

Práctica 2: Perceptrón multicapa para problemas de clasificación

Asignatura: Introducción a los modelos computacionales, 4º Grado
Ingeniería Informática (Escuela Politécnica Superior de Córdoba -
Universidad de Córdoba)

Trabajo realizado por:

-Antonio Gómez Giménez (32730338G)
i72gogia@uco.es



Índice:

1. Descripción de los modelos de redes neuronales utilizados:	2
2. Descripción en pseudocódigo de los pasos del algoritmo de retropropagación:	3
3. Experimentos y análisis de resultados:	5
4. Resumen:	21



1. Descripción de los modelos de redes neuronales utilizados:

Principalmente, los modelos que usamos en estas redes neuronales se basan en el número de capas y el número de neuronas que nos pide el usuario, es decir, aparte de la capa de entrada y la capa de salida (ambas siempre están presentes), el usuario indica cuantas capas oculta desea, como mínimo una capa oculta (sin incluir la capa de salida).

Aparte del número de capas, el usuario proporciona el número de neuronas por capas, es decir, si dice por ejemplo cinco, entonces en todas las capas del modelo (exceptuando capa de entrada y salida, ya que se especifican las neuronas de esta capa en la propia base de datos proporcionada por el usuario) hay cinco neuronas en total, no puede haber capas con cinco neuronas y otras capas con un número distinto de cinco neuronas.

En resumen, la arquitectura de todo el modelo depende de los datos introducidos por el usuario (número de capas y número de neuronas por cada capa, que en este caso siempre es el mismo). El resto de valores introducidos por el usuario son parámetros que se usan dentro de la red pero no modifican la topología de la red, un ejemplo es el uso de la sigmoide o la softmax en la última capa.

Durante toda esta práctica se ha tenido que tener en cuenta las dos versiones del algoritmo, tanto la online como la offline, aparte, en muchas de las funciones como por ejemplo en la propagación hacia delante o la propagación del error hacia atrás, ha sido necesario tener en cuenta si las neuronas de la capa de salida eran de tipo softMax o Sigmoides.

Cabe destacar que el error podía ser tanto de tipo MSE como entropía cruzada, era necesario tener esto en cuenta para ciertas funciones.

Una vez tenemos todos estos parámetros y al fijarnos en el CCR que nos permite ver la cantidad de patrones bien clasificados, podemos observar como va funcionando nuestra red.

En si nuestra red asigna a cada neurona de salida un valor entre 0 y 1 siendo aquella neurona con mayor valor de todas la clase a la que pertenece nuestro patrón de entrada.



2. Descripción en pseudocódigo de los pasos del algoritmo de retropropagación:

En este apartado se van a mostrar los algoritmos que más han variado respecto a la práctica anterior, estos son **forwardPropagate**, **obtainError**, **backpropagateError**, **weightAdjustment** y **performEpoch**:

```
void MultilayerPerceptron::forwardPropagate(){
    for(i=1 hasta nCapas){
        if(Es_softMax == 1 && i==nCapas-1){
            netTotal = calcularNetTotal_de_Ultima_capa();
        }
        for(j=0 hasta nNeuronas){
            if(capa[i].neurona[j].pesos != NULL){
                capa[i].neurona[j].salida = sesgo;
                (sesgo =capa[i].neurona[j].peso[0])
                for(x=1 hasta nNeuronas de Capa anterior + 1){
                    capa[i].neurona[j].salida +=
                    capa[i].neurona[j].peso[x]*
                    capa[i-1].neurona[x-1].peso;
                }
            }else{
                capa[i].neurona[j].salida = 0;
            }
            if(Es_softMax == 1 && i==nCapas-1){
                capa[i].neurona[j].salida = aplicoSoftMax();
            }else{
                capa[i].neurona[j].salida = aplicoSigmoide();
            }
        }
    }
}
```

