



Universidad
Carlos III de Madrid

Grado en Ingeniería Informática

Curso 2020/2021

Redes de Neuronas Artificiales

Problema de Clasificación: Parte I

Clasificación de imágenes del cielo con el Perceptrón Multicapa

Autores:

Alba Reinders Sánchez
Alejandro Valverde Mahou

100383444
100383383

Índice

1. Introducción	3
2. Preparación de los datos	3
2.1. Normalizar	3
2.2. Preparar los datos para Validación Cruzada Estratificada	3
3. Experimentación y análisis de resultados	4
3.1. Experimento 1	4
3.2. Experimento 2	4
3.3. Experimento 3	5
3.4. Experimento 4	5
3.5. Experimento 5	5
3.6. Experimento para intentar reducir el sobreaprendizaje	6
3.7. Análisis del mejor experimento	6
3.8. Comparativa de los mejores experimentos	8
4. Conclusión	9

1. Introducción

El problema planteado consiste en clasificar imágenes del cielo con el objetivo de ayudar a la estimación de la radiación solar que incide en un lugar, ya que la cantidad de esta radiación varía si hay nubes o no, y del tipo de las nubes.

Para ello, se hace una simplificación del problema real reduciendo a tres posibles clases las imágenes:

- Cielo Despejado
- Nube (*sólo un tipo de nube*)
- Multinube (*varios tipos de nube*)

Dado que en esta primera parte de la práctica se trabaja con un **Perceptrón Multicapa**, es necesario transformar la información de la imagen en distintos atributos numéricos.

Los datos que se usan en esta práctica provienen del *grupo MATRAS de la Universidad de Jaén* y contienen estadísticos y transformaciones de imágenes del cielo completo, que permiten al Perceptrón Multicapa realizar la clasificación.

El conjunto de datos tiene **717** instancias con **12** atributos de entrada y **1** atributo de salida, la clase.

- | | |
|--------------------------------------|--------------------------------------|
| 1. Media del canal azul | 7. Diferencia de medias Verde – Azul |
| 2. Media del canal rojo | 8. Entropía del canal azul |
| 3. Desviación típica del canal azul | 9. Energía del canal azul |
| 4. Sesgo del canal azul | 10. Contraste del canal azul |
| 5. Diferencia de medias Rojo – Verde | 11. Homogeneidad del canal azul |
| 6. Diferencia de medias Rojo – Azul | 12. Cobertura |

2. Preparación de los datos

Se debe hacer un preprocesado de los datos para poder realizar de manera más efectiva el entrenamiento de la red. Primero se normaliza el conjunto de datos y después se divide en 4 subconjuntos, ya que se lleva a cabo *validación cruzada estratificada* de 4 hojas.

2.1. Normalizar

Es necesario normalizar los datos de entrada, dado que cada atributo suele tener rangos de valores muy diferentes. Normalizarlos en el intervalo $[0,1]$ evita posibles sesgos generados por esta diferencia de rangos de los atributos durante el aprendizaje de la red.

2.2. Preparar los datos para Validación Cruzada Estratificada

Se realiza esta técnica debido a que las clases están desbalanceadas, por tanto no se puede aplicar la validación cruzada normal y se tiene que llevar a cabo la *estratificada*.

Para realizar la división de los datos en 4 hojas que mantengan la proporción de instancias de las 3 clases, se crean 4 subconjuntos de las instancias de cada clase y después se junta un subconjunto de cada clase en cada hoja.

Una vez se tienen las 4 hojas, se crean 4 parejas de ficheros para componer los datos de entrenamiento y test. Combinando las distintas hojas de distinta forma, dejando siempre una para el test. Por último se aleatorizan los datos de entrenamiento para evitar posibles sesgos.

3. Experimentación y análisis de resultados

La experimentación con el *Perceptrón Multicapa* se basa en ir haciendo experimentos de manera progresiva intentado conseguir la mejor configuración posible, de forma que tanto el error de entrenamiento como el error de test sea mínimo. Además, dado que las clases no están balanceadas, se tiene que tener en cuenta a su vez el porcentaje de aciertos, siendo esta última medida sobre el conjunto de test la que se va a usar como determinante para decidir qué configuración es mejor que otra.

Para ello se van variando los hiperparámetros de la red: la **topología** de la red, la **razón de aprendizaje (RA)** y el **número ciclos máximo**. Los distintos experimentos que se crean mantienen la topología pero se va modificando la RA que se utiliza.

