**CAPÍTULO 4**

4. APLICACIÓN DE REDES NEURONALES AL PRONÓSTICO DE UNA SERIE DE TIEMPO

4.1. Definición

Las Redes Neuronales son una herramienta muy poderosa para la modelación matemática, especialmente cuando se utilizan para modelar problemas no lineales, como es el caso de las series de tiempo. Así, su propiedad de ajustarse a un conjunto de datos de entrenamiento, las hacen muy apropiadas para tratar el problema de la extrapolación o predicción de un conjunto de datos.

La predicción de eventos futuros de una serie de tiempo con ruido, generalmente se lleva a cabo utilizando diversas técnicas estadísticas. Aplicados a un problema formulado apropiadamente, los modelos de redes de neuronales que se utilizan para este fin están fuertemente relacionados con modelos estadísticos y a estimaciones de probabilidad bayesianas *a posteriori.* Los modelos de series de tiempo han tenido gran éxito en varios problemas de reconocimiento de patrones, y fácilmente pueden generalizarse del problema de la predicción de valores de una serie temporal, al reconocimiento de patrones y tendencias. Lo que hacen las redes neuronales es mapear un número de datos previos en el valor a predecir.

En breves rasgos el proceso de predicción de una serie de tiempo utilizando redes neuronales es el siguiente: primero se recopilan la mayor cantidad posible de observaciones ordenadas e indexadas en el tiempo (a intervalos iguales). Luego se analiza el ploteo de estos datos, buscando principalmente tendencias cíclicas, esto se hace con el fin de escoger un correcto número de datos anteriores para predecir el dato actual. A continuación se escoge una arquitectura de red adecuada y definimos la red. Del conjunto de datos escogeremos los valores que conformaran el llamado “grupo de entrenamiento”, y el resto serán el “grupo de simulación”, que nos permitará el correcto cálculo de errores de la red para poderla comparar con otros modelos. Se acostumbra dejar alrededor del 30% de los datos para el grupo de simulación. El siguiente paso es “entrenar” a la red. Esto se hace aplicando los valores del grupo de entrenamiento y realizando los ajustes respectivos según ciertos parámetros definidos previamente en la topología o arquitectura de la red. Una vez ajustados los “pesos” de la red, procedemos a simular la red y a comparar estos datos simulados con los datos pertenecientes al grupo de simulación. Si los resultados son satisfactorios el proceso de modelación de la serie temporal conlcuye.

**4.2. Ventajas y desventajas.-**

Una de las principales ventajas de utilizar las redes neuronales para la predicción de series de tiempo es la escencia heurística intrínseca de las redes neuronales. En este sentido, no necesitamos supuestos teóricos, tales como la normalidad, la independencia del ruido, la media cero, etc; ni tampoco es necesario reconocer o asumir estacionalidad dentro de la serie. Su modelamiento es más sencillo y general que el de modelos estadísticos más utilizados. El investigador tiene mayor libertad en cuanto a manejo de parámetros y con habilidad se puede llegar a ajustar el modelo a la serie.

Pero sin duda la mayor ventaja que presenta la modelación de series de tiempo a través de redes neuronales es la capacidad que éstas poseen para aproximar problemas no lineales, lo que los modelos ARIMA solo logran hacer hasta cierta medida. Las series de tiempo ciertamente son mucho mejor representadas con funciones no lineales, debido a su gran cantidad de factores y variables que estas reflejan.

Como desventajas es de notar que series con ruidos muy volátiles y no estacionales dificultan la predicción utilizando redes neuronales. Otras desventajas son:

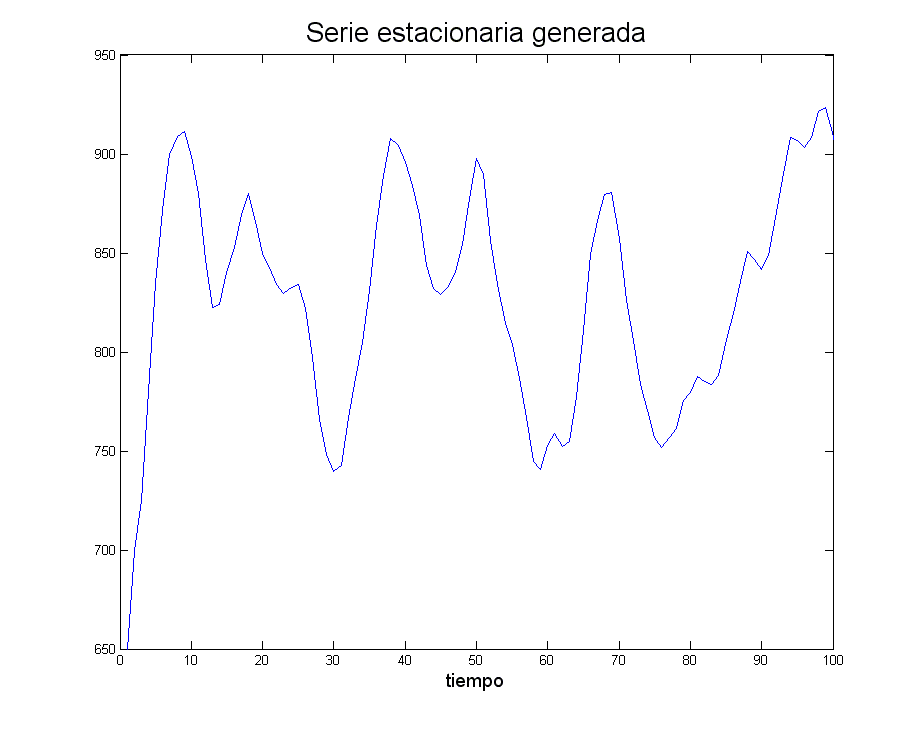
1. El problema del aprendizaje de un conjunto de datos es complejo, ya que existe un número infinito de modelos que se ajustan con la precisión deseada a los datos de entrenamiento, pero pocos de ellos logran generalizar bien esta simulación a datos fuera del conjunto de datos de entrenamiento. Se desea contar con la mayor cantidad de datos de entrenamiento, sin embargo, en series de tiempo no estacionarias el aumento de datos de entrenamiento aportará con características estadísticas que no son muy relevantes para los valores a predecir.
2. Fuertes ruidos y un conjunto de datos de entramiento pequeño hacen que el modelo no sea preciso. Correlaciones al azar entre los datos de entrada y los de salida pueden presentar grandes dificultades. Los modelos no reconocen la relación temporal explícita entre los datos de entrada y los de salida, es decir, no reconocen entre las correlaciones que ocurren en orden temporal y las que no.

