

OPENCOURSEWARE
REDES DE NEURONAS ARTIFICIALES
Inés M. Galván – José M. Valls



Tema 5: Redes de Base Radial

Las Redes de Neuronas de Base Radial (RNBR) son redes multicapa con conexiones hacia adelante que disponen de una única capa oculta con funciones de activación no lineal y una capa de salida con activaciones lineales. Las neuronas de la capa oculta tienen carácter local que se consigue gracias al uso de funciones de activación especiales llamadas funciones de base radial.

En este tema haremos una introducción a las RNBR, estudiaremos su arquitectura, así como los diferentes métodos de aprendizaje y finalmente las compararemos con el Perceptron Multicapa.

5.1 Introducción

Como hemos dicho, las RNBR son redes multicapa con conexiones hacia adelante (feed-forward) y con una única capa oculta. Las neuronas de esta capa oculta poseen un carácter local en el sentido de que cada neurona se especializa en una determinada zona del espacio de entrada. Dicho de otro modo, cada neurona oculta se activa preferentemente con datos pertenecientes a una determinada región del espacio de entrada. Puede verse como que cada neurona oculta define una hiperesfera en el espacio de entrada. Dicho espacio estará particionado en múltiples hiperesferas definidas por las correspondientes neuronas ocultas. Este carácter local se debe al uso de funciones de base radial (FBR). Una de las FBR más utilizadas es la función gaussiana.

Las neuronas de salida realizan una combinación lineal de las activaciones de las neuronas ocultas, es decir, de las múltiples funciones locales no lineales.

Las RNBR, al igual que el Perceptron Multicapa, son aproximadores universales y se han aplicado a gran variedad de problemas: diagnóstico médico, procesamiento de imágenes, predicción de series temporales, etc.

5.2 Arquitectura

Poseen tres capas de neuronas: capa de entrada, capa oculta y capa de salida.

La capa de entrada no realiza procesamiento; simplemente se encarga de transmitir las señales de entrada a la capa oculta. Las conexiones con la capa oculta no tienen pesos asociados.

La capa oculta realiza una transformación local y no lineal de dichas señales. La capa de salida tiene activaciones lineales, por tanto, cada neurona de salida realiza una combinación lineal de las funciones locales no lineales. Las conexiones de la capa oculta a la capa de salida sí tienen pesos asociados. En la transparencia 6 puede verse representación gráfica de una RNBR.

Las FBR y por tanto cada neurona oculta tiene dos parámetros: el centro y la desviación. El centro es un vector con la misma dimensión que el espacio de entrada y la desviación es un número real. Las FBR son funciones del cociente entre la distancia euclídea entre el dato de entrada y el centro de la función y su desviación (ver tr. 8). La FBR más habitual es la gaussiana (ver tr. 11). Podemos ver que los datos cercanos al centro de la neurona producirán activaciones altas y que cuanto más se alejen de dicho centro más pequeña será la activación. En un espacio de entrada bidimensional, podemos imaginar cada neurona oculta como una “campana de Gauss” de altura unidad y situada en el centro de dicha neurona. La amplitud o anchura de cada campana dependerá de la desviación de cada neurona (tr. 11). Como cada neurona de la capa de salida realiza una combinación lineal de las activaciones de las neuronas ocultas, producirá una suma ponderada de gaussianas. Los factores de ponderación son los pesos de las conexiones desde la capa oculta a dicha neurona de salida. En las transparencias 12 a 15 puede verse un ejemplo con una entrada unidimensional y una neurona de salida.

5.3 Aprendizaje

En las RNBR, el aprendizaje consiste en encontrar los valores más adecuados para las desviaciones y centros de cada neurona oculta, así como los pesos de las conexiones a la capa de salida. Los métodos más importantes son el **híbrido** y el **totalmente supervisado**.

Método híbrido

Se separa el aprendizaje en dos fases independientes: Fase no supervisada y fase supervisada.

En la **fase no supervisada**, en primer lugar se determinan los **centros** de las neuronas ocultas de forma no supervisada guiándose por la situación de los datos en el espacio de entrada. Para ello puede usarse cualquier algoritmo de agrupación o clustering que divida el espacio de entrada en clases, siendo el representante de cada clase el centro de la neurona oculta correspondiente. Uno de los métodos más utilizados es K-medias donde K sería el número de neuronas ocultas. Este algoritmo divide el espacio de datos en K clases. El representante de cada clase, el centroide C_i , será el centro de la neurona i .

Una vez situados los centros, hay que determinar las **desviaciones** de forma que cada neurona oculta se active en una región del espacio de entrada y que el solapamiento de las zonas de neuronas adyacentes sea lo más pequeño posible. Una opción simple para determinar las desviaciones es calcular para cada neurona, la media geométrica de las distancias a los dos centros más cercanos.

Una vez calculados los centros y desviaciones se pasaría a la **fase supervisada** donde se calculan los pesos y umbrales de las conexiones de salida que minimizan el error cuadrático medio calculado al comparar la salida de la red y la salida esperada (ver tr. 21). Esto puede hacerse de forma **iterativa** (similar al aprendizaje de Adaline) o bien de forma directa por el método de la **pseudoinversa**, ya que es un problema de optimización cuadrática que tiene un mínimo global (el error es función cuadrática de los pesos). Ambos métodos pueden verse en las transparencias 21-26.

Método totalmente supervisado

En este método todos los parámetros (centros, desviaciones y pesos) se calculan de forma supervisada con el objetivo de minimizar el error cuadrático medio. En las transparencias 29 a 32 se explica detalladamente este método y se muestran las leyes de aprendizaje correspondientes para los pesos, centros y desviaciones.

En este método, no se garantiza que las neuronas ocultas conserven sus características locales si se inicializan todos los parámetros aleatoriamente. Sería más adecuado inicializar los centros y desviaciones de forma que representen a regiones del espacio de entrada y partiendo de estos valores se utilizaría el método totalmente supervisado para limitar la búsqueda a ciertas regiones del espacio de datos.

Según esto, parece conveniente **combinar el método híbrido con el totalmente supervisado**: con el método híbrido se fijan los centros, amplitudes y pesos y se adaptan posteriormente de forma supervisada todos los parámetros. Sería un ajuste fino realizado sobre la red resultante del método híbrido. Para no perder lo aprendido con el método híbrido, las razones de aprendizaje utilizadas en el método totalmente supervisado deberán tomar valores pequeños.

5.4 RNBR vs PM

Ambos son aproximadores universales y construyen transformaciones no lineales de los datos de entrada. PM posee carácter global y RNBR carácter local. Estas características se deben a las diferentes funciones de activación que utilizan.

El carácter local de las RNBR generalmente permite un aprendizaje más rápido ya que el cambio en un peso de la red afecta sólo a la neurona oculta conectada con ese peso y por tanto afecta sólo al grupo de patrones de entrada situados en la región correspondiente a esa neurona.

En muchos casos, para poder construir una aproximación mediante la suma de aproximaciones locales se necesitan muchas neuronas ocultas, lo que puede disminuir la capacidad de generalización de estas redes.

Cuando el número de variables de entrada es muy elevado puede ser necesario un número muy grande de neuronas ocultas para poder realizar la aproximación. En estas situaciones, las RNBR no son la mejor opción.