

PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito parcial para
obtener el Título de INGENIERO DE SISTEMAS

PRONÓSTICO DEL DEFÍCIT DE VIVIENDAS EN EL ESTADO MÉRIDA A TRAVÉS DE REDES NEURONALES ARTIFICIALES

Por

Br. Annjolie A. Gil R

Tutor: Gerardo Colmenares

Cotutora: Josefa Ramoni

Abril 2008



©2008 Universidad de Los Andes Mérida, Venezuela

*A MIS PADRES Y HERMANOS,
POR SER PILAR FUNDAMENTAL EN MI VIDA.
Y A TI MI ADRIANA POR SER FUENTE DE MIS ALEGRÍAS.*

Agradecimientos

Al ver materializada esta meta, logro reafirmar que en compañía de Dios Todopoderoso no hay prueba que no pueda ser superada. Quiero expresar mis más profundos agradecimientos a todas las personas que de una manera u otra formaron parte de esta tarea.

A mis Padres y hermanos por su apoyo incondicional y por ser fuente de sabiduría. Especialmente les agradezco al Profesor Gerardo Colmenares y a la Profesora Josefa Ramoni, por todos esos conocimientos que me permitieron adquirir con ustedes, por toda su paciencia y por haberme hecho sentir parte de este equipo.

Al Instituto de Investigaciones Económicas y Sociales, no sólo por representar una institución para el desarrollo de proyectos y del mío en particular, sino además por haberme brindado en sus espacios la calidez de todo el equipo que allí labora. Gracias a todos.

A Lisandro por su apoyo incomparable.

A todos mis amigos y amigas por siempre estar allí y compartir los momentos gratos y menos gratos de esta experiencia.

**PRONÓSTICO DEL DÉFICIT DE VIVIENDAS EN EL ESTADO MÉRIDA A TRAVÉS DE REDES
NEURONALES ARTIFICIALES**

Br. Annjolie A. Gil Ruiz

Proyecto de Grado — Investigación de Operaciones, 73 páginas.

RESUMEN.

Esta investigación estuvo dirigida a construir un modelo que permitiese realizar el Pronóstico del Déficit de Viviendas en el estado Mérida, mediante la utilización de Redes Neuronales Artificiales con Funciones de Bases Radiales (RBF). Para ello, fue necesario aplicar técnicas estadísticas de Análisis Multivariante, las cuales permitieron construir los indicadores para evaluar el carácter deficitario de las viviendas. Tomando los datos de las Encuestas de Hogares por Muestreo (EHM) correspondientes al período 1994-2005, se construyó un registro de datos que proporcionó los indicadores del Número de Hogares, Tenencia, Hacinamiento, Adecuación y Condición de la vivienda; para así proceder a la construcción de los modelos y, con esto, lograr el objetivo planteado. Los resultados obtenidos fueron utilizados para analizar los índices deficitarios existentes en el estado Mérida y las posibles fluctuaciones encontradas en el período antes mencionado. Entre las conclusiones se destaca que los modelos de redes, usando RBF para el pronóstico demostraron un alto nivel de efectividad y una alta capacidad de adaptación, adecuándose al tipo de problema que se modeló. En general, los resultados obtenidos en el entrenamiento y generalización alcanzaron errores cuadráticos medios muy bajos, demostrándose un alto nivel de acierto para el pronóstico y robustez de los modelos.

Palabras claves: Déficit Cualitativo, Déficit Cuantitativo, Redes Neuronales Artificiales, Indicadores.

Índice general

	Pág.
Dedicatoria	
Agradecimientos	
Resumen.....	i
Índice General.....	ii
Lista de Figuras.....	v
Lista de Tablas.....	vi
Lista de Gráficos.....	vii
Introducción.....	1
CAPITULO I DÉFICIT DE VIVIENDAS.....	4
1.1.- Planteamiento del Problema.....	4
1.2.- Objetivos.....	7
1.2.1.- Objetivo General.....	7
1.2.2.- Objetivos Específicos.....	7
1.3.- Importancia de la Investigación.....	8
1.4.- Alcances y Limitaciones.....	8
CAPITULO II MARCO REFERENCIAL.....	10
2.1.- Consideraciones Generales.....	10
2.2.- Definición del Déficit de Viviendas.....	12
2.2.1.- Déficit Cuantitativo.....	12
2.2.2.- Déficit Cualitativo.....	13
2.3.-Definición de las Encuestas de Hogares por Muestreo (EHM).....	13
2.3.1.-Consideraciones aplicadas en la Encuesta.....	15
2.3.2.-Criterios de selección de la muestra.....	15
2.3.3.- Bondades de la Encuesta.....	16
2.3.4.-Variables Investigadas.....	16

	Pág.
2.4.- Técnicas de Preprocesamiento de Datos.....	17
2.4.1.- Análisis de Correspondencia.....	17
2.4.2.- Análisis de Conglomerados.....	17
2.5.- Selección de las Muestras para entrenamiento.....	18
2.6.- Aspectos generales de Redes Neuronales Artificiales.....	18
2.6.1.- Neurona Biológica.....	19
2.6.2.- Definición de RNA.....	21
2.6.3.-Topología de las RNA.....	21
2.6.4.- Funciones de Activación	24
2.6.5.- Entrenamiento de las RNA.....	24
2.7.- Funciones de Base Radial.....	25
2.7.1.- Arquitectura de la RBF.....	26
2.7.2.-Algoritmo de Bases Radiales(RBF).....	27
CAPITULO III ANALISIS Y EXPLORACIÓN DE LOS REGISTROS	
VIVIENDAS Y HOGARES DE LAS EHM.....	29
3.1- Disposición de la Encuesta.....	29
3.2.- Pasos para obtener el Registro Vivienda_Hogar94-05.....	30
3.3.- Selección de las Observaciones.....	34
3.4.- Selección de las variables.....	36
3.4.1.- Variables Originales.....	38
3.4.2.- Variables Transformadas.....	38
3.4.3.- Construcción de las variables del modelo.....	39
3.4.4.- Variables de salida.....	45
3.5.- Conformación de los bloques de datos 1994-2001 y 2001-2005.	47
CAPITULO IV EXPERIMENTOS Y RESULTADOS.....	48
4.1.- Modelo de la red con bases radiales.....	48

	Pág.
4.1.1.- Red neuronal RBF: pronóstico/clasificación del déficit de vivienda.....	49
4.1.2.- Parámetros empleados para los modelos RBF del período 94_2-01_1 y 01_2-05_2.....	50
4.2.- Entrenamiento de las RBF.....	54
4.3.- Generalización de la Red.....	58
4.4.- Inferencias sobre el déficit de viviendas para cada período estudiado.....	63
CAPITULO V CONCLUSIONES Y RECOMENDACIONES.....	69
5.1.- Conclusiones.....	70
5.2.-Recomendaciones.....	70
REFERENCIAS BIBLIOGRÁFICAS.....	71
ANEXOS	

Lista de figuras

	Pág.
Figura 1. Neurona biológica.....	20
Figura 2. Analogía de neurona artificial y biológica.....	21
Figura 3. Estructura de las RNA.....	22
Figura 4. Arquitectura de RBF.....	27
Figura 5. Arquitectura del aproximador lineal con extractor de característica	28
Figura 6. Agrupación de la información por estado en cada registro.....	31
Figura 7. Extracción de los datos del estado Mérida en cada registro.....	32
Figura 8. Fusión de los registros vivienda y hogar.....	34
Figura 9. Registro vivienda _ hogar 1994-2005.....	35
Figura 10. RBF construida.....	52
Figura 11. Ambiente del NeuroSolutions de la RBF construida.....	54

Lista de tablas

	Pág.
Tabla 1. Funciones de activación más comunes.....	25
Tabla 2. Descripción de las variables de identificación en la EHM.....	30
Tabla 3. Variables para la fusión del registro Viviendas-Hogar. Vivienda.....	32
Tabla 4. Variables para la fusión del registro Viviendas-Hogar. Hogar.....	33
Tabla 5. Selección de las observaciones.....	35
Tabla 6. Descripción de las variables observadas en el ACM.....	41
Tabla 7. Valores iniciales de los conglomerados. Condición de vivienda.....	42
Tabla 8. Valores finales de los conglomerados. Condición de vivienda.....	42
Tabla 9. Valores iniciales de los conglomerados. Adecuación de la vivienda..	44
Tabla 10. Valores finales de los conglomerados. Adecuación de la vivienda..	45
Tabla 11. Descripción de la variable de salida.....	47
Tabla 12. Ejemplo de los patrones de entrenamiento.....	51
Tabla 13. Configuración de la capa de salida.....	53
Tabla 14. Configuración del aprendizaje no supervisado ambos períodos.....	53
Tabla 15. Parámetros del entrenamiento supervisado.....	54
Tabla 16. Configuraciones de RBF para el entrenamiento del período 94_2-01_1.....	55
Tabla 17. Resultados del entrenamiento para el período 94_2-01_1.....	55
Tabla 18. Configuraciones de RBF para el entrenamiento del período 01_2-05_2.....	56
Tabla 19. Resultados del entrenamiento para el período 01_2-05_2.....	56
Tabla 20. Generalización de la mejor red usando patrones no conocidos del período 94_2-01_1.....	59
Tabla 21. Generalización de la mejor red usando patrones no conocidos por del período 01_2-05_2.....	61

	Pág.
Gráfico 1. ACM para la condición de la vivienda.....	40
Gráfico 2. ACM para la definición de la salida.....	46
Gráfico 3. Salida observada-Salida deseada. Ambiente NeuroSolutions (1)...	57
Gráfico 4. MSE alcanzado por la RBF. Ambiente NeuroSolutions (1).....	57
Gráfico 5. Salida observada-Salida deseada. Ambiente NeuroSolutions (2)...	58
Gráfico 6. MSE alcanzado por la RBF. Ambiente NeuroSolutions (2).....	58
Gráfico 7. Resultados de la variabilidad en el desempeño de la red para diferentes tamaños de muestra del período 94_2-01_1.....	60
Gráfico 8. Variabilidad en la generalización del modelo para el período 94_2-01_1.....	60
Gráfico 9. Resultados de la variabilidad en el desempeño de la red para diferentes tamaños de muestra del período 01_2-05_2.....	62
Gráfico 10. Variabilidad en la generalización del modelo para el período 01_2-05_2.....	62
Gráfico 11. Condición deficitaria en el estado Mérida período 94_2-01_1.....	64
Gráfico 12. Incidencia de cada indicador en las viviendas deficitarias período 94_2-01_1.....	65
Gráfico 13. Condición deficitaria en el estado Mérida período 01_2-05_2.....	66
Gráfico 14. Incidencia de cada indicador en las viviendas deficitarias período 01_2-05_2.....	67
Gráfico 15. Resumen de las variaciones de los indicadores de los períodos bajo estudio.....	67

Introducción

El tema del déficit de vivienda es uno de los que más afecta nuestra sociedad hoy en día. No se trata solamente de indagar acerca de aquellas familias que no cuentan con una vivienda, sino también evaluar las condiciones en las cuales se encuentran las viviendas existentes. En la actualidad se habla de un proceso acelerado en materia habitacional para cubrir esta deficiencia, pero es necesario evaluar hasta que punto estos proyectos realmente han logrado cubrir el nivel deficitario existente, ya que para muchos las condiciones que ofrecen estas nuevas viviendas no son las más idóneas.

En este trabajo no se profundiza en este aspecto pero si da una idea del enfoque que tiene el déficit de vivienda. Una vivienda no es sólo un techo donde resguardarse, sino también debe cumplir con un conjunto de requerimientos que realmente les brinde a los individuos que allí habitan el bienestar social y la salud física y mental necesarias para mejorar su calidad de vida.

Esta investigación se enmarca en un análisis del déficit de viviendas en el estado Mérida. Para realizar esta investigación se contó con las EHM realizadas por el Instituto Nacional de Estadística (INE), el análisis se circunscribe a la serie de datos desde el segundo semestre de 1994 hasta el segundo semestre de 2005.

Existen diversos aspectos que son considerados para sustentar la condición de adecuación de la vivienda. En esta investigación se toma en cuenta el basamento jurídico establecido en la Ley de Régimen Prestacional de Vivienda y Hábitat (Gaceta oficial N° 38.192 del 9 de Mayo de 2005) en su Capítulo III. De los objetivos de la Ley. Art. 13 *“Vivienda y Hábitat Dignos”* y de la Constitución de la República Bolivariana de Venezuela en su Artículo 82; así como en el Contenido de Derecho discutido por la Comisión de los Derechos Humanos de las Naciones Unidas en la Observación N° 4, en las que se explica las condiciones en las cuales deben vivir los individuos, la habitabilidad de las viviendas para ser aceptadas por ellos.

Tomando los criterios considerados en los apartados antes mencionados, el incumplimiento de algunas de las condiciones que en ellos se explican permiten realizar la evaluación de la condición deficitaria para las observaciones que se disponen de las EHM. Es necesario aclarar que se evalúan para el estudio sólo aquellas características que son posibles de acuerdo a la información que aportan las encuestas. En este sentido, cabe destacar que se evalúan cinco aspectos que describen las características fundamentales de cada vivienda: El número de hogares que habitan la vivienda, la condición de hacinamiento, la seguridad jurídica de la tenencia, la adecuación y condición de la vivienda.

Dado el carácter de la investigación y en razón de su naturaleza, el presente informe se estructura en los siguientes capítulos.

El primero contiene el planteamiento y formulación del problema, los objetivos de la investigación, importancia de la investigación, y el alcance y las limitaciones que se encontraron en la investigación. El segundo capítulo plantea todos los aspectos teóricos, que enmarcaron la investigación realizada y que le dan fundamento al proyecto que aquí se describe. El tercero está relacionado con las técnicas de preprocesamiento aplicadas para la construcción del conjunto de datos. Tal como están dispuestas las EHM no se obtiene mayor información, y en consecuencia, las técnicas aquí aplicadas enriquecen el estudio dándole sentido a cada una de las variables seleccionadas para el desarrollo de la investigación. Es explicado cómo se

da respuesta a las características ofrecidas por las variables seleccionadas y así, poder evaluar la condición deficitaria o no de las observaciones que son evaluadas en el plazo 1994-2005. En el cuarto capítulo se discute todo lo relacionado con la construcción y prueba del modelo para el pronóstico/clasificación. Se toma en cuenta la Red Neuronal como instrumento de clasificación a través del modelo de red de funciones de base radial en sus nodos ocultos, de acuerdo a los resultados de entrenamiento que se obtengan, se escoge aquel modelo que mejor generalice. Con estos resultados se obtienen los niveles de incidencia de cada indicador que permiten mostrar si existe alguna variación del déficit período a período.

Por último, en el quinto capítulo se señalan las conclusiones a las que se llega con el estudio, las respectivas recomendaciones para dar curso a nuevos proyectos.

Capítulo I

Déficit de viviendas

En este capítulo se desarrollan los aspectos fundamentales de este proyecto como son Planteamiento del Problema; los objetivos de la Investigación, tanto el General como los Específicos; Importancia de la Investigación y los Alcances y Limitaciones encontrados durante el desarrollo del Proyecto.

1.1 Planteamiento del problema

En muchos reportes sobre la situación deficitaria de las viviendas se muestran sólo cifras, pero en ninguna de ellas se refleja información detallada de las causas de tal déficit. Para analizar los problemas de vivienda es necesario identificar cuáles son las causas que provocan el crecimiento del déficit, ya que no sólo se trata que haya hogares que no poseen vivienda, sino además que aquellas existentes pueden requerir de algún tipo de reparación o mejora.

Los entes encargados han deseado crear políticas que permitan mitigar la diferencia entre el número de viviendas existentes y las que se demandan, así como, mejorar las condiciones de aquellas existentes. Sin embargo los estudios que se han realizado, no han concluido con resultados que permitan vincularlos con la realidad, razón por la cual las políticas que se han aplicado para aliviar este fenómeno no han sido efectivas. Los estudios realizados anteriormente relacionados con Déficit de

Viviendas muestran sólo valores generales que no permiten identificar qué déficit está asociado a una vivienda en particular, emplean un criterio distinto al que sigue esta investigación. Es por ello que este estudio adquiere mayor relevancia ya que no sólo se identifica la vivienda como ente físico aislado, sino también la manera en que inciden las características que posee y cuál de las condiciones de habitabilidad requeridas están ausentes.

Asimismo, una vivienda no es deficitaria sólo por cuanto albergue a un mayor número de hogares de los que su capacidad tiene, sino también por cuanto esta vivienda no cuente con cierta adecuación que la califique como un lugar digno para habitar. En tal sentido, del derecho a una "vivienda adecuada" definida en el informe sobre Estrategia Mundial de Vivienda hasta el año 2000¹, se consideró con disponer de un lugar adecuado donde poder aislarse si se desea, en espacio, en seguridad, en iluminación y ventilación, con una infraestructura básica con acceso aceptable a su trabajo y los servicios básicos, todo ello, a un costo razonable. Esta visión básica de la adecuación, fue esencialmente reafirmada y ampliada en la observación general N° 4 del Comité de Derechos Económicos Sociales y Culturales (Comité de DESC)².

Así mismo, en la segunda Conferencia de Naciones Unidas sobre los Asentamientos Humanos (Hábitat II), celebrada en 1996 en Turquía, fue aprobada la declaración de Estambul sobre Asentamientos Humanos, por los gobiernos y delegaciones de los países allí reunidos, que sobre vivienda adecuada estipula entre otros puntos:

“Como el ser humano es el aspecto más importante de nuestras preocupaciones respecto del desarrollo sostenible, es también la base de nuestra acción para dar efecto al programa Hábitat. Reconocemos que las mujeres, los niños y los jóvenes tienen una

¹ NACIONES UNIDAS. COMISIÓN DE DERECHOS HUMANOS: *El derecho a una vivienda adecuada: informe sobre la marcha de los trabajos presentado por el Sr. Rajindar Sachar, Relator Especial*. Documento: E/CN.4/Sub.2/1993/15. 22.06.93. Párrafo 44. Sobre la Estrategia Mundial de la Vivienda hasta el año 2000, aceptada por todos los gobiernos, ver: Documentos Oficiales de la Asamblea General, cuadragésimo tercer período de sesiones, Suplemento No. 8, adición (A/43/8/Add.1).

² NACIONES UNIDAS. COMITÉ DE DERECHOS ECONÓMICOS SOCIALES Y CULTURALES: *Observación general N° 4. El derecho a una vivienda adecuada (párrafo 1 del artículo 11 del Pacto)*. 1991. Documento: E/1991/23.

necesidad especial de vivir en condiciones seguras, salubres y estables [...]. Reafirmamos nuestra voluntad de lograr progresivamente el pleno ejercicio del derecho a una vivienda adecuada, como se ha previsto en los instrumentos de derecho internacional. A tal fin, solicitaremos la activa participación de nuestros copartícipes de los sectores público y privado y de las organizaciones no gubernamentales, a todos los niveles, para brindar a todas las personas y a sus familias garantías jurídicas con respecto a la tenencia, la protección frente a la discriminación y a la igualdad de acceso a una vivienda asequible y adecuada. Ampliaremos la oferta de vivienda asequible, para lo cual velaremos por que los mercados funcionen con eficiencia y de manera social y ambientalmente racional, porque se mejore el acceso a la tierra y al crédito y por que se ayude a los que estén excluidos del mercado de la vivienda”.

En tal sentido en Venezuela a través de la constitución de la República Bolivariana de Venezuela, en su Artículo 82 contempla:

"Toda persona tiene derecho a una vivienda adecuada, segura, cómoda, higiénica, con servicios básicos esenciales que incluyan un hábitat que humanice las relaciones familiares, vecinales y comunitarias. La satisfacción progresiva de este derecho es obligación compartida entre los ciudadanos y ciudadanas y el Estado en todos sus ámbitos. El Estado dará prioridad a las familias y garantizará los medios para que éstas y especialmente las de escasos recursos, puedan acceder a las políticas sociales y al crédito para la construcción, adquisición o ampliación de viviendas."

Con la información expuesta en los párrafos anteriores, se puede comprender en qué dirección se desarrolla está investigación. Ya que se exponen claramente cuáles son los aspectos que se involucran el analizar el déficit de viviendas, para este estudio en particular. En este sentido, se toma la información necesaria de las EHM; las cuales desde el año 1967 atendieron inicialmente con mayor énfasis la fuerza

laboral y sus incidencias. En sus nuevas versiones se han incluido campos complementarios que proporciona información relacionada con la vivienda y hogares, lo cual facilita el análisis del tema específico de Déficit de Vivienda. Las EHM se aplican a la mayoría de las entidades de nuestro país, lo que ocasiona un número muy grande de datos que pueden ser empleados para realizar cualquier estudio. En consecuencia, la investigación se centra en el análisis deficitario de las viviendas específicamente para la ciudad de Mérida. Se realiza la estimación del déficit de vivienda a partir del déficit cuantitativo, que mide la cantidad de hogares que viven en la misma vivienda. Y complementariamente, se estima el déficit cualitativo que se determina conjugando información referida a la tenencia; el hacinamiento existente en la vivienda, cuyo número de individuos que duermen en cada dormitorio es mayor a dos; la adecuación de la vivienda enmarcada en la disponibilidad de los servicios básicos; y por último las condiciones de la vivienda de acuerdo a las características estructurales físicas (paredes, pisos, techo etc.). De ahí, el grado de importancia que tiene la investigación realizada, por cuanto se podrá calificar cuáles son los aspectos que determinan el Déficit de Vivienda y el caso en particular del estado Mérida.

Este proyecto se realiza aprovechando las bondades computacionales para desarrollar minería de datos desde el área de estadística multivariante y las innovaciones que ofrece el área de inteligencia artificial, específicamente las redes neuronales artificiales con funciones de bases radiales.

1.2 Objetivos

1.2.1 Objetivo general

Diseñar un modelo para el pronóstico/clasificación del Déficit de Vivienda.

1.2.2 Objetivos específicos

- Realizar una revisión sobre el uso de las Redes Neuronales Artificiales en aplicaciones de pronóstico/clasificación.
- Realizar una revisión sobre el uso de las técnicas de preprocesamiento de datos.

- Usar técnicas de preprocesamiento de datos para seleccionar las variables que se van a incluir en el modelo de la Red Neuronal Artificial.
- Desarrollar un Modelo de Redes Neuronales con Bases Radiales (RBF) que permita realizar el pronóstico/clasificación.
- Realizar la estimación respectiva del déficit de viviendas, sujeta a los resultados obtenidos por el modelo.
- Obtener la incidencia de los indicadores involucrados en la condición deficitaria de las viviendas.

