjunto difuso en más de una regla; es decir, se tienen reglas continuas, completas y consistentes para tal dominio en el universo de discurso  $U$ .

Una vez definida la base de conocimiento se integra ésta con una arquitectura de sistema difuso con el difusificador, máquina de inferencia y desdifusificador adecuados. La primera arquitectura de sistema difuso a utilizar está compuesta por un difusificador *singleton*, máquina de inferencia producto y desdifusificador centro promedio, tal que se obtiene un sistema difuso equivalente al tipo TSK de orden cero. Tal como se ha planteado anteriormente y de acuerdo con la definición del problema de la sección 1.1, esta primer arquitectura se ampliaría para cada uno de los tres casos de la tabla 4.1 a un sistema difuso TSK de primer orden, esto al ampliar el consecuente  $y$  en términos de  $\bar{y}_l$  a funciones lineales  $f_l$  para cada regla  $l = 1 \dots M$ .

A los sistemas difusos tipo TSK de primer orden y de orden cero se les implementa la funcionalidad de red adaptable, esto al agregarles capacidad de aprendizaje híbrido y de sólo gradiente descendente respectivamente; de esta forma quedan conformadas dos arquitecturas de lo que en forma genérica se conoce como Sistema de Inferencia Difuso basada en una Red Adaptable. Para ambas arquitecturas el proceso de aprendizaje es un procedimiento de entrenamiento en lote o en línea en que se sintonizan los parámetros de estructuras de modelación inicial para los casos de la tabla 4.1.

La comparación de resultados para los diferentes casos de modelación inicial se realizará con los datos de evaluación a través de la medición del error cuadrado medio definido en (3.13), el cual es en principio el que se busca minimizar con el aprendizaje.

Cabe señalar, tal y como se menciona en [3], que a diferencia de entrenamiento en lote, el algoritmo de gradiente descendente versión línea no es un procedimiento para minimizar el error  $E$  de los datos, sin embargo, al basarse en  $E_p$ , este algoritmo de aprendizaje es necesario en la identificación de parámetros en línea para sistemas con características cambiantes. Además del entrenamiento en lote, en este caso de estudio se busca representar el aprendizaje versión en línea, el cual se aproximaría a uno en lote si el parámetro de aprendizaje es pequeño. De esta forma se utilizarían parámetros *alpha* de  $\alpha = 0.01$  y  $\alpha = 0.001$  para comparar los resultados de los datos de evaluación después del entrenamiento. En las siguientes secciones se reporta el error cuadrado medio (3.13) de la identificación difusa para datos de evaluación, lo que en [5] se maneja como error de pronóstico aunque con diferente medición del mismo para otra arquitectura de red adaptable. En [5] se concluye que la medición del error aumenta

con el horizonte de pronóstico, esto sin que sea significativo el esquema de evaluación, por lo que en nuestro caso se explora el error de las diferentes identificaciones difusas de la tabla 4.1 para un mes y para el conjunto de los 13 datos semanales.

### 4.3 Resultados de la Identificación Difusa Sintonizada con un Algoritmo por Gradiente Descendente

La integración de cada uno de los casos de modelación difusa inicial a una arquitectura de sistema TSK de orden cero da como resultado cuatro sistemas difusos que en forma genérica se representan en fórmula compacta por (3.9). Para la sintonización de los parámetros de la modelación difusa inicial, a estos sistemas difusos se le implementa un algoritmo de aprendizaje por gradiente descendente en el que en su versión en línea y en lote, dan como resultado las mediciones de error de la figura 4.2 para los datos de evaluación. Dado que el error cuadrado medio da errores de pronóstico del orden de  $10^{-3}$  o menores, se utiliza mejor el criterio de Porcentaje Error Medio Absoluto, el criterio PEMA, empleado también en [5], se define como

$$PEMA = \frac{1}{T} \sum \left| \frac{y_t - y^*}{y_t} \right| \times 100 \quad (4.2)$$

donde  $T$  es el número de datos utilizados en la medición del error.

De los casos de sistemas difusos con funcionalidad de red adaptable presentados en la figura 4.2, en ninguno se dio la situación de que con el entrenamiento en línea o en lote, los centros  $e_i^l, e_{i+1}^l, \dots, e_n^l$  de funciones de membresía en las variables  $x_i, x_{i+1}, \dots, x_n$  del antecedente de una regla  $l$  se traspasaron o intercalaron con centros  $e_i^{l+1}, e_{i+1}^{l+1}, \dots, e_n^{l+1}$  subsecuentes en tales variables  $x_i, x_{i+1}, \dots, x_n$  respectivamente; de haber acontecido una situación contraria a la anterior descripción, se diría que los conjuntos difusos en variables  $x_i$  conformando reglas  $l$  perdieron su interpretabilidad inicial con el entrenamiento.

Por otra parte, dado que se busca explorar el error para horizontes cortos de pronóstico, en la siguiente tabla 4.2 se muestran los resultados de evaluación para el primer mes exclusivamente, lo que se expresa como  $P + x$  semanas después de datos de entrenamiento.

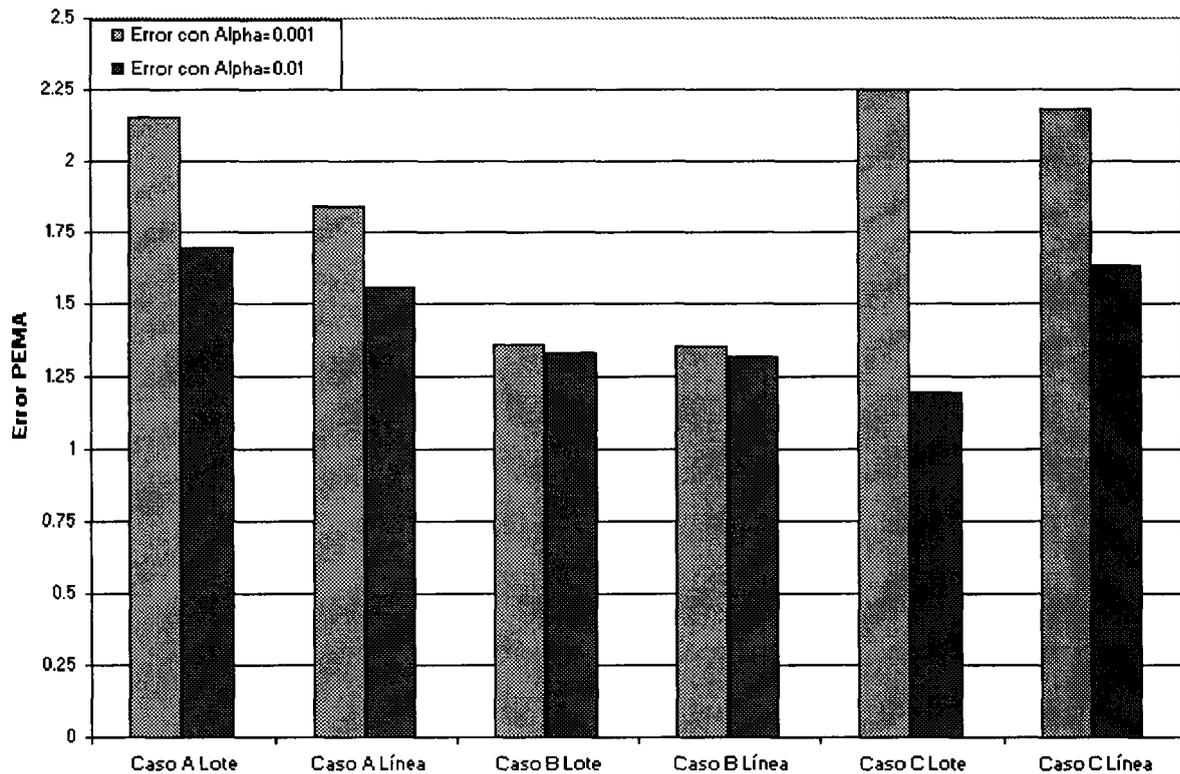


Figura 4.2: Error PEMA para los datos de evaluación, sistema TSK de orden cero

Continuando con la observación de resultados de la modelación difusa utilizando diferente número de reglas, se procede a graficar en la figura 4.3 el caso con la menor medición PEMA para el grupo completo de datos de evaluación

Observando las mediciones de error de pronóstico en el primer mes, se considera que la mejor modelación difusa es la del Caso A entrenado en lote con  $\alpha=0.01$ . Para mostrar la posterior variación en la dinámica de las cotizaciones se grafica en 4.4 tal caso para el grupo completo de datos de evaluación.

