



REDES NEURONALES FUNCIONES DE BASE RADIAL

Práctica 3

Introducción a los Modelos Computacionales
Grado en Ingeniería Informática
Universidad de Córdoba – Escuela Politécnica Superior

Índice

Índice.....	1
Descripción del modelo neuronal.....	2
Experimentos y análisis de los resultados	3
Descripción de las bases de datos	3
Archivos <i>.dat</i>	3
Archivos <i>.csv</i>	3
Descripción de los parámetros considerados.....	3
Resultados obtenidos	4
Comparación de arquitecturas	4
Valores de eta	6
Clasificación como regresión	8
Análisis de los resultados.....	8
Matriz de confusión	9
Bibliografía	10

Descripción del modelo neuronal

En esta ocasión se ha realizado un modelo de red neuronal de funciones de base radial (RBF), con una sola capa oculta. Estas redes se basan en una aproximación local. Cada RBF guarda un centro como referencia, y cada vez que se le presenta un patrón nuevo, se calcula la distancia a dicho centro. Esta distancia se compara con el radio que tiene almacenado dicha neurona también. Si este radio es pequeño, la activación se producirá sólo cuando el patrón esté muy cerca del centro. Si la distancia está dentro del radio la RBF se activa (salida = 1), si no la RBF no se activa (salida = 0).

Se ha usado la función Gaussiana como función RBF:

$$\varphi(\mathbf{x}, \mathbf{c}, \sigma) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2(\sigma)^2}\right)$$

Donde $\|\mathbf{x} - \mathbf{c}\|$ es la norma del vector diferencia entre el centro y el patrón, es decir, la distancia euclídea entre el centro de la RBF y el patrón de entrada.

El ajuste de los parámetros se ha realizado con un entrenamiento híbrido, es decir, una parte no supervisada (clustering) y una parte supervisada (regresión logística o inversión de una matriz). Haciendo esto el coste computacional es menor que con un algoritmo de retropropagación, ya que para implementar éste habría que calcular las derivadas respecto al radio y a los centros, las cuales son complejas y por ende tienen un mayor coste computacional.

En el entrenamiento hay que obtener tres cosas:

- Coordenadas de los centros de las RBF: pesos de capa de entrada a oculta
- Ancho de las RBF: se usa el hueco del sesgo
- Pesos de capa oculta a capa de salida (con sesgo)

El ajuste de los centros de la red se realiza mediante un proceso de *clustering*, usando el algoritmo K-medias. Las coordenadas de los centroides de cada *cluster* serán las coordenadas de los centros de las RBF. Al algoritmo hay que decirle el número de *clusters*, que en este caso será el número de neuronas de la RBF. Los puntos se asignan al *cluster* cuyo centroide esté más cerca e iterativamente se van actualizando los centroides en función de las asignaciones de puntos a *clusters*, hasta que los centroides dejen de cambiar. Los radios se calcularán tomando la mitad de la distancia media al resto de centroides.

Para ambos tipos de problema se necesita la matriz R de salidas, en la que cada fila corresponde a cada una de las salidas de las neuronas RBF para un patrón de entrada. Para simular el sesgo se añade una columna de unos.

- **Clasificación:** se aplica una regresión logística, que aplicatambién la regularización, que se usa para lograr que los parámetros β_{ji} tiendan a valores muy pequeños. El objetivo de la regresión es obtener los valores β_{ji} que maximizan la entropía cruzada.
- **Regresión:** los β_{ji} se calculan a través de una ecuación de matrices. Despejando, la matriz de los β_{ji} es igual a la inversa de la matriz R multiplicada por la matriz de salidas esperadas. Si la matriz R no tiene inversa, se utilizará la pseudo-inversa de Moore Penrose.

Experimentos y análisis de los resultados

Descripción de las bases de datos

Esta vez tenemos dos formatos para las bases de datos, pudiendo utilizar cualquiera de las dos. Uno de los formatos es el mismo que las prácticas anteriores (los archivos *.dat*), el otro es un nuevo formato que está preparado para abrir como una hoja de cálculo (los archivos *.csv*)

Archivos *.dat*

Todos los ficheros que contienen las bases de datos tienen el mismo formato:

- En la primera línea se encuentran 3 enteros que indican el número total de entradas del problema(n), el número total de salidas del problema(k) y el número total de patrones en dicho fichero(p).
- A partir de la segunda línea, cada línea corresponde a un patrón:
 - Los primeros n valores corresponden a las entradas.
 - Los siguientes k valores son las salidas deseadas.

Archivos *.csv*

Cada línea del fichero corresponde a un patrón, separando las variables de los patrones por comas, es decir, las entradas del problema. Las salidas esperadas se corresponden con el último valor de cada fila de los patrones, es decir, la última columna de la matriz que forman.