1.3 Importancia de la investigación

1. No sólo indagar sobre la condición en que se encuentra este indicador social en el estado, sino sentar un precedente en la manera en que se necesita identificar quienes recaen en este problema para ofrecer soluciones efectivas.
2. Demostrar que las técnicas de Inteligencia Artificial puede resultar una herramienta muy productiva para realizar estudios de esta naturaleza donde el componente social es el principal objetivo.
3. Hacer notar que es imprescindible en cualquier estudio una fase de preprocesamiento, para descubrir el nuevo conocimiento que puede aportar cada variable mediante la construcción de variables latentes o nunca vistas.
4. Agregar nuevos conocimientos al área bajo estudio, de modo tal que los resultados obtenidos en este proyecto ayuden al desarrollo de investigaciones enfocadas en la obtención de nuevos métodos que permitan clasificar y estimar el déficit de viviendas.

1.4 Alcances y Limitaciones

El modelo de pronóstico/clasificación del Déficit de Vivienda en el estado Mérida también puede ser utilizado para estimar el déficit en los demás estados, mediante el uso de las EHM o cualquier Encuesta que ofrezca información referente a Viviendas y Hogares.

Una de las mayores limitantes que se encontró durante el estudio fue la necesidad inminente de adquirir nuevos conocimientos relacionados al área de Análisis Multivariante. Además de ser necesario recolectar mucha información en relación al estudio de déficit, se tenía una idea muy limitada al respecto y para comprenderlo un poco más fue necesario la información para que el estudio resultase lo más completo posible. No sólo se trata de contar viviendas sino de evaluar características inherentes a ellas que pueden ser unas más relevantes que otras. No se puede dejar de mencionar que la manera como se encontraba dispuesta la encuesta acarreó otro desafío. No se trataba solamente de ordenar los datos sino también de organizarlos de manera coherente. Sin embargo, este reto fue superado con el desarrollo de un algoritmo que si bien puede ser mejorado, ofrece una amplia posibilidad para manejar esta encuesta. Las EHM recopilan información en un gran número de variables que han cambiado su denominación en el tiempo y en consecuencia dificultan su organización en un conjunto de series de EHM. Por tanto, fue necesario manejar muy bien la información que contiene la encuesta y cada una de las variables, para relacionarlas en cada uno de los años que están bajo estudio.

Capítulo II

Marco referencial

En este capítulo se explican los fundamentos teóricos que fueron aplicados para el desarrollo del presente estudio. En este sentido, se explican las consideraciones a tomar en cuenta para evaluar la condición deficitaria, los aspectos relevantes de las EHM, las técnicas de Análisis Multivariante y finalmente los fundamentos de las redes neuronales artificiales, especialmente para el desarrollo de modelos con Funciones de Base Radial conocidas como RBF.

2.1 Consideraciones generales

Es necesario evidenciar que para medir el Déficit de Viviendas no sólo basta con tener la idea de contar cuántas viviendas hacen falta; es analizar cuáles son las necesidades que requiere sean satisfechas en este sentido. No es simplemente tener la vaga noción que una vivienda es deficitaria por el concepto sesgado y errado de "Falta o Escasez" sino también, es necesario observar la existencia de muchas viviendas en estados deplorables y que simplemente pueden no contar dentro del déficit porque albergan una o más familias. Es indispensable entender

que es lo que mide el Déficit de Vivienda. En este sentido se trata de indagar a gran escala cuáles son los requerimientos que debe satisfacer una vivienda para no ser considerada en condición deficitaria. Al albergar a una familia, la necesidad de vivienda es satisfecha en la medida que esta ofrece mejores y mayores condiciones en el área de salubridad, seguridad, servicios, etc. Adicionalmente, cabe destacar que el conjunto de viviendas que tengan estos requerimientos satisfechos, de algún modo están vinculadas a las redes de servicio público que puede garantizar no estar en condición deficitaria por este aspecto en particular.

Para realizar las estimaciones de la condición deficitaria de la vivienda hace necesario conocer los conceptos que enmarca esta situación, comprender que no se trata únicamente de solucionar este problema creando un plan para construir las viviendas faltantes, sino también planificar la reconstrucción y mejora de las viviendas existentes para ofrecer a las familias mejores condiciones de vida en este sentido.

Esta investigación intenta reconocer a través de su estudio, cuáles son las características o patrones más relevantes en el área de Vivienda en el Estado Mérida, no sólo revisando los resultados que ofrece la aplicación de la metodología para medir el déficit sino también dando una caracterización de la forma de vida de los habitantes. Dicho de otro modo, en una observación que resulte tener un alto nivel de déficit puede reconocerse cuáles son los aspectos que hacen que tenga esa condición deficitaria. Puede ser debido al alto nivel de déficit que puede estimarse cuantitativa o cualitativamente tal como se indicará en las definiciones.

Para llevar a cabo la investigación a través de las EHM, se han observado algunas limitaciones. La más resaltante es que estas encuestas han sufrido modificaciones constantemente al anexarles complementos a las fuentes principales de información fuerza laboral, viviendas y hogar enriqueciendo el estudio específico del Déficit de Vivienda. Esto resulta una limitante ya que se deben hacer búsquedas específicas para obtener datos acordes con la investigación que se realiza.

La vivienda hoy en día no puede contemplarse como una simple estructura física para resguardarse de la intemperie, sino que debe ser un lugar dotado de los requerimientos que contribuyen a alcanzar los niveles adecuados de calidad de vida

y la posibilidad de una vida digna y decorosa. El Comité de expertos de la Organización Mundial de la Salud (OMS) en Higiene de la Vivienda, señala en su primer informe relacionado a la vivienda:

“Es la estructura material que el hombre emplea para cobijarse y sus dependencias, es decir; todos los servicios, instalaciones y dispositivos necesarios o convenientes para el bienestar social y la salud física y mental del individuo y la familia.”

Y así se tiene una descripción más amplia de las consideraciones que deben tomarse en cuenta al momento de evaluar las características de la vivienda.

2.2 Definición de déficit de vivienda

La vivienda juega un decisivo papel en la calidad de vida de las personas y conlleva, en la mayoría de los casos, al acceso a servicios considerados esenciales para alcanzar niveles mínimos de bienestar. El déficit de vivienda se expresa tanto en la carencia absoluta de vivienda, como en la habitación de viviendas de calidad muy deteriorada o que no ofrecen los servicios básicos. Szalachman (1999)

Para determinar hasta qué punto las necesidades de vivienda de la población están satisfechas, se requiere tomar en cuenta dos factores. En primer lugar, se debe estimar en qué medida la cantidad de viviendas existentes es suficiente para albergar a la totalidad de la población. En segundo lugar se requiere que las viviendas cumplan con ciertos estándares mínimos con el objeto de ofrecer una calidad de vida adecuada a sus habitantes.

De acuerdo a lo anterior, se podrían definir dos tipos de déficit conforme a los dos aspectos sobre la situación de déficit de vivienda.

2.2.1 Déficit cuantitativo

Según Szalachman (1999) la definición tradicional de “déficit cuantitativo” de vivienda se basa en la comparación entre el número de hogares y el de viviendas existentes. El momento en el cual la primera de estas cifras supera la segunda es lo que en la mayoría de los textos se designa como déficit cuantitativo. Sin atenuar la importancia de la definición anterior, este trabajo sostiene que éste es tan sólo un primer acercamiento a lo que se desea medir.

2.2.2 Déficit cualitativo

Al igual que en gran parte de estudios relativos al tema, se asocia el déficit cualitativo a las siguientes variables disponibilidad de servicios ó adecuación de la vivienda; condición de la vivienda, indicador de hacinamiento, el cual va a permitir reconocer si la vivienda alberga más individuos de lo que su capacidad le permite (tres personas o más durmiendo en una misma habitación); condición jurídica de la vivienda, es decir la tenencia, que ha sido incluida por primera vez en un estudio de condición deficitaria.

En cuanto a la disponibilidad de servicios, se considera el acceso a servicios públicos tales como agua, luz, aseo y servicio de eliminación de excretas. Se considera que son satisfechas estas necesidades cuando: a) La vivienda recibe el agua por acueducto, e insatisfecho si es por camión cisterna, pila pública, pozo, etc. b) Cuando el servicio de eliminación de excretas es por red de cloacas o pozo séptico e insatisfecho si es por letrina, hoyo o no tiene. c) Si posee servicios de luz y aseo.

Por lo tanto, ya que la información lo permite, en este caso se emplea una definición de disponibilidad de servicios más amplia que la tradicional. La condición de la vivienda, es la variable que resulta de conjugar las características estructurales de la vivienda y asociarla con la percepción que tiene el habitante de la condición de la vivienda definiéndola como buena, mala o regular. Durante unos años en las EHM se mantuvo información acerca de la condición de la vivienda, no obstante esta información fue omitida años más tarde. Para subsanar esta situación se emplean las técnicas de Análisis Multivariante, que permiten construir la nueva variable de la condición de la vivienda relacionando las características estructurales de la misma. En este sentido, se ampliara toda esta información en el Capítulo III.

2.3 Definición de encuesta de hogares por muestreo (EHM)

Las EHM aplicadas en Venezuela desde el año 1967, es una investigación de naturaleza estadística que es realizada con múltiples propósitos, que permite tener información sobre el fenómeno del empleo y desempleo y de las características socioeconómicas de la población. Se realizan encuestas complementarias referidas a

la vivienda, hogares y personas. Señala Seijas (citado en Bolívar, 1997) que a comienzos de la década de los sesenta se expresó lo siguiente: “Es hora de que los organismos públicos y privados que se encargan de la planificación económica y social a nivel nacional y regional, investigadores profesionales, tomen conciencia de la importancia que reviste para un país el establecimiento y mantenimiento de un programa continuo de propósitos múltiples, como el de la Encuesta de Hogares por Muestra”.

Hasta 1967 sólo a través del Censo General de Población y Vivienda se obtenía información de las características de la población activa, migraciones internas, alfabetismo y otros aspectos socio-económicos, pero sólo se producía cada diez años por lo que era imposible realizar un seguimiento de esas variables.

El programa de Encuestas de Hogares por Muestra puesto en funcionamiento en forma permanente a partir del segundo semestre del año 1967, contó con el apoyo incondicional de los gobiernos, garantes del suministro presupuestario que mantiene vivo el proyecto. De igual manera, contaron con el apoyo de recursos humanos de primera calidad formados en universidades e institutos educativos nacionales e internacionales brindando nuevas técnicas y métodos para el mejoramiento continuo del diseño y de la puesta en marcha del proyecto.

Según Seijas (citado en Bolívar, 1997) este programa ha logrado mantenerse por más de treinta años por contar con la flexibilidad para incorporar de manera ágil y sólida los ajustes que se han requerido. En vista de que en una población permanentemente activa existen muchas características que pueden variar de un período a otro, y el programa debe adecuarse a registrar los cambios para que se puedan presentar en la población, a fin de garantizar que la información producida sea fiel reflejo de las realidades socioeconómicas por las que ha atravesado el país. Numerosos han sido los ajustes que han debido experimentar las Encuestas de Hogares por Muestreo durante estos casi cuarenta años. Sin embargo, ha logrado instrumentar esos cambios sin necesidad de interrumpir el programa, convirtiéndose así en el más antiguo de América Latina.

2.3.1 Consideraciones aplicadas en la encuesta

Las EHM tiene como finalidad, proporcionar en forma oportuna y confiable, a los organismos de planificación nacional y regional, información sobre la estructura y evolución de la Fuerza de Trabajo y, en particular, sobre el fenómeno de empleo y el desempleo, como insumo para el análisis que pueda derivar en una toma de decisiones. También a través de encuestas complementarias información sobre la persona, el hogar, y la vivienda, para el análisis de temas específicos a los organismos oficiales y los usuarios que así lo requieran.

2.3.2 Criterios de selección de la muestra

La muestra es una selección aleatoria de viviendas y la entrevista se aplica a todos y cada uno de los hogares pertenecientes a dichas viviendas y dentro de cada hogar, a cada uno de los habitantes habituales. Por lo tanto, los hogares que conforman la muestra son todos aquellos pertenecientes a las viviendas seleccionadas.

En foros realizados por el Instituto Nacional de Estadística, señalan que actualmente, la muestra de cada semestre esta conformada por 47.625 viviendas, considerándose que el tamaño de esta muestra es suficiente para generar resultados para la población total del país.

Sin embargo, no se logra la entrevista para todas las viviendas, por alguno de los siguientes motivos: Viviendas desocupadas, en ruinas, de uso vacacional, utilizadas como almacén o negocio, o porque en aquellas que están ocupadas no está la persona que puede dar información o en caso contrario se niegan a darla.

Las EHM se realiza con periodicidad semestral, pero cada hogar será entrevistado mientras este seleccionada en la muestra la vivienda a la cual él pertenece. Hasta el año 2001 en su primer semestre las viviendas permanecían en la muestra durante seis semestres consecutivos a esto se le conoce como "Rotación de la Muestra", el cual consistía en que cada semestre salía 1/6 de las viviendas de la muestra, que fueron entrevistadas durante seis semestres seguidos; esta proporción es

reemplazada por un nuevo grupo de viviendas (1/6) incorporándose al panel de muestra.

A partir del segundo semestre del 2001 se cambió esta modalidad de rotación por una que es aplicada a aquellas viviendas que sólo hayan sido encuestadas durante los últimos tres años y ya muestren cansancio los informantes, rehusándose a dar información tal que esto implique un aumento en la tasa de no respuesta en dicha entidad.

2.3.3 Bondades de la encuesta

- Las EHM son de cobertura nacional, cuyo tamaño muestral permite generar resultados a nivel de la entidad federal.
- Es una encuesta que se ha realizado ininterrumpidamente desde 1967.
- Se generan los principales indicadores de Fuerza de Trabajo a nivel nacional con periodicidad mensual.
- Sus definiciones están basadas en acuerdos y recomendaciones de organismos nacionales e internacionales.
- Esta basada en un diseño muestral probabilístico, lo que permite hacer inferencia sobre toda la población.

2.3.4 Variables investigadas

De manera descriptiva, las variables investigadas en la encuesta se pueden clasificar en los siguientes grupos:

- Características de la vivienda: se refiere a la tipología constructiva, tenencia y densidad habitacional.
- Características del hogar: se trata de identificar los servicios disponibles para el hogar, características de la Familia.
- Variables sociodemográficas: involucran un conjunto de características referidas a las unidades familiares y a la población.

- Variables de condición de actividad y fuerza de trabajo: interesa identificar aquellos aspectos básicos socio-ocupacionales que se vinculan claramente con las condiciones de vida de los hogares.

2.4 Técnicas de preprocesamiento de datos

Para mejorar el entendimiento del fenómeno bajo estudio, se hace imprescindible el uso de técnicas estadísticas que permiten analizar simultáneamente el conjunto de datos de los que se disponen, estas son las técnicas de Análisis Multivariante. En este sentido, Hair (1999) expresa que sólo a través del análisis multivariante las relaciones múltiples entre las variables podrán ser examinadas adecuadamente para obtener un entendimiento más completo y real del entorno que permita tomar las decisiones más adecuadas.

2.4.1 Análisis de correspondencia

Es una técnica estadística que se utiliza para analizar, desde el punto de vista gráfico, las relaciones de dependencia e independencia de un conjunto de variables categóricas a partir de los datos de una tabla de contingencia. El análisis de correspondencias se desarrolla mediante el trabajo sobre dos tablas de datos, una primera tabla contiene las frecuencias respecto a las modalidades de dos variables, usualmente se denomina análisis de correspondencias binarias o simples; mientras que el segundo tipo de tabla contiene la información sobre varias variables, el análisis se conoce como de correspondencias múltiples (ACM). Díaz (2002) señaló que el ACM se dirige a buscar aquellas variables o factores cercanos (altamente correlacionados) con todos los grupos de modalidades. Y además expresó que el método encuentra asociaciones entre variables de tipo categórico a través de las respectivas modalidades entre éstas.

2.4.2 Análisis de conglomerados "cluster"

Según Hair (1999), el análisis de conglomerados es una técnica multivariante que permite agrupar un conjunto de datos en función del parecido o similitud que

existe entre ellos, permitiendo detectar el número óptimo de los grupos y su composición únicamente a partir de esa similitud, sin asumir ninguna distribución específica para las variables.

El análisis de conglomerados posee algoritmos y métodos para llevar a cabo ese agrupamiento de los datos en cada una de las categorías que son identificadas. La clasificación de los datos en los grupos o categorías se lleva a cabo de manera tal que si el grado de la asociación entre dos objetos es máximo pertenecen al mismo grupo, y de otra manera si es mínimo.

2.5 Selección de las muestras para entrenamiento

Entre los errores más frecuentes que podemos encontrar en la selección de los datos para entrenamiento están los siguientes:

- Selección de muestras con un número excesivo de datos para entrenamiento.
- Mala selección de la muestra usada para el entrenamiento, en donde la misma no representa el dominio del problema medido en las características de las variables en estudio, esto conduce a la creación de modelos inválidos.
- Selección de muestras con valores faltantes.

El procedimiento más común consiste en particionar el total de datos en dos conjuntos: uno para entrenamiento y otro para validación o prueba. Por ejemplo 70-30 (70% entrenamiento y 30% validación), 80-20, etc.

La realización de estas particiones tiene limitaciones evidentes:

- No toman en cuenta la dispersión de los datos de entrada, (variables con elevada dispersión necesitan más datos de entrenamiento), en ese caso se utiliza Muestreo Aleatorio Simple.
- No toman en cuenta la multidimensionalidad de los datos. (diferentes dispersiones pueden sugerir particiones distintas), en ese caso se utiliza Muestreo Aleatorio Estratificado. (Colmenares, 1999).

En cada caso los estudios sugieren rigurosidad en la selección de los datos para entrenamiento y prueba.

2.6 Aspectos generales de redes neuronales artificiales

A continuación se presenta la teoría de un método inspirado en el funcionamiento del Cerebro humano: Redes Neuronales Artificiales. Para dar inicio a los fundamentos que rigen este modelo es necesario hacer referencia a la preocupación que muchos científicos e investigadores han tenido por la construcción y/o diseño de máquinas capaces de realizar tareas o procesos con cierto grado de inteligencia. Hace ya algunos años surgieron un conjunto de métodos que tienen como objetivo principal, imitar computacionalmente destrezas vinculadas de alguna manera con la inteligencia humana. Esta área del conocimiento fue enmarcada como Inteligencia Artificial. Del Brío y Molina (2002)

Se ha establecido una conexión casi directa de esta nueva área con el computador del ser humano: el cerebro. Se ha considerado como un procesador pero con características muy particulares. El cerebro es capaz de procesar a gran velocidad grandes cantidades de información provenientes de los sentidos, realizar combinaciones o comparaciones entre ellas o con alguna otra información que ha sido almacenada anteriormente y dar respuestas adecuadas incluso ante nuevas situaciones. Es desconocida en parte la manera cómo el cerebro aprende a procesar la información, y en ese empeño, de tratar de imitar el funcionamiento del cerebro surge un modelo conocido como Redes Neuronales Artificiales. Con este método, está la construcción de un modelo artificial que puede alcanzar dos cosas, la primera abstraer o emular las características esenciales de las neuronas biológicas con la propiedad de aprender y asociar hechos, quedando claro que la solución a los problemas que requieren de experiencia es la construcción de sistemas que sean capaces de reproducir esta característica humana, y la segunda, implementar el modelo en un computador para simulaciones. Acosta (2000)

La red neuronal como sistema para el tratamiento de información, posee como unidad básica de procesamiento, e inspirada tal como lo muestra la Figura 1, en la célula fundamental del sistema nervioso del ser humano: la neurona. Son componentes relativamente simples del ser humano, pero cuando se tienen por millones conectadas, como están presentes en el cerebro, se hacen poderosas.

2.6.1 Neurona biológica

De la figura 1 se puede observar la neurona y sus partes. Las dendritas son un conjunto de extensiones tubulares, las cuales son las receptoras y transportan señales eléctricas al cuerpo de la célula. El axón difiere de las dendritas en su forma, es la conexión de salida usada por la neurona para emitir señales. Las sinapsis, son las unidades funcionales y estructurales elementales que median entre las interacciones de las neuronas. En las terminaciones de las sinapsis se encuentran unas vesículas que contienen unas sustancias químicas llamadas neurotransmisores que ayudan a la propagación de las señales de una neurona a otra.

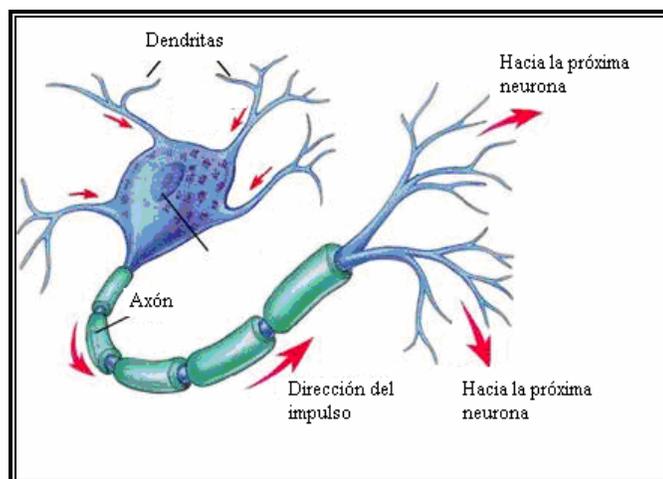


Figura 1 Neurona biológica

El cuerpo celular contiene el núcleo químico, aquí se ejecutan todas las transferencias necesarias para la vida de la neurona (Colina y Rivas 1998)

Según Hilera y Martínez (2000) se estima que el cerebro humano contiene más de cien mil millones de neuronas, estudios sobre la anatomía del cerebro humano concluyen que hay más de 1000 sinapsis a la entrada y a la salida de cada neurona. Las neuronas y las conexiones entre ellas (sinapsis) constituyen la clave para el procesamiento de la información.

