uc3m | Universidad **Carlos III** de Madrid

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

CC BY NC SA

Preguntas y Ejercicios para Evaluación: Tema 5

1. Contestar brevemente a las siguientes cuestiones relacionadas con las Redes de Base Radial:
   a) ¿Qué quiere decir que las neuronas ocultas tienen carácter local?
   b) ¿Cuál es la función de base radial más utilizada? Indicar la ecuación.
   c) ¿Qué hay que aprender en el proceso de entrenamiento de una RNBR?
   d) Explicar las fases del método de entrenamiento híbrido
   e) ¿Qué tipo de aprendizaje puede ser más adecuado para el entrenamiento de las RBR, el método híbrido o el método totalmente supervisado? Razone su respuesta.

Respuesta

a) Que se especializan en zonas del espacio de entrada.
b) La función gaussiana.

$$\phi_i(n) = e^{-\frac{\|X(n)-C_i\|^2}{2d_i^2}}$$

c) Las posiciones de los centros de las neuronas ocultas, las desviaciones y los pesos y umbrales de la capa de salida.
d) Se realiza en dos fases:
   - Fase no supervisada, donde se determinan las coordenadas de los centros de las neuronas ocultas y las desviaciones de las funciones de base radial. Se hace de forma no supervisada de forma que las neuronas representen a los datos en el espacio de entrada.
   - Fase supervisada. Una vez fijados los centros y desviaciones en la fase anterior, se determinan los pesos de la capa de salida de forma supervisada para minimizar el error cuadrático.
e) La combinación de ambos. Mediante el método híbrido se calculan los parámetros para que las neuronas ocultas representen el espacio de entrada. Con los valores obtenidos, se inicializa la red y se aplica el método totalmente supervisado utilizando razones de aprendizaje pequeñas con el objetivo de realizar un ajuste fino de los parámetros y aproximar mejor las salidas deseadas.

2. En un dominio de diagnóstico médico, se ha utilizado una RBR para clasificar los datos de los pacientes en una de las dos clases, C0 y C1, que corresponden a pacientes sanos (clase C0) o enfermos (clase C1). Una vez efectuada la fase de entrenamiento de la red, indique qué operaciones se deben realizar sobre la(s) salida(s) que proporcione la red para evaluar el porcentaje de aciertos de la red sobre dicho problema.

Respuesta:

Se pueden utilizar 2 salidas o 1 salida para la red.
Si se utilizan dos salidas, la codificación de la salida deseada será (1,0) para la clase C0 y (0,1) para la clase C1. Entonces, una vez entrenada la red y debido a que las salidas puede tomar un valor real, habría que calcular el índice para el cual se obtiene el máximo de (y_1, y_2). Si el índice 1, entonces C0 y si es 2, entonces se le asigna la clase C1.
Si se utiliza 1 salida, la codificación de la salida deseada será, por ejemplo, 0 para la clase C0 y 1 para la clase C1. Entonces, una vez entrenada la red y debido a que la salida puede tomar un valor real, para determinar la clase aproximada por la red, bastaría aplicar, por ejemplo, el siguiente criterio: Si salida < 0.5, entonces C0, en otro caso, clase C1.
Conocida la clase que asigna la red a cada patrón de entrada, basta calcular la proporción de patrones respecto del total.

3. Responda a las siguientes preguntas:
   a) Explique las diferencias y similitudes entre el Perceptron Multicapa y las Redes de Base Radial
   b) Dado un Perceptron Multicapa con 2 neuronas de entrada, tres neuronas ocultas (con función de activación sigmoidal) y 1 neurona de salida con función de activación lineal, escriba las ecuaciones para ajustar los pesos de la capa oculta a la capa de salida con el objetivo de minimizar el error $e=1/2(s-o)^2$, siendo s la salida deseada y o la salida de la red.
   c) Dada una Red de Base Radial con 2 neuronas de entrada, tres neuronas ocultas y 1 neurona de salida, escriba las ecuaciones para ajustar los pesos de la capa oculta a la capa de salida utilizando el método del gradiente para minimizar el error $e=1/2(s-o)^2$, siendo s la salida deseada y o la salida de la red.
   d) ¿Qué diferencias y/o similitudes existen entre las leyes de aprendizaje obtenidas en los apartados c) y d)?