## 4.4 Resultados de la Identificación Difusa Sintonizada por un Algoritmo de Aprendizaje Híbrido

A los anteriores casos de modelación difusa inicial integrados a arquitecturas TSK de orden cero se les amplía el consecuente  $y$  en términos de  $\bar{y}^l$  a una función lineal  $f_l = B_1^l * x_1^l + \dots + B_n^l * x_n^l + B_{n+1}^l$  para cada regla  $l = 1 \dots M$ , lo que resulta en casos de sistemas TSK de primer orden. A estos sistemas se les agrega la funcionalidad de red

Tabla 4.2: Error PEMA para cada dato en un mes de evaluación, sistema TSK de orden cero

|             |              | semana P+1 | semana P+2 | semana P+3 | semana P+4 |
|-------------|--------------|------------|------------|------------|------------|
| Alpha=0.001 | Caso A Lote  | 0.121741   | 0.702075   | 1.398381   | 1.250659   |
|             | Caso A Línea | 0.209043   | 0.789992   | 1.486941   | 1.339081   |
| Alpha=0.01  | Caso A Lote  | 0.755379   | 0.180122   | 0.510087   | 0.363659   |
|             | Caso A Línea | 0.621656   | 1.208083   | 1.908679   | 1.75921    |
| Alpha=0.001 | Caso B Lote  | 1.331388   | 0.701186   | 0.036954   | 0.216315   |
|             | Caso B Línea | 1.330799   | 0.697891   | 0.035759   | 0.216099   |
| Alpha=0.01  | Caso B Lote  | 1.270691   | 0.64132    | 0.023614   | 0.155062   |
|             | Caso B Línea | 1.240912   | 0.58727    | 0.054275   | 0.129589   |
| Alpha=0.001 | Caso C Lote  | 0.078445   | 1.886786   | 2.935778   | 2.607789   |
|             | Caso C Línea | 0.420601   | 1.823106   | 2.808286   | 2.169293   |
| Alpha=0.01  | Caso C Lote  | 2.327902   | 0.458742   | 0.533352   | 0.18338    |
|             | Caso C Línea | 2.112126   | 0.347302   | 0.321658   | 1.143967   |

adaptable a través del algoritmo híbrido de aprendizaje, el cual divide los parámetros a sintonizar en dos grupos:  $S_1$ , los del antecedente de las reglas difusas, y  $S_2$  los del consecuente. Los parámetros  $S_1$  se sintonizan por un algoritmo de aprendizaje por gradiente descendente y en la determinación de los parámetros del consecuente  $S_2$  se utiliza un algoritmo por estimación de mínimos cuadrados. De esta manera, el algoritmo híbrido de aprendizaje sintoniza los parámetros  $S_1$  y  $S_2$  en diferentes etapas secuenciales, tal y como se describe en la identificación del problema sección 1.1.

En la determinación de los parámetros  $S_2$  iniciales se utilizan vectores entrada-salida que representan patrones típicos descritos por una regla y se realiza la estimación de mínimos cuadrados (3.31). Entonces, teniendo los valores iniciales tanto para  $S_1$  y  $S_2$  se realiza la sintonización de cada uno de los casos de sistemas difusos a través de  $k = 1 \dots P =$  datos de entrenamiento en línea y en lote por el algoritmo híbrido de aprendizaje para sistemas variantes en el tiempo con factor de atenuación de 0.98. Los resultados del error PEMA (4.2) para cada caso de identificación difusa se presentan en la figura 4.5.

De los casos de sistemas difusos con funcionalidad de red adaptable presentados en la figura 4.5, en ninguno se dio la situación de que con el entrenamiento en línea o en lote, los centros  $e_i^l, e_{i+1}^l, \dots, e_n^l$  de funciones de membresía en las variables  $x_i, x_{i+1}, \dots, x_n$  del antecedente de una regla  $l$  se traspasaron o intercalaron con centros  $e_i^{l+1}, e_{i+1}^{l+1}, \dots, e_n^{l+1}$  subsecuentes en tales variables  $x_i, x_{i+1}, \dots, x_n$  respectivamente; es decir, al igual que en el sistema TSK de orden cero, los conjuntos difusos no perdieron la interpretabilidad inicial con el entrenamiento.

De los casos presentados en la figura 4.5, para obtener pronósticos con cotizaciones

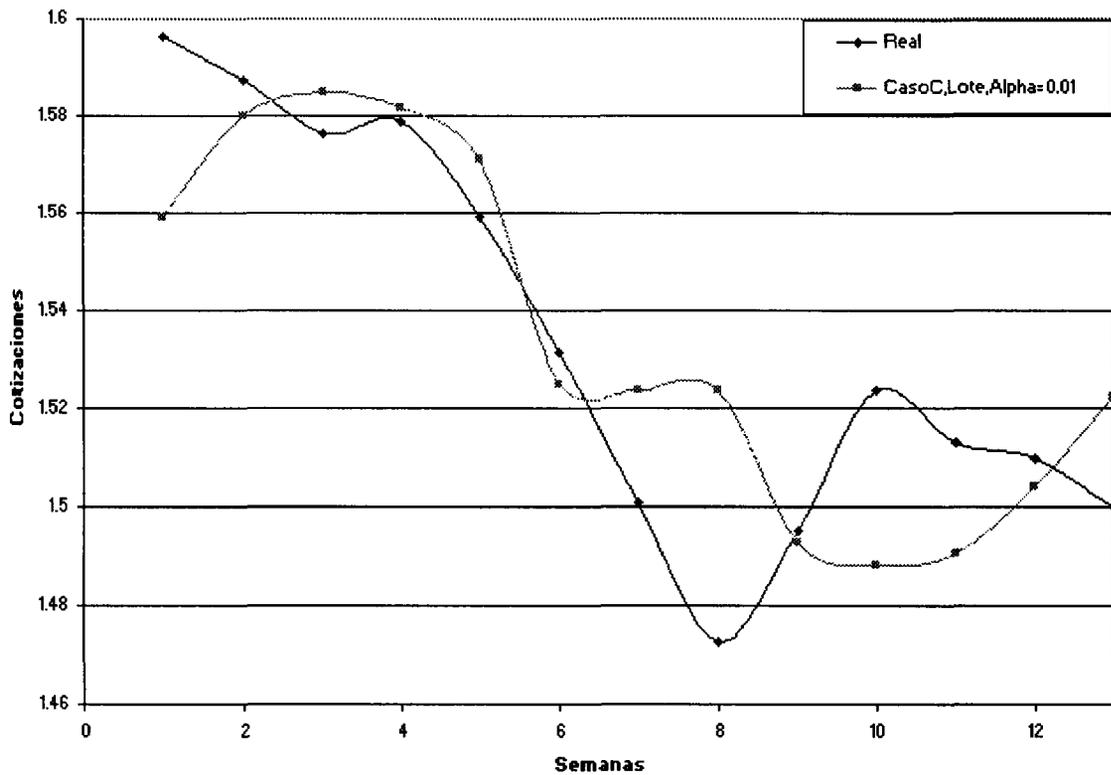


Figura 4.3: Caso C del sistema TSK de orden cero utilizando entrenamiento en lote con un parámetro de aprendizaje de 0.01

factibles en el caso C utilizando entrenamiento en línea, se eliminó la primera regla describiendo patrones de datos en el rango inferior de cotizaciones, lo que dió resultados semejantes al mismo caso con el grupo completo de reglas utilizando entrenamiento en lote con  $\alpha = 0.001$ ; aumentando el parámetro de aprendizaje, el error con una modelación inicial inadecuada aumenta significativamente. Por otra parte, los errores PEMA para horizontes cortos de pronóstico se presentan en la tabla 4.3.

Continuando con la observación de resultados de la modelación difusa utilizando diferente número de reglas, se procede a graficar en la figura 4.6 el caso con la menor medición de error PEMA para el grupo completo de datos de evaluación.

Observando las mediciones de error de pronóstico en el primer mes, se considera que la mejor modelación difusa es la del Caso B utilizando entrenamiento en lote con  $\alpha = 0.001$ . Para mostrar la posterior variación en la dinámica de las cotizaciones se grafica este caso para el grupo completo de datos de evaluación en la figura 4.7.

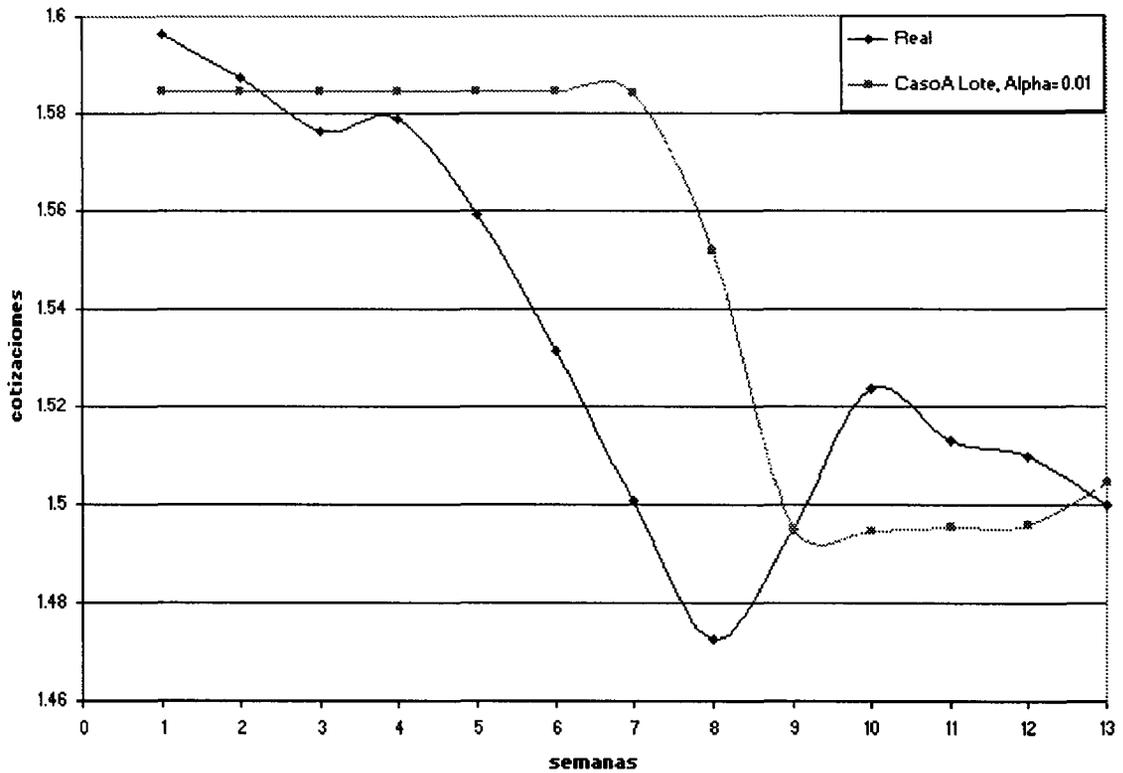


Figura 4.4: Caso A utilizando entrenamiento en lote con un parámetro de aprendizaje de 0.01