Respecto a **obtainError**, es similar al anterior pero si la función de error es Entropía cruzada la principal diferencia la encontramos que en el acumulador vamos almacenando el valor esperado por el logaritmo de las salidas de la capa de salida.



```
void MultilayerPerceptron::backpropagateError() {
    if(si_Sigmoide){
        if(si_MSE){
            calcular_derivadas_neuronas_de_salida_MSE();
        }else if(si_EntropiaCruzada){
            calcular_derivadas_neuronas_de_salida_EntropiaCruzada();
        }
    }else if(si_softMax){
        if(si_MSE){
            calcular_derivadas_neuronas_de_salida_MSE();
        }else if(si_EntropiaCruzada){
            calcular_derivadas_neuronas_de_salida_EntropiaCruzada();
        }
    }
    for(i=nCapas-2 hasta 0){
        for(j=0 hasta nNeuronas+1){
            calculamos_derivadas_de_las_capas_ocultas();
            //se calcula igual que la práctica anterior,
            tener en cuenta la neurona que puede apuntar a null
        }
    }
}
```

Respecto a **performEpoch**, es similar a la de la práctica anterior pero hay que tener en cuenta que cuando se aplica el algoritmo de forma online no se llama en esta función a `weightAdjustment()` ya que se llama en el `train()`, lo mismo ocurre a la hora de copiar los `deltaW` a `lastDeltaW` justo antes de limpiar los `deltaW`.

Por último la función **weightAdjustment**, ha sufrido ciertos cambios, la parte en concreto donde se realiza el cambio es a la hora donde el algoritmo se aplica online o offline.

La principal diferencia que nos encontramos es que cuando el algoritmo se ejecuta de manera offline, es necesario que se divida por el número total de patrones de entrenamiento tanto F (tasa de aprendizaje) * peso de la neurona en la capa que nos encontremos como cuando multiplicamos el factor de momento por F (tasa de aprendizaje) * peso de la neurona en la capa que nos encontremos de la iteración anterior.



3. Experimentos y análisis de resultados:

Para los experimentos se han utilizado las siguientes bases de datos:

-Problema XOR: Esta base de datos representa el problema de clasificación no lineal del XOR. Se utiliza el mismo fichero para entrenamiento y para test. Como puede verse, se ha adaptado dicho fichero a la codificación 1-de-k, encontrándonos en este caso con dos salidas en lugar de una. Consta de dos entradas, dos salidas y cuatro patrones.

-Base de datos divorce: divorce contiene 127 patrones de entrenamiento y 43 patrones de test. La base de datos contiene la respuesta a una serie de preguntas de un conjunto de encuestas en las que se pretende predecir si se va a producir un divorcio en la pareja. Las respuestas a las preguntas de la encuesta se proporcionan en escala de Likert con valores de 0 a 4. Todas Las variables de entrada se consideran numéricas. La base de datos tiene un total de 54 preguntas (por lo tanto, 54 variables de entrada) y dos categorías (0 no hay divorcio, 1 hay divorcio).

-Base de datos noMNIST: esta base de datos está compuesta por 900 patrones de entrenamiento y 300 patrones de test. Está formada por un conjunto de letras (de la a A la F) escritas con diferentes tipografías o simbologías. Están ajustadas a una rejilla cuadrada de 28×28 píxeles. Las imágenes están en escala de grises en el intervalo $[-1,0; +1,0]$. Cada uno de los píxeles es una variable de entrada (con un total de $28 \times 28 = 784$ variables de entrada) y las clases se corresponden con la letra escrita (a,b,c,d,e y f, con un total de 6 clases).

Se han realizado 3 experimentos:

En el **primer experimento** se buscaba encontrar la mejor arquitectura para nuestras bases de datos, por ello no se ha incluido la base de datos XOR ya que se va a proceder a usar la mejor arquitectura dada en la práctica 1 para esta base de datos. Para las otras dos bases de datos, utilizaremos la entropía cruzada y la función de activación SoftMax para la capa de salida, con el algoritmo configurado como off-line. No se usa validación y el factor de decremento (F) es igual a 1.

Se probaron las siguientes arquitecturas donde varía el número de capas ocultas de 1 a 2, el número de neuronas varía de 4, 8, 16 a 64, la tasa de aprendizaje varía de 0.5, 0.7 y 0.9 y por último, el momento varía de 0.7, 1, 1.5.



Ciertos valores al volcar los resultados del script se han perdido, estos valores se van a ignorar ya que por inferencia podemos intuir que no eran la mejor arquitectura y calcular todas las arquitecturas de nuevo es un gasto de tiempo innecesario, aparte esta sola prueba ha tardado un día, no sería viable volver a realizarla.