Respuesta:

a) Ambas redes construyen transformaciones no lineales de los datos de entrada y son aproximadores universales. Poseen neuronas ocultas. Se pueden utilizar para problemas de regresión, clasificación y predicción.
   Las neuronas del PM tienen un carácter global, mientras que las de las RBR poseen un carácter local, debido a las funciones de activación utilizadas en cada caso
   El carácter local hace que generalmente el aprendizaje de las RBR sea más rápido
   El carácter local puede provocar una mala generalización de las RBR cuando el número de neuronas ocultas es excesivo
b) Sea w1, w2, y w3 los pesos de la capa oculta a la capa de salida. Siguiendo la regla delta generalizada, la modificación de los pesos se realiza teniendo en cuenta el delta de la neurona de salida y la activación de la que procede el peso. Como la función de activación de salida es la lineal, el delta viene dado por el error medido en la salida. Por tanto, los pesos se modifican siguiendo al siguiente expresión

$$w_i(n) = w_i(n-1) + \alpha(s-o)a_i(n)$$

   siendo $a_i(n)$ la activación de la neurona oculta i para el patrón de entrada n, es decir:

$$a_i(n) = f(v_{1i}x_1(n) + v_{2i}x_i(n) + u_i)$$

   v_{ij} son los pesos de la capa de entrada a la capa oculta

c) Sea w1, w2, y w3 los pesos de la capa oculta a la capa de salida, C1, C2, y C3 los centros de las 3 neuronas ocultas y d1, d2, y d3 las amplitudes. Aplicando el método del descenso del gradiente para minimizar el error en la salida, los pesos wi se modifican siguiendo la siguiente ley:

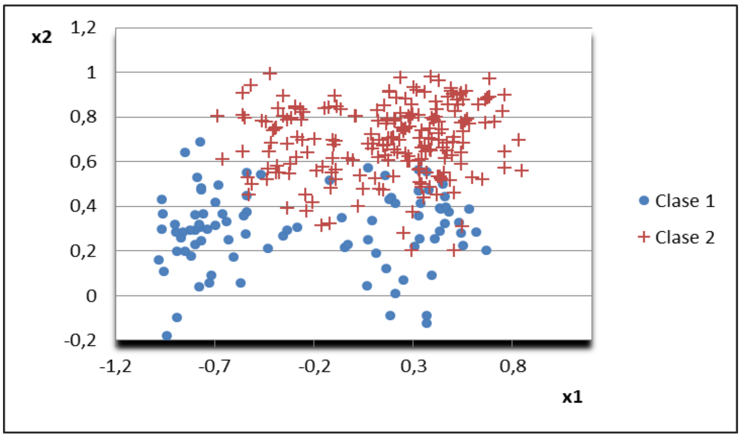
$$w_i\,(n) = w_i\,(n-1) + \alpha(s-o)\phi_i\,(n)$$

siendo $\phi_i(n)$ la activación de la neurona oculta i para el patrón de entrada n. Para las RBR, esta activación viene dada por:

$$\phi_i(n) = e^{-\frac{\|X(n)-C_i\|^2}{2d_i^2}}$$

d) Las leyes obtenidas son las mismas, tienen en cuenta el error que se mide en la salida y la activación de la neurona oculta. Esto es debido a que en ambas redes esos pesos se calcular para minimizar el error en la salida y que el PM tiene función de activación lineal en la salida (al igual que las RBR). La diferencia está en el modo de calcular la activación de las neuronas ocultas en cada red

4. Considérese el problema de clasificación que se muestra en la figura, conociendo la clase deseada para cada patrón.



Supóngase que se entrena un mapa de Kohonen con 2x6 neuronas en la capa de competición utilizando sólo las entradas al problema, de manera que cada neurona del mapa representará una zona del espacio de variables de entrada. Con los centros de dichas neuronas se pretende construir una red de base radial para abordar el problema. Se pide:

a) Indicar la arquitectura de red de base radial a utilizar (entradas, ocultas y salidas)

b) Explicar esquemáticamente los pasos para llevar a cabo el aprendizaje híbrido de la red de base radial