Tabla 4.3: Error PEMA para cada dato en un mes de evaluación, sistema TSK de primer orden

|             |              | semana P+1 | semana P+2 | semana P+3 | semana P+4 |
|-------------|--------------|------------|------------|------------|------------|
| Alpha=0.001 | Caso A Lote  | 0.499745   | 1.164837   | 0.069319   | 0.817655   |
|             | Caso A Linea | 0.589453   | 1.228778   | 0.060187   | 0.831271   |
| Alpha=0.01  | Caso A Lote  | 7.084101   | 6.980064   | 6.011517   | 5.585575   |
|             | Caso A Linea | 0.650809   | 0.549592   | 0.095013   | 0.695736   |
| Alpha=0.001 | Caso B Lote  | 0.6552     | 0.186856   | 0.116922   | 0.056019   |
|             | Caso B Linea | 0.6629     | 0.197748   | 0.119596   | 0.044295   |
| Alpha=0.01  | Caso B Lote  | 0.655961   | 0.178557   | 0.12893    | 0.056924   |
|             | Caso B Linea | 0.658838   | 0.154627   | 0.145656   | 0.054475   |
| Alpha=0.001 | Caso C Lote  | 1.894363   | 0.763338   | 0.241241   | 0.495935   |
|             | Caso C Linea | 0.767549   | 0.520782   | 0.377649   | 0.614422   |
| Alpha=0.01  | Caso C Lote  | 1.551896   | 0.504165   | 0.130153   | 0.590502   |
|             | Caso C Linea | 5.149662   | 4.085834   | 1.486491   | 1.509158   |

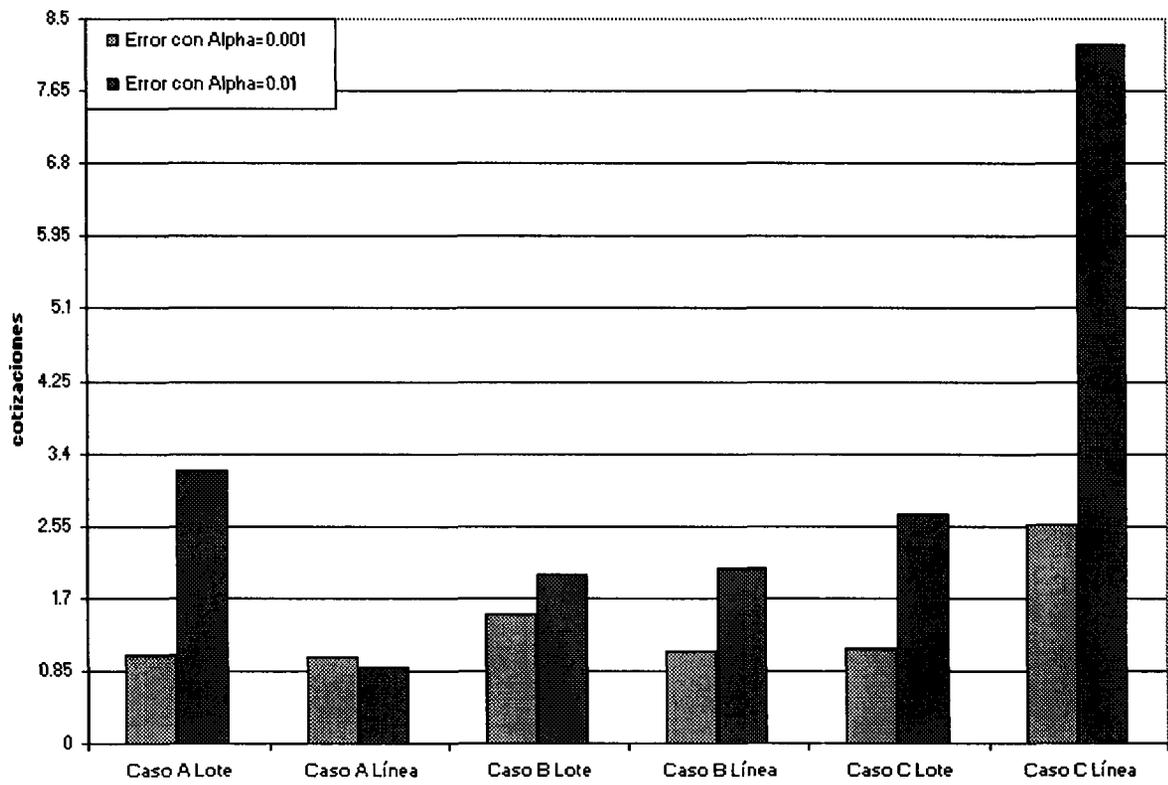


Figura 4.5: Error PEMA para los datos de evaluación, sistema TSK de primer orden

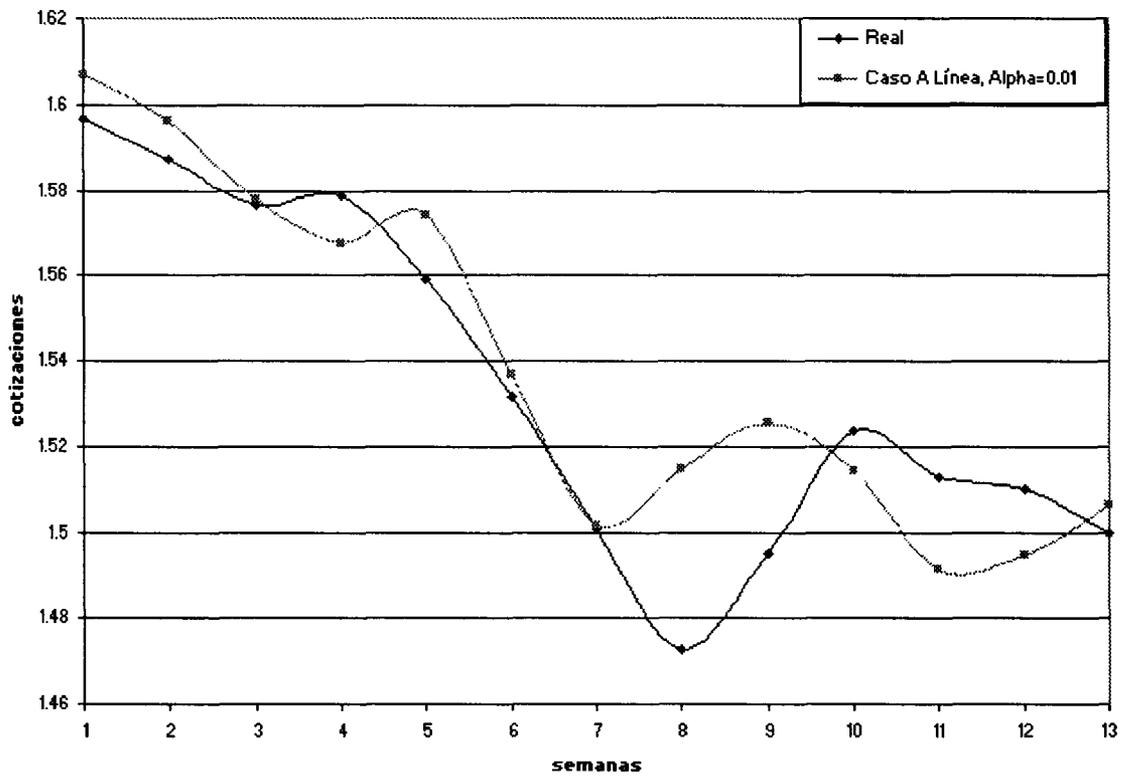


Figura 4.6: Caso A del sistema TSK de primer orden utilizando entrenamiento en línea con un parámetro de aprendizaje de 0.01

