

UNIVERSIDAD AUTÓNOMA DE MADRID

Neurocomputación

Práctica 2 - Backpropagation

ENRIQUE CABRERIZO FERNÁNDEZ
GUILLERMO RUIZ ÁLVAREZ

3 DE ABRIL DE 2016

Índice

1. Tarea 1: Implementación de la red neuronal.	2
2. Tarea 2: Chequeo.	2
3. Tarea 4: Predicción en un problema complejo	2
3.1. Tarea 5	3
3.2. Tarea 6	4

1. Tarea 1: Implementación de la red neuronal.
2. Tarea 2: Chequeo.
3. Tarea 4: Predicción en un problema complejo

Para resolver este problema se han elegido los siguientes parámetros para el aprendizaje de la red:

- Conjunto de entrenamiento: 70 %
- Conjunto de test: 30 %
- Tasa de aprendizaje $\alpha = 0,01$.
- Número de nodos de la capa oculta: 7.

Con estos datos, se han obtenido los siguientes resultados:

- Tasa de acierto en el conjunto de **entrenamiento**: 80,2 %
- Tasa de acierto en el conjunto de **test**: 60,3 %

En la figura 1 se puede observar la evolución del error cuadrático medio en función del número de la época de entrenamiento. Se puede observar que converge a un valor para el error cuadrático medio de 1,2.

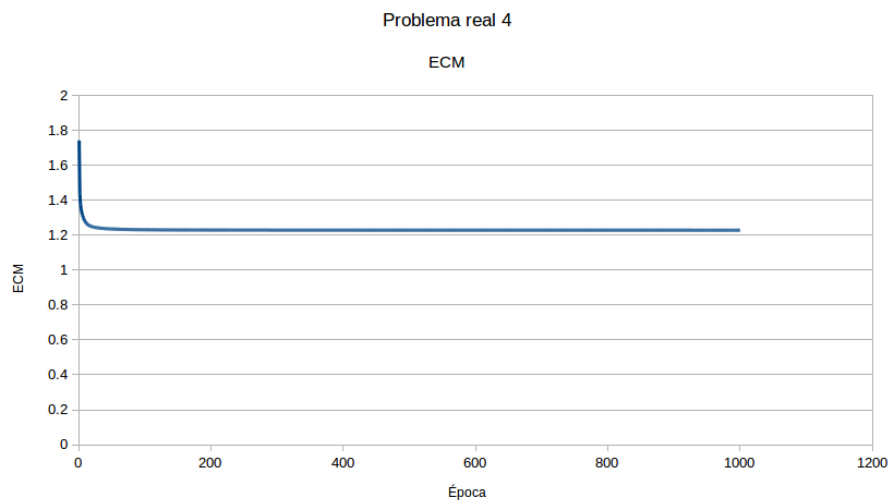


Figura 1: Evolución del error cuadrático medio

Vemos que la tasa de acierto en ambos conjuntos no es muy elevada, sin ni siquiera acercarse al 90 %. Esto se debe, principalmente, a cómo están distribuidos los parámetros utilizados para el entrenamiento de la red, es decir, el

conjunto de atributos. Se han calculado las medias y las desviaciones estándar de todos los atributos del problema, obteniéndose para varios de ellos, medias y desviaciones típicas muy grandes, incluso del orden de 10^3 .

Normalmente, la convergencia es más rápida si las medias de los atributos son cercanas a cero. Esto se debe a que, en caso de que esto no ocurra, el descenso por gradiente del algoritmo de *backpropagation* hará zig-zag en lugar de tomar la dirección de máximo decrecimiento, obteniéndose una convergencia más lenta.

La convergencia también es más rápida si los valores son reescalados y tienen desviación estándar 1. Esto se debe a que, si reescalamos los valores de entrada, se balancea la tasa a la que los pesos conectados a los nodos de entrada aprenden.

Por tanto, si se normalizan los datos haciendo que el conjunto de los mismos tenga media 0 y desviación típica 1, se obtendrá una convergencia más rápida, y por tanto mejores resultados para el mismo número de épocas.

3.1. Tarea 5

Para resolver este problema, se ha realizado un programa que normaliza los parámetros de los atributos de entrada. De esta manera, al restarle a cada atributo la media del conjunto y dividirlo por la desviación típica del mismo, obtenemos que el conjunto de los atributos adquiere una distribución de media 0 y desviación típica 1. Se han utilizado los mismos parámetros que en la tarea anterior, con el fin de poder realizar un análisis comparativo:

- Conjunto de entrenamiento: 70 %
- Conjunto de test: 30 %
- Tasa de aprendizaje $\alpha = 0,01$.
- Número de nodos de la capa oculta: 7.

Con estos datos, se han obtenido los siguientes resultados:

- Tasa de acierto en el conjunto de **entrenamiento**: 99,6 %
- Tasa de acierto en el conjunto de **test**: 94,3 %

En la figura 2 se puede observar la evolución del error cuadrático medio en función del número de la época de entrenamiento.

En este caso, se puede observar que la velocidad de convergencia es mucho mas rápida. El error desciende en todo momento, y aunque se puede notar que en las primeras épocas se ralentiza este descenso, finalmente parece converger a cero.

Si comparamos los resultados con el del apartado anterior, podemos confirmar que la normalización de los datos mejora notablemente la velocidad de convergencia, obteniéndose así unos resultados mucho mejores tanto para el conjunto de entrenamiento (99,6 %) como para el conjunto de test (94,3 %), frente a los valores anteriores, que eran de un 80,2 % para el conjunto de entrenamiento y de un 60,3 % para el conjunto de test.

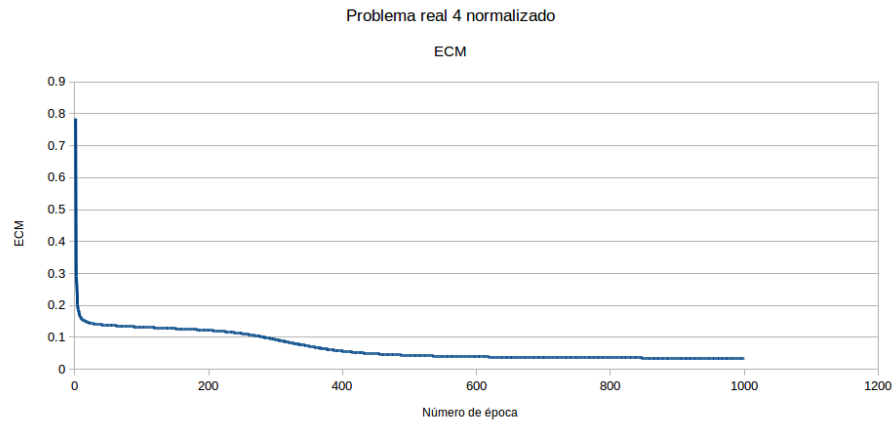


Figura 2: Evolución del error cuadrático medio

3.2. Tarea 6

Para realizar esta tarea, se ha realizado un programa que toma un fichero para entrenar la red y otro para clasificar, obteniéndose un fichero de salida con la clasificación obtenida. El fichero *predicciones_nnet.txt* se adjunta con la práctica.