Respuesta:

a) 2 neuronas en la capa de entrada. 12 neuronas ocultas que se corresponden con las neuronas del mapa de Kohonen. 1 neurona de salida

b) **Fase no supervisada.** Se realiza el aprendizaje del mapa de Kohonen para calcular los centros de las funciones de base radial, obteniéndose los 12 centros. Los pasos son:

Paso 1: Se inicializan aleatoriamente los pesos de las neuronas del mapa

Paso 2: Se presenta un patrón de entrada y se propaga hasta la capa de competición, calculando la distancia euclidea del patrón de entrada a las neuronas del mapa

Paso 3: Se selecciona la neurona ganadora, la más cercana

Paso 4: Se modifican los pesos entre la capa de entrada y la neurona, así como las de su vecindad, según su grado de vecindad, siguiendo la ley de aprendizaje

Paso 5: se repiten los pasos 2, 3 y 4 para todos los patrones de entrada

Paso 5. Se decrementa el valor de la razón de aprendizaje. Si α por encima de cierto umbral volver al paso 2, en caso contrario FIN.

Se calculan las desviaciones como la media geométrica de los dos centros más cercanos.

Fase supervisada: Se determinan los pesos y umbrales de la capa de salida siguiendo el siguiente proceso iterativo:

Paso 1: Se inicializan aleatoriamente los pesos y umbrales

Paso 2: Se toma un patrón X(n), se calcula la salida de la red Y(n) y se evalúa el error e(n) cometido por la red para dicho patrón

Paso 3: Se modifican los pesos y umbrales utilizando las leyes de aprendizaje

Paso 4: Se repiten los pasos 2 y 3 para todos los patrones de entrenamiento

Se repiten los pasos 2, 3 y 4 hasta conseguir la convergencia, es decir hasta que la suma de los errores para todos los patrones se estabilice.

5. Disponemos de un conjunto de 1000 datos de viviendas de una determinada ciudad. Cada uno de estos datos (o patrones) se compone de 14 atributos numéricos y del precio actual de la vivienda en miles de euros. A modo de ejemplo, se citan algunos de los atributos:
   - Número de habitaciones
   - Índice de contaminación de óxido nítrico (en partes por millón) de la zona
   - Distancia al centro sanitario más próximo en kilómetros
   - Índice de criminalidad del barrio
   - Etc..

Se desea utilizar estos datos para construir un modelo basado en Redes de Neuronas de Base Radial (RNBR) que permita estimar el precio de otras viviendas similares.

Responder a las siguientes preguntas:

a) ¿Qué tipo de problema queremos resolver? (clasificación, predicción de series temporales, etc……. Supervisado/no supervisado).

b) Los datos están ordenados por el precio de la vivienda. ¿Sería necesario realizar algún preproceso de los datos? Justifique la respuesta.

c) Queremos realizar un esquema de validación cruzada de 10 hojas. ¿De cuántos datos de entrenamiento dispondríamos en cada hoja o partición?

d) ¿Cuántas neuronas de entrada y de salida tendría la red? ¿Es necesario normalizar la variable de salida entre 0 y 1? Justificar la respuesta.

e) Queremos que la red tenga características locales. ¿Qué significa esto?

f) Dentro de los tipos de aprendizaje que pueden utilizarse en RNBR ¿cuál se aconsejaría para mantener las características locales de la red? Justificar la respuesta.

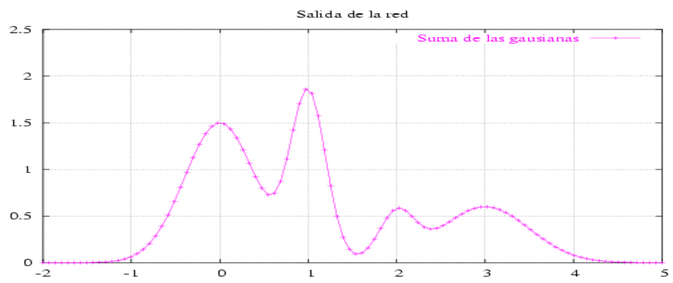
g) ¿Qué algoritmo(s) podría utilizar para para calcular los centros de la red?

h) Indique si alguno de los otros modelos estudiados en la asignatura se puede aplicar para resolver este problema.