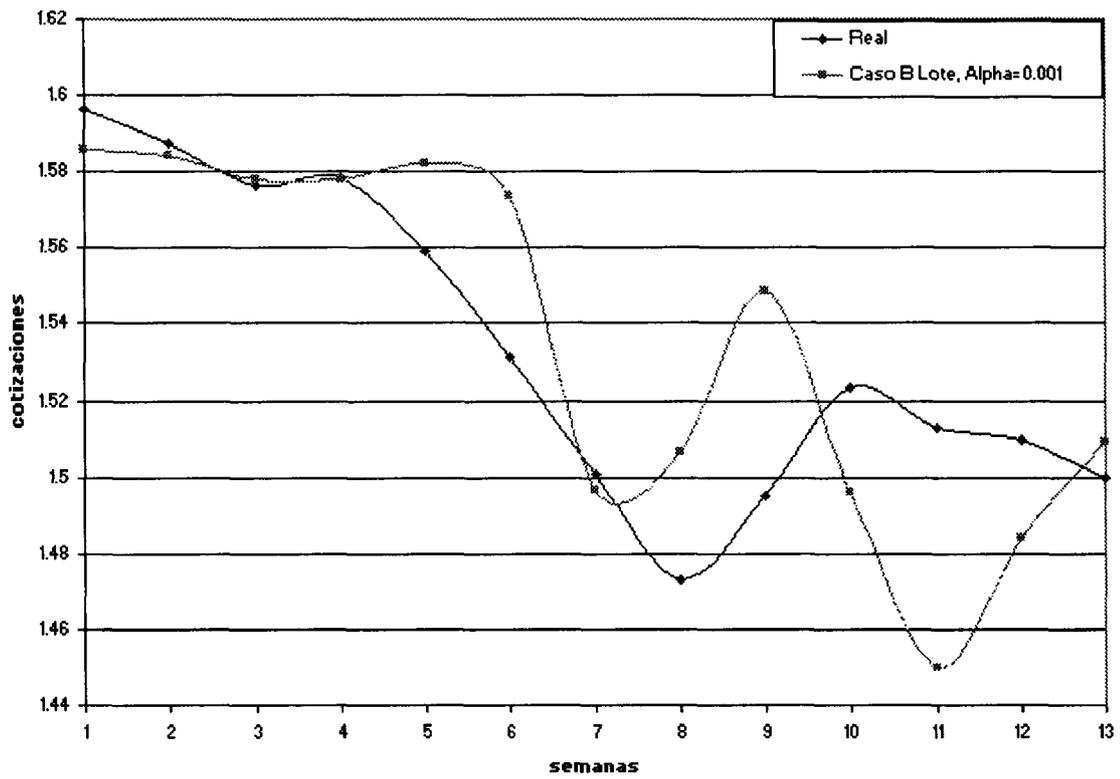


Figura 4.7: Caso B del sistema TSK de primer orden utilizando entrenamiento en lote con un parámetro de aprendizaje de 0.001

## Capítulo 5

### Conclusiones

Partiendo del planteamiento de la definición del problema e hipótesis de investigación, en esta tesis se estudiaron los fundamentos de sistemas difusos, la igual que dos arquitecturas y los correspondientes procedimientos de aprendizaje detrás de un sistema de inferencia difusa basado en una red adaptable para la identificación de sistemas, en específico para la modelación difusa de la cotización del tipo de cambio entre la libra esterlina y el dolar americano.

A la modelación difusa inicial integrada a arquitecturas de sistemas difusos TSK de orden cero y primer orden se le agregó la funcionalidad de un sistema con aprendizaje supervisado, lo que conforma instancias de un sistema de inferencia difusa basado en una red adaptable.

Al sistema difuso TSK de orden cero se le aplicó un algoritmo de aprendizaje por gradiente descendente para entrenamiento en línea y en lote. Este método de aprendizaje minimiza la suma del error cuadrado medio  $E$  para datos de entrenamiento cuando se aplica en lote, lo que no resulta ser equivalente cuando se basa en el error  $E_p$  con entrenamiento en línea, sin embargo dan aproximadamente el mismo resultado en la identificación de las primeras semanas cuando el parámetro de aprendizaje es pequeño,  $\alpha = 0.001$ , tal y como se muestra con el criterio PEMA, lo que equivaldría a un error cuadrado medio del orden de  $10^{-5}$  por predicción para el primer mes; cabe señalar que el error aumenta dada la dinámica del sistema en los dos subsecuentes meses de datos de evaluación. La variación de la medida del error con un mayor parámetro de aprendizaje dió resultados mixtos para los datos de evaluación según el caso de modelación inicial y el horizonte de tiempo en que se realiza la medición del error.

Una discusión semajante a la anterior gira en torno al número de reglas pues en la tercer hipótesis de evaluación se maneja óptimos en el error para datos de evaluación cuando lo que se minimiza es el error en los datos de entrenamiento. El número de reglas repercute en el sentido en que agrupa los conjuntos difusos, los cuales cada uno debe tener una activación significativa y diferenciada por un patrón entrada-salida para

poder determinar los parámetros iniciales del sistema TSK de primer orden.

Lo que podría perder la interpretabilidad inicial en los sistemas difusos adaptables con el aprendizaje sería la definición de valores en el dominio por funciones de membresía que se traslapen e intercalan cuando corresponden a conjuntos difusos subsecuentes que guardan un orden jerárquico, tal como en reglas difusas en que el antecedente se conforme por términos cardinales u ordinales. Otro punto diferente sería que las variables lingüísticas perdieran la interpretabilidad inicial con el aprendizaje, lo cual no es coherente, sino más bien correspondería a crear una diferente modelación difusa.

Desde el punto de vista de la última hipótesis en que se cuestiona ahora sí la interpretabilidad de los conjuntos difusos del antecedente de las reglas, habría que señalar que en los resultados en el caso de estudio en ninguna ocasión los centros  $e_i^l, e_{i+1}^l, \dots, e_n^l$  de funciones de membresía en las variables  $x_i, x_{i+1}, \dots, x_n$  del antecedente de una regla  $l$  se traspasaron o intercalaron con centros  $e_i^{l+1}, e_{i+1}^{l+1}, \dots, e_n^{l+1}$  subsecuentes en tales variables  $x_i, x_{i+1}, \dots, x_n$  respectivamente.

En el caso de que para una variable lingüística  $x_i$  uno de los centros  $e_i^l$  se haya traspasado con uno subsecuente  $e_i^{l+1}$  en el proceso de aprendizaje, se aseveraría, en términos de las hipótesis, que se perdió la interpretabilidad del arreglo inicial de conjuntos difusos en la modelación difusa, situación que no acontece por lo que se puede concluir que para cada caso se tenía una modelación inicial adecuada.

## Capítulo 2

# Fundamentos Teóricos de Sistemas Difusos

## 2.1 Introducción

La teoría difusa tiene sus inicios en la década de 1960 a partir de las publicaciones de Lofti A. Zadeh quien formaliza el término en el artículo “Fuzzy Sets” [9]. En sus publicaciones Zadeh introduce una teoría cuyos objetos -conjuntos difusos- son conjuntos cuyas fronteras no son precisas. La pertenencia o membresía de un conjunto difuso no es exclusiva a la afirmación o negación, sino más bien a cuestiones de grado [4]. Como principio se busca que los conjuntos difusos transformen las descripciones vagas en lenguaje natural a términos matemáticos precisos, esto por medio de funciones de membresía que definen y especifican conocimiento humano [7].

Al arreglar los conjuntos difusos en enunciados *Si-Entonces* se conforman las llamadas reglas difusas que son la base de la lógica difusa. El objetivo de la lógica difusa es proveer fundamentos para la realización de razonamiento aproximado; es decir, se estudian métodos y principios para obtener nuevas proposiciones a partir de los enunciados existentes. Tales procedimientos se generalizan a través de una máquina de inferencia, que en conjunto con una base de conocimiento, conforma un sistema difuso.

## 2.2 Conceptos sobre Lógica Difusa

En nuestra vida diaria, las palabras son comúnmente usadas para describir variables. En la teoría de sistemas difusos, una variable de este tipo toma valores que son caracterizados por conjunto difusos definidos en el universo de discurso en el cual tal variable es definida [7]. Los conjuntos difusos son una forma básica de agrupación del conocimiento cualitativo sobre un sistema, esto a través de funciones de membresía que mapean valores del dominio físico del universo de discurso a palabras. En esta sección se plantean estos conceptos básicos de sistemas difusos para dar paso a la conformación de proposiciones difusas y a la descripción de operaciones de lógica en *conjuntos difu-*

sos. De esta manera se describen las reglas difusas cuya relación entre proposiciones en las premisas y en el consecuente se da en términos de implicaciones. Por último se introduce el principio de inferencia deductivo Modus Ponens, el cual en su forma generalizada es la base del proceso de razonamiento aproximado en la teoría de lógica difusa.

## 2.2.1 Variables Lingüísticas y Conjuntos Difusos

Una variable lingüística se considera como el elemento fundamental en la representación del conocimiento ya que toma palabras en el lenguaje natural como sus valores. En la teoría difusa cada valor lingüístico o palabra que conforma una variable lingüística es un objeto -conjunto difuso- cuyas fronteras no son precisas y cuya definición está caracterizada por un función de membresía en un universo de discurso  $U$ .