Los resultados que se muestran en las distintas tablas son las medias de resultados de las 4 hojas de la *validación cruzada estratificada*. Destacar que el número de ciclos máximos se fija a 10000 en un primer momento, y se probará a modificarlo más adelante.

3.1. Experimento 1

Es el caso base desde el que se parte, se tiene **una única capa oculta con 10 neuronas**. A continuación en la tabla se muestra los resultados obtenidos para las distintas razones de aprendizaje.

RA	Error entrenamiento	% de aciertos entrenamiento	Error test	% de aciertos test
0,001	0,3344952944	0,7629008897	0,4235530262	0,6738700565
0,0007	0,3549707553	0,3823117112	0,4343040342	0,3219868173
0,01	0,2505371572	0,839641527	0,4613691504	0,6432438795
0,005	0,2792817569	0,7991775295	0,4438048242	0,6474811676

Los resultados que se obtienen no son muy prometedores, partiendo del primer subexperimento (se denominará como el original) con RA de 0,001, se intenta ir aumentando o disminuyendo su valor en los siguientes subexperimentos para encontrar la mejor RA para esta topología. Después de probar con diferentes valores, se concluye con que la mejor RA es la original: 0,001.

Se ha podido observar que cuando se disminuye la RA los errores son similares pero el porcentaje de aciertos, tanto en entrenamiento como en test disminuye prácticamente a la mitad. Esto puede deberse a que en la mayoría de casos está prediciendo la clase mayoritaria (**nube**), lo cuál hace que el modelo tenga un error que no es del todo malo, pero el porcentaje de aciertos, como es de esperar, es malo.

Por otro lado, al aumentar la RA, en entrenamiento el error disminuye y el porcentaje de aciertos aumenta, lo que puede ser una buena señal. Pero en test el error aumenta y el porcentaje de aciertos disminuye, lo cuál indica un sobreaprendizaje, haciendo que estos modelos sean peores que el original. Aún así, el modelo original muestra también un claro sobreaprendizaje, lo que se intentará reducir en los siguientes experimentos.

3.2. Experimento 2

Al igual que en el experimento anterior, se tiene una única capa oculta pero se aumenta el **número de neuronas a 100** para comprobar si esto afecta positivamente a los resultados. Los resultados obtenidos para las distintas razones de aprendizaje son:

RA	Error entrenamiento	% de aciertos entrenamiento	Error test	% de aciertos test
0,001	0,3410726584	0,7643001241	0,4513887673	0,6627354049
0,005	0,2828611825	0,8001163873	0,4455312415	0,6613700565
0,0008	0,3488769075	0,7582428098	0,4522441869	0,6669256121
0,0006	0,3566408764	0,7521984275	0,4524900392	0,5525188324

Tal y como se ve en la tabla, el aumentar el número de neuronas no ha hecho que los resultados mejoren. Primero se aumenta la RA, y en este caso, los valores son muy similares al caso original con 0,001 de RA.

Dado que aumentar la *RA* no genera mejores resultados, se prueba a disminuirla. El tercer subexperimento proporciona mejores resultados en el conjunto de test, por lo que se prueba a disminuir todavía más, pero esto empeora los resultados.

Por tanto, el mejor subexperimento para la topología de este experimento es el que tiene una *RA* de *0,0008*. Sin embargo, no se observa una mejora respecto al mejor del **Experimento 1**, por lo que se puede concluir que aumentar el número de neuronas con una única capa no genera un cambio significativo.

3.3. Experimento 3

En este caso se aumenta el **número de capas ocultas a 2, con 10 neuronas cada una**. Se intenta probar si una topología algo más compleja obtiene mejores resultados. Los resultados obtenidos para las distintas razones de aprendizaje son:

<i>RA</i>	Error entrenamiento	% de aciertos entrenamiento	Error test	% de aciertos test
0,001	0,3760381966	0,3692892613	0,4430659371	0,2372175141
<i>0,01</i>	<i>0,2229600868</i>	<i>0,851228533</i>	<i>0,5159168119</i>	<i>0,6447975518</i>
0,03	0,08668297523	0,9432547072	0,662564028	0,6235640301
0,008	0,2439436857	0,8307702255	0,4893574402	0,6433145009

Al principio, los resultados son muy malos, pues tanto el porcentaje de aciertos en entrenamiento como en test es bajísimo para una *RA* de 0,001. Por lo que se intenta probar a aumentarla de forma drástica a 0,01. Esto hace que se obtengan mejores resultados en entrenamiento, pero en test, aunque también mejora, se sigue apreciando bastante sobreaprendizaje.