**4.3. Simulación de la serie y Desarrollo del modelo**

Para propósitos prácticos utilizaremos una serie generada. El modelo utilizado fue un AR(4), más precisamente la fórmula utilizada para generar la serie fue  , donde  es un ruido blanco con media 0 y varianza 75. Se puede comprobar fácilmente que la serie es estacionaria, para ver este resultado encontramos las raices del polinomio de resagos y verificando que efectivamente todas se encuentran fuera del círculo unitario, suficiente para la estacionalidad de la serie. En la figura 4.1 podemos ver el gráfico de la serie.

Figura 4.1

Gráfico de la serie estacionaria generada



Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

La serie que generamos se extiende por 100 observaciones, lo cual, para series de variables económicas no es tan fácil de conseguir. A menos que las observaciones sean mensuales, esto implicaría datos estadísticos de demasiados años.

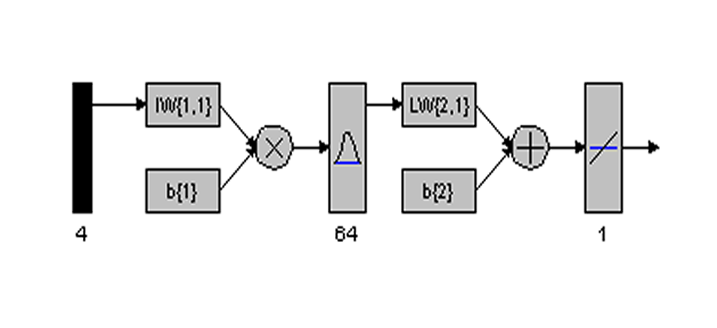
Para el modelamiento de esta serie con redes neuronales utilizaremos tres arquitecturas diferentes, para analizar las diferencias en cuanto a calidad de predicción y eficiencia de los métodos utilizados. En todos los casos respetaremos el principio de parsimonia, tratando de reducir al mínimo el número de parámetros utilizados.

**4.3.1. Red 1**

La primera red que presentaremos se conoce como Radial Basis (exact fit) y no representa un modelo muy refinado. No trabajamos con ninguna regla de aprendizaje, de hecho se puede ver a esta red como un método que utiliza la fuerza bruta para la predicción. Topológicamente la red tiene dos capas: una escondida y la capa de salida. Recibe cuatro valores de entrada,  , ,  y  y entrega como salida . Dado que vamos a utilizar el 70% de los datos de la serie para entrenamiento y el restante 30% para evaluar el poder de predicción fuera del conjunto de entrenamiento, la capa intermedia o escondida tiene 64 neuronas (una para cada uno de las entradas), un número un poco exagerado. Esto se debe a que la red busca ajustarse perfectamente a los datos de entrenamientos y de ahí, predecir los valores fuera de estos. La función de transferencia de la última capa es una función lineal, ya que las funciones *sigmoid* y la función *hardlimit* tienen como imagen a [-1,1], por lo tanto no podríamos predecir los valores de una serie de tiempo. En la figura 4.2 podemos ver el diseño de esta red.

Figura 4.2

Esquema de la red neuronal de 64 nodos en la capa oculta.