Las neuronas biológicas según Colina y Rivas (1998):

“Son células nerviosas que constituyen los elementos primordiales del sistema nervioso central. En general, las neuronas son capaces de recibir señales

provenientes de otras neuronas, procesar estas señales, generar pulsos nerviosos, conducir estos pulsos y transmitirlos a otras neuronas”

Lo anteriormente explicado es lo que de algún modo justifica el amplio campo en el cual puede ser aplicada las Redes Neuronales Artificiales. Entre las aplicaciones más exitosas se pueden mencionar: Procesamiento de imágenes y voz, reconocimiento de patrones, predicción, control y optimización, filtrado de señales.

2.6.2 Definición de RNA

Tal como lo define Colina y Rivas (1998), una RNA es un modelo diseñado para emular algunas de las características computacionales del ser humano. Este tipo de modelo incluye tanto las características funcionales como configuraciones topológicas.

La operación de la neurona es usualmente explicada como un proceso donde la célula ejecuta una suma de señales que llegan por sus dendritas. Cuando esta suma es mayor que cierto umbral, la neurona responde transmitiendo un pulso a través de un axón. Si la suma es menor que el umbral, la neurona permanece inactiva.

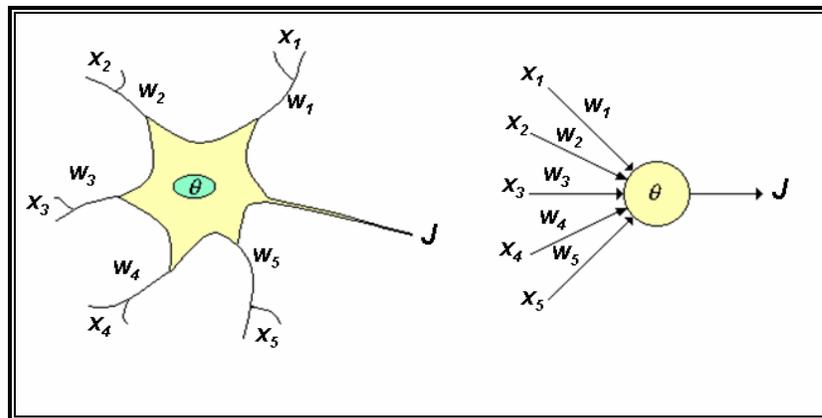


Figura 2 Analogía de la neurona artificial y biológica

2.6.3 Topología de las RNA

En un modelo de red neuronal se pueden especificar los siguientes elementos:

- Entradas o Nodos de Entrada: son los patrones o ejemplo de entrada medidos que se le suministran a la red y dependen del dominio de problema en estudio.

- Salidas o Nodos de Salida: son los nodos donde se representa resultado del problema en estudio.
- Pesos: valores numéricos que representa la fortaleza de conexión que tiene cada entrada sobre la neurona.
- Punto de Suma de Entradas Ponderadas: este punto no es más que la sumatoria de todas las señales de entrada a la neurona multiplicadas por sus correspondientes pesos.
- Función de Activación: esta función limita el rango de la salida de la neurona y la misma puede ser lineal o no lineal. La determinación del tipo de función de activación depende del problema que se quiera resolver.
- Sesgo: es equivalente al peso de una entrada fija igual a 1. Este valor permite que haya una cierta flexibilidad en la salida de cada neurona, generando un mejor ajuste de la salida obtenida con la salida deseada.

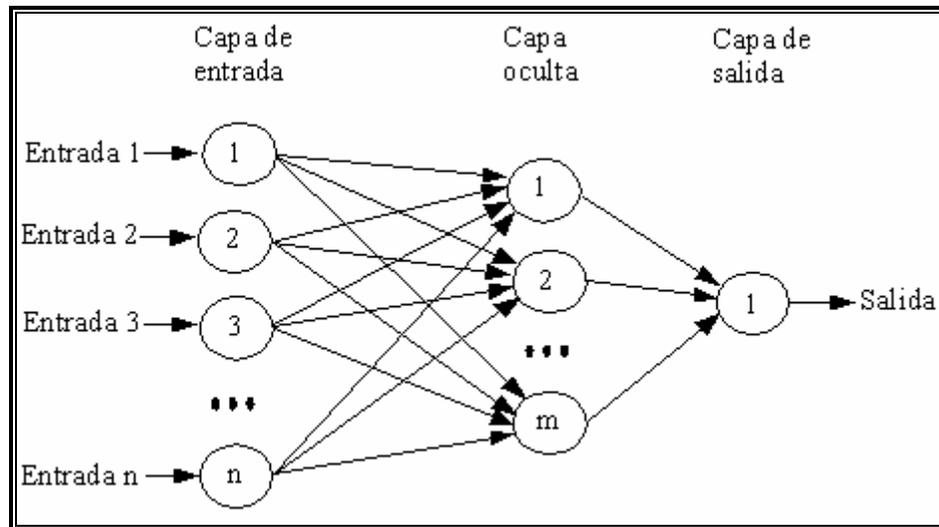


Figura. 3 Estructura de RNA

Tal como se describe en la figura 3, la característica funcional de la red neuronal es como sigue.

- Capa de entrada: está formada por los nodos que reciben directamente la información proveniente de las fuentes externas a la red. Esta información puede requerir una etapa de preprocesado para mejorar el conjunto de datos originales

a fin de lograr un mejor desempeño de la red. El número de entradas se determina de acuerdo al problema en estudio.

- **Capas Ocultas:** esta ubicada entre la capa de entrada y de salida, no tienen contacto con el exterior y son las encargadas de extraer, procesar y capturar la información. El número de capas ocultas debe ser especificado en la arquitectura de la red. Depende en gran parte del problema en estudio, pero en general, un problema podrá representarse bastante bien con una o dos capas (Hilera y Martínez 2000). De igual manera el número de neuronas por capa es variable, una de las formas de determinar este número es por ensayo y error, aunque también se puede tomar como un criterio útil el promedio entre el número de entradas y salidas de la red (Collantes 2001).
- **Capa de Salida:** esta formada por los nodos que transfieren la información de la red al exterior. El número de nodos o neuronas de salida dependerá del problema en estudio.
- **Conexiones:** la conexión entre los nodos de una red neuronal está relacionada con la forma en que las salidas de unas neuronas se canalizan para convertirse en entradas de otras neuronas. Las conexiones que unen a los nodos que forman una red neuronal artificial tienen asociado un valor numérico identificado como peso, que es el que hace que la red adquiera conocimiento. De acuerdo al tipo de conexión se pueden reconocer las siguientes :
 - **Totalmente conectadas:** cada nodo o neurona de una capa está conectada a todas las neuronas de un nivel o capa superior
 - **Parcialmente conectadas:** la salida de las neuronas de una capa son entradas de algunas de la siguiente capa.
- **Dirección:** es la forma en que fluye la información de una capa a otra. Las redes según la dirección se pueden clasificar en:
 - **Redes de Propagación hacia adelante:** son aquellas donde las salidas de las neuronas de una capa sólo se propagan a las neuronas de la capa siguiente. La información fluye de la entrada hacia la salida.
 - **Redes de Propagación hacia atrás:** son aquellas donde las salidas de las neuronas de una capa pueden estar conectadas como entradas de

neuronas de capas previas. Un caso particular de estas redes son las Redes recurrentes son redes de propagación hacia atrás que forman lazos cerrados; es decir la salida de una neurona es entrada a la misma neurona. Las redes de propagación hacia atrás que tienen lazos cerrados son sistemas recurrentes.

- Redes de Alimentación Lateral: son redes en las cuales las salidas de las neuronas pueden ser entradas de neuronas de la misma capa.

2.6.4 Funciones de activación

Existen una multiplicidad de funciones de activación y la elección de cualquiera de ellas depende del problema en estudio y del investigador. En este apartado se mencionan algunas de las que son utilizadas para los problemas de pronóstico y se amplía un poco más la empleada para el problema que está planteado en este proyecto, que son las funciones de bases radiales (RBF). En la tabla 1 se observan algunas de las funciones a emplearse con redes, y la selección de una de ellas depende de la naturaleza del problema bajo estudio.

2.6.5 Entrenamiento de las RNA

De acuerdo a Brío y Molina (2002), en general en el entrenamiento de una red, requiere de la presencia repetida de un número relativamente amplio de patrones que permiten que la red aprenda, haciendo modificaciones por medio de alguna regla de aprendizaje. Los tipos de aprendizaje pueden ser clasificados en distintos tipos entre ellos tenemos:

- a) Aprendizaje Supervisado: se presenta a la red un conjunto de patrones, junto con la salida deseada e iterativamente esta ajusta sus pesos hasta que su salida tiende a ser la deseada, utilizando para ello información detallada del error que comete en cada paso.
- b) Aprendizaje No Supervisado o Auto-organizados: En este tipo de aprendizaje se presenta a la red multitud de patrones sin adjuntar la respuesta que deseamos.
- c) Reforzados: Se sitúa a medio camino entre el supervisado y el autoorganizado. Como en el primero de los citados se emplea información sobre el error cometido,

pero en este caso existe una única señal de error, que representa un índice global del rendimiento de la red (Indicando lo bien o mal que esta actuando, sin mas detalles). Como en el caso del no supervisado, no se suministra explícitamente la salida deseada.

d) Híbridos: Coexisten en la red los dos tipos básicos de aprendizaje, el supervisado y el no supervisado los cuales tienen lugar normalmente en distintas capas de neuronas.

Es importante mencionar que en este estudio se hace uso del "Aprendizaje Híbrido". Ya que este es aplicado a los modelos de redes de contra-propagación y Funciones de bases radiales o conocidas como RBF, siendo utilizado este último modelo para el presente estudio.

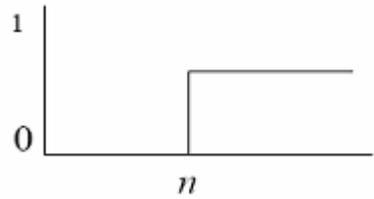
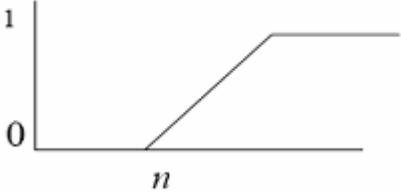
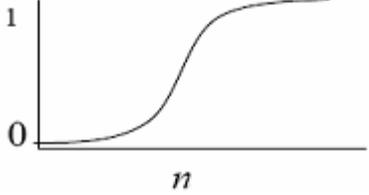
Función umbral	
Función continua saturada	
Función Sigmoide	

Tabla 1. Funciones de activación más comunes

2.7 Funciones de Base Radial

Conocido también como modelos RBF y que aunque es de reciente introducción, cada vez cuenta con más aplicaciones prácticas, gracias a su simplicidad, generalidad y rapidez de aprendizaje. Se trata de un modelo que a menudo se estudia junto al MLP (Multilayer Perceptron) por ser una red unidireccional para aproximación funcional, pero considerado de tipo híbrido por incorporar aprendizaje supervisado y no supervisado.

Una de las características más resaltantes de las RBF es que permite modelar con cierta facilidad sistemas no lineales, requiriendo de un tiempo más reducido para su entrenamiento.

2.7.1 Arquitectura de la RBF

Este tipo de redes se han asociado tradicionalmente con una arquitectura simple de tres capas, donde cada capa está conectada unidireccionalmente con su capa consecutiva. Este modelo tiene la caracterización de poseer una sola capa oculta.

En la capa de entrada como las emplean habitualmente las RNA, se encargan de enviar la información del exterior a las neuronas de la capa oculta. La capa oculta está compuesta de un conjunto de nodos a los cuales se le asocia una función de activación de tipo radial, denominada Funciones de Base Radial. Las funciones radiales reciben como entradas todos y cada uno de los atributos de los patrones, y se caracterizan por estar centradas en un punto del espacio de entrada.

En la capa de salida de la red se combina linealmente la salida de la función radial de la capa oculta, mediante unas ponderaciones.

Las RBF no son comúnmente utilizadas en aplicaciones que posean un alto volumen de patrones de entrenamiento. Sin embargo se les reconoce por tener una eficiencia en la fase de entrenamiento. Enfocando un poco la idea del entrenamiento de las RBF se pueden describir como el aprendizaje por etapas, ya que en una primera fase se entrenan las neuronas ocultas y luego las de salida.

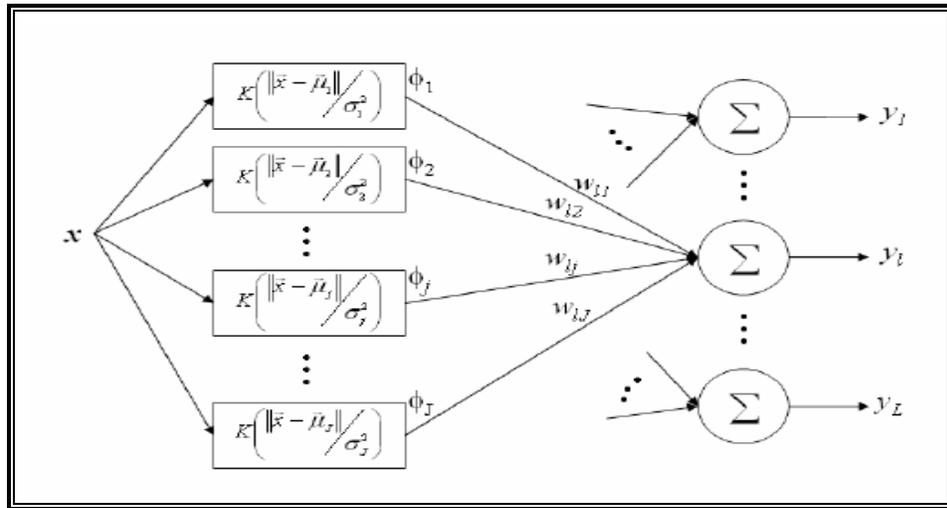


Figura 4. Arquitectura de RBF

2.7.2 Algoritmo de las RBF

Es necesario considerar, que así como en las redes multicapa BP, las RBF requieren de la selección de un número de nodos que para este caso son denominados "*Nodos Radiales*" que son los que cubren el espacio de entrada. Estos "*Nodos Radiales*" se ven influenciados y crecen exponencialmente de acuerdo al número de patrones que estén presentes en el modelo. Se puede iniciar con un número reducido de nodos pero una vez sea detectada la ausencia de algún patrón de entrada, se tiene la comodidad de poder incrementar los nodos radiales.

Una vez definida la topología de las RBF, se procede a encontrar los mejores parámetros para la arquitectura. El aprendizaje de las RBF se podría describir en dos etapas: la primera constituida por el entrenamiento en la capa oculta y la subsiguiente con las salidas de la capa oculta se entrena la capa de salida.

Para la primera etapa deben encontrarse los valores de los centroides utilizando cualquier algoritmo no supervisado de agrupamiento, tomando en cuenta el número de grupos que se desea encontrar y que corresponde con el número de nodos radiales. En un caso en particular, luego de determinar los centros, los anchos de los campos receptivos pueden determinarse por uno de varios criterios heurísticos. Teóricamente una red RBF es un aproximador universal si todas las unidades de la

primera capa tienen el mismo ancho. Este ancho podría fijarse como el promedio de las distancias entre cada unidad y su vecino más cercano:

$$\sigma = \frac{\sum_{j=1}^J \|\mu_j - \mu_j^*\|}{J} \quad (2.1)$$

Donde μ_j^* es el centro más cercano a μ_j

Después de haber entrenado la primera capa, los pesos de las unidades lineales de la capa de salida pueden ajustarse con descenso de gradiente según la siguiente expresión:

$$\begin{aligned} \Delta w_{ij} &= \alpha \cdot (d_i - y_i) \cdot \frac{\partial}{\partial w_{ij}} y_i = \alpha \cdot (d_i - y_i) \cdot \frac{\partial}{\partial w_{ij}} \left(\sum_{j=1}^J w_{ij} \cdot \phi_j(x) \right) \\ \Delta w_{ij} &= \alpha \cdot (d_i - y_i) \cdot \phi_j(x) \end{aligned} \quad (2.2)$$

o de la siguiente manera:

$$w_{kj}(t+1) = w_{kj}(t) + \varepsilon(t_z - z_k) \phi(r_j) \quad (2.3)$$

Siendo la expresión de salida:

$$z_k = \sum_j w_{kj} \phi(r_j) + \theta_k \quad (2.4)$$

θ_k : representa los umbrales.

Obsérvese que el procedimiento descrito para entrenar las redes RBF utiliza tanto técnicas del aprendizaje supervisado como del aprendizaje no supervisado. Se entrena por separado la etapa de extracción de características y la del aproximador lineal, como se muestra en la figura 5.

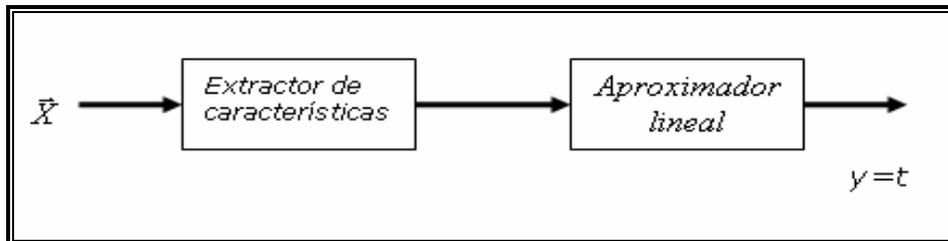


Figura 5. Arquitectura de aproximador lineal con extractor de características

Capitulo III

Análisis y exploración de los registros de viviendas y hogares de las EHM.

En este capítulo se presentan las características de los datos contenidos en las EHM. Se describen los pasos a seguir para obtener el registro con toda la información relacionada a Viviendas y Hogares. Así mismo, se realiza el análisis descriptivo de cada una de las variables utilizadas para el estudio, y de la aplicación de técnicas estadísticas y de análisis multivariante aplicadas a las variables tomadas de las encuestas, para así obtener las nuevas variables que serán empleadas para estudio del déficit de viviendas.

3.1 Disposición de la encuesta

La información en las EHM se encuentra dividida en tres registros (Vivienda, Hogares, y Personas) que contienen una estrecha interrelación. Es a través de las variables CONTROL, ÁREA, LÍNEA contenidas en las encuestas que se alcanza la fusión de estos registros. Como descripción para el caso de un período se puede observar con un ejemplo cómo se establece esta interrelación a través de dichas variables. En la tabla 2 se observa que cada uno de los registros de vivienda, hogar y persona contienen los campos CONTROL, ÁREA, LÍNEA y son únicos para cada encuestado.

REGISTRO	CONTROL	AREA	LINEA
VIVIENDA	088	003	001
HOGAR	088	003	001
PERSONA	088	003	001

Tabla 2. Descripción de las variables de identificación en las EHM.

Sin embargo, es necesario resaltar que en los casos de viviendas donde habitan más de un hogar hay información repetida en estos campos, debido a que esta información permite identificar estrictamente la vivienda que está siendo encuestada. Por otro lado, cuando se fusionan estos registros, se puede identificar claramente que una vivienda puede estar habitada por dos o más núcleos familiares. Esto se debe a que existe información única sobre las características de la vivienda e información múltiple para cada hogar.

La descripción de cómo obtener el Registro de Viviendas y Hogares 1994-2005 para este proyecto en particular se hizo a través de SPSS (Versión 13.0)³. A pesar de poder mejorarse la selección de vivienda y hogares, este criterio de emplear las variables CONTROL, ÁREA, LÍNEA fue el más idóneo; donde CONTROL identifica un segmento/sector de la muestra seleccionada en cada entidad a ser encuestada, ÁREA identifica las subdivisiones del sector de la muestra y LÍNEA identifica la vivienda encuestada.

3.2 Pasos para obtener el registro de viviendas y hogares 1994-2005:

Las EHM recogen la información para cada encuesta los registros de Vivienda, Hogar, y Persona separadamente. Para el estudio del Déficit de Viviendas se deben agrupar estos registros. Dicha actividad fue realizada como se describe a continuación.

1. Agrupación de la información de cada entidad en los registros. La información de cada estado fue almacenada en los registros Vivienda, Hogar y Personas de manera arbitraria y en consecuencia, es necesario agruparla, para facilitar la búsqueda de la información relacionada a un estado específico cuando así se requiera.

³*Statistical Package for the Social Sciences (SPSS)*. El SPSS es un programa estadístico informático muy usado en las ciencias sociales y las empresas de investigación de mercado, creado en 1968 por Norman H. Nie, C. Hadlai (Tex) Hull y Dale H. Bent. Se han desarrollado múltiples versiones para cumplir con las exigencias de los usuarios.

En la siguiente figura se observa cómo están almacenados los datos en las EHM en cualquier registro y así mismo, cómo resulta al aplicar esta nueva agrupación.

Datos almacenados en las EHM

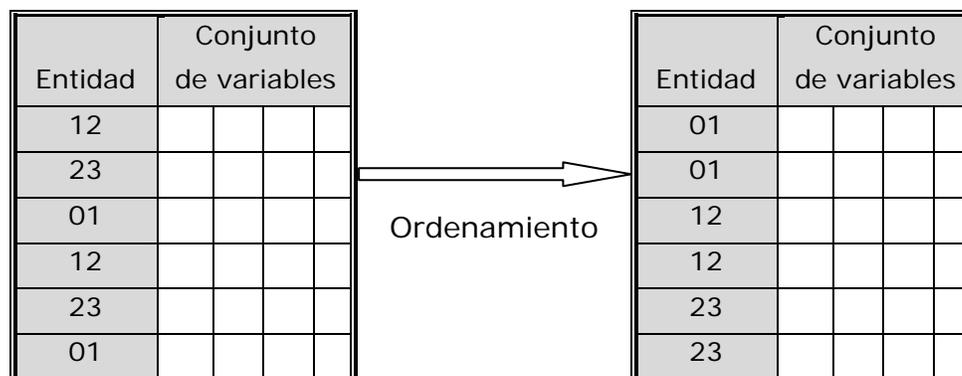


Figura 6. Agrupación de la información por estado en cada registro

2. El registro Personas no aportaba información relevante. Sin embargo, de este se extrae sólo lo referido al número de individuos pertenecientes a cada hogar, debido a que en los otros registros no se tiene la información completa. Para obtener esta información se desarrolla un algoritmo programado para contabilizar el número de individuos correspondientes a un hogar en particular, esto debido al elevado número de observaciones existentes. Esta nueva información fue incorporada en cada registro Vivienda-Hogar perteneciente a cada período.
3. Luego de ordenar los datos, se extrae los pertenecientes al estado o entidad en particular al cual se refiere el estudio, como se observa en la figura 7. Es necesario aclarar que cada uno de estos registros de Vivienda, Hogar y Personas se encuentran separados físicamente. Además, el conjunto de estos registros componen la base de datos de las encuestas recolectadas semestralmente durante el período 1994-2005. A este momento, se excluye el componente Personas de las EHM y solamente se trabajará con Vivienda-Hogar para la construcción de las variables que se emplearán en el estudio. En las tablas 3 y 4 se observan las variables correspondientes a cada registro vivienda y hogar.
4. Se ordenan los datos de acuerdo al número correlativo de vivienda y al sector que pertenece de cada registro Vivienda y Hogar correspondientes a todos los períodos mencionados. Este ordenamiento se realiza a través de las variables CONTROL y LÍNEA.