Después se intenta ajustar la *RA* un poco más, tanto aumentándola como disminuyéndola, pero no se obtienen mejores resultados, por lo que el mejor subexperimento para la topología de este experimento es el que tiene una *RA* de *0,01*.

3.4. Experimento 4

Ya que los resultados que se obtienen con 2 capas ocultas todavía no han conseguido mejorar los valores del caso base, se prueba con una topología ligeramente más compleja: **3 capas ocultas** con 10 neuronas cada una. Los resultados obtenidos para las distintas razones de aprendizaje son:

<i>RA</i>	Error entrenamiento	% de aciertos entrenamiento	Error test	% de aciertos test
0,001	0,4362842418	0,06694599628	0,4363116735	0,06694915254
<i>0,01</i>	<i>0,2587682627</i>	<i>0,8191754604</i>	<i>0,4534598606</i>	<i>0,6962806026</i>
0,008	0,2977744114	0,6734171322	0,4198154056	0,5252118644

El primer subexperimento no era muy prometedor. Sin embargo, al aumentar la *RA* a *0,01* se consiguen los mejores resultados hasta el momento, mejorando incluso los valores del **Experimento 1**.

Dado que aumentar la *RA* parece positivo para esta topología, se prueba a aumentar más su valor. Pero al hacer esto se observa que la evolución del error oscila mucho demostrando así que la *RA* es demasiado alta y por ello no se puede aumentar más allá de 0,01.

Se intenta probar a disminuirla con el objetivo de encontrar la *RA* óptima, pero los mejores resultados siguen siendo los obtenidos con el experimento que tiene una *RA* de *0,01*. Ya que es con el que mayor porcentaje de aciertos en test genera.

3.5. Experimento 5

Como con una topología más compleja se consiguen mejores resultados, se decide seguir aumentando esta complejidad con: **4 capas ocultas** con 10 neuronas cada una. Los resultados obtenidos para las distintas razones de aprendizaje son:

RA	Error entrenamiento	% de aciertos entrenamiento	Error test	% de aciertos test
0,001	0,4363243875	0,06694599628	0,4362894737	0,06694915254
0,01	0,436545285	0,06694599628	0,4363050425	0,06694915254
0,02	0,365291525	0,1727989861	0,4416298473	0,1692561205

Los resultados de los distintos subexperimentos no son buenos en ningún caso. Si se intenta introducir una *RA* mayor a 0,02, los valores de error varían demasiado, indicando que es demasiado alta. Por debajo de 0,01 se estancan en valores muy similares a los obtenidos con 0,01 y 0,001, por tanto, se puede decir que esta topología no es buena para resolver el problema planteado.

A pesar de que su valor es muy inferior al de resto de experimentos, el mejor resultado es el que se obtiene con una *RA* de 0,02.

3.6. Experimento para intentar reducir el sobreaprendizaje

Como que se aprecia que todos los experimentos sufren de bastante sobreaprendizaje, se intenta reducirlo *limitando el número de ciclos en el entrenamiento*. Se limita al número de ciclos en el que obtienen su error más bajo en el conjunto de test. Esto se va a realizar sobre los mejores subexperimentos de cada experimento, listados en la siguiente tabla resumen:

Nombre Exp.	Error entrenamiento	% de aciertos entrenamiento	Error test	% de aciertos test
Exp. 1-1	0,3344952944	0,7629008897	0,4235530262	0,6738700565
Exp. 2-3	0,2229600868	0,851228533	0,5159168119	0,6447975518
Exp. 3-2	0,2229600868	0,851228533	0,5159168119	0,6447975518
Exp. 4-2	0,2587682627	0,8191754604	0,4534598606	0,6962806026
Exp. 5-3	0,365291525	0,1727989861	0,4416298473	0,1692561205

Para llevar a cabo esta tarea se realiza el mismo acercamiento que se hizo en la primera práctica para poder quedarnos con el número de ciclos óptimo que minimiza el error. Se repiten los subexperimentos de la anterior tabla pero esta vez aplicando esta técnica.