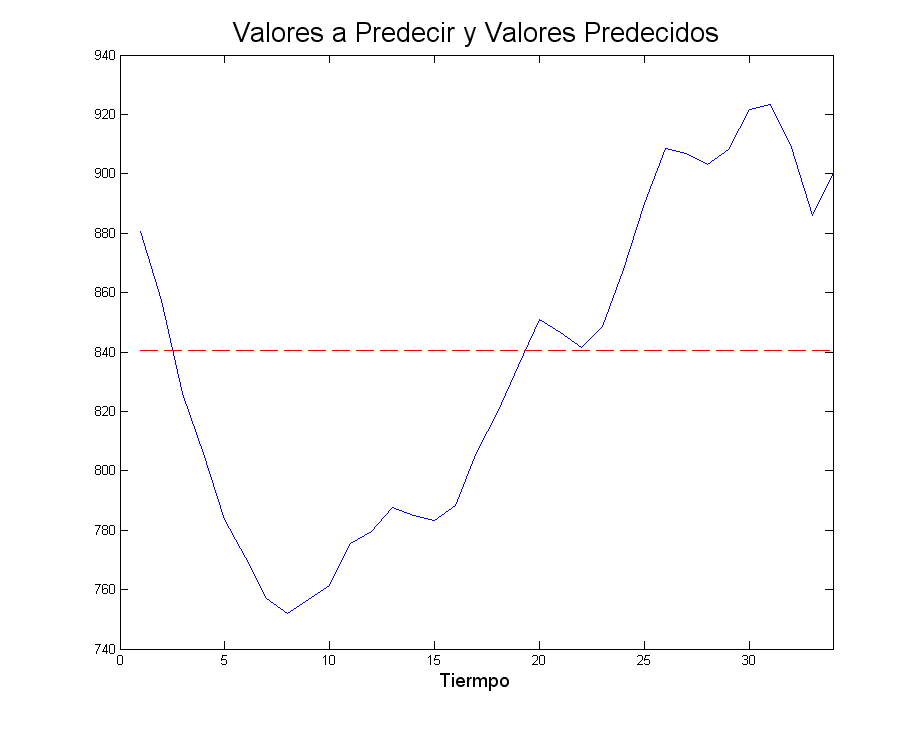
Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Por lo tanto cuando simulamos la red utilizando los datos con los que fue entrenada la red, el vector de errores que obtenemos es casi un vector de ceros. Lo que nos interesa es la predicción del 30% de los datos que guardamos como datos para verificar el poder de predicción de la red, y su consiguiente vector de errores. En la figura 4.3 podemos ver, en línea contínua los valores a predecir, y en línea entrecortada los valores predecidos por la red neuronal, y comprobamos que aunque el modelo predice casi a la perfección los datos de entrenamiento, su poder de predicción para valores fuera del conjunto de entrenamiento es muy pobre.

Figura 4.3

Valores a predecir y valores predichos por la red



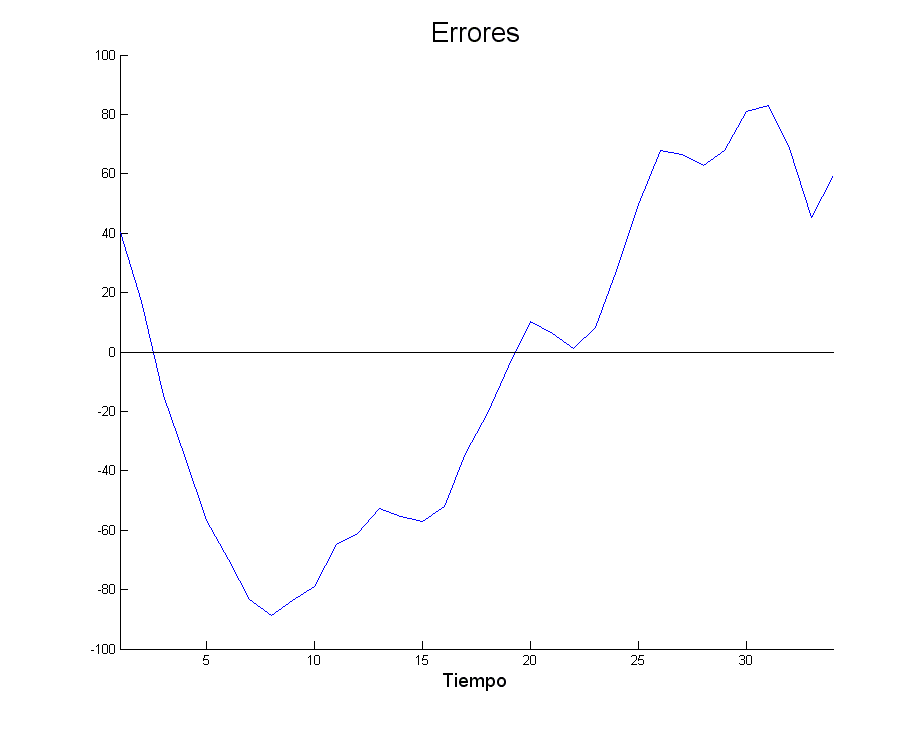
Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

El modelo predice los valores como la media de los datos de entrenamiento. Analizando el vector de errores comprobamos lo que era de esperarse. El gráfico de estos errores tiene la misma forma de los datos a predecir pero desplazada y centrada alrededor del cero, esto se debe a que lo que hizo el modelo fue entregar como salida el valor medio para cada una de las predicciones.

Figura 4.4

Gráfico de los errores



Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Finalmente calcularemos la suma cuadrática de estos errores, vemos que su valor es SSE=105530. Una cantidad muy grande.

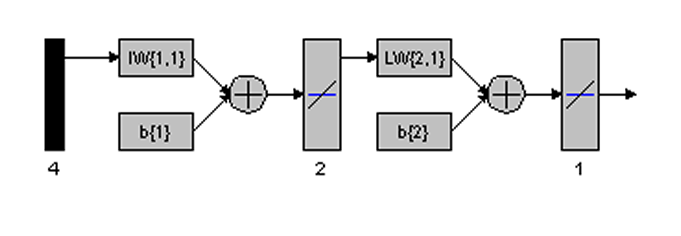
En conclusión vemos que este modelo no cumple con la función de predecir datos en una serie de tiempo, trataremos de refinar este modelo para así poder ser más precisos en nuestros objetivos.

**4.3.2. RED 2**

Esta segunda red que analizaremos se denomina Feed Forward con Back-Propagation debido a que cada capa recibe los datos de la capa inmediata anterior, no existen conexiones de neuronas hacia atrás ni tampoco dentro de una misma capa y los ajustes se hacen comenzando por la capa de salida de atrás hacia delante y según lo influyente que es cada neurona al error total de la red. Como vemos en la figura 9 la red tiene 2 capas, una escondida y la capa de salida. Para predecir el valor , la red recibe como entrada los cuatro valores previos de la serie,  , ,  y . La función de transferencia de la capa de entrada a la capa escondida es una función lineal. En la capa intermedia o escondida tenemos dos neuronas. Finalmente la función de transferencia de esta capa a la de salida es también la función lineal.