Descripción de los parámetros considerados

Los parámetros que se han considerado en esta práctica para el estudio del comportamiento de la red RBF son los siguientes:

- Semilla: Es un entero que indica el número con el que se iniciará la semilla para el cálculo de los números aleatorios. Es importante variarla ya que la inicialización de los centros de los *clusters* para el algoritmo de K-medias es aleatoria y determinante para la precisión del modelo.
- Nº de neuronas en capa oculta: Es un entero que indica el número de neuronas que tendrá la capa oculta. Esto es igual al número de RBFs del que estará compuesta la red, o lo que es lo mismo, el número de clusters que habrá en el algoritmo del K-medias. Mientras mayor sea el número, más compleja será la red.
- Valor de η (η): Es un valor real entre 0 y 1 que indica el factor de importancia de la regularización en problemas de clasificación.

Resultados obtenidos

En este apartado se explicarán los diferentes experimentos que se han realizado con el programa desarrollado y sus resultados, con el fin de estimar la configuración de los parámetros óptima para cada uno de los problemas, es decir, encontrar la serie de parámetros con la que obtendremos los mejores resultados posibles. Se harán experimentos con la arquitectura de la red neuronal utilizada y con el cambio en el factor de importancia que se le da a la regularización. Finalmente se mostrará la configuración con mejores resultados que será la que se analizará en el apartado siguiente.

Comparación de arquitecturas

Se va a lanzar el script modificando el número de neuronas de capa oculta y se utilizará el parámetro $\eta = 10^{-5}$. El número de neuronas será igual al 5%, 10%, 25% y 50% del número de patrones de la base de datos.

SIN

	Entrenamiento		Test	
Neuronas	Media MSE	Desviacion Típica MSE	Media MSE	Desviacion Típica MSE
5%	0,014019985	0,00020752	0,02246506	0,000250051
10%	0,012546304	9,34E-05	0,035542283	0,004298083
25%	0,010265067	0,000188164	1,890850582	0,984461068
50%	0,010246715	0,000155226	3,531489103	3,384123597

Como se puede observar en la tabla, el error cometido por el modelo realizado es bajo, pero no hay una diferencia demasiado grande con los resultados obtenidos por la regresión lineal realizada con Weka. La mejor configuración sería la que usa un 5% del número de patrones en neuronas. Esto se debe a que el problema es bastante simple y no hace falta realizar un modelo demasiado complejo para simular la función, ya que como se aprecia, podría llegar a ser contraproducente, pues habría demasiados pesos a ajustar y el modelo no responde correctamente.

CPU

	Entrenamiento		Test	
Neuronas	Media MSE	Desviacion Típica MSE	Media MSE	Desviacion Típica MSE
5%	0,00517516	0,002300331	0,006600704	0,002414999
10%	0,003686503	0,000901278	0,004506535	0,000378401
25%	0,001256396	0,001014002	0,00331014	0,001175001
50%	0,000271825	0,000025923	0,002886968	0,000532859

Al contrario que en la anterior, a medida que se incrementa el número de neuronas RBF, más preciso es el modelo. Esto se debe a que el problema es bastante más complejo que el anterior, por lo que necesita un modelo más grande para que sea más preciso. El mejor modelo es el que usa un número de neuronas igual al 50% del número de patrones, aunque llega a sobrentrenar, ya que obtiene mejores resultados en entrenamiento que en test.

IRIS

	Entrenamiento			
Neuronas	Media MSE	Desviacion Típica MSE	Media CCR	Desviación Típica CCR
5%	0,035714286	0	96,42857143	0
10%	0,017857143	0	98,21428571	0
25%	0,016071429	0,003571429	98,39285714	0,357142857
50%	0,007142857	0,006681531	99,28571429	0,668153105

	Test			
Neuronas	Media MSE	Desviacion Típica MSE	Media CCR	Desviación Típica CCR
5%	0,068421053	0,035696474	93,15789474	3,56964736
10%	0,052631579	0	94,73684211	0
25%	0,084210526	0,019692934	91,57894737	1,969293362
50%	0,094736842	0,012892051	90,52631579	1,289205128

Como se puede apreciar, para un problema de clasificación, si el modelo es más complejo, mejor va a clasificar en el entrenamiento, pues el tamaño del modelo es proporcional al de la base de datos, pero en test no ocurre esto, por la misma razón que con el problema del seno. El modelo sobrentrena porque el problema es bastante simple y basta con un porcentaje bajo del número de patrones como número de neuronas. En este caso un 10% del número de patrones es el óptimo para tener el mejor modelo.