Nombre Exp.	Error entrenamiento	% de aciertos entrenamiento	Error test	% de aciertos test	Núm. ciclos
Exp. 1-1	0,3658924266	0,4649493068	0,4014796408	0,5585922787	7180,25
Exp. 2-3	0,3983796585	0,2805891786	0,4175185209	0,2322033898	3754
Exp. 3-2	0,3545379908	0,3487766398	0,3796682922	0,33173258	2142,75
Exp. 4-2	0,3486060242	0,3504034761	0,3776217959	0,3294020716	4918,75
Exp. 5-3	0,3860130839	0,1727989861	0,4110949435	0,1692561205	8145

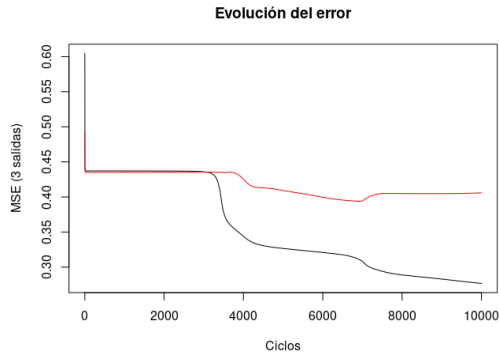
Se puede apreciar en la tabla anterior que hacer esto no genera mejores resultados. De hecho, genera peores. Esto se debe a que minimizar el error que comete la red no es lo mismo que aumentar el porcentaje de aciertos, dado que lo que se busca es maximizar este porcentaje de aciertos, usar esta técnica no resulta apropiada en este caso.

3.7. Análisis del mejor experimento

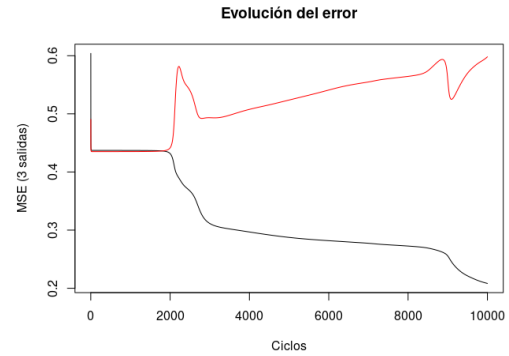
El experimento que genera mejores resultados es el **Experimento 4-2**, con un porcentaje de aciertos sobre el conjunto de test de **0,6962806026**, una topología de 3 capas ocultas con 10 neuronas cada una, una *RA* de 0,01 y 10000 ciclos.

Nombre Exp.	Error entrenamiento	% de aciertos entrenamiento	Error test	% de aciertos test
Exp. 4-2	0,2587682627	0,8191754604	0,4534598606	0,6962806026

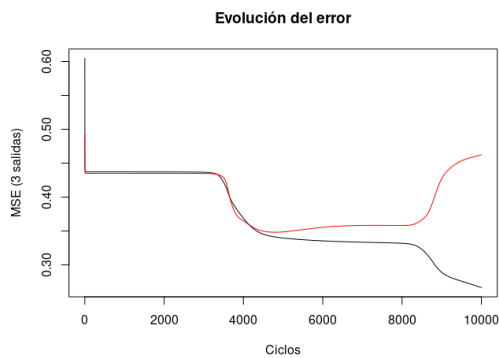
La evolución del error sobre el conjunto de test y el conjunto de entrenamiento a lo largo del aprendizaje se puede ver en las siguientes gráficas para cada una de las hojas:



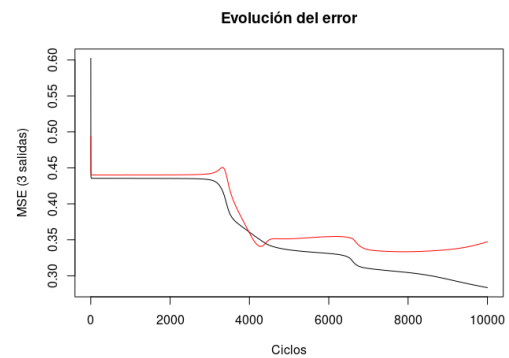
Hoja 1



Hoja 2



Hoja 3



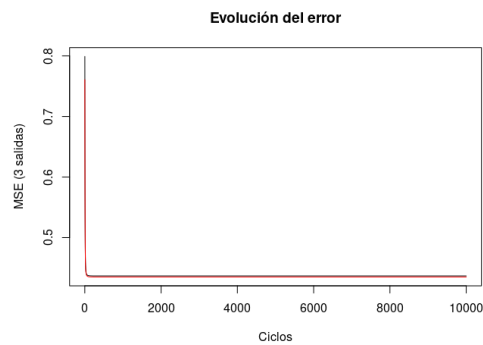
Hoja 4

Se observa claramente que, a pesar de que este sea el mejor experimento conseguido, sigue produciendo mucho sobreaprendizaje en el error. En algunos casos más grande que en otros.

Por ejemplo, la hoja 1 y la hoja 4 son las que sufren menor sobreaprendizaje. Mientras que la hoja 2 y la hoja 3 sufren más sobreaprendizaje, especialmente la 2.

