

# **Introducción a las Redes Neuronales Evolutivas de Unidades Sigmoides y Producto**

*César Hervás*

*Dpto. de Informática y A. Numérico*

*Universidad de Córdoba*

*chervas@uco.es*

El análisis de métodos de regresión no lineal es una de las áreas de mayor interés en la actualidad en el campo del modelado de sistemas dinámicos reales en diferentes áreas de investigación. Asociado a este análisis se han desarrollado en los últimos años diferentes modelos de redes neuronales artificiales para regresión y clasificación, redes de tipo perceptrón multicapa (MLP) basadas en la utilización de unidades de base sigmoide, US, como funciones de transferencia en los nodos de la capa oculta de la red; funciones de base radial (RBF), también llamadas funciones de ventana radial, redes neuronales de regresión generales (GRNN) desarrolladas por Specht y redes multiplicativas, y dentro de ellas, redes de unidades de base producto, UP. Por otra parte, a pesar del carácter de aproximadores universales de las redes neuronales basadas en unidades de diferente tipo (MLP, RBF y UP), la velocidad de aprendizaje y la complejidad de la red final pueden diferir de forma significativa de un caso a otro. La velocidad de convergencia y la complejidad de las redes utilizadas se convierte en uno de los principales problemas a resolver. Este problema ha sido abordado desde diferentes puntos de vista.

Investigadores de diferentes áreas de conocimiento ha sugerido utilizar redes neuronales artificiales, RNA, como una alternativa a los modelos de Superficie de Respuesta, MSR, (Han et al., 1994; Miel and May, 1991; Carpenter and Bartheley, 1993; Hervás et al 2000). Una RNA se ha definido inicialmente como una estructura que se diseña para resolver cierto tipo de problemas asociados a emular la forma en la que el cerebro humano puede resolverlos, de esta manera su interpretación biológica es la de similar la activación de las neuronas cerebrales (McCulloch and Pitts 1943). La forma general de una RNA es la de un modelo de “caja negra” en el sentido de que la interpretabilidad de las relaciones no lineales causa-efecto no se obtiene fácilmente y donde las características de entrada pertenecen a un espacio de una alta dimensión.

Las RNA multicapa con activación hacia adelante las vamos a considerar, en nuestro caso, como modelos predictivos asociados al análisis estadístico multivariante donde se utilizan  $k$  variables predictoras,  $\mathbf{x}$ , y una sola variable de respuesta. La representación

del modelo se estructura en varias capas, donde la estructura más común consta de tres capas, la capa de entrada, que está formada por las variables predictoras, la capa oculta, que consta de un conjunto de variables obtenidas a partir de transformaciones no lineales de sumas ponderadas de las variables de la capa de entrada; y la capa de salida, formada, en este caso, por la variable y de respuesta. Unos buenos libros de consulta son (Cheng, et al. 1994; Haykin, 1994; Bishop, 1995; Ripley, 1994).

De esta forma el modelo lineal con funciones de base sigmoides se puede escribir en la forma

$$f(\mathbf{x}) = \beta_0 + \sum_{j=1}^p \beta_j B_j(\mathbf{x}, \mathbf{w}_j)$$

Es claro que la estimación paramétrica estándar debe de ser sustituida por métodos de optimización computacionales, puesto que no es posible estimar los parámetros  $\beta_j$  de forma independiente de los parámetros  $\mathbf{w}_j$ . Las RNA han sido utilizadas en los últimos años para la resolución de este problema en muchos casos en los que los métodos tradicionales de regresión son incapaces de encontrar tal relación. La justificación de esta capacidad de las redes neuronales se encuentra en el siguiente hecho: cualquier función continua puede ser aproximada sobre un conjunto compacto por una red neuronal con una sola capa intermedia, con tal de que ésta tenga un número suficiente de nodos, aunque esto no implica el aprendizaje de la red. Sin embargo, en la práctica, el número de nodos ocultos o de funciones de base necesarios para realizar la tarea de aproximación suele ser muy alto, lo que provoca la escasa interpretabilidad de los resultados obtenidos. A este hecho hay que unir que, con frecuencia, en los algoritmos de entrenamiento de búsqueda local, la superficie de error está plagada de mínimos locales y superficies planas que complican la búsqueda del mínimo global. En este sentido Rumelhart y su equipo del PDP, (Rumelhart et al. 1986) describen un algoritmo de aprendizaje, búsqueda de valores óptimos, de gradiente descendente, para la estimación de los parámetros del modelo  $\{\beta, \mathbf{w}\}$ ; pero este algoritmo denominado de retropropagación “backpropagation” presenta serios problemas de quedar atrapado en óptimos locales y de ser lento en la obtención de los estimadores.

Además nos queda un problema adicional y es la determinación del número de funciones de base y del número de parámetros significativamente distintos de cero necesarios para poder obtener una buena función de ajuste, este problema se define como el diseño de la arquitectura de la red que en lenguaje de RNA es tanto como decir

diseñar el número de nodos de la capa oculta y el número de conexiones entre la capa de entrada y la capa oculta,  $w_j$ , y el número de conexiones entre la capa oculta y la capa de salida,  $\beta_j$ . Una alternativa a estos métodos de estimación es por tanto utilizar algoritmos basados en poblaciones (Deb 2003), y dentro de ellos utilizaremos los algoritmos evolutivos, algoritmos que, por una parte, permitan obtener expresiones analíticas simples del modelo que puedan ser interpretables, y por otra, no queden atrapados en mínimos locales durante el proceso de búsqueda. Los algoritmos evolutivos son un tipo de algoritmos de búsqueda estocástica que permiten encontrar la solución en espacios complejos. Dichos algoritmos se han utilizado con éxito en el campo de las redes neuronales para encontrar la estructura adecuada de la red. Incluso, también se han utilizado, para calcular el valor de los pesos de las conexiones evitando quedar atrapado en mínimos locales, generalmente derivados de un sobreentrenamiento.

En el proceso de evolución, son esenciales los operadores de selección, replicación y mutación, para introducir, por una parte, presión selectiva, y, por otra, diversidad, de forma tal que el algoritmo de búsqueda de la mejor red presente un compromiso entre su capacidad de explotar la localización de las mejores soluciones y explorar otras localizaciones del espacio, con el fin de obviar el problema de la obtención de óptimos locales. La no inclusión del operador de cruce se debe al problema del engaño asociado a este tipo de codificación del cromosoma, lo que hace que sea en general poco eficiente.