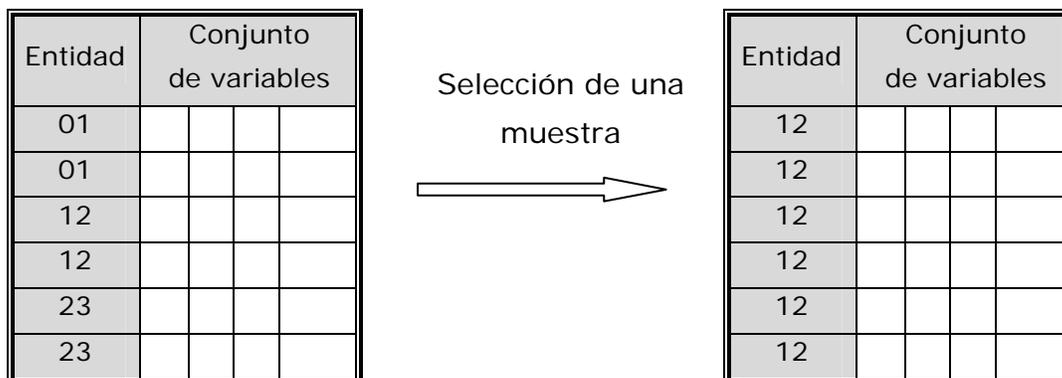


Figura 7. Extracción de los datos del estado Mérida en cada registro

	VARIABLES	DESCRIPCIÓN
Vivienda	entidad	Identifica el estado al que pertenece la observación
	control	Identifica un segmento/sector de la muestra
	área	Identifica las subdivisiones del sector de la muestra
	línea	Identifica la vivienda encuestada
	num_hogare	Número del hogar según sea el principal (1) o no (2 y más)
	ocupantes	Número de habitantes en la vivienda
	pv1	Tipo de Vivienda
	pv2	Paredes
	pv3	Techo
	pv4	Piso
	pv5a	Condición de la vivienda
	pv5b	Condición de las paredes
	pv5c	Condición del techo
	pv5d	Condición del piso
	pv6	Total de cuartos, contando sala, comedor y otros
	pv7	Número de cuartos utilizados para dormir
	pv8	A esta vivienda llega el agua por:
	pv9	Servicio de eliminación de excretas
	pv10	Número de pocetas
	pv11	Número de baños con ducha o regadera
pv12a	Servicio eléctrico público	
pv12b	Recolección directa de basura	
pv12c	Container de basura	
pv12d	Servicio telefónico (Telefónico fijo, anexo 2003)	
	SEMESTRE	Indica el semestre que fue realizada la encuesta

Tabla 3. Variables para la fusión del registro Vivienda- Hogar. Vivienda.

	Variables	DESCRIPCIÓN
Hogar	ph1	Número de cuartos para dormir
	ph2	¿Tiene uso exclusivo de baños con ducha o regadera?
	ph3	Número de baños con ducha o regadera
	ph4a	¿ Posee nevera ?
	ph4b	¿Posee Lavadora?
	ph4c	¿ Posee televisor ?
	ph4d	¿ Posee Cocina a gas ó eléctrica ?
	ph4e	¿Posee cocina de 'Kerosene u otros ?
	ph4f	¿ Posee secadora de Ropa ?
	ph4g	¿ Posee calentador de agua ?
	ph4h	¿ Posee aire acondicionado?
	ph5	Numero de automóviles tiene ese hogar
	ph6a	Tenencia
	ph6b	Alquiler
	ph7	Gastos del hogar dependen de...
	ph8a	Posee beca
	ph8b	Ingresos por concepto de beca
	ph9	Nº de mujeres de 12 a 49 años embarazadas
	num_person	Identifica el nexos de cada individuo dentro del núcleo familiar
	PH14I	¿Posee filtro de agua? (sólo a partir de 2003)
	PH14J	¿Posee radio? (sólo a partir de 2003)
	PH14K	¿Posee horno microondas? (sólo a partir de 2003)
	PH14L	¿Posee teléfono móvil celular? (sólo a partir de 2003)
	PH14M	¿Posee televisión por cable? (sólo a partir de 2003)
	PH14N	¿Posee computadora? (sólo a partir de 2003)
	PH14O	¿Posee Acceso a Internet? (sólo a partir de 2003)
	pv11e	Numero de automóviles tiene ese hogar
ph14p	¿Ninguno? (sólo a partir de 2003)	

Tabla 4. Variables para fusión del registro Vivienda- Hogar. Hogar.

5. Se fusionan las variables ÁREA Y LÍNEA de los registros Vivienda y Hogar. Se inserta al registro de vivienda las variables pertenecientes al registro de Hogar que no son parte integrante del registro de vivienda así se obtiene el registro Vivienda-Hogar para cada período semestral. Esta acción se realizó en SPSS a través de la opción DATA, con el comando MERGE FILES y escogiendo Add Variables.

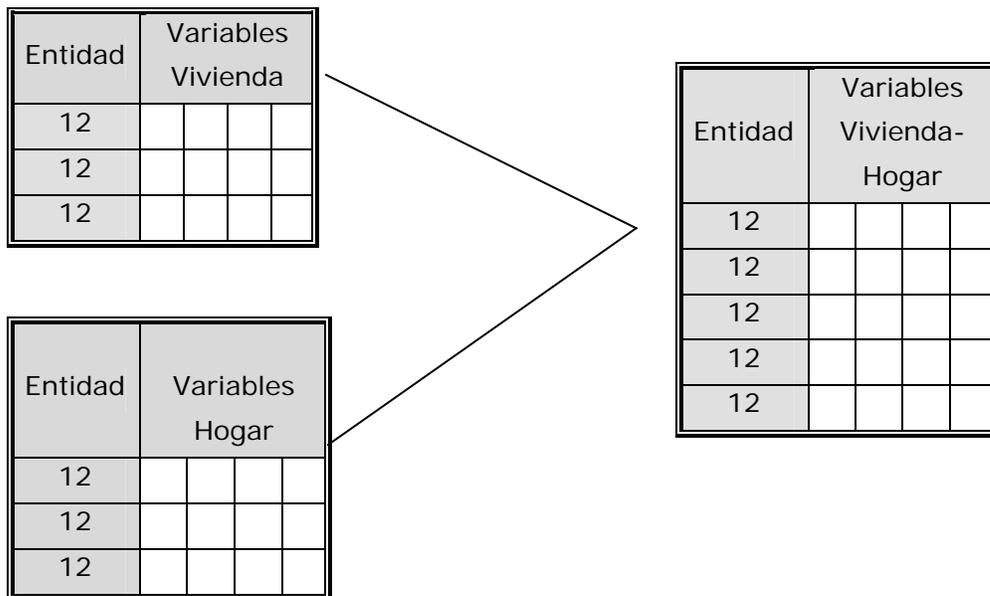


Figura 8. Fusión de los Registros Vivienda y Hogar

6. Se crea una nueva variable denominada SEMESTRE para identificar el semestre al que corresponde el registro Vivienda-Hogar. A partir del segundo semestre del año 1994 se codifica esta variable con uno y así sucesivamente hasta veintitrés que representa el segundo semestre de 2005. Finalmente se integran todos los registros Vivienda-Hogar desde el segundo semestre de 1994 hasta el segundo semestre de 2005 en un solo conjunto de datos, obsérvese Figura 9. Esta acción la permite realizar el SPSS a través de la opción DATA seleccionar MERGE FILES, Add Cases.

3.3 Selección de las observaciones

Considerando que el registro construido, recoge información de las viviendas que albergan más de un hogar, lo que indica que la información de esta vivienda es la misma cuando se encuesta todos los hogares; se hace necesario reconocer aquellas observaciones donde el Número de Hogares sea mayor que uno. Es necesario dejar en el registro el número mayor de los hogares que se albergan en una vivienda y eliminar los demás para evitar repetición de la información de esa vivienda. En la tabla 5 ilustra la información de aquellas observaciones que forman parte del conjunto de datos.

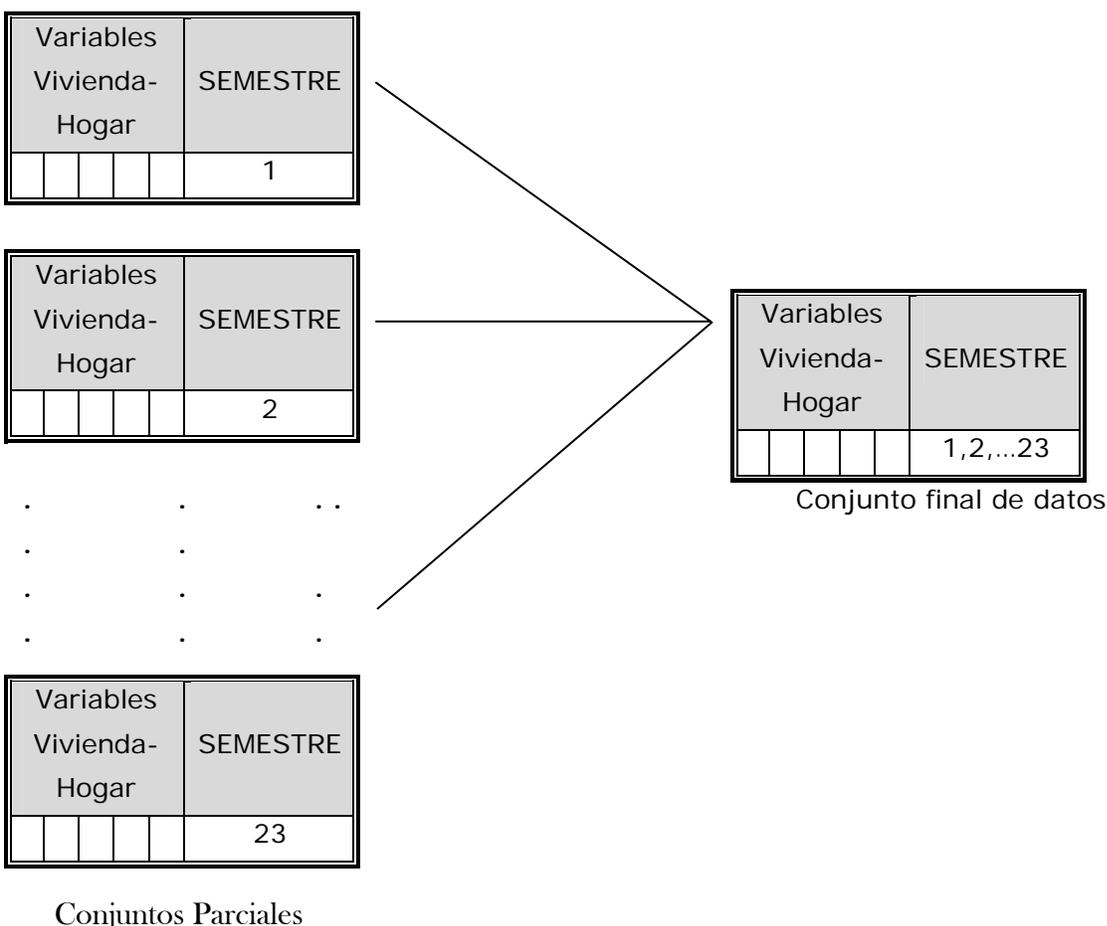


Figura 9 Registro Vivienda-Hogar 1994-2005

ID	Num_Hogares	Variables Viviendas-Hogares			
193221	1				
878221	2				
878221	1				
145131	2				
145131	1				

Tabla 5. Selección de Observaciones

Para aclarar cuales viviendas son tomadas en cuenta para formar parte del conjunto de datos con el que se evalúa la condición deficitaria. El conjunto de observaciones sombreadas, en la tabla anterior, corresponde a las observaciones cuyo ID es el mismo y está relaciona con los hogares o núcleos familiares que comparten una

vivienda, por lo que quedará para el conjunto de observaciones aquel registro donde el indicador de Número de Hogares sea el mayor, esto para reconocer las viviendas con más de un hogar y por tanto medir el déficit cuantitativo.

3.4 Selección de las variables

Una vez construido el Registro Vivienda-Hogar 1994-2005, se evalúa la importancia de aquellas variables que se relacionan con el estudio del Déficit de Vivienda. Antes de realizar este proceso de selección se debe enmarcar el dominio del problema para determinar las variables originales que se desean incluir de las encuestas EHM. El Déficit de Vivienda se orienta a reconocer aspectos muy puntuales de las características de las viviendas existentes y la necesidad que reporta la falta de vivienda para algunas familias. Es útil contar con una ubicación referencial de la vivienda que ha sido seleccionada y/o encuestada. Esta información la contiene la variable ID que corresponde con la identificación de la vivienda y que fue construida de acuerdo a la agrupación de las variables CONTROL, ÁREA, LÍNEA y SEMESTRE (Ver Anexo 1). Con esta información se podrá ubicar en cuál sector y a cual vivienda exactamente pertenece cada caso observado. Sin embargo, esta no es más que una información referencial y de control que se utiliza para el manejo de las observaciones.

Las variables consideradas de las EHM para realizar la investigación, ayudan a conformar el concepto relacionado al Déficit de Viviendas, obtenido a través del Déficit Cualitativo y Cuantitativo.

El Déficit Cualitativo, tal como se indicó, hace referencia a la deficiencia básicamente de las cualidades y características de la vivienda. Partiendo de la idea que una vivienda no es sólo un techo para resguardarse. Es evidente que para que una vivienda sea un lugar apto para habitar debe contar con ciertas especificaciones, las cuales han sido mencionadas en los apartados anteriores.

En general, de acuerdo a regulaciones universales establecidas por la Organización Naciones Unidas se considera una amplia gama de características para que una

vivienda pueda ser considerada como adecuada⁴. Sin embargo en esta encuesta alguna de estas características no fueron consideradas.

Tal como se mencionó, el Déficit Cualitativo se relaciona con las características de la vivienda. Para abarcar la mayor información en este sentido, se consideraron las variables que hacen referencia a condiciones de la vivienda (referidas a su estructura: tipo de vivienda, paredes, techo, piso); a la seguridad jurídica de la tenencia; a la calificación de una vivienda como adecuada, aceptable o inadecuada en términos de habitabilidad, que se mide por condiciones de higiene y seguridad al acceso a servicios tales como agua, aseo, electricidad, eliminación de excretas. De acuerdo a la distribución espacial de la vivienda y al uso de estos espacios, el hacinamiento se relaciona con la cantidad de individuos que habitan en dicha vivienda y el número de cuartos disponibles para dormir. Según Molina y Gonzáles en su estudio realizado en 1998 acerca de los *Principales Indicadores de Pobreza*, señalan que el indicador de hacinamiento es satisfecho si duermen hasta 2 personas por cuarto e insatisfecho si son 3 o más. El hacinamiento en una vivienda se puede observar a través de la relación de las variables Ocupantes y Número de Cuartos para dormir (NumCD). Cuando la relación Ocupantes/NumCD es mayor a dos existe condición de Déficit por concepto de Hacinamiento.

En resumen el Déficit Cualitativo está medido por la contribución numérica de una o todas las variables Tenencia (T), Hacinamiento (H), Adecuación (A), Condición (C) de acuerdo a

$$D_{CL} = f(T, H, A, C) \quad (3.1)$$

El Déficit Cuantitativo, por su lado, se mide de acuerdo al hecho que todo grupo familiar tiene derecho a disfrutar de una vivienda cuantifica el déficit para esos núcleos familiares.

De ahí

$$D_{CT} = f(Num_Hogares) \quad (3.2)$$

⁴ Vivienda adecuada: se le incorpora un grupo de características relacionadas a la vivienda. Para determinar en qué medida las diversas formas de vivienda son adecuadas: a) seguridad jurídica de la tenencia; b) disponibilidad de los servicios, facilidades e infraestructura; c) Gastos soportables; d) Habitabilidad; e) Asequibilidad; f) Lugar.

Es evidente que para la medición del Déficit Cualitativo es necesario construir variables que contengan la información referida anteriormente. Este proceso será descrito más adelante con mayor detalle.

Finalmente, la condición deficitaria de vivienda depende de la consideración de ambas mediciones:

$$\text{Condición Deficitaria} = f(D_{CL}, D_{CT}) \quad (3.3)$$

3.4.1 Variables originales

Una sola variable no ha recibido ningún tipo de tratamiento para ser utilizadas en este estudio. Se corresponde con Num_Hogares, y representa el número de hogares que viven en una vivienda en particular. A través de esta variable se puede medir la diferencia que existe entre el número de viviendas existentes y el número de viviendas necesarias para albergar el número de hogares ó familias existentes.

3.4.2 Variables transformadas

a) Identificación: Se corresponde con ID y contiene los valores de: Control, Área, Línea, Semestre

Para obtener esta variable fue necesario concatenar los campos antes mencionados, con lo que se hace posible identificar si se requiere una vivienda en particular.

b) Hacinamiento: Mide la relación que existe entre las variables número de cuartos y el número de personas que ocupan la vivienda.

Esta variable se obtiene de acuerdo a:

$$\text{Hacinamiento} = \text{Ocupantes/Num_CuartosPD} \quad (3.4)$$

La expresión anterior permite obtener el número de individuos para cada encuesta que en promedio utilizaran un cuarto para dormir; para efectos del estudio no se desea sólo tener esta información sino también si cierta vivienda se encuentra en estado de Hacinamiento. Para obtener esa información se categoriza la variable de acuerdo a los valores excluyentes que indiquen si hay o no hacinamiento. Es decir, mayor que dos han sido categorizadas como hacinada asignándole un uno y las que están fuera de esta condición como no hacinada asignándole un cero.

c) Tenencia: Esta variable se redefinió en una nueva de dos categorías, con información del estado de propiedad de la vivienda. Esta variable toma valor uno

para identificar aquellas viviendas que se encuentran en estado de Alquiler y cero para las que son Propias.

3.4.3 Construcción de las variables del modelo

Los requerimientos del modelo para la estimación de la Condición Deficitaria a partir de las variables explicativas que se consideraron en los puntos previos, deben formularse de tal modo que tales variables incluyan la información pertinente para la construcción del modelo de clasificación. Estas variables son estimadas empleando técnicas multivariante, tales como, Análisis de conglomerados y Análisis de Correspondencia Múltiple. La variable Condición de la Vivienda es tratada de modo que complete su serie en el periodo considerado y se complementa con las características físicas que posee la vivienda. Adecuación de la vivienda es definida en una nueva variable cualitativa de origen latente. La Condición Deficitaria es construida a partir de las variables explicativas como variable observada para obtener valores respuestas que induzcan a un pronóstico del modelo.

A continuación se explica la aplicación de estas técnicas y sus resultados.

a) Condición de la vivienda: La información en las encuestas acerca de la condición de las viviendas está limitada a sólo los primeros semestres que son utilizados para los que puede obtenerse esta información directamente. Durante el período desde 1994 hasta 1998, fue incluida en las EHM la variable Condición de la Vivienda que caracterizaba el estado en que se encontraban las viviendas en términos de su estructura física (techo, paredes, piso). Adicionalmente, en la encuesta se dispone de información relacionada con el tipo de materiales con que estaba construida la vivienda. Al observar que se generaban patrones, entre la condición de la vivienda y la descripción de los materiales con que estaba construida, se considera pertinente para resolver el problema de la falta de información, la aplicación de las técnicas estadísticas y de análisis multivariante.

Para este caso particular, en primer lugar se realiza el análisis de correspondencia múltiple a fin de observar la asociación lineal, entre la variable condición de la vivienda que tenía como respuesta si era buena (1), regular (3), mala (2) y la variable que ofrecía información sobre los materiales con que fue construida.

Efectivamente se observó que existe un patrón entre los materiales que se emplearon para la construcción de la vivienda y la condición de la vivienda, tal como puede observarse en el siguiente gráfico con los resultados obtenidos de este primer análisis.