Figura 4.5

Esquema de la red con 2 neurona en la capa escondida



Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

La arquitectura escogida para esta red tiene como función de entrenamiento a la función TRAINLM o función de Levenberg-Marquardt backpropagation. La función TRAINLM recibe su nombre debido a que esta ajusta los pesos y la constante de las neuronas según la regla de optimización propuesta por Levenberg-Marquadt. Esta función puede se aplicada a cualquier red neuronal que posea como funciones de transferencia a funciones derivables, como ciertamente es el caso de la función lineal.

Inicializamos los pesos de la primera capa con los valores 0.068158, -0.38142, 0.13614 y 0.40548 para la neurona 1 y 0.45423, 0.67699, -0.25917 y 0.093142 para la neurona 2 respectivamente y para la constante de la primera capa neurona 1, -0.11024 y para la neurona 2 0.38913. Estos resultados se resumen en la tabla I.

Tabla I

Pesos iniciales de la capa 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CAPA 1 | | | | | |
|  | W1 | W2 | W3 | W4 | B |
| Neurona 1 | 0.068158 | -0.38142 | 0.13614 | 0.40548 | -0.11024 |
| Neurona 2 | 0.45423 | 0.67699 | -0.25917 | 0.093142 | 0.38913 |

Fuente: Pronóstico de Ventas: Comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Así mismo para la segunda capa o capa de salida tenemos 0.24262 y 0.58964 respectivamente. Nótese que en esta segunda capa solo tenemos dos datos de entrada. Su constante es 0.91369. Se resumen en la tabla II.

Tabla II

Pesos iniciales de la capa 2

|  |  |  |  |
| --- | --- | --- | --- |
| CAPA 2 | | | |
|  | W1 | W2 | B |
| Neurona 1 | 0.24262 | 0.58964 | 0.91369 |

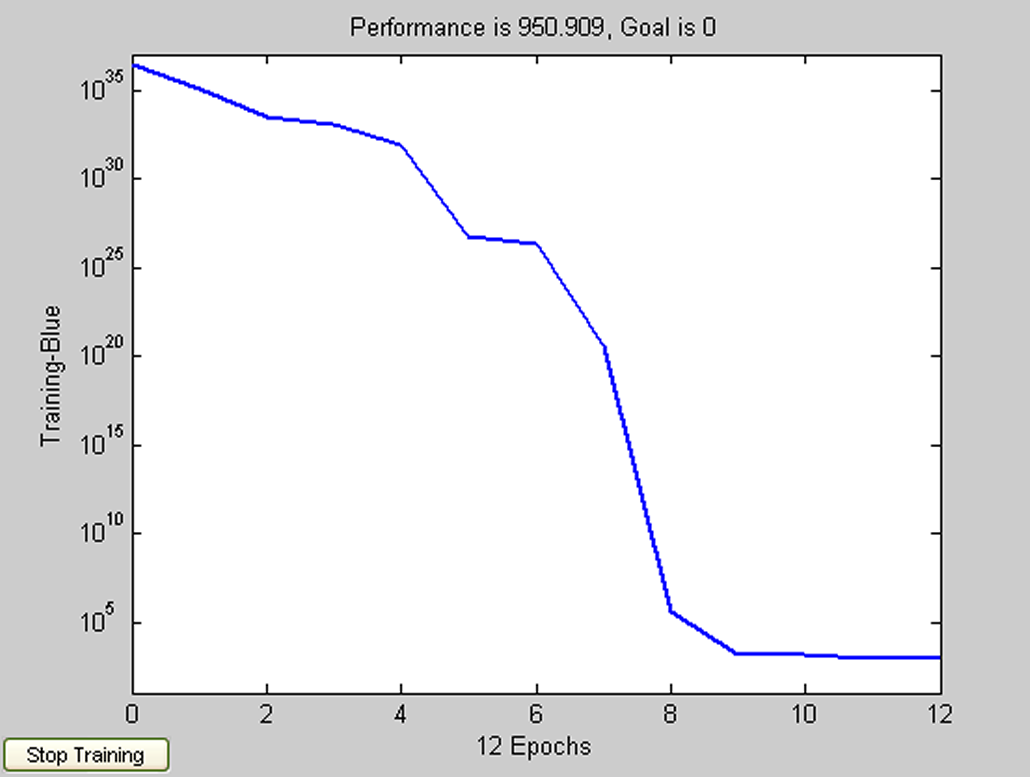
Fuente: Pronóstico de Ventas: Comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Estos valores fueron asignados de manera aleatoria como valores iniciales. Al realizar el entranamiento con el 70% de los datos de la serie tenemos la siguiente función de aproximación en la figura 4.6. Esta función nos muestra como se fue ajustando las salidas de la red a nuestros valores de entrenamiento.

Figura 4.6

Función de aproximación

****

Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Vemos que el algoritmo de entrenamiento alcanzó su criterio de salida en 12 iteraciones. Los valores ajustados para los pesos y las constantes son resumidos en la tabla III.