Un universo de discurso contiene todos los posibles elementos concernientes a un contexto en particular o aplicación; de esta forma, cada palabra de una variable lingüística está asociada a valores numéricos del universo de discurso de manera que existe una relación semántica entre el correspondiente valor lingüístico y su conjunto difuso en el dominio físico.

Una manera de representar la relación entre un valor lingüístico de una variable lingüística  $X$  y un conjunto difuso  $A$  en  $U$  es por medio de pares ordenados de los elementos genéricos  $x$  y su valor de membresía  $\mu_A(x)$  a  $A$ :

$$A = \{(x, \mu_A(x)) | x \in U\} \quad (2.1)$$

De manera general, un conjunto difuso  $A$  en un universo de discurso  $U$  continuo es descrito como:

$$A = \int_U \mu_A(x)/x \quad (2.2)$$

En el caso discreto, el conjunto difuso se representa como:

$$A = \sum_U \mu_A(x)/x \quad (2.3)$$

En un dominio continuo las funciones de membresía describen figuras geométricas o de campana con altura máxima de 1 y mínimo de cero, como las que muestran a continuación.

Función gaussiana (2.4) y su representación en la figura 2.1 sobre un universo de discurso  $U$ :

$$\mu_{A_i}(x) = \exp \left\{ - \left( \frac{x - \bar{x}_i}{\sigma_i} \right)^2 \right\} \quad (2.4)$$

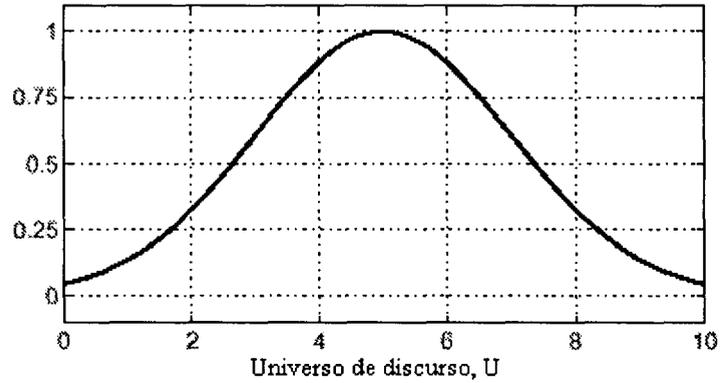


Figura 2.1: Función de membresía gaussiana

Función campana (2.5) y su representación en la figura 2.2 sobre un univeso de discurso  $U$ :

$$\mu_{A_i}(x) = \frac{1}{1 + \left[ \left( \frac{x - \bar{x}_i}{\sigma_i} \right)^2 \right]^{b_i}} \quad (2.5)$$

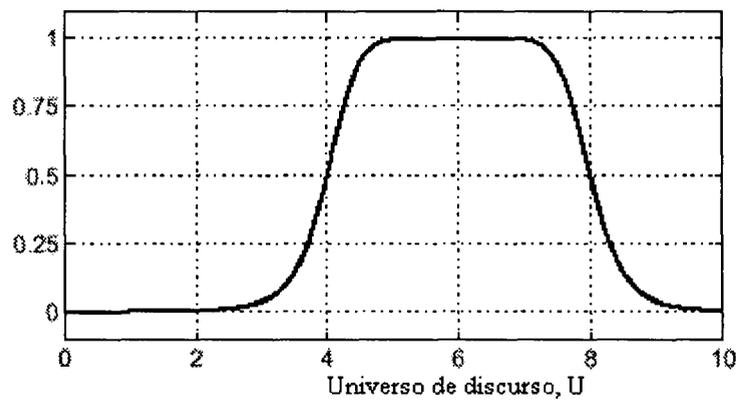


Figura 2.2: Función de membresía de campana gaussiana

Un corte  $\alpha$  de un conjunto difuso se define como los elementos de  $A$  cuyo valor de membresía es mayor o igual que  $\alpha$ , es decir:

$$A_\alpha = \{x \in U | \mu_A(x) \geq \alpha\} \quad (2.6)$$

En el caso de utilizar más de un conjunto difuso para cubrir el universo de discurso  $U$ , si existe una intersección entre un par de funciones de membresía, al valor del dominio  $U$  en que ocurre se le llama punto de cruce.

El centro  $e_i$  de un conjunto difuso es el valor medio de  $U$  cuyo valor de membresía es máximo para ese conjunto. Los elementos del universo de discurso cuyo valor de membresía es diferente de cero, se les llama el soporte del conjunto difuso, es decir:

$$supp(A) = \{x \in U | \mu_A(x) > 0\} \quad (2.7)$$

Una proposición difusa FP atómica es un enunciado de la forma

$$x \text{ es } A$$

donde  $x$  es una variable lingüística y  $A$  es un valor lingüístico de  $x$ ; en otras palabras,  $A$  es un conjunto difuso definido en el dominio físico de  $x$  por medio de la función de membresía  $\mu_A(x)$ .

### 2.2.2 Proposiciones Difusas Compuestas y Operaciones en Conjuntos Difusos

La composición de dos o más proposiciones difusas atómicas para formar una proposición difusa FP compuesta se da a través de conectivos "Y", "O" y "NO", de manera que en lógica difusa tal intersección, unión y negación respectivamente tienen su propia definición.

La negación es una operación de complemento difuso  $c : [0, 1] \rightarrow [0, 1]$ , el cual es un mapeo que transforma las funciones de membresía de un conjunto difuso  $A$  en la función de membresía del complemento de  $A$ :  $c[\mu_A(x)] = \mu_{\bar{A}}(x)$ . La operación complemento más comúnmente utilizada para la negación es

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (2.8)$$

La unión es una operación Norma-S ( $\oplus$ ),  $s : [0, 1] \times [0, 1] \rightarrow [0, 1]$ , el cual es un mapeo que transforma las funciones de membresía de los conjunto difusos  $A$  y  $B$ , en una función de membresía de unión de  $A$  y  $B$  tal que  $s[\mu_A(x), \mu_B(x)] = \mu_{A \cup B}(x)$ . La norma-S más comúnmente utilizada es la operación máximo

$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] \quad (2.9)$$

La intersección es una operación Norma-T ( $\otimes$ ),  $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ , el cual es un mapeo que transforma las funciones de membresía de los conjunto difusos  $A$  y  $B$ , en una función de membresía intersección de  $A$  y  $B$ :  $t[\mu_A(x), \mu_B(x)] = \mu_{A \cap B}(x)$ . Las operaciones norma-T más comúnmente utilizadas son el producto algebraico

$$\mu_{A \cap B}(x) = \mu_A(x) * \mu_B(x) \quad (2.10)$$

y el mínimo

$$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)] \quad (2.11)$$

La proposición difusa FP compuesta es, en otras palabras, una relación difusa de dos o más conjuntos difusos (i.e.  $A, B$ ) comprendidos en variables lingüísticas (i.e.  $x, y$ ) de manera tal que la relación de las proposiciones atómicas está expresada en términos de las operaciones norma-s, norma-t y complemento definidas anteriormente; así una forma de representar la proposición difusa FP compuesta:  $x$  es  $A$  Y  $y$  es  $B$ , sería:  $\mu_{A \cap B}(x, y) = t[\mu_A(x), \mu_B(y)]$ .

### 2.2.3 Reglas Difusas y Razonamiento Aproximado

Las reglas difusas son enunciados condicionales de la forma: SI  $\langle$  *proposición difusa*  $FP_1$  $\rangle$ , ENTONCES  $\langle$  *proposición difusa*  $FP_2$  $\rangle$  donde las proposiciones difusas de las premisas están relacionadas con el consecuente a través de implicaciones como las siguientes:

La implicación Dienes-Rescher está conformada por los operadores lógicos de negación y unión definidos en la sección anterior, para así obtener una relación difusa  $Q_D$  en  $U \times V$  con la siguiente función de membresía

$$\mu_{Q_D}(x, y) = \max[1 - \mu_{FP_1}(x), \mu_{FP_2}(y)] \quad (2.12)$$

La implicación Zadeh es una relación difusa  $Q_Z$  en  $U \times V$  expresada por la función de membresía

$$\mu_{Q_Z}(x, y) = \max[\min(\mu_{FP_1}(x), \mu_{FP_2}(y)), 1 - \mu_{FP_1}(x)] \quad (2.13)$$

Las implicaciones Mamdani son implicaciones locales en el sentido que pueden ser interpretadas como SI  $\langle FP_1 \rangle$  ENTONCES  $\langle FP_2 \rangle$  DE LO CONTRARIO  $\langle NADA \rangle$ . La relación difusa  $Q_{MM}$  ó  $Q_{MP}$  en  $U \times V$  del anterior enunciado estaría expresado por las funciones de membresía (2.14) y (2.15) respectivamente:

$$\mu_{Q_{MM}}(x, y) = \min[\mu_{FP_1}(x), \mu_{FP_2}(y)] \quad (2.14)$$

$$\mu_{Q_{MP}}(x, y) = \mu_{FP_1}(x) * \mu_{FP_2}(y) \quad (2.15)$$

Al representar una regla o enunciado condicional por medio de alguna de las anteriores implicaciones se obtiene una relación que a su vez es una premisa en un proceso de inferencia deductivo llamado Modus Ponens. En lógica clásica, la regla de inferencia Modus Ponens declara que para dos proposiciones  $p$  y  $p \rightarrow q$ , se infiere la verdad de la proposición  $q$ ; esto se expresa como:

Premisa 1, proposición atómica  $p$ :  $x$  es  $A$

Premisa 2,  $p \rightarrow q$ : SI  $x$  es  $A$  ENTONCES  $y$  es  $B$

Conclusión  $q$ :  $y$  es  $B$

En lógica difusa se sigue un principio de inferencia análogo usando proposiciones imprecisas representadas por conjuntos difusos. Tal proceso de razonamiento aproximado se realiza con la regla Modus Ponens generalizada, la cual declara que para dos proposiciones difusas  $x$  es  $A'$  y SI  $x$  es  $A$  ENTONCES  $y$  es  $B$ , se infiere una nueva proposición difusa  $y$  es  $B'$  de manera tal que lo más cercano esté  $A'$  de  $A$ , lo más cercano estará  $B'$  de  $B$ ; es decir:

Premisa 1:  $x$  es  $A'$

Premisa 2: SI  $x$  es  $A$  ENTONCES  $y$  es  $B$

Conclusión:  $y$  es  $B'$

La conclusión del proceso de razonamiento aproximado es el resultado de la composición del conjunto difuso de la Premisa 1 con la implicación de la Premisa 2,

la cual se expresa como una relación difusa en el producto cartesiano de los dominios de  $x$  y  $y$ . La composición es el resultado de integrar dos relaciones binarias (i.e.  $P(U,V)$  y  $Q(V,W)$ ), o un conjunto difuso con una relación binaria difusa (i.e.  $\mu_{A'}(x), \mu_Q(x,y)$ ), a partir de un dominio o variable que tengan en común; en el primero de los casos el resultado de la composición es una relación difusa y en el segundo es un conjunto difuso. Una expresión general de la composición  $P(U,V) \circ Q(V,W)$  sería:  $\mu_{P \circ Q}(x,z) = \sup_{y \in V} t[\mu_P(x,y), \mu_Q(y,z)]$  donde  $(x,z) \in U \times W$ , y  $t$  es alguna norma- $t$ . De esta manera, según la norma- $t$  utilizada, se podría tener alguna de las siguientes dos composiciones:

La *composición max-min* de las relaciones difusas  $P(U,V)$  y  $Q(V,W)$  es una relación  $P \circ Q$  en  $U \times W$  definida por la función de membresía

$$\mu_{P \circ Q}(x,z) = \max_{y \in V} \min[\mu_P(x,y), \mu_Q(y,z)] \quad (2.16)$$

donde  $(x,z) \in U \times W$ .

La *composición max-producto* de las relaciones difusas  $P(U,V)$  y  $Q(V,W)$  es una relación  $P \circ Q$  en  $U \times W$  definida por la función de membresía

$$\mu_{P \circ Q}(x,z) = \max_{y \in V} [\mu_P(x,y) * \mu_Q(y,z)] \quad (2.17)$$

donde  $(x,z) \in U \times W$ .

La regla Modus Ponens generalizada en términos de la composición del conjunto difuso  $A'$  (la cual expresa la premisa  $x$  es  $A'$ ) y de la relación difusa  $A \rightarrow B$  en  $U \times V$  (la cual representa premisas de la forma SI  $x$  es  $A$  ENTONCES  $y$  es  $B$ ) quedaría expresada como

$$\mu_{B'}(y) = \sup_{x \in U} t[\mu_{A'}(x), \mu_{A \rightarrow B}(x,y)] \quad (2.18)$$

donde el conjunto difuso  $\mu_{B'}(y)$  es la nueva proposición “ $x$  es  $B$ ” inferida del presente proceso de razonamiento aproximado.

## 2.3 Sistemas Difusos

En los conceptos de lógica difusa se explicó el proceso de razonamiento Modus Ponens usando un sólo enunciado condicional difuso y alguna proposición correspondiente al antecedente de la implicación, lo que nos daba como resultado el consecuente de la regla.

Al describir un sistema dinámico a través de múltiples reglas difusas se debe considerar cada una de ellas en el proceso de razonamiento para obtener el próximo estado del sistema; este proceso es desarrollado por una máquina de inferencia difusa, la cual en conjunto con una base de reglas conforma un sistema difuso.

### 2.3.1 Base de Reglas Difusas

La base de reglas está compuesta por enunciados condicionales difusos que en su forma básica se expresan como

$$R_u^{(l)} : \text{SI } x_1 \text{ es } A_1^l \text{ Y } \dots \text{ Y } x_n \text{ es } A_n^l , \text{ENTONCES } y \text{ es } B^l \quad (2.19)$$

donde  $A_i^l$  y  $B^l$  son conjuntos difusos en  $U_i \subset R$  y  $V \subset R$  respectivamente y  $\mathbf{x} = [x_1, \dots, x_n]^T \in U$  y  $y \in V$  son variables lingüísticas de entrada y salida. En otras palabras, cada premisa está formada de  $i = 1, 2, \dots, n$  variables lingüísticas, las cuales a su vez tienen de  $j = 1, 2, \dots, m$  valores diferentes que son representados de manera difusa por medio de funciones de membresía. La cantidad de funciones de membresía para cada variable puede ser diferente dependiendo del dominio cuantitativo real que represente la variable.

En la definición de la estructura de las funciones de membresía que cubren el dominio de una variable lingüística habría que considerar la condición de anchura, representada en la versión 1 de la figura 2.3 la cual establece que la anchura izquierda de  $\mu_{A_2}$  debería ser igual a la anchura derecha de  $\mu_{A_1}$  y ambas son iguales a la longitud del intervalo entre el centro de dos funciones de membresía adyacentes. La segunda versión de las funciones de membresía no cumple la condición de anchura por lo que hay una región de valores en el dominio de la única variable en el antecedente  $x_1$  que activaría sólo una regla, esto evita transiciones suaves a la salida dado que no se pondera la pertenencia de un valor en la frontera de los dos conjuntos difusos  $\mu_{A_1}$  y  $\mu_{A_2}$ .

Además de lo anterior se desprende que el número de variables lingüísticas y valores de las mismas repercuten en el número de las reglas que se podrían formar; por

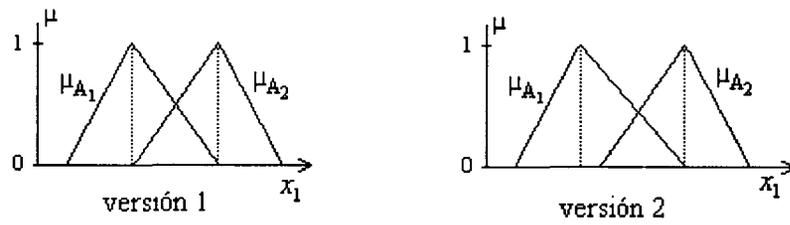


Figura 2.3: Condición de anchura para un sistema difuso con una variable en el antecedente de las reglas

ejemplo, teniendo dos variables:  $x_1$  con tres funciones de membresía  $\mu_{A_1^1}$ ,  $\mu_{A_2^1}$  y  $\mu_{A_3^1}$ ; y  $x_2$  con dos funciones de membresía  $\mu_{A_1^2}$  y  $\mu_{A_2^2}$  se tendría el siguiente grupo completo de enunciados *Si-Entonces* difusos:

1. SI  $x_1$  es  $\mu_{A_1^1}$  Y  $x_2$  es  $\mu_{A_1^2}$  ENTONCES  $y$  es  $B^1$
2. SI  $x_1$  es  $\mu_{A_1^1}$  Y  $x_2$  es  $\mu_{A_2^2}$  ENTONCES  $y$  es  $B^2$
3. SI  $x_1$  es  $\mu_{A_2^1}$  Y  $x_2$  es  $\mu_{A_1^2}$  ENTONCES  $y$  es  $B^3$
4. SI  $x_1$  es  $\mu_{A_2^1}$  Y  $x_2$  es  $\mu_{A_2^2}$  ENTONCES  $y$  es  $B^4$
5. SI  $x_1$  es  $\mu_{A_3^1}$  Y  $x_2$  es  $\mu_{A_1^2}$  ENTONCES  $y$  es  $B^5$
6. SI  $x_1$  es  $\mu_{A_3^1}$  Y  $x_2$  es  $\mu_{A_2^2}$  ENTONCES  $y$  es  $B^6$

Otra manera de ver las anteriores reglas es mediante la siguiente tabla 2.4.

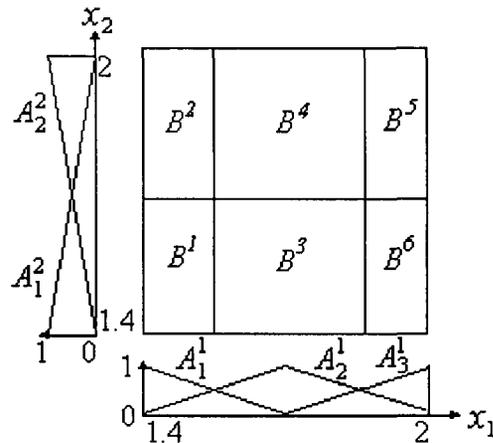


Figura 2.4: Ejemplo de funciones de membresía para un sistema difuso de dos entradas

Cuando se forman reglas con todas las posibles combinaciones de funciones de membresía de cada variable, se dice que se tiene una base de reglas completa. Una desventaja de este tipo de base de reglas es que su tamaño aumenta exponencialmente con la dimensión de espacio de entradas  $U_i$ . En general, en la mayoría de los procesos a modelar, las entradas no toman valores arbitrarios por lo que para tener completés de las reglas no sería necesario combinar todas las funciones de membresía para cada variable, sino más bien se busca tener una base de reglas en la que para vectores factibles de entrada  $\mathbf{x} \in U$  exista al menos una regla  $Ru^{(l)}$  de la forma (2.19) tal que  $\mu_{A_i^l}(x_i) \neq 0$  para toda  $i = 1, 2, \dots, n$ .