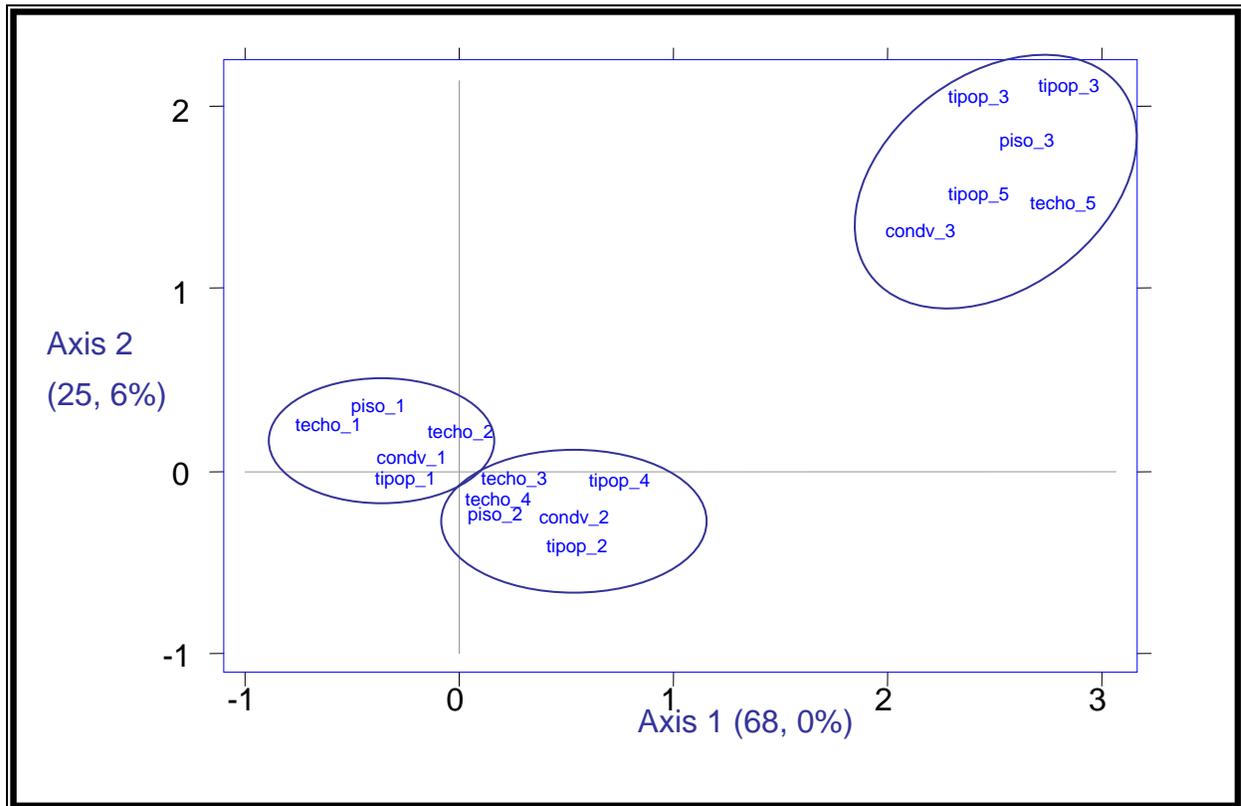


Gráfico 1. Análisis de correspondencia múltiple para la condición de la vivienda

Análisis de los resultados obtenidos con el ACM. De acuerdo al gráfico 1 las variables consideradas en el ACM son las que se describen en la tabla 6. Que derivan del gráfico, donde claramente se aprecian tres grupos. El primer grupo (a la izquierda) se refiere a viviendas en buenas condiciones (1), asociado a las viviendas que se caracterizan por tener paredes y pisos tipo 1 y techos tipo 1 y 2. El grupo del centro, se refiere a viviendas en condiciones regulares (2). Con características en paredes tipo 2 y 4, techos tipo 3 y 4 y piso tipo 2.

Variables	Descripción	Valores
CondV	Condición de la vivienda	Buena (1); regular (2); mala (3).
TIPOP	Tipo de paredes	Bloque o ladrillo frisado, concreto (1); bloque o ladrillo sin frisar (2); madera, palma, tablas, caña (3); adobe, tapia o bahareque frisado (4); adobe, tapia o bahareque sin frisar (5).
TECHO	Tipo de techo	Platabanda (1); teja (2); cemento y similares (3), láminas de zinc y similares (4); palmas, tablas, otros (5).
PISO	Tipo de piso	Mosaico, granito, cerámica, ladrillo, etc (1); cemento (2); tierra u otros (3).

Tabla 6. Descripción de las variables observadas en el ACM.

Finalmente, el grupo de la derecha describe las viviendas en malas condiciones (3), asociado a paredes tipo 3 y 5, pisos tipo 3 y techos tipo 5. Debido al alto volumen de observaciones y a las ventajas que proporcionan los paquetes computacionales para dar respuesta a la condición de vivienda se aplica el análisis de conglomerados para completar la construcción de la variable en el resto de los períodos. Al comparar este análisis con los obtenidos usando Análisis de Correspondencias Múltiples se puede observar coincidencia en los resultados. La aplicación de conglomerados ofrece la ventaja de clasificar automáticamente cada observación de acuerdo al conglomerado que le corresponde.

A continuación se muestra en detalle la conformación de 3 conglomerados que son los que se emplean para obtener la nueva variable Condición de la vivienda.

Centros Iniciales de los Conglomerados: Representan los valores de las variables de clasificación Tipo de Vivienda; Paredes; Techo; Piso; correspondientes a observaciones particulares representadas por: el primer conglomerado Tipo de Vivienda Rancho Campesino; Paredes de caña, palma, tablas, y otros; Techo de Fibrocemento, cemento, ligero y similares y el Piso de Tierra. Para el segundo conglomerado Tipo de Vivienda otro tipo; Paredes de adobe, tapia y bahareque sin frisar; Techo de palma, tabla y similares y el Piso de Tierra. Finalmente, para el tercer conglomerado Tipo de Vivienda Casa; Paredes boque o ladrillo sin frisar;

Techo de Fibrocemento, cemento, ligero y similares; y el Piso de cemento. Se emplean como centros de partida asignando cada observación al conglomerado que le corresponde, y de esta manera iniciar el proceso iterativo de ubicación de nuevos centros y resignación de los datos.

	Conglomerados		
	1	2	3
Tipo de Vivienda	7	8	2
Paredes	6	5	2
Techo	4	7	4
Piso	3	3	2

Tabla 7. Valores Iniciales de los Conglomerados. Condición de vivienda

Centros Finales de los Conglomerado: Finalizado el proceso iterativo antes mencionado, se obtienen los nuevos centros que corresponden con el conjunto de valores donde no se consigue un nuevo desplazamiento de centros.

	Conglomerados		
	1	2	3
Tipo de Vivienda	6,12	8	2,12
Paredes	5,82	6	1,18
Techo	4,04	7	2,73
Piso	2,56	4	1,65

Tabla 8. Valores Finales de los Conglomerados. Condición de vivienda

De acuerdo a las tablas anteriores el Conglomerado 1 define la condición de la vivienda Regular. Las características asociadas son: Tipo de vivienda: Rancho Campesino, Vivienda Rustica; Paredes: Adobe, tapia, Bahareque frisado y sin Frisar; Techo: Laminas Metálicas de zinc; Piso: Cemento.

El Conglomerado 2 define las viviendas en Malas condiciones. Con las siguientes características asociadas: Tipo de vivienda: Vivienda Colectiva u otro tipo de Vivienda; Paredes: Caña, Palma, Tablas; Techo: Palma, Tablas y similares; Piso: Tierra, Otros.

El conglomerado 3 define que la condición de la vivienda es Buena. Con las siguientes características: Tipo de vivienda: Quinta, Casa, Apartamento; Paredes: Bloque o ladrillo frisado; Techo: Tejas; Piso: Mosaico, Granito, vinil, cerámica, ladrillo, terracota, parquet, alfombra y similares.

Como se puede observar, a pesar de la buena aproximación del Análisis de Correspondencia Múltiples para la asociación de las variables, la técnica de conglomerados mostró ser más poderosa a los efectos de preprocesar la nueva variable.

b) Adecuación de vivienda

Tal como se explica en la página 13 la adecuación viene determinada en parte por características tales como la disponibilidad de servicios, materiales empleados, infraestructura disponible, requerimientos de acceso permanente a recursos naturales y a los servicios básicos como agua potable, luz, aseo. Asociadas a estas características, se debe contemplar la necesidad de contar con un nivel mínimo de salubridad y de instalaciones sanitarias, esta se vincula con la eliminación de excretas. Adicionalmente, se consideran igualmente importantes para la adecuación de la vivienda considerar los factores sociales, económicos, culturales, climatológicos, ecológicos (Bedoya 2006).

Se crea la nueva variable Adecuación de la vivienda (A), de tipo categórico. Evaluada de acuerdo a adecuada ó inadecuada. Para llevar a efecto esta transformación en este nuevo patrón, se describen las variables con las que se dispone para la construcción de la variable antes mencionada: Servicio de agua: Acueducto (1), Pila Pública (2), Camión (3), Otros medios (4); Servicio eléctrico: Si Tiene (1) o No Tiene (2); Recolección directa de basura: Si tiene (1) o No Tiene (2); Container de basura: Si Tiene (1) o No Tiene (2); Eliminación de excretas: Poceta a Cloaca (1), Poceta a Pozo séptico (2), Excusado a Hoyo o Letrina (3), No tiene poceta o excusado (4).

De acuerdo al análisis de conglomerados que se aplicó a los datos de las EHM los grupos que se obtuvieron se muestran en las siguientes tablas.

Centros Iniciales de los Conglomerado: Representan los valores de las variables de clasificación Servicio de Agua; Servicio de eliminación de Excretas; Servicio Eléctrico; Recolección de Basura; Container de Basura. El primer conglomerado

corresponde a las viviendas cuyo servicio de agua es por Pila pública; el servicio de eliminación de excretas es de poceta a cloaca; cuentan con servicio eléctrico; con recolección de basura periódicamente; y que no cuentan con un container de basura. El segundo conglomerado corresponde a las viviendas cuyo servicio de agua es por camión; el servicio de eliminación de excretas es de excusado a hoyo o letrina; no cuentan con servicio eléctrico; no hay recolección de basura periódicamente; y cuentan con un container de basura. Estos son los valores utilizados como centros de partida asignando cada observación al conglomerado que le corresponde, y de esta manera iniciar el proceso iterativo de ubicación de nuevos centros y resignación de los datos.

	Conglomerados	
	1	2
Serv.Agua	2	3
ServExcretas	1	4
Serv. Elec.	1	2
RecolBasura	1	2
ContainerBas	2	1

Tabla 9. Valores Iniciales en el Análisis de Conglomerados. Adecuación de la vivienda

Centros Finales de los Conglomerados: Finalizado el proceso iterativo antes mencionado, se obtienen los nuevos centros que corresponden con el conjunto de valores donde no se consigue un nuevo desplazamiento de centros.

Con los resultados obtenidos en el análisis de conglomerados se puede clasificar las viviendas de acuerdo a las características particulares de los grupos formados. A continuación se muestra cómo queda definido cada conglomerado de acuerdo a los resultados mostrados en las tablas 9 y 10

	Conglomerados	
	1	2
Serv.Agua	1	4
ServExcretas	2	3
Serv. Elect	1	1
RecolBasura	2	2
ContainerBas	1	2

Tabla 10. Valores finales en el Análisis de Conglomerados. Adecuación de la vivienda

La vivienda es considerada Adecuada cuando se caracteriza de acuerdo al conglomerado 1 por: Servicio de agua a través de Acueducto, Pila Pública, tiene servicio Eléctrico, Recolección de Basura o Container de basura y la eliminación de excretas es de Poceta a cloaca, Poceta a Pozo Séptico. La vivienda es Adecuada.

Finalmente es inadecuada la vivienda cuando se caracteriza por: Servicio de agua a través de Camión u otros medios, tiene o no servicio Eléctrico, No tiene Recolección de Basura o Container de basura y la eliminación de excretas es por Excusado a Hoyo o Letrina, No tiene poceta o excusado. La vivienda es Inadecuada.

De este modo la nueva variable Adecuación resulta del agregado de las variables originales Servicio de agua, Servicio eléctrico, Recolección directa de basura, Container de basura, Eliminación de excretas.

3.4.4 Variable de salida

A fin de tener respuesta de la condición deficitaria de las viviendas, y debido a que los datos que aportan las EHM no disponen de información observada para que sea reconocida como una variable de salida para los modelos de pronóstico/clasificación, nuevamente se aplica el Análisis Multivariante a las variables independientes, para construir la respuesta a emplear como variable dependiente.

Nuevamente se aplica el Análisis de Correspondencia Múltiple, que debido a sus propiedades permite observar como lo muestra el gráfico 2 una tendencia clara, que permitió definir la variable dependiente de la siguiente manera y que se describe a continuación.

En el Gráfico 2, la zona con categorías (1, 5, 7, 9,11) de las variables consideradas (Número de Hogares, Tenencia, Hacinamiento, Adecuación y Condición), no registra ningún cambio en el comportamiento de ellas y se induce que define en la nueva variable de salida que sea calificadas como no deficitaria. Esto debido a la información que contienen estos campos la cual describe aquellas viviendas en las que: El número de hogares que habitan la vivienda es uno sólo, la tenencia de la vivienda es Propia, no hay hacinamiento, y la adecuación y condición de la vivienda es adecuada y buena respectivamente. Por otro lado, como puede observarse estas condiciones se cumplen todas a la vez; en consecuencia, en ausencia de alguna de ellas la observación será calificada como vivienda deficitaria, tal como lo describe los valores restantes que se observan (2,3,4,6,8,10,12) en el gráfico y que corresponden con la información de aquellas viviendas en las que habitan dos, tres, cuatro o más núcleos familiares, donde la tenencia es alquilada, existe hacinamiento, y la adecuación y condición es inadecuada y mala respectivamente.

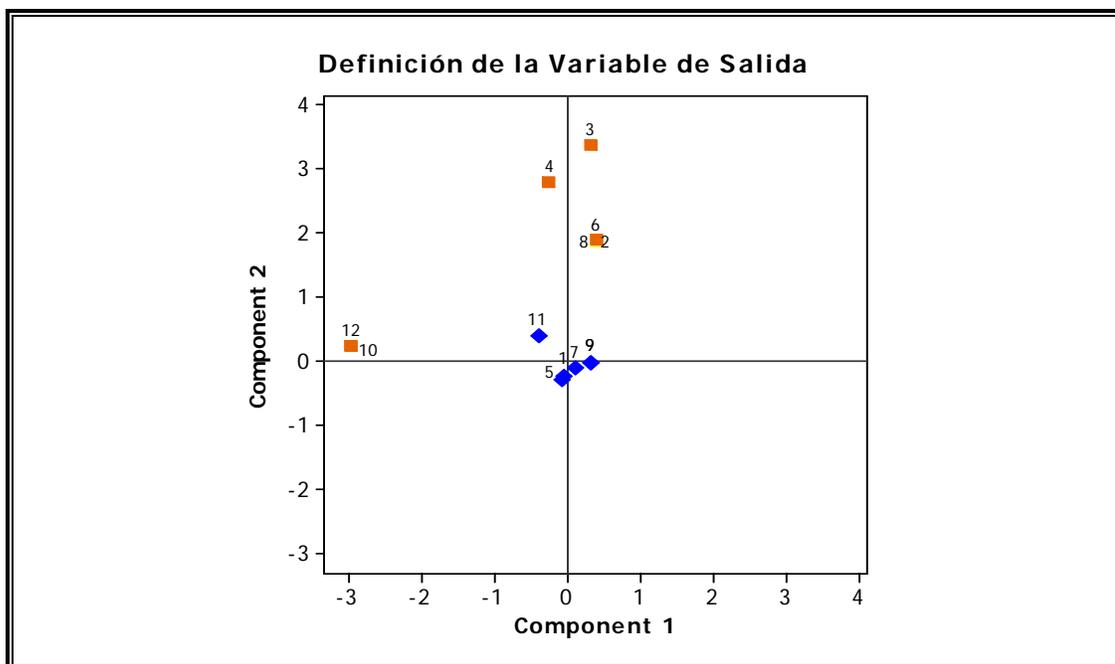


Gráfico. 2 ACM para definición de la Salida

De acuerdo al gráfico anterior, la variable de salida queda definida según la descripción que se da en la tabla siguiente, donde se especifican las categorías que se generaron y que serán asignadas a cada observación según sea el caso.

VARIABLE DE SALIDA	DESCRIPCIÓN
<i>Vivienda No Deficitaria:</i> Representada en el conjunto de datos con cero. Incluye todas las observaciones que cumplan con todas y cada una de las condiciones para considerar una vivienda adecuada, señaladas en los apartados anteriores.	Viviendas en las que número de hogares que habitan la vivienda es sólo uno, la tenencia de la vivienda es Propia, no hay hacinamiento, y la adecuación y condición de la vivienda es adecuada y buena respectivamente, para este caso se deben cumplir todas y cada una.
<i>Vivienda Deficitaria:</i> Dentro del conjunto de datos registrada con un uno; todas las observaciones que no cumplan con al menos una de las condiciones de adecuación que han sido señaladas en los apartados anteriores.	Número de Hogares dos, tres, cuatro o más familias; tenencia alquilada, hay hacinamiento, adecuación inadecuada y condición mala.

Tabla 11. Descripción de la Variable de salida

3.5 Conformación de los bloques de datos 1994-2001 y 2001-2005

Las EHM ha sufrido modificaciones en cuanto al registro y el tamaño de la muestra que ha particularizado el desarrollo de esta investigación. No sólo se trata de variables que no han sido incluidas hasta un año en particular, sino que durante varios períodos subsiguientes se mantiene (segundo semestre del año 1994 hasta el primer semestre del año 2001), y se incrementa el tamaño de la muestra a partir de este momento hasta el 2005. Dadas estas circunstancias, se puede tener sospecha que un estudio bajo estas características, podría estar influenciado por la masa creciente de datos presentes en el último período antes mencionado, y que a su vez reducir estas observaciones a un número acorde con las muestras que se observan en el primer período originaría una pérdida de información. Por lo que en adelante se hará referencia a dos bloques o períodos mencionados como 94_2-01_1 y 01_2-05_2, correspondientes al segundo semestre de 1994 hasta el primer semestre de 2001 y el segundo semestre del 2001 hasta el segundo semestre de 2005.

CAPITULO IV

Experimentación y resultados

En el presente capítulo, se presentan los experimentos para la construcción del modelo de redes con Funciones de Bases Radiales para el pronóstico/clasificación del déficit de viviendas. Se construye un modelo para cada período bajo estudio (94_2-01_1, 01_2-05_2). Se describen los resultados del entrenamiento/prueba y generalización. Adicionalmente, se realizan estimaciones respecto a cada conjunto de datos para observar la variabilidad deficitaria e inferir con base en estos resultados sobre la aplicabilidad del modelo. Es necesario destacar que en la construcción del modelo se empleó como herramienta computacional el NeuroSolutions (NS)⁵. Cabe destacar que la mención de los comandos del NS, empleados para la construcción del modelo serán resaltados con letra cursiva.

4.1 Modelo de la red con bases radiales (RBF)

Al considerar el hecho que las RBF poseen una arquitectura muy simple, no puede desconocerse la potencialidad que poseen en el proceso de reconocimiento de nuevos patrones. Su topología es bien influenciada por el número de patrones a

⁵ El NeuroSolutions es un software empleado para el diseño, formación y despliegue de redes neuronales con aprendizaje supervisado y no supervisado. Pueden construirse modelos para realizar una amplia variedad de tareas como la minería de datos, clasificación, función de aproximación, regresiones con variables múltiples y predicción de series temporales. La versión empleada es 5.0 de uso libre disponible bajo versión beta en la WEB.

emplear durante el entrenamiento, específicamente en su capa oculta. En este sentido, durante el entrenamiento correspondiente a ambos conjuntos de datos (94_2-01_1, 01_2-05_2), se observan características particulares producto de la diferencia existente en el tamaño de las muestras.

4.1.1 Red neuronal RBF: pronóstico/clasificación del déficit de vivienda.

En este apartado se describe el procedimiento a desarrollar para la construcción del modelo con RBF, considerando cada uno de los parámetros requeridos por este tipo de red que son descritos a continuación.

- **Definición de las entradas.** Corresponde a las variables que se obtuvieron en el preprocesamiento y descritas en el capítulo anterior. A continuación, se definen cada una de las variables con sus respectivos valores de las categorías.

Num_Hogares	Tenencia	Hacinamiento	Adecuación	Condición
[1,2,3,4]	[0,1]	[0,1]	[0,1]	[0,1]

- **Definición de la salida.** Para la RNA se define una sola variable de salida la cual puede tomar las categorías enumeradas [0,1] indicando cada una, estado deficitario o no de la vivienda: Déficit= 1; No Déficit = 0.

- **Selección de los patrones de entrenamiento.** Con la información recopilada del conjunto de datos a emplear para el estudio, se procede a seleccionar los patrones de entrenamiento y de prueba. Para la selección de las observaciones para los modelos RBF se hace la selección mediante la técnica de muestreo aleatorio estratificado. Se tomaron diez muestras estratificadas de cada bloque de datos para el entrenamiento⁶. Los tamaños fueron de trescientos setenta y ocho observaciones correspondientes al período 94_2-01_2 y novecientos setenta y ocho para el período 01_2-05_2. El resto de las observaciones fueron empleadas para las pruebas de los modelos.

⁶ Para ver en detalle el algoritmo, refiérase al Anexo 3

- **Configuración de las capas del modelo RBF.** Debido a los fundamentos teóricos que señalan que este tipo de redes está constituido por una sola capa oculta, sólo es necesario configurar el número de centros de los nodos en la capa oculta del modelo. Así mismo, tomando en cuenta que el objetivo del estudio es realizar el pronóstico/clasificación del déficit de viviendas, el número de salidas es una sola. Por lo tanto, sólo debe configurarse la función de activación en la capa de salida a ser empleada en el modelo y que por especificaciones también teóricas se trata de una función lineal.
- **Fases de entrenamiento/prueba y validación de la red.** Con los patrones seleccionados anteriormente se procede a entrenar la red. Recordando que se busca la mejor red, se toman un total de diez redes las cuales muestran el mejor desempeño. Una vez entrenada cada red se procede a la validación de la mejor red entrenada con el conjunto de datos no usado por la fase de entrenamiento.

4.1.2 Parámetros empleados para los modelos RBF del período 94-01_1 y 01_2-05

A continuación se muestran los pasos que se siguieron para la construcción del modelo con NeuroSolutions. Como se ha mencionado anteriormente, se tomaron dos bloques de datos para los cuales se lleva a cabo las mismas instrucciones requeridas por el NS en sus primeras fases; es decir, la selección del modelo, los patrones de entrenamiento relacionadas a las variables de entrada y salida del modelo. Los patrones de entrenamiento que se introducen son las muestras seleccionadas por estratificación para cada bloque de datos. En la tabla 12 se da a conocer la estructura de los datos con una pequeña submuestra para ambos períodos, recordando que se tomaron 10 muestras para cada uno de ellos.

En la tabla siguiente, los campos identificados con I, II, III, IV, V, OUT_Y⁴, corresponden a cada una de las variables que se describieron en los apartados anteriores. Ellas son Num_Hogares, Tenencia, Hacinamiento, Adecuación, Condición y la variable de salida respectivamente.

⁴ Se emplea esta nomenclatura para dar referencia a las variables de entrada y salida descritas en los apartados anteriores

Periodo 94_2-01_1						Periodo 01_2-05_2					
I	II	III	IV	V	OUT_y	I	II	III	IV	V	OUT_y
1	1	0	0	0	1	1	1	1	0	0	1
1	1	1	0	0	1	1	1	0	1	1	1
1	0	1	1	1	1	1	1	1	0	0	1
1	0	1	0	0	1	1	1	1	0	0	1
1	0	1	0	0	1	1	0	1	1	1	1
1	0	0	0	0	0	1	0	0	1	1	1
1	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0
Total datos					378	Total datos					978

Tabla 12. Ejemplos de los patrones de entrenamiento⁷

En la figura 10 se observa la topología de la red. La arquitectura consta de cinco nodos de entrada, nodos ocultos o radiales que varían de acuerdo al nivel de agrupación y un nodo de salida. De acuerdo a las características de las redes RBF, estas siguen una topología de las *redes en cascada* por cuanto están conectadas unidireccionalmente.