Tabla III

Pesos ajustados de la capa 1 y 2

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CAPA 1 | | | | | | | | |
|  | W1 | | W2 | W3 | | W4 | | B |
| Neurona 1 | -6222926.51 | | 3010643.54 | -4867641.42 | | 5353776.97 | | 260316.55 |
| Neurona 2 | -10936654.07 | | 5291138.65 | -8554770.85 | | 9409143.20 | | 457500.51 |
| CAPA 2 | | | | | | | | |
|  | | W1 | | | W2 | | B | |
| Neurona 1 | | -42934511.242 | | | 24429620.497 | | -2230.769 | |

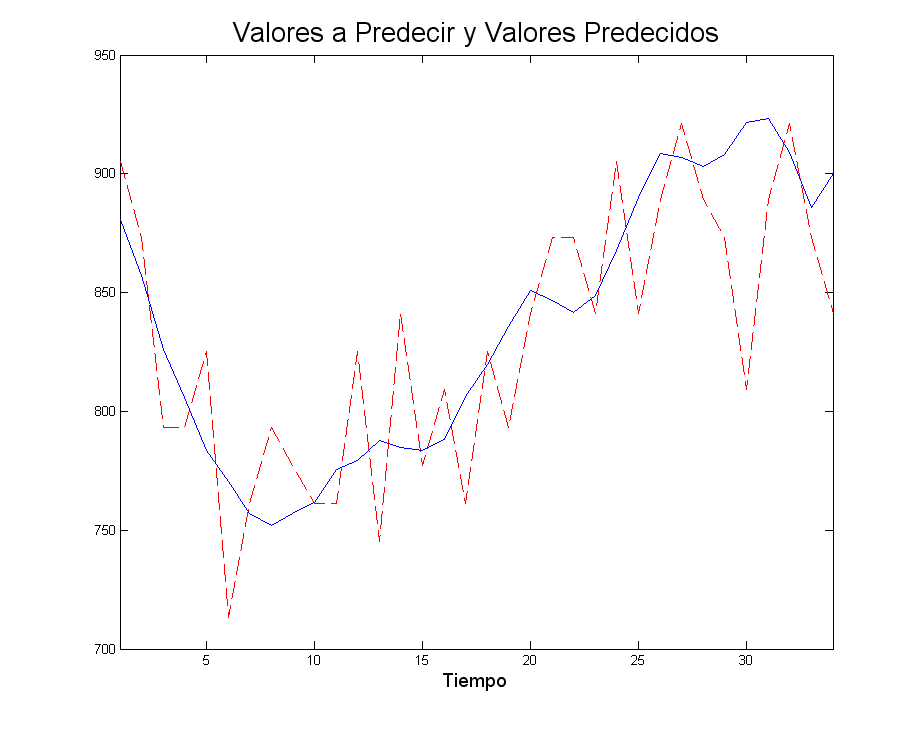
Fuente: Pronóstico de Ventas: Comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Una vez entrenada la red procedemos a verificar que tan bien se ajustan las predicciones que obtenemos, a los datos de la serie. Simulamos la red con el 30% de datos que asignamos a este propósito y obtenemos un conjunto de datos simulados. En la figura 4.7 vemos que los datos pronosticados por esta segunda red se ajustan mucho mejor que los de la red anterior. Vemos que la tendencia central de la serie temporal es simulada por la red, pero la simulación es algo errática. En el gráfico la línea entrecortada representa los datos simulados por la red, mientras que la línea contínua representa los datos reales de la serie.

Figura 4.7

Valores a predecir y valores predecidos por la red



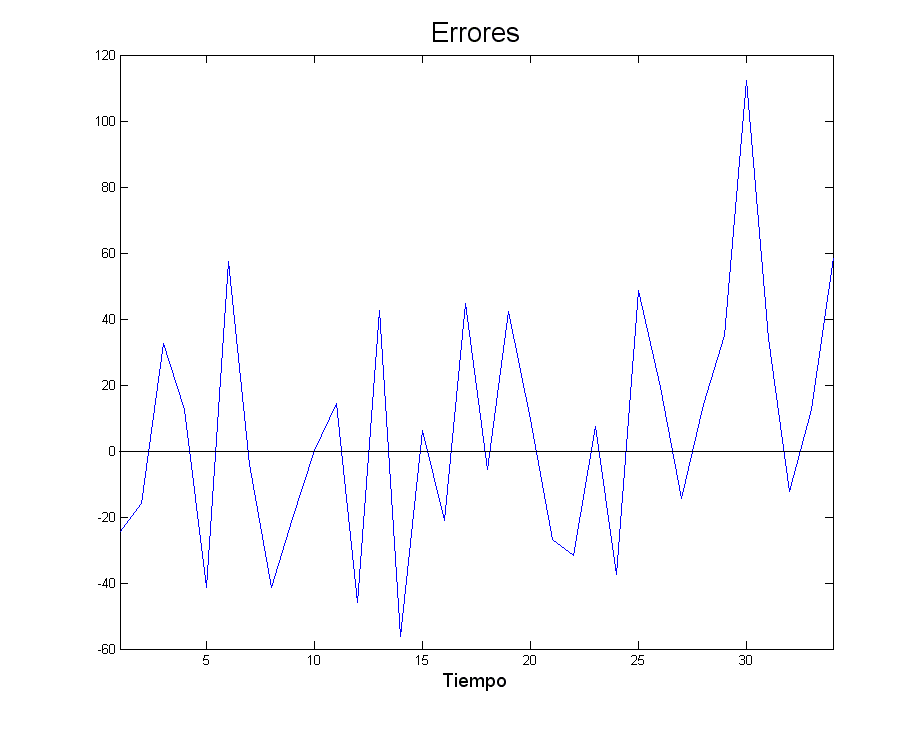
Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Restando los datos simulados por la red de los datos reales de la serie obtenemos un vector de errores. Como es de esperarse queremos que cada elemento de este vector sea tan pequeño como podamos. En la figura 4.8 vemos el ploteo de estos errores.