Respuesta:

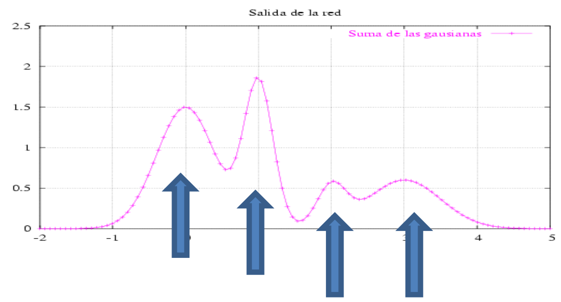
a) Problema de regresión. Aprendizaje supervisado.
b) Es necesarios desordenar los datos para que en los ejemplos de entrenamiento haya ejemplos representativos de todos los datos.
c) 900
d) 14 neuronas de entrada y una neurona de salida. No es necesario porque las neuronas de salida de una RNBR pueden producir cualquier valor real.
e) Que cada neurona oculta represente a una zona del espacio de entrada.
f) Aprendizaje híbrido. Con el aprendizaje totalmente supervisado se pueden perder las características locales de la red.
g) Algoritmos de agrupación no supervisados, como: K-medias y mapas de Kohonen
h) Perceptron Multicapa

6. Dada la función que se muestra en la figura (una variable de entrada y una de salida). ¿Podría aproximarse por una red de base radial? En ese caso, indique cuantas neuronas ocultas sería conveniente utilizar y valores aproximados para los centros, amplitudes y pesos de la red

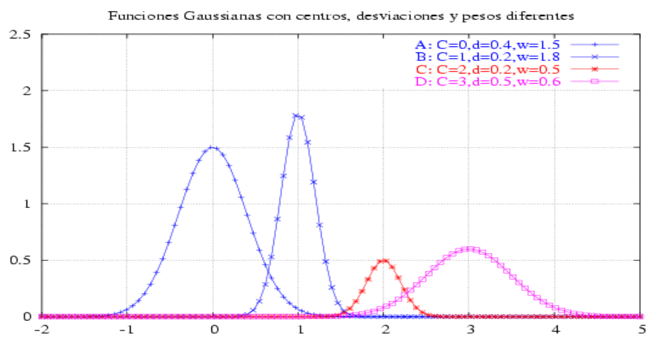


Respuesta

Sí, es una función no lineal. En principio convendría utilizar 4 neuronas ocultas para especializarse en las zonas marcadas en la siguiente figura.



Sería conveniente que las neuronas estuviesen centradas en las zonas marcadas y que cada gausiana estuviese ponderada por un determinado valor para que tuvieran alturas diferentes. La primera debe tener altura 1.5, la segunda 1.8, la tercera 0.5 y la cuarta 0.5 o 0.6. Valores aproximados para los centros, amplitudes y pesos se muestran en la siguiente figura:



Realizando la suma de estas gausianas se obtiene la función que se desea aproximar

7. Se dispone de un conjunto de datos de entrenamiento compuesto por cinco puntos bidimensionales (ver tabla 1) con los que se quiere entrenar, mediante el método híbrido, una Red de Base Radial con dos neuronas ocultas, como puede verse en la figura 1:

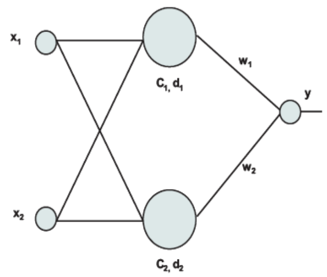


Figura 1

	x1	x2	y
p1	0.1	0.2	0.2
p2	0.3	0.5	0.5
p3	0.9	0.6	0.1
p4	0.5	0.4	0.7
p5	0.7	0.5	0.9

Tabla 1

Los centros y pesos iniciales de la red son los siguientes: $C_1 = (0.5, 0.7)$, $C_2 = (0.9, 0.8)$, $w_1=0.7$, $w_2 = 0.2$, $u=-0.3$

Se pide:

a) Determinar los centros de las neuronas ocultas utilizando el algoritmo K-medias. Realizar dos iteraciones como máximo. Representar gráficamente la situación de los patrones y las distintas posiciones de los centros.

b) Suponiendo que el valor de las activaciones de las neuronas ocultas para el patrón de entrada p1 es $\varphi_1=0.87$ y $\varphi_2=0.32$, respectivamente, determinar los nuevos valores de los pesos (w_1 y w_2) y el umbral tras la presentación del primer patrón (p1). La tasa de aprendizaje utilizada será $\alpha=0.2$.