La cantidad de patrones de entrenamiento para cada bloque de datos correspondientes a los períodos 94-2_01-1 y 01_2-05_1 es de trescientos setenta y ocho, y de novecientos setenta y ocho respectivamente. Considerando que el número de nodos radiales crece en función del número de ejemplos en la muestra, la formación de centros (más patrones implican más nodos radiales) crecerá en función de estos nodos. En este sentido, se observan varias configuraciones para el entrenamiento correspondiente a cada periodo. En cuanto a la Regla de Competencia, que tiene como función controlar la frecuencia con que es ubicada una observación en un centro particular, evitando así la incidencia de un número muy grande de observaciones en algún nodo y muy poco en otros, a través del comando *ConscienceFull*, y se utiliza la medida euclidiana para determinar la

⁷ La estructura de los patrones de Prueba es la misma, la diferencia radica en el tamaño de la muestra.

distancia radial entre el vector de entrada y el centro de gravedad de ese mismo nodo, para cada una de las N observaciones.

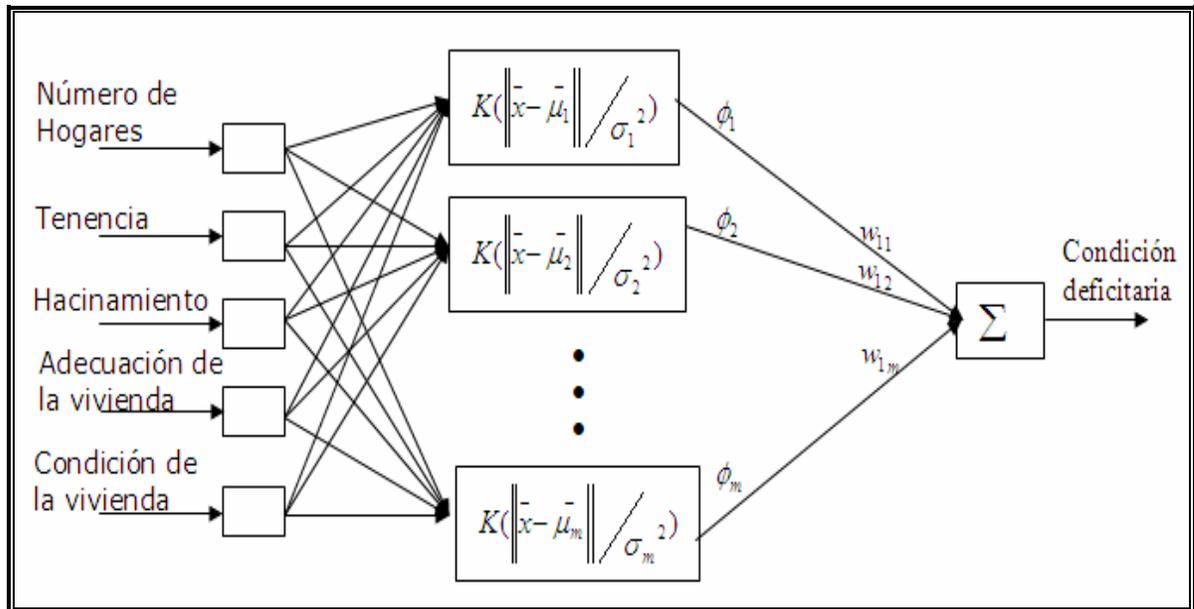


Figura 10. RBF Construida

Para el modelo construido, se definen cinco variables de entrada y m corresponde a los centros o nodos en la capa oculta. Para el caso bajo estudio se emplearon diferentes valores para m en cada modelo. En la capa de salida, la respuesta será la clasificación del vector de entrada, y así de las N observaciones.

Una vez conseguida la configuración de la capa oculta sólo queda configurar la capa de salida. Para ello se utiliza la función de activación *LinealAxon* debido a que admite cierta flexibilidad en la transformación que se genera sobre los datos que entran a esta capa. Es decir la salida de los nodos radiales. Para ambos grupos de datos se mantiene la misma regla de aprendizaje, y el valor de los parámetros de la regla que se muestran en la tabla 13.

Función de Activación	<i>LinealAxon</i>
Regla de Aprendizaje	<i>Momentum</i>
Tamaño del Paso	1.0000
Variación de los Pesos	0.7000

Tabla 13. Configuración de la Capa de Salida

Las redes RBF tienen un aprendizaje híbrido. En su primera fase el aprendizaje es no supervisado y es desarrollado en la capa oculta o capa de nodos radiales. En lo sucesivo se describe la configuración de esta capa que se usó durante los entrenamientos para el aprendizaje no supervisado. En esta fase del modelo se indica el número de ciclos o *epochs* que se requieren para el entrenamiento del proceso no lineal realizado en la capa de nodos radiales. Para ambos bloques de datos se utilizaron 100 ciclos para el aprendizaje no supervisado. Para el modelo de ambos bloques de datos (94-2_01-1; 01_2-05), de acuerdo a NS la tasa de aprendizaje se activa en 0.01 hasta llegar a 0.001.

Tasa de Aprendizaje			
Periodo	START	DECAY	CICLOS
94-01_1	0.01	0.001	100
01_2-05	0.01	0.001	100

Tabla 14 Configuración para el Aprendizaje no Supervisado para ambos períodos

De igual modo que para el aprendizaje no supervisado, también se configura lo que corresponde con el supervisado. En este caso, se requiere conocer el número de ciclos requeridos como parámetro de parada; en este mismo sentido, el MSE (error cuadrático medio) se usa como factor de culminación del entrenamiento.

En ambos modelos se especifica el mismo número de *epochs*; para esta fase se fijó este parámetro en 1000 y el factor de terminación empleado del MSE teniendo como umbral un valor de 0.00001.

Error Teórico(94_01-2)	Error Teórico(01_05-2)	Ciclos
0,00001	0,00001	1000

Tabla 15. Parámetros del entrenamiento supervisado

Una vez asignados los valores a los parámetros correspondientes, *NeuroSolutions* ofrece una amplia gama de opciones para mostrar diferentes gráficos para observar la respuesta que ofrece cada uno de los modelos construidos. Una vez cumplidas todas las fases de construcción de cada modelo se puede ver el ambiente que muestra *NeuroSolutions* con la arquitectura de la red que se muestra a manera de ejemplo en la figura 17. Entiéndase que para ambos periodos se observa las mismas condiciones de entrada y salida pero con diferencias topológicas.

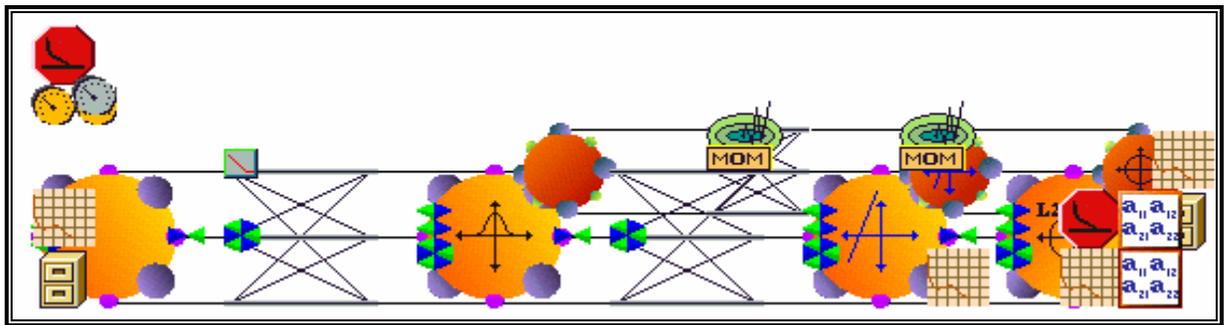


Figura 11. Ambiente del *NeuroSolutions*, de la red con RBF construida

4.2 Entrenamiento de las RBF.

Se tienen dos bloques de datos; el bloque 94_2-01_1 está formado por un total de trescientos setenta y ocho patrones para el entrenamiento y cinco mil ciento cuarenta y cuatro datos para la validación de la red y el bloque de 01_2-05_2 cuenta con un total de novecientos setenta y ocho para el entrenamiento y de catorce mil quinientos cincuenta y cuatro para la prueba. Se construyen diez modelos con diferentes características topológicas, para cada periodo y así escoger entre ellos el que mejor aproxima los datos. En total son veinte redes RBF para escoger las dos mejores, una para cada periodo.

A continuación se muestra en resumen las configuraciones empleadas para cada modelo; así mismo, su respuesta en la fase de entrenamiento.

Parámetros de la RBF						
Nº RED	Nº de entradas	Nº salidas	Nodos capa oculta	Nº obs.	Epochs propuesto	MSE propuesto
1	5	1	5	378	1100	0,0001
2	5	1	8	378	1100	0,0001
3	5	1	12	378	1100	0,0001
4	5	1	15	378	1100	0,0001
5	5	1	18	378	1100	0,0001
6	5	1	20	378	1100	0,0001
7	5	1	21	378	1100	0,0001
8	5	1	23	378	1100	0,0001
9	5	1	25	378	1100	0,0001
10	5	1	30	378	1100	0,0001

Tabla 16. Configuraciones de la RBF para el entrenamiento con el periodo 94_2-01_1

Entrenamiento				
Nº RED	MSE	RMSE	Epochs alcanzados	Orden de importancia
1	0,0001005	0,0001251	153	1
2	0,00018866	0,00023292	343	8
3	0,0001864	0,0002301	178	7
4	0,0001933	0,0002386	212	10
5	0,000178	0,00021975	225	3
6	0,00019093	0,00023572	277	9
7	0,00018532	0,0002288	283	5
8	0,0001861	0,00022983	218	6
9	0,000177587	0,00021924	221	2
10	0,00018347	0,00022651	230	4

Tabla 17. Resultados del entrenamiento para el periodo 94_2-01_1

En las tablas anteriores la fila sombreada especifica la red mejor entrenada para el periodo 94_2-01_1. Lo mismo sucede en las tablas que corresponden al periodo 01_2-05_2. Sin embargo, hasta el momento no se puede decidir cual será la mejor red, ya que es necesario efectuar la fase de prueba en la que se evalúa la capacidad de generalización de la red; es decir, la capacidad de realizar pronósticos con datos que la red no conoció durante la fase de entrenamiento.

A continuación se describen las configuraciones propuestas para el conjunto de datos correspondientes al período 01_2-05_2.

Parámetros de la RBF						
Nº RED	Nº de entradas	Nº salidas	Nodos capa oculta	Nº obs.	Epochs propuesto	MSE propuesto
1	5	1	40	978	1100	0,0001
2	5	1	43	978	1100	0,0001
3	5	1	45	978	1100	0,0001
4	5	1	48	978	1100	0,0001
5	5	1	50	978	1100	0,0001
6	5	1	53	978	1100	0,0001
7	5	1	55	978	1100	0,0001
8	5	1	58	978	1100	0,0001
9	5	1	60	978	1100	0,0001
10	5	1	65	978	1100	0,0001

Tabla 18. Configuraciones de la RBF para el entrenamiento con el periodo 01_2-05_2

Con las configuraciones descritas anteriormente, se logra observar el comportamiento de las redes en su fase de entrenamiento. Se muestra el resumen en la siguiente tabla.

Entrenamiento				
Nº RED	MSE	RMSE	Epochs alcanzados	Orden de importancia
1	0,00018293	0,0002432	234	6
2	0,0001974	0,000249	274	10
3	0,00019291	0,0002433	185	7
4	0,00017981	0,0002267	178	1
5	0,00019316	0,0002436	225	8
6	0,00019345	0,000244	245	9
7	0,00018363	0,00023251	294	2
8	0,00018913	0,0002385	179	4
9	0,00019033	0,00024	179	5
10	0,0001889	0,0002382	193	3

Tabla 19. Resultados del entrenamiento para el periodo 01_2-05_2

En términos generales puede observarse que las redes entrenadas para ambos periodos muestran un buen desempeño. En las siguientes figuras puede observarse

la actividad de las redes mejor entrenadas para cada periodo, con su respectivo error de entrenamiento. Ver gráficos desde la 3 a 6.

La siguiente figura representa el entrenamiento de la red que muestra el mejor entrenamiento para el período 94_2-01_1.

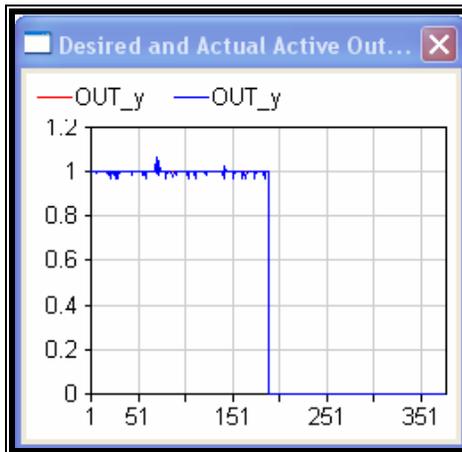


Gráfico 3. Salida observada-Salida deseada
Ambiente del NeuroSolutions

Como puede observarse en la figura anterior, la red clasifica correctamente cada uno de los patrones de entrenamiento y esto se refleja además en la matriz de confusión que se muestra en el experimento. En este sentido, puede observarse el grado de convergencia de la red al alcanzar rápidamente el error propuesto para la finalización del entrenamiento; y así mismo, puede expresarse que no necesita iterar para todos los *epochs* especificados en la configuración de la red. Para el caso del período 94_2-01_1 la mejor red de esta fase converge a los 153 *epochs*, y esta actividad puede observarse en la Figura 19.

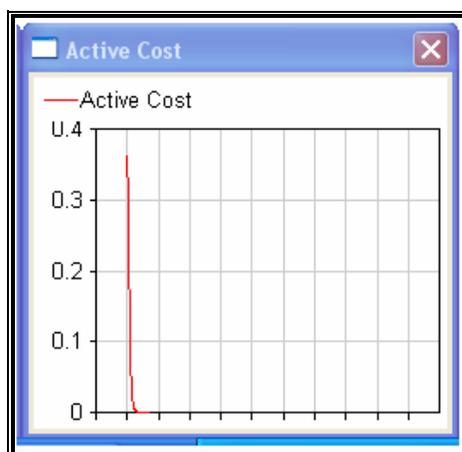


Gráfico 4. MSE alcanzado por la RBF Ambiente del NeuroSolutions

Seguidamente, se muestra la actividad de la red mejor entrenada para el periodo 01_2-05_2.

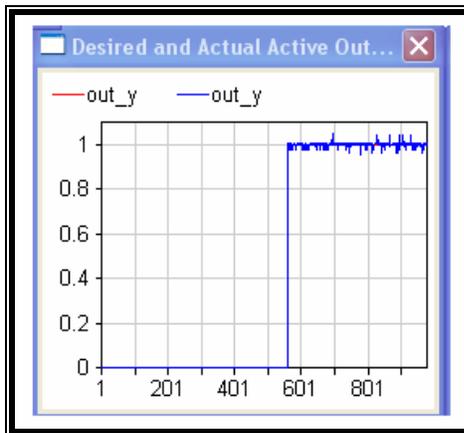


Gráfico 5. Salida observada-Salida deseada
Ambiente del NeuroSolutions

En la siguiente figura se observa una vertiginosa caída en el error cometido por esta red y es alcanzado en un número mínimo de *epochs*, específicamente para este caso de 178 *epochs*. Ver los resultados de la tabla 19.

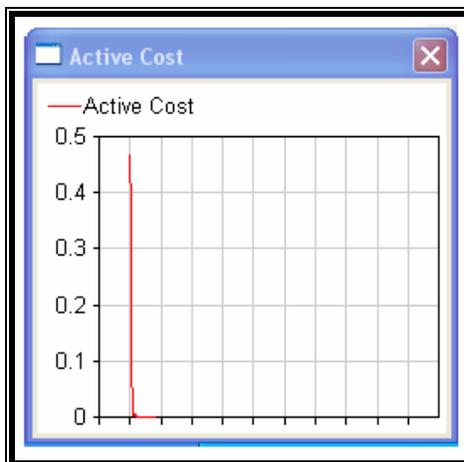


Gráfico 6. MSE alcanzado por la RBF
Ambiente del NeuroSolutions

4.3 Generalización de la Red

Una vez que las redes han logrado aprender la relación entrada-salida de los patrones utilizados durante el entrenamiento, se verifica la capacidad de generalización que estas poseen al utilizar un conjunto de datos no conocidos por

ellas. Las salidas obtenidas en esta fase son el pronóstico del fenómeno bajo estudio y determinan la capacidad de generalización de la red.

Para esta fase se toma en cuenta la muestra complementaria para probar la red, tomándose diferentes tamaños para efectuar cada prueba. En total se realizaron diez pruebas para cada red con tamaños de muestra totalmente aleatorios; ciento cincuenta, cien, cincuenta, veinticinco, y cinco patrones que no había conocido la red en la fase de entrenamiento. De esta manera, en las siguientes tablas se presentan los resultados obtenidos para la generalización de la mejor red, considerando no sólo el desempeño durante el entrenamiento, sino también observando la consistencia del modelo que se está probando al medir la raíz del error cuadrático medio MSE en cada prueba.

En la siguiente tabla se muestran los resultados de generalización de la red mejor entrenada, usando el conjunto de datos con los tamaños antes mencionados y que corresponden al periodo 94_2-01_1.

PRUEBA	RECM1 N=150	RECM2 N=100	RECM3 N=50	RECM4 N=25	RECM5 N= 5	VARIANZA
1	0,00133761	0,00237377	0,0144132	0,0099511	0,01931605	5,95E-05
2	0,0016341	0,00356647	0,01109312	0,00629567	0,01999269	5,38E-05
3	0,00116909	0,0095067	0,00795795	0,00599511	0,01931605	4,45E-05
4	0,00292453	0,0031962	0,00885315	0,00446081	0,0109265	1,30E-05
5	0,00353675	0,0052103	0,00660124	0,00793759	0,0232337	6,33E-05
6	0,00106113	8,25E-03	0,00512315	0,00532166	0,05661096	5,41E-04
7	0,00193813	0,00412589	0,00362957	0,00568948	0,07595649	1,04E-03
8	0,00112549	0,00856986	0,00899511	0,00256482	0,01352699	2,58E-05
9	0,00117815	0,00654213	0,01446081	0,04582135	0,02042711	3,02E-04
10	0,00126585	0,00213568	0,00254112	0,00152366	0,01485934	3,40E-05
VARIANZA	7,2705E-07	7,3368E-06	1,6883E-05	0,00016817	0,00045374	

Tabla 20. Generalización de la mejor red usando patrones no conocidos del periodo 94_2-01_1

En general, al observar la varianza registrada de los valores de RMSE entre las diferentes muestras para cada prueba, la variabilidad es mínima con la aplicación de diferentes patrones, y en consecuencia, considerando que se emplearon muestras de tamaños diferentes, se puede observar un alto nivel de consistencia en el modelo

y un alto nivel de desempeño que se puede apreciar en el siguiente gráfico, al generalizar el modelo escogido como mejor entrenado.

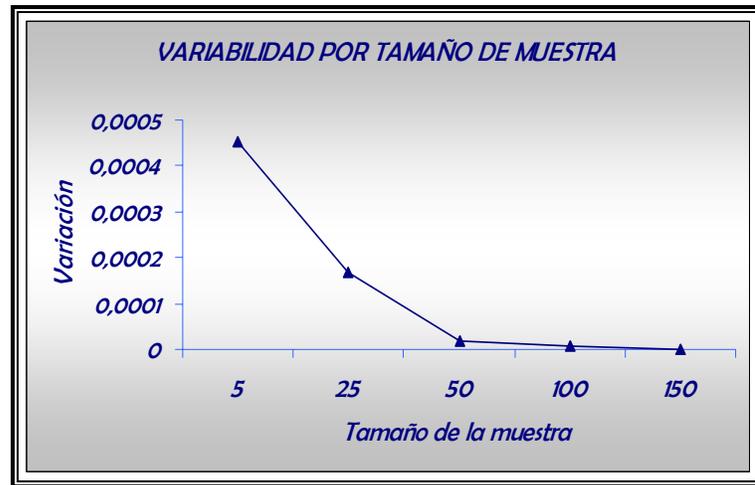


Gráfico 7. Resultado de la variabilidad en el desempeño de la red para diferentes tamaños de muestras del periodo 94_2-01_1

Del gráfico anterior puede resaltarse como la red muestra una mínima variabilidad en esta fase cuando el número de patrones es grande. Así mismo, se puede observar que para cada prueba el modelo muestra un nivel de consistencia aceptable.

En el siguiente gráfico se observa la variabilidad en las generalizaciones de este modelo después de haber realizado diez pruebas por cada tamaño de muestra. Es decir cincuenta pruebas.

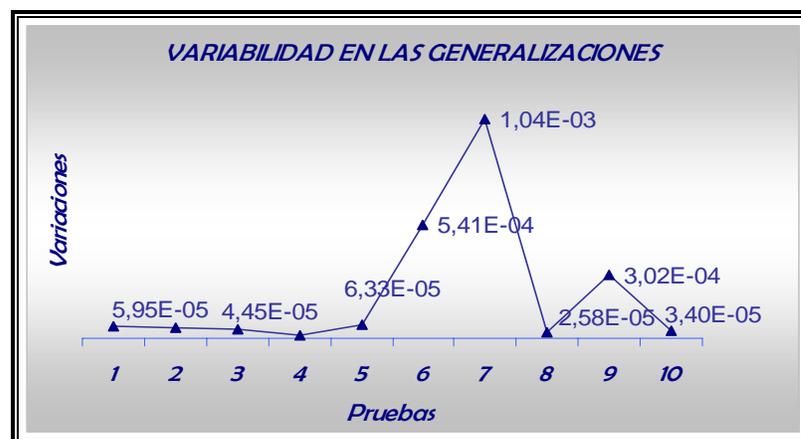


Gráfico 8. Variabilidad en la generalización del modelo para el período 94_2-01_1

Después de calcular para cada prueba la varianza, se observa en el gráfico anterior que en las pruebas seis, siete y nueve la red muestra una mayor variación respecto a las otras. Sin embargo, al observar los valores bastante bajos, no varían demasiado, por lo que puede asegurarse que si a la red se le aplican un mayor número de pruebas con este modelo, los resultados serían similares y en consecuencia, podría hacerse referencia a esta red como un modelo robusto.