Figura 4.8

Gráfica de los errores para la segunda red



Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Como indicador del tamaño total de error tomaremos la suma cuadrática de los errores. La SSE es simplemente la suma del cuadrado de todos los elementos del vector de errores, siendo en este caso 46003.

Como conclusión de este segundo modelo tenemos que se ajusta muy bien al pronóstico de los datos de la serie. La suma cuadrática de los errores en esta segunda red fue mucho menor que la de la primera, de hecho fue menos de la mitad. La figura 4.8 nos muestra que el ajuste a la serie real fue muy buena. Y vemos que el principio de parsimonia, que recomienda que la menor cantidad de parámetros necesarios predecirá mejor al modelo, nos condujo a un mejor resultado, ya que en la primera red teníamos cerca de 70 neuronas y en esta segunda red tenemos solamente 3.

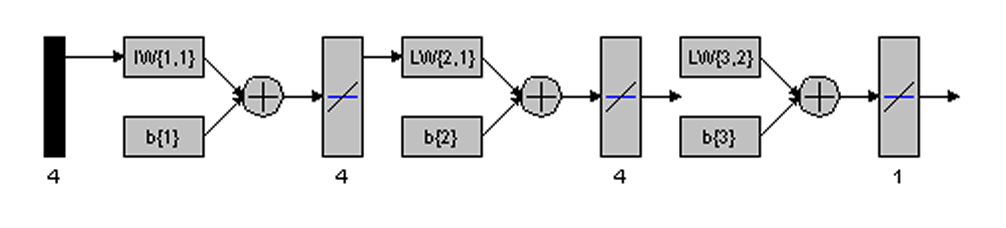
**4.3.3. Red 3**

Una vez que hemos mejorado el primer modelo y hemos encontrado un modelo que responde bastante bien a nuestros propósitos de predicción procederemos a refinar más aún este modelo para así encontrar la arquitectura de red que más se ajuste al problema de predecir los datos de una serie de tiempo. Asumimos que los datos son trimestrales (esto está implícito en el hecho que escogemos los cuatro periodos anteriores como datos de entrada). Es de recalcar que siendo las redes neuronales un método heurístico, la habilidad del investigador en encontrar la mejor topología de red para cada problema específico es muy importante.

Vimos que el modelo feed-forward con back-propagation se ajustó bastante bien al problema de predicción, así lo que haremos ahora es refinar este mismo modelo, aumentando dos neuronas en la capa intermedia existente e incrementando una capa intermedia más, con cuatro neuronas. Utilizaremos la misma regla de aprendizaje, TRAINLM, y la función de transferencia de esta nueva capa es así mismo PURELIN o función lineal. La arquitectura de la capa de entrada se mantendrá igual, así como la arquitectura de la capa de salida. En la figura 4.9 podemos ver el esquema de esta tercera red.

Figura 4.9

Esquema de una red neuronal con dos capas intermedias



Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

La iniciación de los pesos se hizo aleatoriamente. Por lo general esta es la manera de hacer esta inicialización ya que es muy difícil o hasta imposible interpretar los pesos de una red, por lo tantos no se puede tener criterios *a priori* de estos valores. La tabla IV presenta estos valores iniciales de los pesos para las capas 1 ,2 y 3 respectivamente.

Tabla IV

Pesos iniciales para las capas 1, 2 y 3

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CAPA 1 | | | | | | | | | | | |
|  | W1 | | W2 | | W3 | | W4 | | | B | |
| Neurona 1 | 0.44396 | | -0.066475 | | -0.43673 | | 0.97232 | | | 0.60903 | |
| Neurona 2 | -0.38758 | | -0.97066 | | -0.47636 | | -0.053315 | | | 0.65773 | |
| Neurona 3 | -0.77567 | | 0.3281 | | 0.41694 | | 0.80564 | | | -0.66746 | |
| Neurona 4 | -0.11342 | | 0.44812 | | 0.56772 | | -0.097882 | | | -0.21219 | |
| CAPA 2 | | | | | | | | | | | |
|  | W1 | | | W2 | W3 | | W4 | | B | | |
| Neurona 1 | 0.041515 | | | 0.10939 | -0.39493 | | 0.11588 | | 0.95418 | | |
| Neurona 2 | 0.43625 | | | -0.82451 | 0.70369 | | -0.97153 | | -0.55618 | | |
| Neurona 3 | 0.13838 | | | -0.11303 | 0.51896 | | 0.19235 | | 0.40737 | | |
| Neurona 4 | -0.078388 | | | -0.2674 | 0.89952 | | 0.63241 | | 0.044122 | | |
| CAPA 3 | | | | | | | | | | | |
|  | | W1 | | W2 | | W3 | | W4 | | | B |
| Neurona 1 | | 0.86579 | | 0.42671 | | -0.54392 | | -0.10072 | | | -0.6556 |