DIGITS

	Entrenamiento			
Neuronas	Media MSE	Desviacion Típica MSE	Media CCR	Desviación Típica CCR
5%	0,583987441	0,055666933	97,34693878	0,207078587
10%	0,005651491	0,011302983	99,98430141	0,031397174
25%	0	0	100	0
50%	0	0	100	0

	Test			
Neuronas	Media MSE	Desviacion Típica MSE	Media CCR	Desviación Típica CCR
5%	1,278996865	0,191935564	92,9153605	0,807780484
10%	1,248902821	0,246217042	94,29467085	0,802899591
25%	1,262695925	0,153475731	94,85893417	0,250783699
50%	1,297178683	0,103730971	94,35736677	0,343399723

El problema de digits es el más complejo de todos los que se han tratado, por lo que con una red más compleja los resultados serán mejores, aunque siempre sobrentrena. Como se puede apreciar, con un número de neuronas/clusters igual al 25% y 50% del número de patrones del conjunto de entrenamiento, el CCR es del 100% para el entrenamiento, por lo que entrena perfectamente, aunque no lo consigue para el conjunto de test. En test los resultados son ligeramente mejores para el de 25%, por lo que se elegirá como modelo para seguir con los experimentos, ya que es un modelo que requiere menos coste computacional.

Valores de eta

Aquí se irá modificando el valor de eta para los problemas de clasificación, de forma que se pueda apreciar cómo influye la regularización a la hora de clasificar con la regresión lineal que se usa de capa oculta a la salida.

IRIS

Eta	Entrenamiento			
	Media MSE	Desviación Típica MSE	Media CCR	Desviación Típica CCR
0	0,0625	0	93,75	0
10e-1	0,030357143	0,004374089	96,96428571	0,437408883
10e-2	0,035714286	0	96,42857143	0
10e-3	0,021428571	0,004374089	97,85714286	0,437408883
10e-4	0,017857143	0	98,21428571	0
10e-5	0,016071429	0,003571429	98,39285714	0,357142857
10e-6	0,016071429	0,008748178	98,39285714	0,874817765
10e-7	0,007142857	0,008748178	99,28571429	0,874817765
10e-8	0,003571429	0,007142857	99,64285714	0,714285714
10e-9	0,003571429	0,007142857	99,64285714	0,714285714
10e-10	0,003571429	0,007142857	99,64285714	0,714285714

Eta	Test			
	Media MSE	Desviación Típica MSE	Media CCR	Desviación Típica CCR
0	0,02631579	0	97,36842105	0
10e-1	0,005263158	0,010526316	99,47368421	1,052631579
10e-2	0,036842105	0,012892051	96,31578947	1,289205128
10e-3	0,052631579	0	94,73684211	0
10e-4	0,052631579	0	94,73684211	0
10e-5	0,052631579	0,016643567	94,73684211	1,664356663
10e-6	0,052631579	0,016643567	94,73684211	1,664356663
10e-7	0,042105263	0,026836945	95,78947368	2,683694481
10e-8	0,047368421	0,030689221	95,26315789	3,06892205
10e-9	0,063157895	0,021052632	93,68421053	2,105263158
10e-10	0,042105263	0,026836945	95,78947368	2,683694481

Como se puede apreciar en la tabla, mientras más alto es el valor de *eta* más preciso es el modelo en test, esto es porque la regularización genera modelos más simples, es decir, mientras más importancia se le dé (valor de *eta* más alto), más simple va a ser el modelo de regresión lineal que se usa en la capa de salida. Por el contrario, con el conjunto de entrenamiento el modelo llega a ser más preciso mientras menos importancia se le dé a la regularización. Este comportamiento se debe a que el modelo se ha entrenado con dichos patrones, por lo que mientras más cercana sea la regresión lineal al modelo creado por las RBFs más preciso será para este conjunto. Sin embargo, lo que se busca en estos problemas es que la generalización sea lo mejor posible, más que el entrenamiento, por lo que el valor de *eta* = 10e-1 es el óptimo para este problema.

DIGITS

Entrenamiento				
Eta	Media MSE	Desviación Típica MSE	Media CCR	Desviación Típica CCR
0	1,665620094	0,075017239	93,01412873	0,210618648
10e-1	0,606593407	0,073232521	97,22135008	0,2920106
10e-2	0,082417582	0,06041407	99,71742543	0,127535925
10e-3	0	0	100	0
10e-4	0	0	100	0
10e-5	0	0	100	0
10e-6	0	0	100	0
10e-7	0	0	100	0
10e-8	0	0	100	0
10e-9	0	0	100	0
10e-10	0	0	100	0