Respuesta

a) Determinación de los centros. Se aplica el algoritmo k-medias:

Iteración 1

Se calcula la distancia de cada patrón a los centros:

$$d(p_i, c_j) = \sqrt{(p_{i1} - cj1)^2 + (p_{i2} - c_{j2})^2)}$$

dando como resultado:

	distancia c1	distancia c2
p1	0.6403	1
p2	0.2828	0.6708
p3	0.4123	0.2
p4	0.3	0.5657
p5	0.2828	0.3606

Por tanto, p1, p2, p4 y p5 pertenecen a la región cuyo centro es c1 y p3 a la región cuyo centro es c2.

A continuación se calculan los centros de cada una de las regiones, hallando la media aritmética de las coordenadas x1 y x2 de los puntos pertenecientes a cada región

	x1	x2
c1	0.4	0.4
c2	0.9	0.6

media de p1,p2,p4 y p5

coindice con p3

Iteración 2

Se vuelven a calcular las distancias desde cada patrón a cada centro:

	distancia c1	distancia c2
p1	0.3606	0.8944
p2	0.1414	0.6083
p3	0.5385	0
p4	0.1	0.4472
p5	0.3162	0.2236

Ahora se observa que p1, p2 y p4 pertenecen a la región correspondiente a c1 y p3 y p5 a c2.

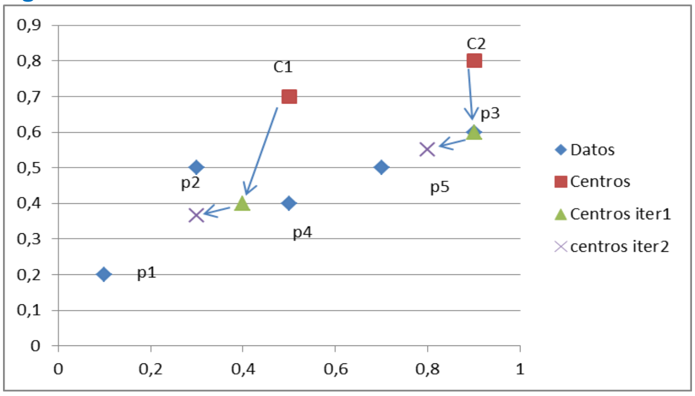
Se vuelven a calcular los centros de las regiones

	x1	x2
c1	0.3	0.367
c2	0.8	0.55

Estos serán ya los centros definitivos.

(Sólo se han pedido dos iteraciones, pero además la composición de las regiones no variaría y por tanto los centros geométricos de cada región serán los mismos y k-medias finalizará)

Representación gráfica



b) La ley de aprendizaje para los pesos viene dada por:

$$w_i(n) = w_i(n) + \Delta w_i$$

Donde el incremento para el patrón p1, viene dado por:

$$\Delta w_1 = \alpha(s - y)\phi_1(p_1)$$

$$\Delta w_2 = \alpha(s - y)\phi_2(p_1)$$

La salida de la red será:

$$y = w_1\phi_1(p1) + w_2\phi_2(p1) + u = 0.7 \times 0.87 + 0.2 \times 0.32 - 0.3 = 0.373$$

Por tanto,

$$\Delta w_1 = \alpha(s - y)\phi_1(p1) = 0.2 \times (0.2 - 0.373) \times 0.87 = -0.03$$

$$\Delta w_2 = \alpha(s - y)\phi_2(p1) = 0.2 \times (0.2 - 0.373) \times 0.32 = -0{,}011$$

$$\Delta u = \alpha(s - y) = 0.2 \times (0.2 - 0.373) = -0{,}0346$$

Los valores de los pesos serán:

$$w_1 = 0.7 - 0.03 = 0.67$$

$$w_2 = 0.2 - 0.011 = 0.09$$

$$u = -0.3 - 0.0346 = -0.3346$$