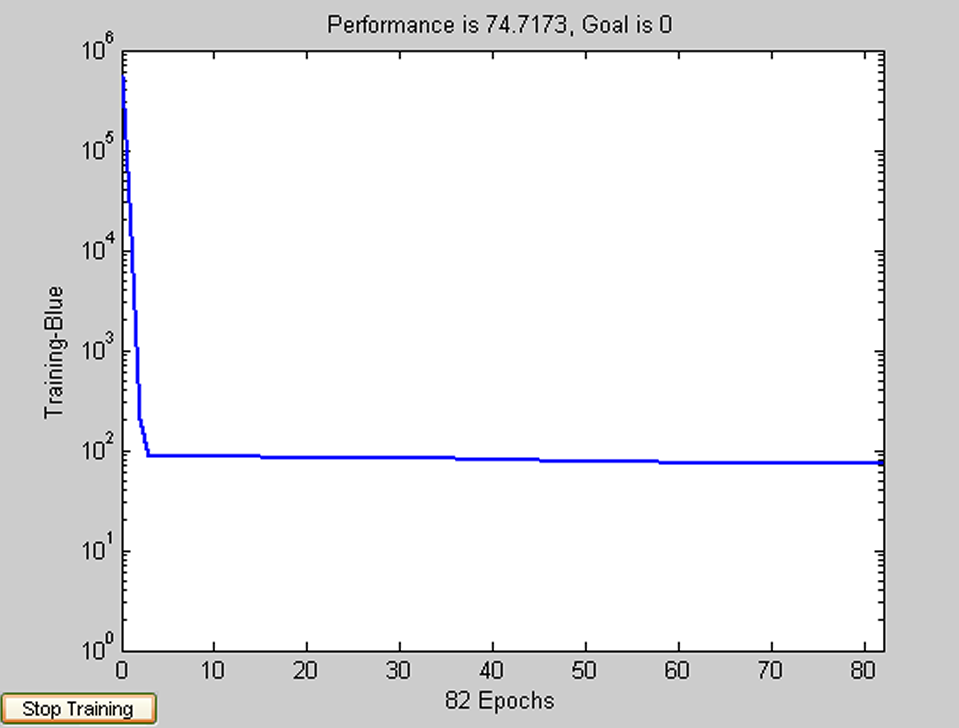
Fuente: Pronóstico de Ventas: Comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Después de inicializar los pesos procedemos a entrenar la red con el conjunto de datos que separamos para este propósito. Para ver con que velocidad alcanzó los criterios de ajuste, ploteamos la función de aproximación y la mostramos en la figura 4.10. Vemos que en este caso esta se acerca mucho más al cero que en la red anterior, y converge asintóticamente hacia un valor por debajo de , mientras que la red 2 convergía a un valor entre  y . Aún cuando la segunda red toma más epochs o corridas para alcanzar su convergencia, ésta es mucho más cercana al cero que en la anterior.

Figura 4.10

Función de aproximación para la red neuronal



Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Después de este entrenamiento los pesos y constantes han sido modificados para adaptarse a estos datos, y poder predecir valores fuera del conjunto de entrenamiento. Los nuevos valores están presentados en la tabla V, correspondiente a una de las neuronas.

Tabla V

Pesos ajustados para la capa 1, 2 y 3

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CAPA 1 | | | | | | | | | |
|  | W1 | | W2 | | | W3 | | W4 | B |
| Neurona 1 | -1.6147 | | -0.45804 | | | -0.62535 | | 0.62688 | 1.7845 |
| Neurona 2 | 0.66942 | | 0.73682 | | | 1.2249 | | 0.20999 | -0.63122 |
| Neurona 3 | 1.1026 | | 0.011882 | | | -0.52073 | | -0.66404 | 1.0938 |
| Neurona 4 | 0.45904 | | 0.38124 | | | 0.56477 | | 0.14161 | 3.1298 |
| CAPA 2 | | | | | | | | | |
|  | W1 | | W2 | | W3 | | | W4 | B |
| Neurona 1 | -0.72571 | | 0.58185 | | -1.2023 | | | -2.5696 | -2.3085 |
| Neurona 2 | 0.63536 | | 0.36378 | | -0.047503 | | | 0.51849 | 2.6831 |
| Neurona 3 | -0.47288 | | 1.0381 | | 0.056528 | | | -1.4685 | -1.2709 |
| Neurona 4 | -0.80673 | | -0.41315 | | -0.47819 | | | -0.83425 | -0.71164 |
| CAPA 3 | | | | | | | | | |
|  | | W1 | | W2 | W3 | | W4 | | B |
| Neurona 1 | | -3.2448 | | 2.333 | -2.0007 | | -0.80901 | | 4.0984 |

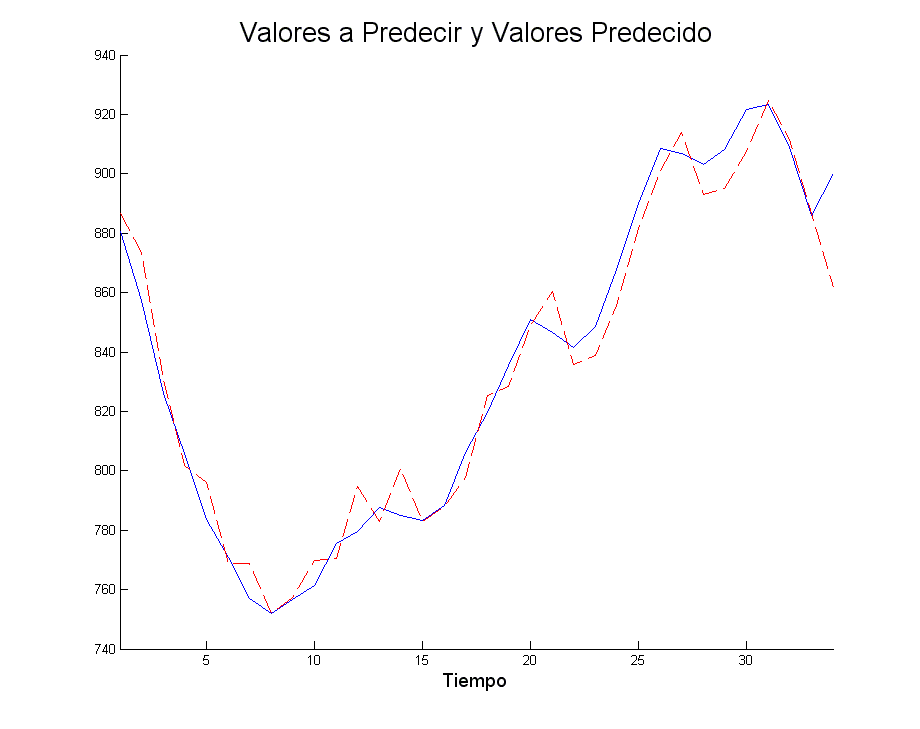
Fuente: Pronóstico de Ventas: Comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Con los pesos ajustados podemos predecir o hacer una simulación de la red con los datos del conjunto a predecir. En la figura 4.11 podemos ver que el ajuste de los valores simulados por la red, graficados con la línea entrecortada, es bastante cercana a la serie original, graficada con la línea continua.