Test				
Eta	Media MSE	Desviación Típica MSE	Media CCR	Desviación Típica CCR
0	1,369905956	0,037929739	90,90909091	0,198261922
10e-1	1,350470219	0,113605985	93,54231975	0,319687744
10e-2	1,344200627	0,087424291	94,54545455	0,153573025
10e-3	1,194984326	0,077205137	94,98432602	0,280384699
10e-4	1,262695925	0,153475731	94,85893417	0,250783699
10e-5	1,238871473	0,139720509	94,79623824	0,376175549
10e-6	1,334796238	0,142271407	94,48275862	0,376175549
10e-7	1,262695925	0,077357727	94,48275862	0,153573025
10e-8	1,310971787	0,064683237	94,10658307	0,639375488
10e-9	1,478996865	0,165063023	93,9184953	0,757557741
10e-10	1,487147335	0,152243986	93,85579937	0,645494053

En esta ocasión ocurre exactamente igual que con el anterior problema, sólo que, al ser este problema bastante más complejo, la importancia de la regularización debe ser media, ya que se necesita un modelo más complejo para obtener unos resultados óptimos. El valor de $\eta = 10e-3$ es el más adecuado para este problema, ya que proporciona el mayor CCR en test, siendo el CCR de entrenamiento prácticamente irrelevante porque para la mayoría de los valores de η se consigue un 100%.

Clasificación como regresión

Ahora se ejecutará el script con un problema de clasificación con las mejores configuraciones, pero simulando que es un problema de regresión, para lo que el CCR se ha calculado redondeando las predicciones al entero más cercano.

Problema	Test			
	Media MSE	Desviación Típica MSE	Media CCR	Desviación Típica CCR
iris	0,030357143	0,004374089	96,96428571	0,437408883
iris	0,03827216	0,0160274	95,35714286	1,731314235
digits	0	0	100	0
digits	1,063203255	0,049921089	48,94819466	1,399377375

Problema	Entrenamiento			
	Media MSE	Desviación Típica MSE	Media CCR	Desviación Típica CCR
iris	0,005263158	0,010526316	99,47368421	1,052631579
iris	0,037614361	0,01247332	94,73684211	2,882750303
digits	1,194984326	0,077205137	94,98432602	0,280384699
digits	1,695832598	0,030209719	36,86520376	1,712413829

En las tablas, el primer resultado de cada problema corresponde al resultado obtenido considerando el problema como un problema de clasificación y el segundo resultado corresponde al resultado obtenido si se considera como un problema de regresión.

Como se puede observar, para el problema de iris, los resultados de la regresión son bastante parecidos a los obtenidos en la clasificación. Esto indica que el problema es linealmente separable, por lo que con una simple regresión logística se puede clasificar. Sin embargo, con el problema digits no ocurre lo mismo, así que se puede llegar a la conclusión de que este problema es más complejo y no es linealmente separable y necesita ser tratado como un problema de clasificación.

Análisis de los resultados
















En este último apartado se mostrarán las gráficas de convergencia del perceptrón multicapa para cada uno de los problemas con los valores de los parámetros acordados anteriormente. En las gráficas se representarán en el eje de abscisas el número de iteraciones y en el de ordenadas el error cometido en la estimación. El error cometido en entrenamiento se representa con la línea azul y el cometido por el de test se representa por una línea naranja.

Matriz de confusión

Aquí se presenta la matriz de confusión de la mejor ejecución del problema digits para la mejor configuración que se ha obtenido a lo largo de los experimentos realizados. En la matriz de confusión encontramos en la diagonal el número de patrones que corresponden a ese número que han sido bien clasificados. En las columnas tenemos las salidas esperadas y en las filas las salidas estimadas.

$$\begin{pmatrix} 32 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 29 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 32 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 31 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 31 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 30 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 32 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 29 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 30 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 & 0 & 2 & 1 & 28 \end{pmatrix}$$

A continuación, se podrán ver los patrones que han fallado, junto con lo que se cree que el modelo ha interpretado de ellos. Puede que la interpretación dada no sea la correcta, pero se podría tomar como válida, pues dichos patrones son bastante confusos. Se llega a la conclusión por tanto que es normal que el modelo falle en estos números.

-  Sería el 0 que se ha tomado como un 9.
-  Sería el 8 que se ha tomado como un 9.
-  Sería el 5 que se ha tomado como un 2.
-  Sería el 3 que se ha tomado como un 9.
-  Sería uno de los 1 que se ha tomado como un 9.
-  Sería el 1 que se ha tomado como 2.
-  Sería el otro 1 que se ha tomado como un 9.
-  Sería el 1 que se ha tomado como un 4.
-  Sería el 4 que se ha tomado como un 7.
-  Sería el 9 que se ha tomado como 2.
-  Sería el 9 que se ha tomado como 7.
-  Sería el 9 que se ha tomado como 3.
-  Sería uno de los 7 que se ha tomado como 9.
-  Sería el otro de los 7 que se ha tomado como 9.
-  Sería el 7 que se ha tomado como 4.

Bibliografía

- Apuntes, transparencias y material referente a esta práctica de la asignatura en la plataforma Moodle.
- Documentación de las librerías scikit-learn, NumPy y Pandas.