Se cree que en los primeros ciclos, cuando el error se mantiene estable, quiere decir que la red ha aprendido a determinar que todos, o casi todos, los ejemplos pertenecen a la clase mayoritaria. Es fundamental que un buen modelo sea capaz de superar ese umbral porque si no es capaz de hacerlo, su porcentaje de aciertos en test no podrá superar el valor de 0,07, el cual es muy bajo. Esto ocurre, por ejemplo, en el experimento 4-1, 5-1 y 5-2.

La evolución del error de todos estos experimentos 'malos' tienen una forma muy parecida, que es la siguiente:



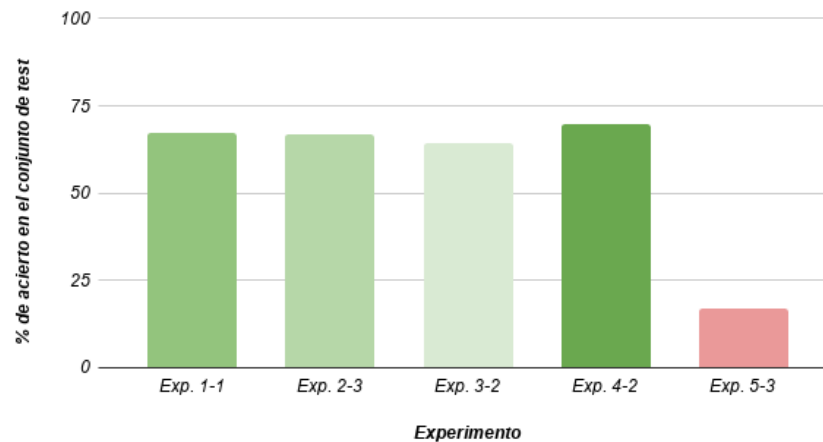
Hoja que no supera el umbral

3.8. Comparativa de los mejores experimentos

Tal y como se mencionaba al principio de la sección de *Experimentación y análisis de los resultados*, la medida que se ha usado como determinante a la hora de decidir qué experimento es mejor que otro es el porcentaje de aciertos en el conjunto de test.

Por ello, se muestra en la siguiente gráfica una comparativa de estos resultados para los mejores experimentos:

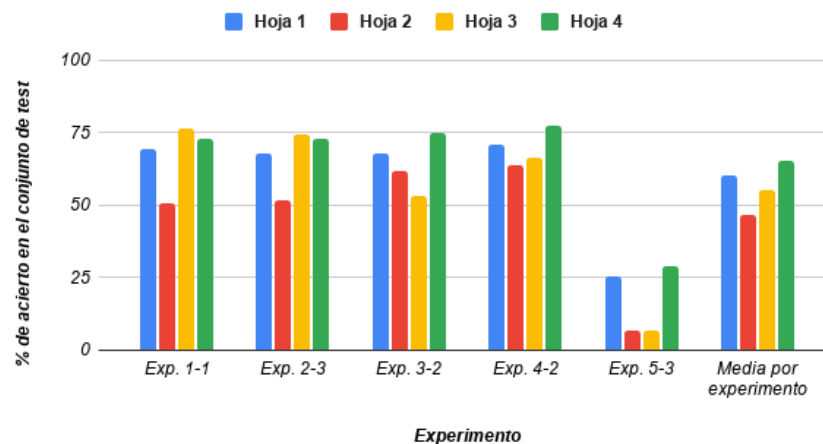
Porcentaje de acierto medio en conjunto de test



Gracias a esta gráfica se puede ver cómo de buenos son los mejores experimentos obtenidos. Se ve que la diferencia entre el mejor (Experimento 4-2) y los siguientes 3 mejores no es muy grande, pues sus porcentajes son ligeramente menores. Salvo el Experimento 5-3, ya que obtiene un porcentaje muy pequeño.

A continuación se va a mostrar una gráfica comparativa de los resultados para cada experimento dividido por hojas, así como la media de hojas:

Porcentaje de acierto para cada hoja en conjunto de test



Como se puede apreciar por los distintos experimentos, y se confirma con la media por experimento, la *Hoja 2* es la que tiene más dificultad para clasificar correctamente a los individuos, mientras que la *Hoja 4* es la que mejor lo hace.

Esto se debe a la generación y organización aleatoria de las hojas, que hacen que unas sean más diferentes respecto a los conjuntos de entrenamiento, y otras sean más similares.

4. Conclusión