A continuación se muestran los resultados de generalización de las diez pruebas a la red mejor entrenada para el periodo 01_2-05_2. Del mismo modo, se observan los resultados referidos a la actividad de la red para los distintos tamaños de muestra que se emplearon en esta fase de la experimentación.

PRUEBA	RECM1 N=150	RECM2 N=100	RECM3 N=50	RECM4 N=25	RECM5 N= 5	VARIANZA
1	0,003140754	0,0052719	0,00110764	0,0489608	0,0847134	1,38E-03
2	0,001093556	0,00390021	0,0019424	0,02404012	0,01395014	9,73E-05
3	0,00386561	0,0022597	0,00106443	0,0673843	0,02814114	8,11E-04
4	0,00268964	0,00346457	0,00209372	0,0175277	0,05210346	4,58E-04
5	0,001023541	0,001846457	0,02093721	0,01210346	4,52E-02	3,30E-04
6	0,001245343	0,005938557	0,00108864	0,01925602	0,01236546	6,08E-05
7	0,003325896	0,005658807	0,02078674	0,03811283	0,08564253	1,14E-03
8	0,00236259	0,001343351	0,00107596	0,03127277	0,06102355	7,06E-04
9	0,001856426	0,006659783	0,00210371	0,03252378	0,07452619	9,74E-04
10	0,002235469	0,001209324	0,00514739	0,05567296	0,01856943	5,26E-04
VARIANZA	9,7352E-07	4,15684E-06	6,504E-05	0,00032192	0,00081677	

Tabla 21. Generalización de la mejor red usando patrones no conocidos del periodo 01_2-05_2

De acuerdo a la tabla anterior se observa un nivel de variabilidad más homogéneo respecto al periodo anterior. Así mismo, se observa que se mantiene el comportamiento de crecimiento de la variabilidad del modelo cuando el número de patrones es reducido. Para mejor detalle de este aspecto se construye el siguiente gráfico correspondiente a las pruebas realizadas en la mejor red entrenada para el periodo 01_2-05_2.

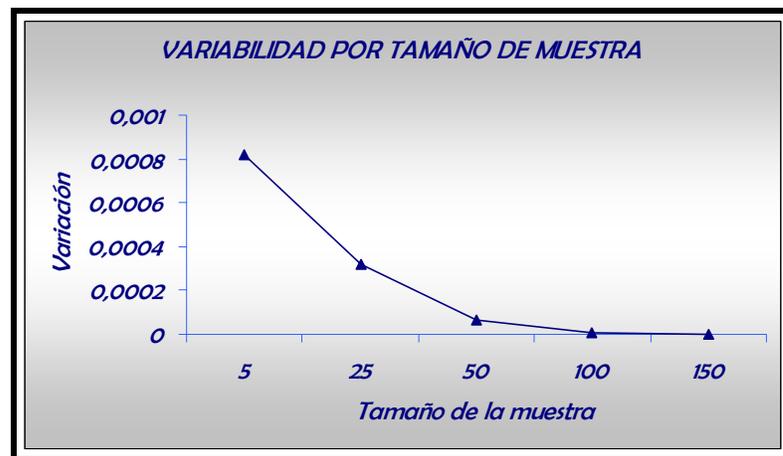


Gráfico 9. Resultado de la variabilidad en el desempeño de la red para diferentes tamaños de muestras del periodo 01_2-05_2.

Adicionalmente, puede observarse como el conjunto de pruebas no varía la varianza en mayor grado y se mantiene en un rango de variabilidad que al igual que el modelo anterior permite considerar que es un modelo consistente en cuanto a la mínima varianza que ofrece en sus pruebas y que a su vez genera robustez si estas pruebas tienden a infinito.

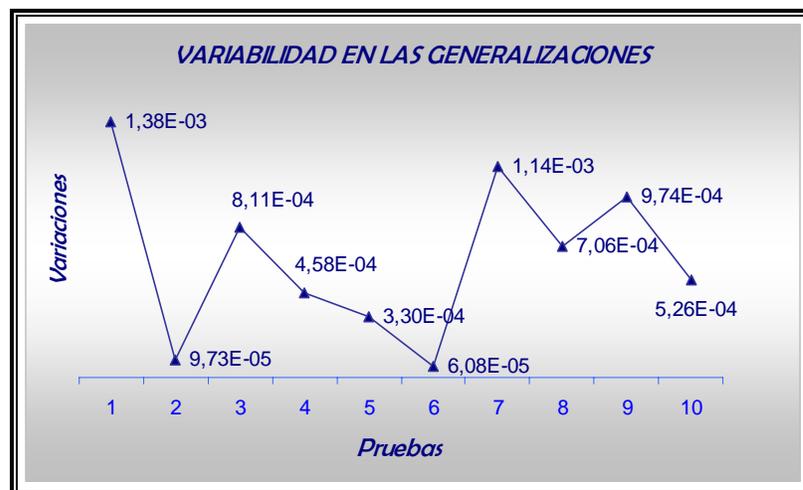


Gráfico 10. Variabilidad en la generalización del modelo para el período 01_2-05_2

En resumen, puede decirse que los modelos construidos para su periodo correspondiente demuestran un buen desempeño en ambas fases de la experimentación. Con lo anterior, se puede concluir que los modelos de red mejor entrenados, también puede considerarse que alcanza buenos niveles de

generalización para el conjunto de datos que no han sido conocidos por la red. Así mismo, estos modelos pueden ser considerados como modelos robustos.

4.4 Inferencias sobre el déficit de viviendas para cada período estudiado.

Las redes al ser comparadas con la herramientas de Análisis Multivariante empleadas, permiten observar claramente el comportamiento de los patrones obtenidos, pudiéndose destacar en ambas técnicas que los índices más elevados los aporta el conjunto de patrones referidos a las viviendas deficitarias. De ahí que sea oportuno aprovechar estos resultados para inferir unos nuevos y analizar como se ve afectada la población bajo estudio por el efecto deficitario de viviendas.

Partiendo del hecho que las observaciones son calificadas como deficitarias dada la ausencia de al menos una de las condiciones establecidas por los distintos entes encargados de velar por el bienestar de los individuos, y así corroborados por las técnicas aquí aplicadas, es concluyente verificar cuál de estas condiciones origina la inclusión de las observaciones al índice deficitario.

Por esta razón se evalúan los resultados obtenidos del análisis antes mencionado y que fuera realizado para ambos bloques de datos. Este análisis permite realizar comparaciones del índice deficitario para el estado Mérida y observar cual de los indicadores utilizados para evaluar el déficit tiene mayor incidencia.

En el siguiente gráfico se muestra la condición deficitaria en el estado Mérida durante el periodo 94_2-01_1, notándose que no hay una diferencia muy notoria entre las viviendas que se encuentran dentro de los requerimientos no deficitarios (53%) y aquellos que no satisfacen alguno de los indicadores y en consecuencia, son deficitarios (47%).

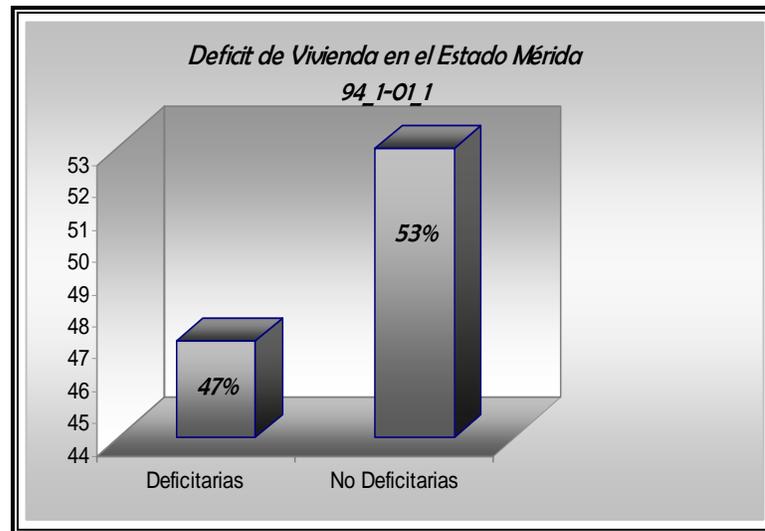


Gráfico 11. Condición deficitaria en el Estado Mérida período 94_2-01_1

De los resultados anteriores se puede analizar específicamente el índice de las viviendas deficitarias. El gráfico 12 permite indagar en este sentido, ya que proporciona los subíndices de cada uno de los indicadores deficitarios. Si se examina el conjunto de datos que incide en el índice deficitario, se obtienen los niveles que alcanza cada indicador. Esto se logra determinando la frecuencia de cuál factor (Número de Hogares, Tenencia, Hacinamiento, Adecuación y Condición de la Vivienda) se le asigna de acuerdo a su categoría al incluir a cada observación como deficitaria, siempre teniendo en cuenta que se considera una categoría por vez. De ahí que los resultados fueron los que se muestran en el gráfico 12 para la población analizada en el periodo 94_2-01_1. Es notoria la incidencia que tiene el indicador Hacinamiento en el índice del déficit de viviendas: 43% de las viviendas para este periodo no cuentan con el espacio requerido para albergar a sus habitantes. Por otro lado, también puede notarse el peso que tiene el indicador Tenencia ya que en un 30% de las viviendas que se encuentran en condición deficitaria se debe a que los habitantes viven en condición de Inquilinos. Y aunque este indicador, en otros estudios no es considerado como un aspecto relevante debido a que se está garantizando un techo para vivir, en este caso se le considera con la misma relevancia que los demás indicadores. Bajo estas condiciones las familias no tienen garantizado el bienestar de tener una vivienda propia, sin riesgos a tener que buscar otro sitio para reubicarse.

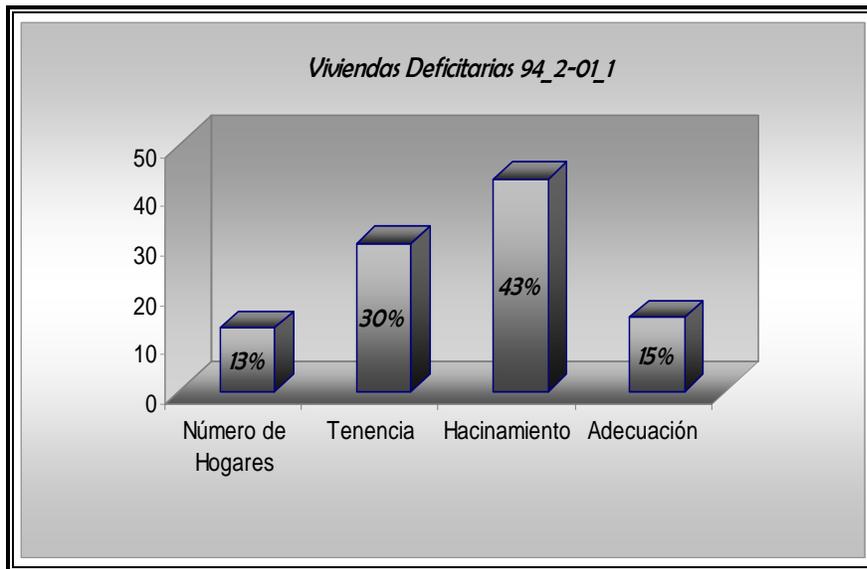


Gráfico 12. Incidencia de Indicadores en las viviendas deficitarias

Por otro lado, el indicador de Adecuación de la Vivienda reúne un 15% del índice deficitario indicando así que esa porción de las viviendas no poseen los servicios básicos, sin que se pueda determinar exactamente cual de ellos esté ausente. Se puede afirmar que es de los servicios indispensables para garantizar el bienestar e higiene en la vivienda. Por último y no menos importante, está el indicador de Número de Hogares, el cual señala que un 13% de las viviendas deficitarias están habitadas por más de un núcleo familiar, aun cuando no se tenga información de la capacidad de la vivienda, ya que puede tratarse de una vivienda que pueda albergar más de un hogar. Se pone de manifiesto el derecho que tiene cada familia de contar con su vivienda.

A continuación se explican los resultados observados en el período 01-2_05-2, en el gráfico 13 se observa como se distribuye la condición deficitaria en el estado Mérida durante este periodo, notándose que no hay una diferencia muy notoria entre las viviendas que se encuentran dentro de los requerimientos establecidos como no deficitarios y aquellas que lo incumplen por cualquiera de los indicadores, y en consecuencia son deficitarias. En este período puede observarse una mínima disminución en el índice deficitario, en razón de un 4% se redujo el nivel deficitario.



Gráfico 13. Condición Deficitaria en el Estado Mérida periodo 01_2-05_2

Podría sospecharse al comparar estos porcentajes que pudo haber una migración de viviendas en condición deficitaria a no deficitaria en un orden del 4 % debido a la corrección del tamaño de la muestra y a la inclusión de nuevas variables para categorizar la condición deficitaria. Sin embargo, esto podría ser una especulación y de ahí que se prosiga a realizar el análisis del índice deficitario para este período. En el Gráfico 14 puede observarse que en el indicador Hacinamiento recae la mayor porción de viviendas que son calificadas como deficitarias, recogiendo el 41,86%. En este periodo varían las proporciones en relación al periodo anterior. El indicador que sigue como mayor índice, es el Número de Hogares, con un 26%; lo que quiere decir que para este período se duplicó el índice anterior. En relación al indicador de Tenencia disminuyó en un 9% el número de familias que vive en condición de inquilinato al ser comparado con el porcentaje del período anterior. También se observa que el indicador de Adecuación de la Vivienda muestra una disminución de un 3% de las viviendas que no contaban con los servicios básicos indispensables para una mejor vida en el hogar.

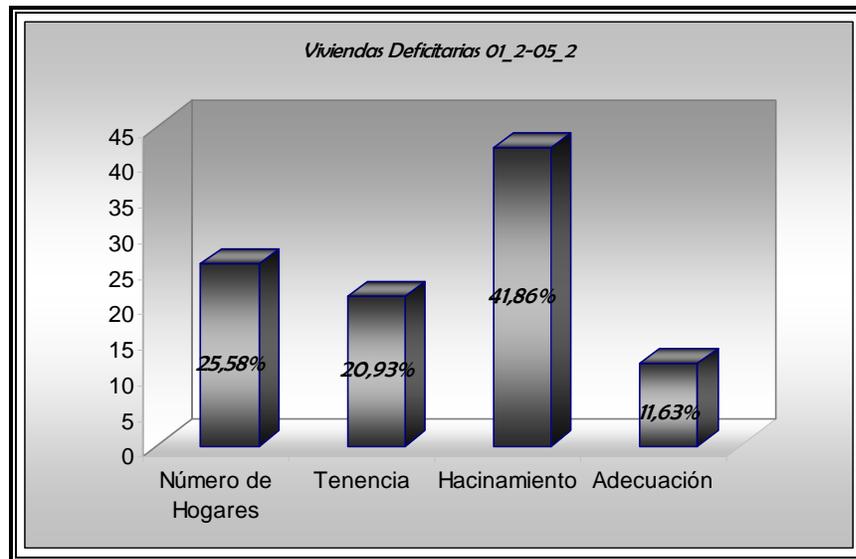


Gráfico 14. Incidencia de Indicadores en las viviendas deficitarias (2)

La información del análisis anterior, se puede observar recopilada en el Gráfico 15, donde se aprecian las variaciones que se suscitaron entre los dos periodos.

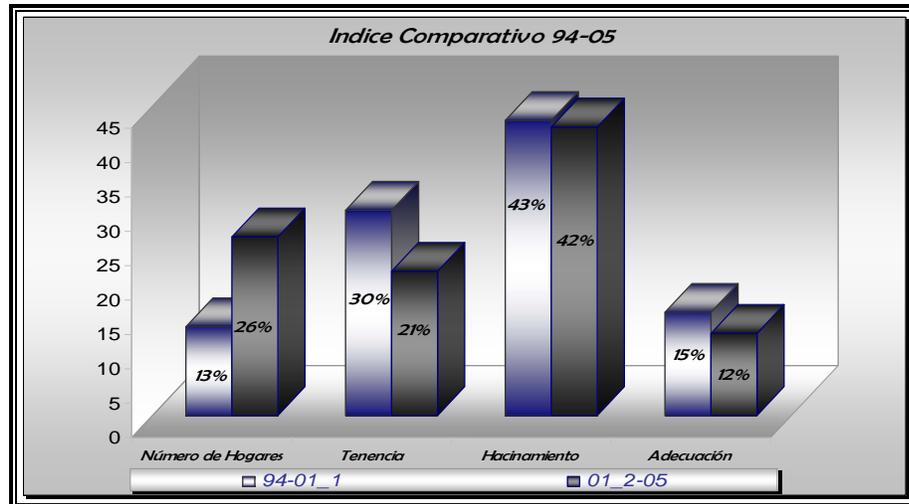


Gráfico 15. Resumen de variaciones en los indicadores de los periodos bajo estudio

Con este gráfico puede observarse claramente las variaciones que se suscitan entre los dos bloques de datos, desprendiéndose de ello, que para el último periodo se observa un incremento del 13% en lo que se ha definido como el Déficit Cuantitativo referenciado como la diferencia entre el número de viviendas existentes y las necesarias. Se observa que para el último periodo hay mayor

número de viviendas que alberga más de un hogar o núcleo familiar. Respecto a los demás indicadores que describen el déficit cualitativo se observa una disminución respecto al periodo anterior, Tenencia con un 21% disminuyó un 9%, Hacinamiento con 42% redujo su incidencia en 1% y finalmente Adecuación en 12% reduciendo su índice respecto al período anterior en 3%.

CAPITULO V

Conclusiones y recomendaciones

En este capítulo se presentan las conclusiones inherentes al estudio y las recomendaciones para dar continuidad a nuevas investigaciones.

5.1 Conclusiones

- El uso de las técnicas de preprocesamiento de datos aplicadas en el estudio, demostraron ser una herramienta útil para lograr organizar de manera coherente toda la información a ser empleada en el estudio. Además, es de resaltar que la utilización de estas técnicas permitió construir la variable de salida, que era fundamental para darle una definición coherente al modelo.
- La construcción de los modelos de redes usando RBF para pronóstico demostró un alto nivel de efectividad y una alta capacidad de adaptación adecuándose al tipo de problema que se modeló. En general, los resultados obtenidos en el entrenamiento y generalización alcanzaron errores cuadráticos medio muy bajos. Se demuestra un alto nivel de acierto para el pronóstico y robustez de los modelos.

- La construcción de los modelos utilizando muestras de tamaños diferentes permitió observar su nivel de consistencia. Asimismo, se observó un incremento en la velocidad de convergencia que tienen las RBF en la medida que el tamaño de las muestras es grandes.
- Es necesario destacar la importancia que ofrece el aprovechamiento en conjunto de las técnicas estadísticas multivariante y de inteligencia artificial. Ya que al obtener nueva información a partir de un conjunto de datos previo a la construcción del modelo, se puede asegurar que se tienen datos confiables y una fuente de información requerida para la construcción de los llamados modelos híbridos.

5.2 Recomendaciones

- Hacer uso del algoritmo empleado en este estudio para la construcción del Registro Único 94_05, y desarrollar un programa computacional amigable que permita realizar este trabajo de manera más sencilla, que pueda ser aplicado en cualquier otro estudio que emplee las EHM.
- Extender esta aplicación, incorporando al pronóstico obtenido la razón por la cual cada una de las viviendas es calificada como deficitaria.
- Aplicar otras técnicas del área de inteligencia artificial, como Algoritmos Genéticos y Máquinas de Vectores Soporte, que permitan comparar la bondad del modelo propuesto en este estudio.

Referencias Bibliográficas

Acosta, M., (2000). *Tutorial de Redes Neuronales*. Universidad Tecnológica de Pereira, Facultad de Ingeniería Eléctrica. Extraído el 20 de julio del 2007 desde: <http://www.ohm.utp.edu.co/neuronales.main.htm>

Bedoya, C., (2006). *El derecho a una vivienda adecuada*. Extraído el 2 de Abril de 2007 de: <http://www.pidhdd.org.pe/boletin/junio2006/5.htm>

Bastidas, R., (2005). *Salud y Ambiente*. Universidad de los Andes. Mérida. Segunda edición. Pág. 23-29.

Bolívar, M., (1997). *30 años de la Encuesta de Hogares por Muestreo*. Oficina Central de Estadística e Informática. Pág. 9-41.

Colina, E., & Rivas, F. (1998). *Introducción a la Inteligencia Artificial*, Universidad de los Andes, Mérida, Venezuela.

Collantes, J. (2001). *Predicción con redes neuronales: Comparación con las metodologías de box y jenkis*, Universidad de los Andes, Facultad de Ciencias Económicas y Sociales, Mérida, Venezuela.

Colmenares, G., (2000). Stratified/pca: Un método de procesamiento de datos y variables para la construcción de modelos de redes neuronales. *Revista Economía: Nueva Etapa* No. Pág 49-79, Universidad de Los Andes, Mérida, Venezuela.

Dagli, C., Kumara, S. & Shin, Y. (1991). *Intelligent Engineering Systems Through Artificial Neural Networks*. Ed. ASME Press.

Del Brío, B., & Molina, A. (2002). *Redes Neuronales y Sistemas Difusos*. Alfaomega. Segunda edición.

Díaz, L. (2002). *Estadística multivariada: inferencias y métodos*. Primera edición. Departamento de Estadística. Facultad de Ciencias. Universidad Nacional de Colombia. Bogotá D.C.

Flores, M., (2004). *Evolución del déficit de viviendas en honduras, 1999 – 2003*.
Extraído el 30 de Mayo de 2007 de:
http://poblacion.rds.hn/analisis_investigaciones/Evolucion_Deficit_Viviendas1999_2003

Hair, J. (1999). *Análisis Multivariante*. Quinta edición. Prentice Hall. Iberia

Haykin, S. (1999). *Neural Networks a Comprehensive Foundation*. Prentice Hall. Segunda edición. Pág. 256-294.