Figura 4.11

Valores a predecir y valores predicho por la red



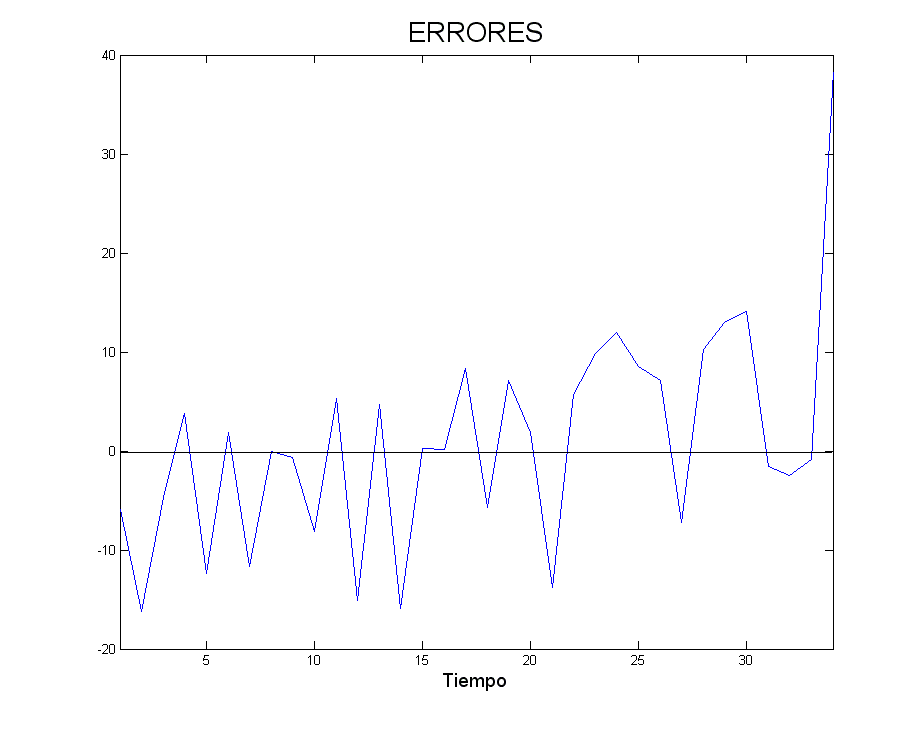
Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Como podemos ver, la línea entrecortada (que representa la predicción hecha por la red neuronal) esta casi encima en cada punto de la serie original. Haciendo un gráfico de los errores vemos que estos oscilan muy cercanos al cero. Podemos ver el plot en la figura 4.12.

Figura 4.12

Gráfico de los errores



Fuente: Pronóstico de ventas: comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Y como indicador de estos errores procedemos a calcular la suma cuadrática de los errores o SSE. Y obtenemos, como era de esperarse, que es un valor bastante pequeño, mucho menor a la SSE de las otras dos redes. Y es 3973.9.

En conclusión tenemos que esta tercera red tiene un poder de predicción mucho mayor a las otras dos expuestas y es por eso que esta será la topología de red que utilizaremos para el análisis de series reales, dada por las cuentas trimestrales del Ecuador. Realmente la serie es predecida de manera muy precisa por esta tercera red. Aquí vemos todo el poder de predicción del que son capaces las redes neuronales.

**4.3.4.** **Comparación de las tres redes expuestas**

Ahora veremos en la tabla VI un cuadro comparativo de las principales características, tanto de arquitectura como de desempeño de estas tres redes.

Tabla VI

Cuadro comparativo de las principales características de las 3 redes

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Red 1** | **Red 2** | **Red 3** |
| **# de capas** | 2 | 2 | 3 |
| **# total de neuroanas** | 65 | 3 | 9 |
| **# total de pesos y constante** | 385 | 13 | 45 |
| **Regla de aprendizaje** | - | TRAINLM | TRAINLM |
| **Función de transferencia** | PURELIN | PURELIN | PURELIN |
| **Iteraciones en entrenamiento** | - | 13 | 85 |
| **SSE en datos de predicción** | 105530 | 46003 | 3973.9 |

Fuente: Pronóstico de Ventas: Comparación de la precisión de la predicción con diferentes métodos, TESIS 2005

Elaboración: Andrés Guillermo Abad Robalino

Analizando el cuadro comparativo vemos que la magnitud de la suma cuadrática de los errores es significativamente menor en la red 3. Aunque el número total de parámetros en la red 3 casi triplica al de la red 2 este incremento se justifica con la mejora en predicción de la red 3. También tenemos que el número de parámetros de la red 1 es inmanejable. En conclusión la red 3 es la que cumplió mejor con la función de predicción que buscábamos.