Se define a un grupo de reglas como consistente si no existe alguna regla que tenga el mismo antecedente pero diferente consecuente. Así también, se dice que un grupo de enunciados *Si-Entonces* difusos es continuo si los conjuntos difusos del consecuente de reglas adyacentes dan como resultado salidas con transiciones suaves.

### 2.3.2 Máquina de Inferencia Difusa

En los sistemas difusos se manejan dos perspectivas generales de la manera en que el proceso de razonamiento aproximado debería utilizar el grupo de reglas para inferir nuevos estados. Por una parte, en una máquina de inferencia basada en composición, cada relación difusa  $R_u^{(l)}$  representando un enunciado condicional en el espacio de entrada-salida en  $U \times V$  es combinada para formar una sólo relación difusa  $Q_M$  en  $U \times V$  que es interpretada como una sólo regla difusa; de esta manera, al tener un conjunto difuso  $A'$  como premisa de entrada a la máquina de inferencia, se aplicaría el principio de razonamiento Modus-Ponens generalizado para a partir de la relación difusa  $Q_M$  obtener un conjunto difuso resultante  $\mu_{B'}(y)$ . Por otra parte, en una máquina de inferencia basada en reglas individuales, el principio de razonamiento Modus-Ponens generalizado se aplica a cada regla por separado, por lo que al tener un nuevo conjunto difuso de entrada  $A'$  en  $U$ , el proceso de inferencia generaría un conjunto difuso  $\mu_{B'}(y)$  por cada regla, los cuales tendrían que ser combinados por alguna norma-t  $\otimes$  ó norma-s  $\oplus$ .

La máquina de inferencia basada en composición realiza el proceso de inferencia de manera interpretable, por lo que tiene sentido considerar la combinación de las reglas como la unión de las relaciones difusas que representan cada regla. De esta manera,  $M$  número de reglas de la forma (2.19) se interpretarían como una sólo relación  $Q_M$  en  $U \times V$  definida por

$$Q_M = \bigcup_{l=1}^M Ru^{(l)} \quad (2.20)$$

La unión  $\cup$  en la anterior combinación llamada Mamdani, se realiza usando alguna norma- $s \oplus$  por lo que la relación difusa se podría reescribir como  $\mu_{Q_M}(x, y) = \mu_{R_u^{(1)}}(x, y) \oplus \dots \oplus \mu_{R_u^{(M)}}(x, y)$ . El resultado de la máquina de inferencia al introducir un conjunto difuso  $A'$  en  $U$  está expresado como  $\mu_{B'}(y) = \sup_{x \in U} t[\mu_{A'}(x), \mu_{Q_M}(x, y)]$ .

En una máquina de inferencia basada en reglas individuales, el mismo conjunto difuso de entrada  $A'$  es procesado por cada una de las  $M$  relaciones difusas que representan cada implicación. De esta manera, en una primera etapa, el resultado del proceso de razonamiento aproximado es:

$$\mu_{B'_l}(y) = \sup_{x \in U} t[\mu_{A'}(x), \mu_{R_u^l}(x, y)] \quad \text{para } l = 1, 2, \dots, M. \quad (2.21)$$

La salida de cada una de las reglas es combinada posteriormente usando una norma- $s \oplus$  obteniendo  $\mu_{B'}(y) = \mu_{B'_1}(y) \oplus \dots \oplus \mu_{B'_M}(y)$ , ó a través de una norma- $t \otimes$  lo que daría  $\mu_{B'}(y) = \mu_{B'_1}(y) \otimes \dots \otimes \mu_{B'_M}(y)$ .

### Desdifusificador

Los desdifusificadores convierten la salida de una máquina de inferencia difusa de un conjunto difuso  $\mu_{B'}(y)$  en  $V \subset R$  a un valor nítido  $y^*$  que represente la distribución de valores y la forma de tal conjunto difuso expresado generalmente como funciones de membresía combinadas. Existen varios métodos de mapear la salida difusa a un sólo valor, como lo es por medio del desdifusificador centro de gravedad o el desdifusificador centro promedio, siendo el más utilizado este último por su simplicidad computacional, además de obtener valores cercanos al primer método más exacto pero computacionalmente caro [7].

El desdifusificador centro de gravedad realiza una integración de los valores de salida de la máquina de inferencia y sus valores de membresía obteniéndose con esto una área ponderada que al ser dividida por el área de tal conjunto difuso  $\mu_{B'}(y)$  da como resultado un valor nítido  $y^*$ . En el caso discreto, el centro de gravedad  $y^*$  de la salida se expresa como :

$$y^* = \frac{\sum_{i=1}^l y_i \cdot \mu_{B'}(y_i)}{\sum_{i=1}^l \mu_{B'}(y_i)} \quad (2.22)$$

En el caso continuo esto resulta en

$$y^* = \frac{\int_V y \cdot \mu_{B^l}(y) dy}{\int_V \mu_{B^l}(y)} \quad (2.23)$$

En un desfusificador centro promedio se toman cada uno de los componentes o subconjuntos difusos  $B^l$  de la función de membresía combinada  $\mu_{B^l}(y)$ , y se obtiene la sumatoria de la multiplicación de la altura  $w$  del subconjunto difuso  $l$  por su centro  $\bar{y}^l$  para luego dividir tal cantidad por la sumatoria de las alturas  $w$  de los  $l$  subconjuntos difusos para obtener así un valor nítido  $y^*$ ; esto se expresa también como:

$$y^* = \frac{\sum_{l=1}^M \bar{y}^l \cdot w_l}{\sum_{l=1}^M w_l} \quad (2.24)$$

donde  $\bar{y}^l$  es el centro del subconjunto difuso  $l$  y  $w_l$  el grado de membresía de tal punto.

## Difusificar

En su forma general, una máquina de inferencia tiene como entrada un conjunto difuso  $A'$ , es por esto que si en el sistema a modelar se manejan valores nítidos, entonces debe haber un mapeo entre tal valor físico  $x^*$  y el conjunto difuso  $A'$ ; tal mapeo se conoce como difusificación.

Un difusificador conocido por simplificar los cálculos en un proceso de inferencia es el difusificador *singleton*, el cual mapea un punto nítido  $x^*$  a un *singleton* difuso  $A'$  en  $U$ , el cual tiene valor de membresía 1 en  $x^*$  y 0 en cualquier otro punto, es decir:

$$\mu_{A'}(x) = \begin{cases} 1 & \text{si } x = x^* \\ 0 & \text{de otra manera} \end{cases} \quad (2.25)$$

El difusificador gaussiano mapea  $x^* \in U$  a un conjunto difuso en  $A'$ , el cual está expresado por la siguiente función de membresía gaussiana:

$$\mu_{A'}(x) = \exp\left(-\frac{x_1 - x_1^*}{a_1}\right)^2 \otimes \dots \otimes \exp\left(-\frac{x_n - x_n^*}{a_n}\right)^2 \quad (2.26)$$

donde  $a_i$  son parámetros positivos y la norma-t  $\otimes$  es alguno de los operadores (2.10) ó (2.11) definidos.

### 2.3.3 Tipos de Sistemas Difusos

Hay tres tipos de sistemas difusos que son comúnmente usados en la literatura [7]:

1. Sistemas difusos puros.
2. Sistemas difusos con difusificador y con desdifusificador.
3. Sistemas difusos Takagi-Sugeno-Kang (TSK).

Los sistemas difusos puros conforman la forma básica de los sistemas difusos consistentes en una base de reglas difusas, una máquina de inferencia difusa y conjuntos difusos de entrada y salida. En este tipo de sistemas difusos, la máquina de inferencia toma como entrada un valor o palabra de una variable lingüística para a partir del razonamiento aproximado realizado por la máquina de inferencia obtener un conjunto difuso de salida.

Los sistemas difusos del segundo tipo le agregan un difusificador y desdifusificador al sistema puro. El difusificador transforma un valor numérico nítido de entrada a un conjunto difuso que es combinado con alguna(s) regla(s) difusa(s) para obtener, por medio de una máquina de inferencia, un conjunto de salida(s) difusa(s). El desdifusificador convierte esa salida(s) a un valor nítido. En este tipo de sistemas difusos se pueden tener diferentes combinaciones de difusificador, máquina de inferencia y desdifusificador, lo que da como resultado clases de sistemas difusos.