Hilera, J., & Martínez, V. (2000). *Redes Neuronales Artificiales*. Ed. Alfaomega.

Hopfield, J. & Tank, D. (1985). *Neural Computations of Decision in Optimization Problems*. *Biology Cybern.* Vol 52, Pág. 141-152. 1985.

Ley del Régimen Prestacional de vivienda y Hábitat. Gaceta Oficial N° 38.182 del 9 de Mayo de 2005. Artículos 12 y 13. Extraído el 23 de Mayo de 2007 desde: http://www.mintra.gov.ve/legal/leyesorganicas/regimen_prestacional_vivienda.pdf

Molina, E., & González, S. (1998). *Principales Indicadores de Pobreza*. Extraído de: <http://mipagina.cantv.net/jbhuerta/indicadores.htm>

Salazar, H., Acosta, M. & Zuluaga, C. (2000). *Tutorial de Redes Neuronales*.
Extraído el 27 de junio 2007 desde: <http://ohm.utp.edu.co/neuronales/main.htm>

Salvador, M. (2000). *Introducción al Análisis Multivariante*, Extraído de:
<http://www.5campus.com/leccion/anamul>.

Sachar, R. (1992). *Documento de trabajo sobre la II Conferencia de Naciones Unidas sobre Asentamientos Humanos*. Declaración de Estambul sobre los Asentamientos Humanos y Programa de HABITAT. Estambul. 1996. Extraído el 23 de Mayo 2007 de:
<http://www.acude.udg.mx/jalisciencia/diagnostico/354humanos.pdf>

Sachar, R. (1992). *Documento de Trabajo sobre el Derecho a una Vivienda Adecuada*; documento de las Naciones Unidas E/CN.4/Sub.2/1992/15. Extraído el 13 de Mayo 2007 de:
http://www.unhchr.ch/spanish/html/menu6/2/fs21_sp.htm#elementosdelderechoalavivienda.

Szalachman, R., (2000). *Perfil de déficit y políticas de vivienda de interés social: Situación de algunos países de la región en los noventa*" Serie Financiamiento del desarrollo 103, Santiago de Chile, CEPAL, .

Szalachman, R., (1999). *Un perfil de déficit de vivienda en Colombia 1994*. Serie Financiamiento del desarrollo 83, Santiago de Chile, CEPAL.

Szalachman, R., (1999). *Un perfil de déficit de vivienda en Bolivia 1992*. Serie Financiamiento del desarrollo 79, Santiago de Chile, CEPAL.

ANEXOS

Anexo 1

Diccionario de variables empleadas para el pronóstico del déficit de vivienda en el estado Mérida

INDICADORES	VARIABLES	SIGNIFICADO	VALORES
Identificación ID	Entidad	Identifica el estado al que pertenece la observación	Caso Mérida (12,14)
	Control	Identifica un segmento/sector de la muestra	1-999
	Área	Identifica las subdivisiones del sector de la muestra	1-999
	Línea	Identifica la vivienda encuestada	1-999
	Semestre	Corresponde al semestre en que fue incluida la observación en la EHM iniciando en el segundo semestre de 1994 al segundo semestre del 2005	1 al 23
Condición	Tipo de Vivienda	Indica el tipo de vivienda	1---9
	Paredes	Material de las paredes	1---6
	Techo	Material del Techo	1---6
	Piso	Material del Piso	1---4
	Condiciones de la Vivienda	Indica como califica el encuestado su vivienda	1,2,3 (Buena, mala, regular)
Hacinamiento	Numero de Cuartos	Total de cuartos empleados para dormir	1---20
	Total Personas	Total de Personas que viven en la vivienda	No definido

Adecuación	Servicio de Agua	La vivienda recibe el servicio por medio de:	1, 2, 3,4
	Servicio de Eliminación de excretas	Condición de higiene, cuenta con el servicio para la eliminación de excretas	1, 2, 3, 4
	Recolección de Basura	Se recoge periódicamente la basura	Si(1),No(2)
	Container de Basura	Cuenta con un container de basura	Si(1),No(2)
	Servicio Eléctrico	Cuenta con el servicio	Si(1),No(2)
Tenencia	Tenencia	Condiciones de Tenencia para cada encuestado	1---8

Descripción de las categorías de cada variable

Tipo de Vivienda	Quinta (1) Casa(2) Apartamento en edificio (3) Apartamento en casa o quinta(4) Casa Vecindad(5) Vivienda Rustica(6) Rancho Campesino(7) Otro tipo (8) Colectividad(9)
Material de Paredes	Bloque o Ladrillo Frisado(1) Bloque o ladrillo sin frisar (no acabado)(2) Madera aserrada (3) Adobe - tapia - bahareque frisado(4) Adobe - tapia - bahareque sin frisar(5) Otros (caña, palma, tablas, etc.)(6)

Techo	Platabanda(1) Teja(2) Láminas asfálticas (3) Fibrocemento, cemento, ligero y similares(4) Láminas metálicas (Zinc y similares)(5) Asbesto y Similares (6) Otros (Palmas, tabla y similares)(7)
Piso	Mosaico, granito, vinil, cerámica, ladrillo, terracota, parquet, alfombra y similares (1) Cemento(2) Tierra(3) Otros (4)
Servicio de Agua	Acueducto(1) Pila pública(2) Camión(3) Otros medios(4)
Servicio de Eliminación de excretas	Poceta a cloaca(1) Poceta a pozo séptico(2) Excusado a hoyo o letrina(3) No tiene poceta o excusado(4)
Tenencia	Propia pagada totalmente(1) Propia pagándose(2) Alquilada(3) Alquilada parte de la vivienda(4) Cedida por razones de trabajo(5) Cedida por familiar o amigo(6) Tomada(7) Otra forma (8)

Anexo 2

Programa generador de muestras

```
% Generador de muestras
% Fuente original de datos

clear all
format short g
load 'poblacion.txt'
%
XY=poblacion;
[fil,columna]=size(XY)
load 'muestras.txt'

% Indicar el vector de muestra (de 1 a 30)

n=input ('Numero de la muestra: ')
muestradatos= input ('tamaño de la muestra: ')

muestrasel=muestras(:,n);
m=1;
p=1;
for i=1:muestradatos
    for k=m:muestrasel(i,1)-1
        complem(p,:)= XY(k,:);
        p=p+1;
    end
    m=muestrasel(i,1)+1;
    entren(i,:)=XY(muestrasel(i,1),:);
end
for q=m:fila
    complem(p,:)=XY(q,:);
    p=p+1;
end

size(entren)
size(complem)
save entren.txt -ascii entren
save complem.txt -ascii complem
```

Anexo 3

Construcción del modelo RBF con NeuroSolutions

La interfaz que ofrece este paquete computacional, tiene múltiples Demos para la construcción de cualquier modelo que pueda figurar entre las áreas de Inteligencia Artificial. Es necesario hacer un reconocimiento de las fases a cumplir para la construcción de cualquier modelo y en particular el de redes con RBF.

Para la construcción de un modelo con RBF se parte del hecho de conectar tres capas, una de ellas que se encarga de llevar a cabo el proceso no-supervisado, en el cual se van a obtener los centros y anchos necesarios que permitan conglomerar de la mejor manera las observaciones, para proceder a la segunda fase de proceso supervisado en el que se da una respuesta a cada observación implicada en el estudio.

A continuación se muestra el procedimiento para construcción del modelo desarrollado en este proyecto.

Paso 1 Seleccionando el NeuralBuilder se muestran los Demos que ofrece este software para la construcción de modelos. Se selecciona el correspondiente a modelos con RBF.

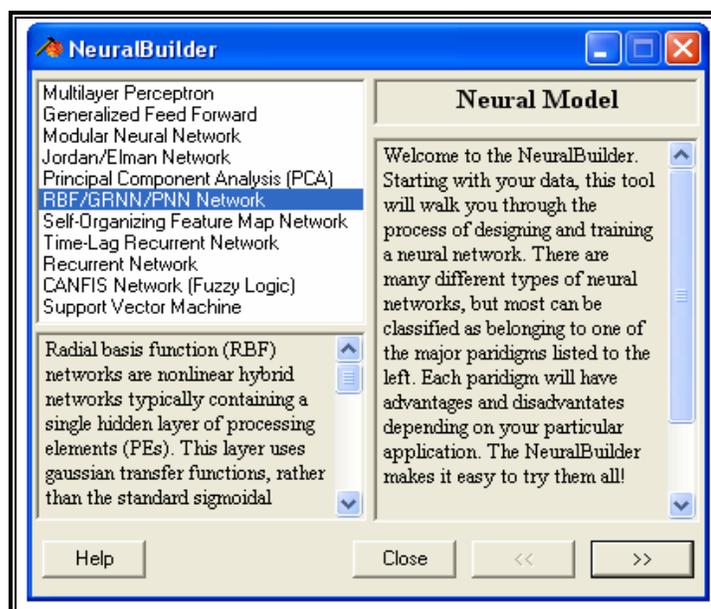


Figura 1. Selección del modelo a construir

Paso 2: Teniendo separados los patrones de entrada y salida en archivos en formato *.txt*, en primer lugar se cargan los datos correspondientes a los patrones de entrada. Finalmente, se cargan los patrones de respuesta, es decir los valores deseados.

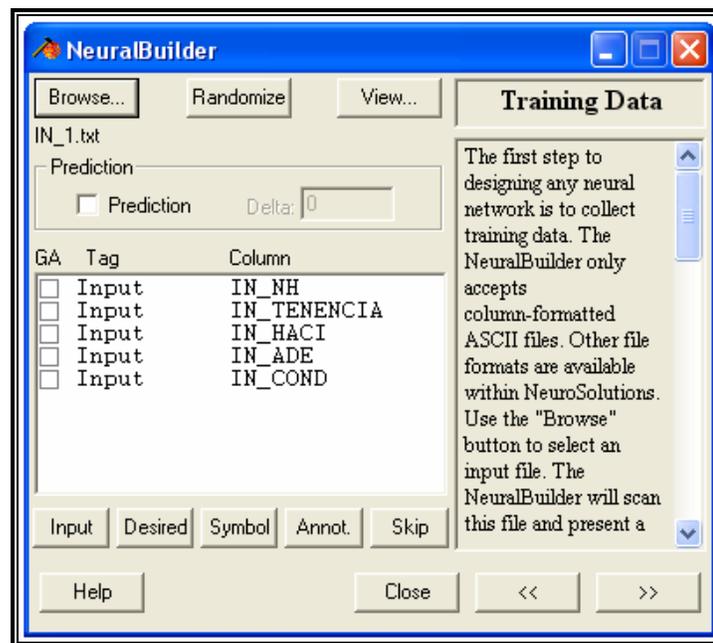


Figura 2. Cargar los datos de entrada

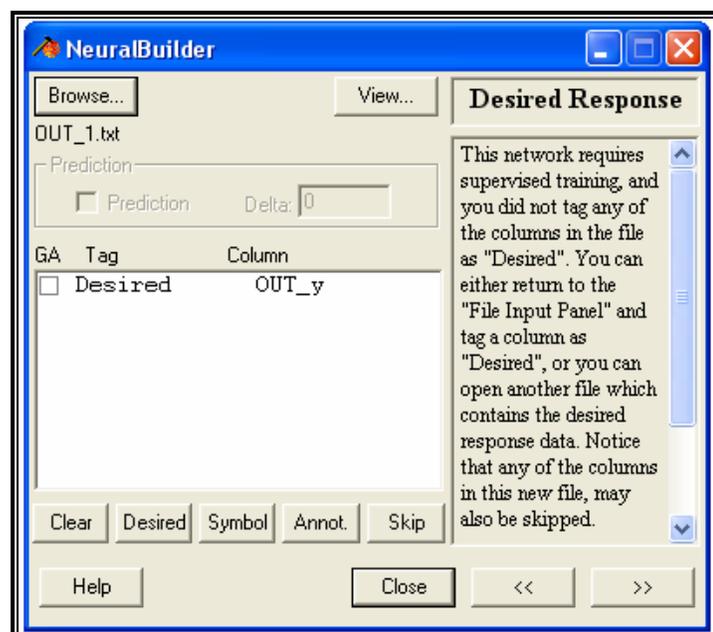


Figura 3. Cargar los patrones de la respuesta deseada

Paso 3: En este paso se configura la distribución porcentual para la validación y prueba del modelo, durante la fase de entrenamiento. Este paso es opcional, ya que el NeuroSolutions ofrece un proceso separado para la prueba del modelo.

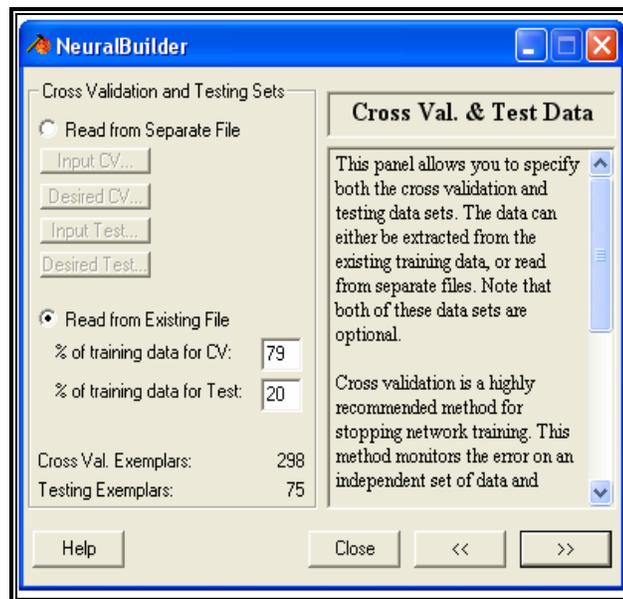


Figura 4. Configuración para la fase de entrenamiento.

Paso 4: Se configura la capa oculta, asignando los nodos radiales que se utilizarán para el modelo. También se especifican, la regla de competencia que tiene como función controlar la frecuencia con que es ubicada una observación en un centro particular evitando así la incidencia de un número muy grande de observaciones en algún nodo y muy poco en otros se escoge el ConscienceFull, y se utiliza la medida euclidiana para determinar la distancia radial entre el vector de entrada y el centro de gravedad de ese mismo nodo, para cada una de las n observaciones.

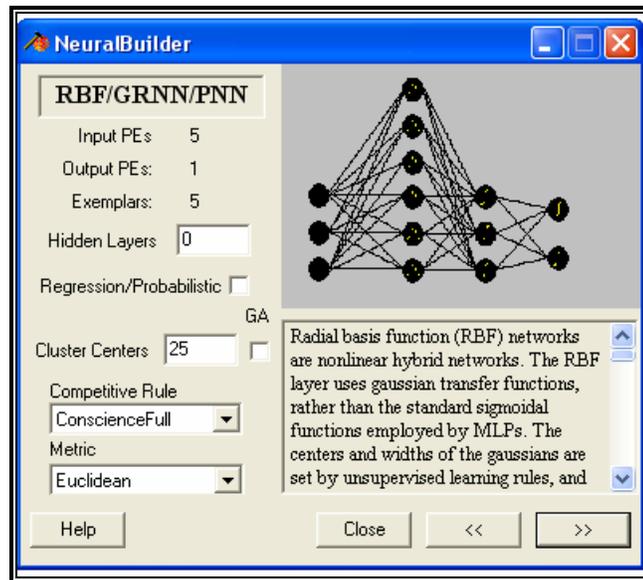


Figura 5. Configuración para la fase de entrenamiento

Paso 5: Se selecciona la función de activación pertinente para la capa de salida, que para el caso de las RBF corresponde a una función lineal.

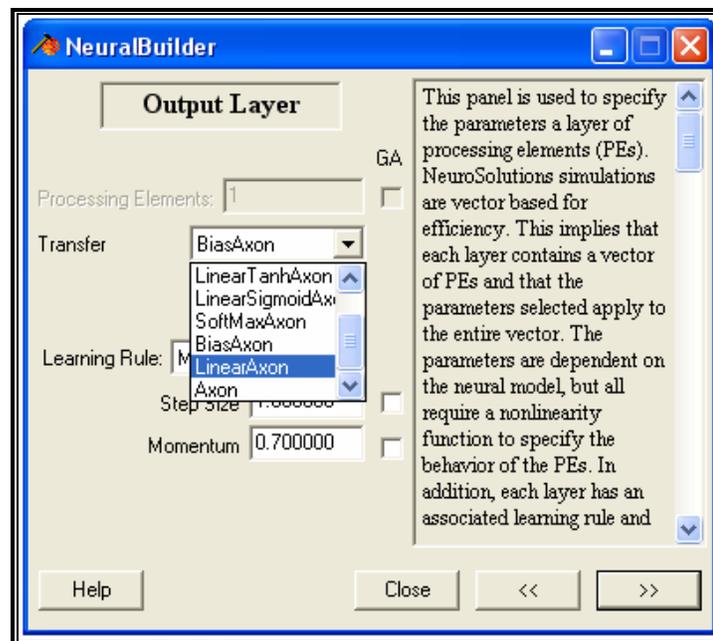


Figura 6. Configuración de la capa de salida

Paso 6: Se configura el aprendizaje a realizar la capa oculta. Como se llevará a cabo el aprendizaje no supervisado. Especificando el número de epochs, y la tasa de

aprendizaje. Además, puede considerarse otra regla de parada especificada por la variación de los pesos sinápticos.

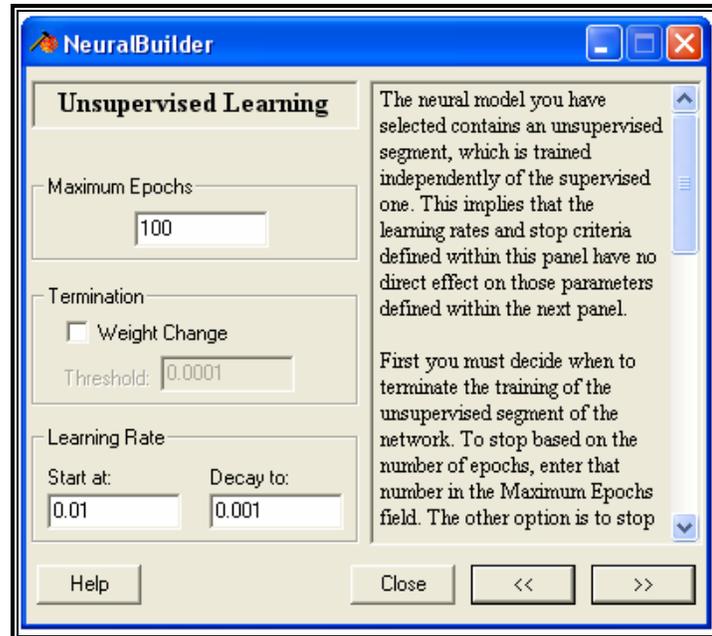


Figura 7. Configuración para el aprendizaje No-supervisado. Capa oculta

Paso 7: Se configura el control del aprendizaje supervisado que se lleva a cabo en la capa de salida. Se especifica el número de epochs o ciclos en los cuales llevara a cabo dicho el entrenamiento. Se asigna como regla de parada un máximo del error cuadrático medio a ser alcanzado por la red.

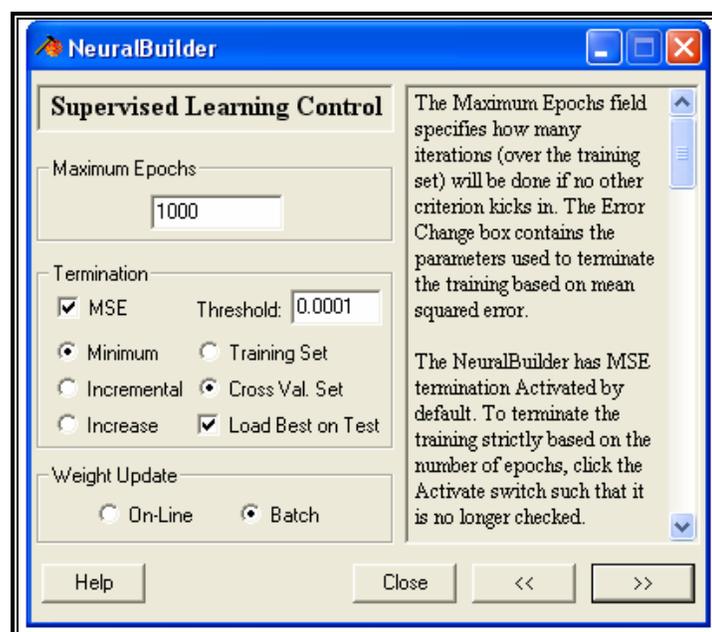


Figura 8. Configuración para el aprendizaje supervisado. Capa salida

Paso 8: Se selecciona la forma gráfica como se desea observar la actividad de la red, así como la efectividad de clasificación que ella realiza. En este paso se procede a construir la red con *Build*.

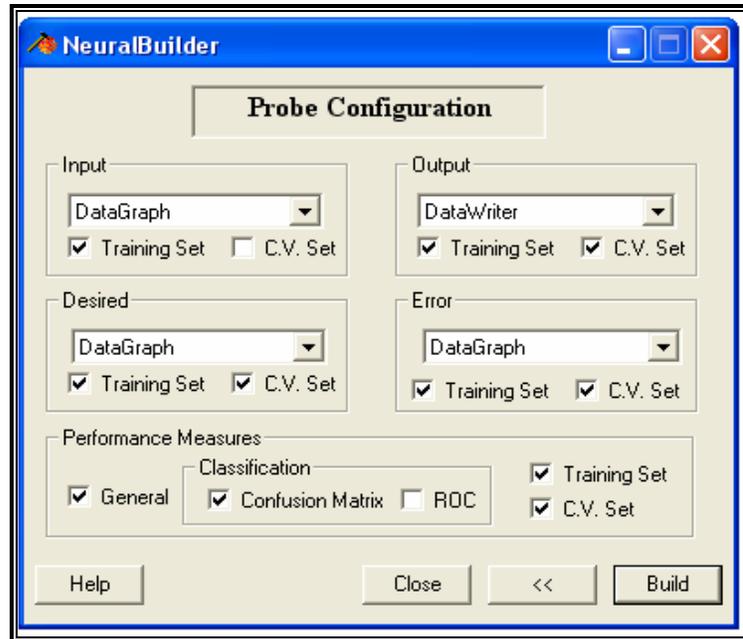


Figura 9. Configuración para observar la actividad de la red