Hay clases de sistemas difusos que se pueden representar en forma de fórmulas compactas, como las siguientes obtenidas con el desdifusificador centro promedio:

Se define un sistema difuso con una máquina de inferencia mínimo como aquel basado en reglas individuales (2.21) combinadas usando el operador máximo (2.9), composición max-min (2.16) e implicación mínimo de Mamdani (2.14) de reglas del tipo (2.19), en donde los conjuntos difusos del consecuente  $B^l$  son de forma triangular con centro  $\bar{y}^l$ , entonces tal sistema con difusificador *singleton* y desdifusificador centro promedio se describe con la siguiente formula

$$y^* = f(\mathbf{x}) = \frac{\sum_{l=1}^M \bar{y}^l \cdot \left( \min_{i=1}^n \mu_{A_i}^l(x_i^*) \right)}{\sum_{l=1}^M \left( \min_{i=1}^n \mu_{A_i}^l(x_i^*) \right)} \quad (2.27)$$

donde  $x^* \in U \subset R^n$  es la entrada considerando el difusificador *singleton*, y  $y^* \in V \subset R$  es la salida nítida del sistema.

De manera análoga, se define un sistema difuso con una máquina de inferencia producto como aquel basado en reglas individuales (2.21) combinadas usando el operador máximo (2.9), composición max-prod (2.17) e implicación producto de Mamdani (2.15) de reglas del tipo (2.19), en donde los conjuntos difusos del consecuente  $B^l$  son de forma triangular con centro  $\bar{y}^l$ , entonces tal sistema con difusificador *singleton* y desdifusificador centro promedio se describe con la fórmula

$$y^* = f(\mathbf{x}) = \frac{\sum_{l=1}^M \bar{y}^l \cdot \left( \prod_{i=1}^n \mu_{A_i^l}(x_i^*) \right)}{\sum_{l=1}^M \left( \prod_{i=1}^n \mu_{A_i^l}(x_i^*) \right)} \quad (2.28)$$

Dentro de los sistemas difusos con difusificador y desdifusificador como los presentados anteriormente, algunas de las configuraciones pueden aproximar en algún valor arbitrario  $\varepsilon > 0$  cualquier función continua real  $g(x)$  en  $U$ , esto es

$$\sup_{x \in U} |f(x) - g(x)| < \varepsilon \quad (2.29)$$

Tales sistemas deben cumplir el teorema Stone-Weierstrass, el cual declara que para un dominio  $D$  en un espacio compacto  $U$  de  $\aleph$  dimensiones se podría tener un grupo  $\mathfrak{S}$  de funciones continuas en  $D$  que satisfagan los siguientes criterios:

1. Función identidad: la constante  $f(x) = 1$  está en  $\mathfrak{S}$
2. Separabilidad: Para cualquiera dos puntos  $x_1 \neq x_2$  en  $D$ , hay un  $f$  en  $\mathfrak{S}$  tal que  $f(x_1) \neq f(x_2)$ .
3. Cerradura algebraica: Si  $f$  y  $g$  son dos funciones en  $\mathfrak{S}$  entonces  $fg$  y  $af + bg$  están en  $\mathfrak{S}$  para cualquiera dos números reales  $a$  y  $b$ .

De esta manera, para cualquier función continua real  $g(x)$  en  $U$  y un valor arbitrario  $\varepsilon > 0$ , existe una función  $f \in \mathfrak{S}$  tal que  $|f(x) - g(x)| < \varepsilon$  para toda  $x \in D$ .

Los sistemas difusos con funciones de membresía en forma de campana que al ser multiplicadas por si mismas conserven la forma de campana, son candidatos a formar funciones en  $\mathfrak{S}$  y más en específico, tal y como destaca Wang [6], los sistemas como (2.27) y (2.28) que utilicen la función de membresía gaussiana (2.4) para describir los conjuntos difusos de las premisas, cumplen con los criterios del teorema de Stone-Weierstrass por lo que son aproximadores universales.

Los sistemas difusos tipo Takagi-Sugeno-Kang (TSK) son semejantes a la máquinas de inferencia de los sistemas difusos con difusificador y desdifusificador pero tienen como entrada exclusivamente valores nítidos a través de un difusificador *singleton* y el consecuente de las reglas está conformado por *singletons* o por funciones lineales usualmente en términos de los valores numéricos de entrada [2].

Se dice que un sistema difuso TSK es de orden cero cuando el consecuente de cada regla de la forma (2.19) es una constante. Este tipo de sistema TSK tiene las mismas características que los sistemas difusos con difusificador y desdifusificador considerando exclusivamente un difusificador *singleton* y reglas cuya función de membresía del consecuente  $B^l$  es un triángulo cuyo centro  $\bar{y}^l$  es igual al valor constante de la correspondiente regla equivalente del tipo TSK de orden cero,  $f_l = \bar{y}^l$  para  $l = 1 \dots M$ . De esta manera, los sistemas descritos por las fórmulas (2.27) y (2.28) pueden ser interpretados como sistemas difusos TSK de orden cero con máquinas de inferencia mínimo y producto respectivamente.

Al agregarle al consecuente de las reglas de un sistema difuso TSK de orden cero una función lineal en términos de entradas singleton, se obtiene un sistema TSK de primer orden; la forma de las reglas de este sistema sería entonces

$$R_u^{(l)} : \text{Si } x_1^l \text{ es } A_1^l \text{ Y } \dots \text{ Y } x_n^l \text{ es } A_n^l, \text{Entonces } f_l = B_1^l * x_1^l + \dots + B_n^l * x_n^l + B_{n+1}^l \quad (2.30)$$

donde  $A_i^l$  son conjuntos difusos en el antecedente mientras que  $\mathbf{B}^l = [ B_1^l \dots B_n^l B_{n+1}^l ]$  es un vector de parámetros a definir.

El consecuente de cada regla  $R_u^{(l)}$  del tipo (2.30) se puede considerar como una función  $f_l$  en términos de *singletons* paramétricos  $B_i^l$  de la regla y de la entrada *singleton* a la máquina de inferencia que ocupa esa regla  $l$ , es decir,  $f_l = B_1^l * x_1^l + \dots + B_n^l * x_n^l + B_{n+1}^l$ ; por lo tanto, el proceso de inferencia de los sistemas difusos TSK está basado en reglas individuales a las que se obtiene un peso ponderado  $w_l$  que se combina con  $f_l$  a través de la norma-t producto. Para un sistema con dos variables lingüísticas en el antecedente de la regla, el sistema difuso TSK de primer orden con normas-t producto se resume en la figura 2.5.

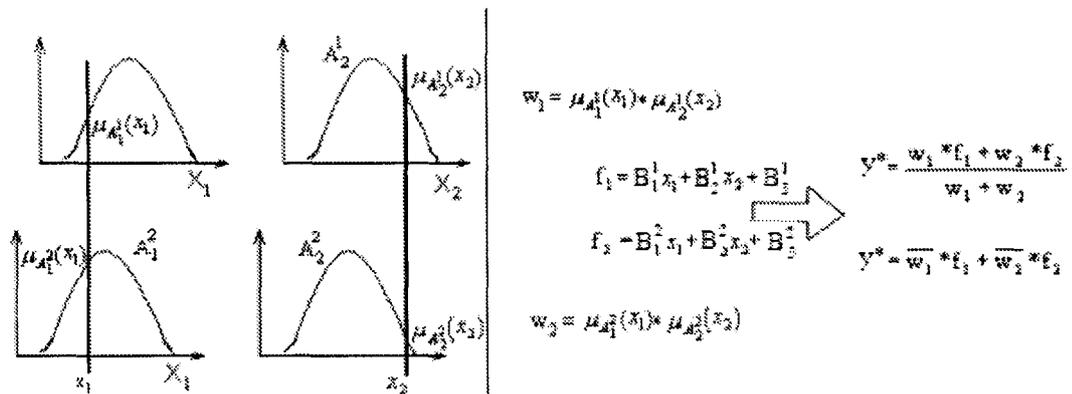


Figura 2.5: Ejemplo de un sistema difuso del tipo TKS de primer orden.

## Bibliografía

- [1] D. Driankov. *An Introduction to Fuzzy Control Second Edition*. Springer, Berlin, 1993.
- [2] C.-T. Sun J.-S. R. Jang and E. Mizutani. *Neuro-Fuzzy and Soft Computing: a computational approach to learning and machine intelligence*. Prentice Hall PTR, New Jersey, 1997.
- [3] J. S. R. Jang. Anfis: Adaptive-network-based fuzzy inference system. *IEEE Trans on Systems, Man, and Cybernetics*, pages 665–685, 1993.
- [4] G. J. Klir and Bo Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall PTR, New Jersey, 1995.
- [5] C. X. Jiang M. Y. Hu, G. Zhang and B. E. Patuwo. A cross-validation analysis of neural network out-of-sample-performance in exchange rate forecasting. *Decision Sciences*, 1:197–216, 1999.
- [6] L-X Wang. Fuzzy systems are universal approximators. *Proc. IEEE Int. Conf. Fuzzy Systems*, 1992.
- [7] Li-Xin Wang. *A Course in Fuzzy Systems and Control*. Prentice Hall PTR, New Jersey, 1997.
- [8] P. Werbos. *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. Ph.D. dissertation, Harvard University, Cambridge, MA, 1974.
- [9] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

Centro de Información-Biblioteca



30002006165161