Para mayor claridad en las tablas se va a dividir en partes:

Base de datos noMNIST:

1 capa oculta:

4 neuronas:

Eta	Mom ento	Media (Train)	Desviació n típica (Train)	Media (Test)	Desviació n típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,0726348	0,0136757	0,127479	0,0128193	86,7333	4,0323442	78,8	4,1718581
0,5	1	0,0693666	0,0130409	0,130525	0,013053	87,7556	3,7721479	78,1333	3,958674
0,5	1,5	0,0647285	0,0124229	0,134794	0,0116516	88,8	3,4486664	78,6667	3,8872998
0,7	0,7	0,0662415	0,0123393	0,13325	0,0134996	88,5111	3,4502464	78,4	3,9855238
0,7	1	0,0626921	0,0119441	0,132234	0,0166271	88,7111	4,0353934	78,3333	4,0496913
0,7	1,5	0,0597662	0,0110188	0,132135	0,0141398	89,7333	3,1837556	78,5333	3,7570733
0,9	0,7	0,0617547	0,0118682	0,13014	0,0152736	89,3111	3,4835757	78,5333	4,2037008
0,9	1	0,0598571	0,0115978	0,132654	0,0147687	89,6222	3,5497183		3,6998513
0,9	1,5	0,0524311	0,0086024	0,132281	0,0121608	91,6667	1,2629568	80,9333	1,6786244

8 neuronas:

Eta	Mome nto	Media (Train)	Desviació n típica (Train)	Media (Test)	Desviació n típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviació n típica (Test CCR)
0,5	0,7	0,0456799	0,0068533	0,108056	0,0015119	91,7333	1,8916316	82,2	0,8326662
0,5	1	0,0417417	0,0063442	0,108861	0,0033907	92,6	1,3637265	82,3333	0,5962852
0,5	1,5	0,0357578	0,0047121	0,109627	0,0049793	93,6	1,1602457	83,3333	1,011049
0,7	0,7	0,0371625	0,0055413	0,108912	0,0045813	93,2667	1,2384789	83,4667	0,9333333



0,7	1	0,0338722	0,0043972	0,11029	0,0061727		0,9375564	83,1333	1,3920417
0,7	1,5	0,0294563	0,0045604	0,112801	0,0091695	94,7333	1,0125167	83,4	1,6786244
0,9	0,7	0,0323774	0,0046506	0,111516	0,005708	94,2889	0,9903237	83,2667	1,7307686
0,9	1	0,0286565	0,0047471	0,113007	0,009677	94,7778	0,9838196	83,6	1,8184251
0,9	1,5	0,0254675	0,0038255	0,117271	0,0133394	95,1333	1,0550308	84,3333	0,8692272

16 neuronas:

Eta	Mome nto	Media (Train)	Desviació n típica (Train)	Media (Test)	Desviació n típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,0308174	0,0033861	0,11357	0,0072467	94,5778	0,6102016	82,0667	2,1643567
0,5	1	0,0272158	0,0038661	0,118655	0,0104402	95,5111	1,0051716	81,6	2,1333331
0,5	1,5	0,0234156	0,003997	0,121638	0,0113895	96,1111	0,6516832	82,4667	1,8208679
0,7	0,7	0,0242322	0,0038178	0,12148	0,0111302	96,0667	0,6347937	81,9333	1,6384261
0,7	1	0,0216153	0,0040152	0,124262	0,0117235	96,3556	0,7111111	82,4	1,8306037
0,7	1,5	0,0184405	0,0039208	0,126922	0,0125211	96,9333	0,896082		1,966385
0,9	0,7	0,0207054	0,004056	0,125119	0,0142726	96,4667	0,7678737	82,4667	2,1561285
0,9	1	0,018161	0,0039308	0,127105	0,0129632	97,0444	0,9338624		2,5033318
0,9	1,5	0,0158313	0,003967	0,131982	0,0106001	97,4	0,8680899	82,2	2,0396078

64 neuronas:

Eta	Mome nto	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación n típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,013012	0,0012658	0,120521	0,0052498	98,5778	0,4466609	81,8667	1,3759833
0,5	1	0,0111305	0,0013419	0,122113	0,008019	98,8	0,2572407	82,6667	1,4452993
0,5	1,5	0,00901289	0,0010733	0,125574	0,0092507	99,0444	0,2775554	82,6667	0
0,7	0,7	0,00941314	0,0010148	0,124948	0,0089355	98,9778	0,2757706	83,2	1,6411368
0,7	1	0,00793475	0,0010602	0,126883	0,0092761	99,1556	0,2060803	82,7333	1,5114364



0,7	1,5	0,00634702	0,0011695	0,13189	0,0101128	99,3778	0,2863133	82,4	1,6248077
0,9	0,7	0,00727774	0,0011883	0,128802	0,0090366	99,2	0,3095198	82,0667	1,6248077
0,9	1	0,00620649	0,0011652	0,13221	0,0104626	99,4222	0,2154525	82,4	1,6786244
0,9	1,5	0,00518435	0,0010512	0,135536	0,0119189	99,4444	0,2721656	81,7333	2,0374273

2 capa oculta:

4 neuronas:

Eta	Mome nto	Media (Train)	Desviació n típica (Train)	Media (Test)	Desviació n típica (Test)	Media (Train CCR)	Desviació n típica (Train CCR)	Media (Test CCR)	Desviació n típica (Test CCR)
0,5	0,7	0,0998801	0,0222772	0,142279	0,0211633	81,4222	6,3016347	72,5333	8,4685064
0,5	1	0,0905732	0,0178189	0,138324	0,0169981		4,3529415	75,4	4,2708313
0,5	1,5	0,0782675	0,0130908	0,138805	0,0156888	86,6889	3,11777	76,8	2,6381812
0,7	0,7	0,0815472	0,0146264	0,138923	0,0150989	85,8889	3,3518055	76,6667	2,5298221
0,7	1	0,0741515	0,0111877	0,136762	0,018018	87,8889	2,3317461	77,3333	2,9814242
0,7	1,5	0,0688483	0,0105197	0,146367	0,0182424	88,4889	2,5245455	76,2667	2,7194779
0,9	0,7	0,0719207	0,0107009	0,142967	0,0152441	88,3778	1,9583309	76,6667	2,951459
0,9	1	0,0676609	0,0103038	0,146086	0,0179345	88,7556	2,4729658	75,9333	2,8237086
0,9	1,5	0,0631008	0,0106379	0,152453	0,0202626	88,6222	3,3478202	75,9333	5,0039984

8 neuronas:

Eta	Mome nto	Media (Train)	Desviació n típica (Train)	Media (Test)	Desviació n típica (Test)	Media (Train CCR)	Desviació n típica (Train CCR)	Media (Test CCR)	Desviació n típica (Test CCR)
0,5	0,7	0,0466122	0,0060653	0,125566	0,0077881	91,8667	1,356101	1,95556	0
0,5	1	0,0415823	0,0048243	0,129398	0,0085208	92,6	0,9521906	80,6	1,218376
0,5	1,5	0,0339311	0,0025298	0,131369	0,0117951	93,7556	0,8085957	80,6667	2,2803509
0,7	0,7	0,035669	0,0026912	0,131926	0,0113649	93,5111	0,7356868	80,8	1,8808987
0,7	1	0,0324039	0,0023105	0,131441	0,0115658	94,2222	0,8344435	81,0667	2,4073969



0,7	1,5	0,0278372	0,0015339	0,140776	0,0090295	95,0222	0,7006169	81,0667	3,0214053
0,9	0,7	0,029998	0,0031059	0,135398	0,0103202	94,5333	0,8055365	81,7333	1,356466
0,9	1	0,0276661	0,0013601	0,143222	0,0111017		0,5305478	81,1333	1,8690452
0,9	1,5	0,0231733	0,0022266	0,155435	0,0095181	95,8	0,6791428	80,6667	1,8378738

16 neuronas:

Eta	Momento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,0285682	0,0026448	0,112153	0,0085202	95,2444	0,735015	82,2	1,1469786
0,5	1	0,0244853	0,002767	0,113654	0,0111154	95,8667	0,8356267	82,9333	1,878534
0,5	1,5	0,0199447	0,0028334	0,11912	0,0125965	96,7778	0,958394	83,3333	1,5055464
0,7	0,7	0,0209397	0,0030286	0,119106	0,0128402	96,4889	0,9952974	83,3333	1,5202335
0,7	1	0,0179086	0,0029624	0,121086	0,0141794	97,1778	0,8793935	83,3333	1,0327972
0,7	1,5	0,0141074	0,0032353	0,132642	0,0163666	97,9556	0,7083276	82,6	1,1234856
0,9	0,7	0,0167552	0,0032828	0,125674	0,0141762	97,3111	0,9464682	82,4667	2,0720352
0,9	1	0,0139109	0,0030341	0,132468	0,0163276		0,6440613	82,7333	1,3232951
0,9	1,5	0,0122818	0,0031336	0,140552	0,0148842	98,0222	0,5937954	81,6667	2,0763213

64 neuronas:

Eta	Momento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,00904609	0,0014431	0,127384	0,0144529	98,9111	0,4181088	82,0667	1,5405616
0,5	1	0,00754852	0,0014845	0,131229	0,0148861	99,0667	0,38873	82,0667	1,3888448
0,5	1,5	0,00584826	0,0015242	0,135395	0,0158394	99,3333	0,2981424	82,1333	1,9160141
0,7	0,7	0,00617214	0,0014921	0,135181	0,0154369	99,2667	0,3624334	81,9333	1,4360432
0,7	1	0,00509146	0,0018074	0,136504	0,0184343	99,3111	0,493914	81,9333	1,5691463
0,7	1,5	0,00410427	0,0019457	0,145798	0,0236341	99,5111	0,5096597	82,0667	2,1228919



0,9	0,7	0,00473891	0,0019014	0,139585	0,0187128	99,3778	0,5862627	82,1333	1,3920417
0,9	1	0,00404611	0,0019364	0,146821	0,0222595	99,5111	0,5144813	82,2667	1,7307686
0,9	1,5	0,00294528	0,0012641	0,154306	0,0257124	99,6889	0,3095198	81,7333	1,9482197

Base de datos divorce:

1 capa oculta:

4 neuronas:

Eta	Momento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,00174592	0,0007135	0,0562878	0,0094843	100	0	97,6744	0
0,5	1	0,00144452	0,0006079	0,0578513	0,0095067	100	0	97,6744	0
0,5	1,5	0,00111333	0,0004913	0,060237	0,0097729	100	0	97,6744	0
0,7	0,7	0,00117865	0,0005153	0,0596844	0,0096857	100	0	97,6744	0
0,7	1	0,000977469	0,0004384	0,0615337	0,0100035	100	0	97,6744	0
0,7	1,5	0,000760908	0,0003413	0,0640949	0,0104528	100	0	97,6744	0
0,9	0,7	0,000883616	0,0003985	0,062553	0,0101871	100	0	97,6744	0
0,9	1	0,000737791	0,0003301	0,0644403	0,010522	100	0	97,6744	0
0,9	1,5	0,000571301	0,0002642	0,0673032	0,0111713	100	0	97,6744	0

8 neuronas:

Eta	Momento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,000949661	0,0001488	0,0598939	0,0142478	100	0	97,6744	0
0,5	1	0,000767904	0,0001118	0,0617803	0,0144764	100	0	97,6744	0
0,5	1,5	0,000582315	8,07E-05	0,0642931	0,0148309	100	0	97,6744	0



0,7	0,7	0,000618192	8,63E-05	0,0637464	0,0147502	100	0	97,6744	0
0,7	1	0,000508505	6,98E-05	0,0655393	0,0150169	100	0	97,6744	0
0,7	1,5	0,000392173	5,37E-05	0,0679477	0,0153804	100	0	97,6744	0
0,9	0,7	0,000458105	6,27E-05	0,0665044	0,0151624	100	0	97,6744	0
0,9	1	0,000379728	5,2E-05	0,0682485	0,0154249	100	0	97,6744	0
0,9	1,5	0,0002951	4,09E-05	0,0706068	0,0157687	100	0	97,6744	0

16 neuronas:

Eta	Momento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,000675822	0,0001214	0,0587859	0,0131324	100	0	97,6744	0
0,5	1	0,000547128	9,3E-05	0,0604756	0,0135371	100	0	97,6744	0
0,5	1,5	0,000412852	6,52E-05	0,0627678	0,0140425	100	0	97,6744	0
0,7	0,7	0,000438982	7,04E-05	0,0622622	0,0139378	100	0	97,6744	0
0,7	1	0,000358989	5,46E-05	0,0639199	0,0142774	100	0	97,6744	0
0,7	1,5	0,000274001	3,86E-05	0,0661798	0,0146899	100	0	97,6744	0
0,9	0,7	0,000322164	4,75E-05	0,0648186	0,0144506	100	0	97,6744	0
0,9	1	0,00026494	3,7E-05	0,0664617	0,0147392	100	0	97,6744	0
0,9	1,5	0,000203521	2,63E-05	0,0687069	0,0150829	100	0	97,6744	0

64 neuronas:

Eta	Momento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
0,5	0,7	0,000397933	1,73E-05	0,072122	0,0114625	100	0	97,6744	0
0,5	1	0,000330356	1,39E-05	0,0737343	0,0115213	100	0	97,6744	0
0,5	1,5	0,000256187	1,03E-05	0,0759526	0,0116017	100	0	97,6744	0
0,7	0,7	0,000270941	1,1E-05	0,0754588	0,0115845	100	0	97,6744	0



0,7	1	0,000225286	8,9E-06	0,0770757	0,0116438	100	0	97,6744	0
0,7	1,5	0,000175066	6,6E-06	0,0792982	0,0117288	100	0	97,6744	0
0,9	0,7	0,000203752	7,9E-06	0,0779557	0,011677	100	0	97,6744	0
0,9	1	0,000169603	6,4E-06	0,0795749	0,0117393	100	0	97,6744	0
0,9	1,5	0,000131981	4,8E-06	0,0818039	0,0118311	100	0	97,6744	0

2 capa oculta:

4 neuronas:

Eta	Mom ento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviació n típica (Train CCR)	Media (Test CCR)	Desviació n típica (Test CCR)
0,5	0,7	0,00103019	0,00015	0,0622613	0,0060544	100	0	97,6744	0
0,5	1	0,000840919	0,0001148	0,0643587	0,0059658	100	0	97,6744	0
0,5	1,5	0,000642169	8,13E-05	0,0672251	0,005838	100	0	97,6744	0
0,7	0,7	0,000680996	8,76E-05	0,0665944	0,0058658	100	0	97,6744	0
0,7	1	0,000561937	6,9E-05	0,0686738	0,0057727	100	0	97,6744	0
0,7	1,5	0,00043449	5,1E-05	0,0715104	0,0056541	100	0	97,6744	0
0,9	0,7	0,000506861	6,1E-05	0,0698049	0,005723	100	0	97,6744	0
0,9	1	0,000420826	4,92E-05	0,0718664	0,0056398	100	0	97,6744	0
0,9	1,5	0,000327575	3,74E-05	0,0746795	0,0055467	100	0	97,6744	0

8 neuronas:

Eta	Mome nto	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviació n típica (Train CCR)	Media (Test CCR)	Desviació n típica (Test CCR)
0,5	0,7	0,00060863	8,08E-05	0,0711573	0,0075856	100	0	97,6744	0
0,5	1	0,000498205	6,69E-05	0,073253	0,0075851	100	0	97,6744	0
0,5	1,5	0,000380835	5,2E-05	0,076114	0,0075767	100	0	97,6744	0
0,7	0,7	0,000403864	5,49E-05	0,0754847	0,0075795	100	0	97,6744	0



0,7	1	0,000333071	4,59E-05	0,0775582	0,0075717	100	0	97,6744	0
0,7	1,5	0,000256837	3,59E-05	0,0803837	0,0075627	100	0	97,6744	0
0,9	0,7	0,000300171	4,16E-05	0,0786854	0,0075683	100	0	97,6744	0
0,9	1	0,00024864	3,48E-05	0,0807382	0,0075626	100	0	97,6744	0
0,9	1,5	0,000192743	2,72E-05	0,0835317	0,007562	100	0	97,6744	0

16 neuronas:

Eta	Mome nto	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviació n típica (Train CCR)	Media (Test CCR)	Desviació n típica (Test CCR)
0,5	0,7	0,000462943	4,94E-05	0,0695238	0,300882	100	0	97,6744	0
0,5	1	0,000374774	3,84E-05	0,071562	0,0095477	100	0	97,6744	0
0,5	1,5	0,000282352	2,73E-05	0,0743051	0,009621	100	0	97,6744	0
0,7	0,7	0,000300381	2,95E-05	0,073704	0,0096017	100	0	97,6744	0
0,7	1	0,000245174	2,31E-05	0,0756699	0,0096767	100	0	97,6744	0
0,7	1,5	0,000186409	1,65E-05	0,078305	0,0098333	100	0	97,6744	0
0,9	0,7	0,000219726	2,02E-05	0,0767256	0,0097315	100	0	97,6744	0
0,9	1	0,000180141	1,59E-05	0,0786298	0,0098598	100	0	97,6744	0
0,9	1,5	0,000137644	1,14E-05	0,0811745	0,0101079	100	0	97,6744	0

64 neuronas:

Eta	Mome nto	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviació n típica (Train CCR)	Media (Test CCR)	Desviació n típica (Test CCR)
0,5	0,7	0,000299786	2,24E-05	0,0746148	0,0111517	100	0	97,674 4	0
0,5	1	0,000244268	1,71E-05	0,0765977	0,0114148	100	0	97,674 4	0
0,5	1,5	0,000185018	1,17E-05	0,0792806	0,0117962	100	0	97,674 4	0
0,7	0,7	0,000196661	1,27E-05	0,0786915	0,011709	100	0	97,674 4	0



0,7	1	0,000160894	9,6E-06	0,0806238	0,0119967	100	0	97,674 4	0
0,7	1,5	0,000122455	6,6E-06	0,0832361	0,0124066	100	0	97,674 4	0
0,9	0,7	0,000144293	8,3E-06	0,0816677	0,0121568	100	0	97,674 4	0
0,9	1	0,000118338	6,3E-06	0,0835617	0,0124578	100	0	97,674 4	0
0,9	1,5	9,03E-05	4,4E-06	0,086122	0,0128771	100	0	97,674 4	0

Como podemos observar en las tablas, una vez realizado este experimento, llegamos a las siguientes dos conclusiones:

-La mejor arquitectura para la base de datos **noMNIST** es una capa oculta con 8 neuronas en capa oculta, utilizando como parámetros un Eta (tasa de aprendizaje) de 0,9 y momento de 1,5.

-La mejor arquitectura para la base de datos **divorce** es una capa oculta con 4 neuronas en capa oculta, utilizando como parámetros un Eta (tasa de aprendizaje) de 0,5 y momento de 0,7.

Vamos a razonar porque nuestro programa nos da estos resultados respecto a las dos bases de datos.

Respecto a la arquitectura dada en noMIST parece extraño que no sea mejor alguna arquitectura de 2 capas ocultas y 64 nodos pero si nos fijamos en el error medio de entrenamiento este es mucho mejor y aparte el CCR es casi perfecto acertando todos los patrones pero tiene un gran inconveniente y es que los patrones clasificados de test correctamente son peores que el caso mejor, tampoco hay mucha diferencia pero si tenemos una red más sencilla que generaliza mejor donde clasifica correctamente más patrones nos vemos forzados a elegir esta arquitectura. Por tanto lo que está ocurriendo es ese caso de la arquitectura, es que nuestra red está comenzando a sobreentrenar. Respecto a los valores de tasa de aprendizaje como momento se ha comprobado simplemente que combinación era la mejor, en este caso ha sido el mayor momento y la mayor tasa de aprendizaje, estos valores han conseguido mejorar el CCR medio de Test.

Respecto a la base de datos divorce, es mucho más sencilla, hasta tal punto que, todas las arquitecturas clasifican correctamente el 100% de los patrones en entrenamiento y se acercan mucho a esta cifra en el Test.

Por ello, nos debemos fijar en el error medio de Test para encontrar el mejor. Si nos fijamos en el de entrenamiento siempre disminuye pero esto no nos sirve ya que



puede ser que este sobreentrenando y de hecho lo hace ya nuestro mejor caso se encuentra en una capa con 4 neuronas en capa oculta, es decir, a medida que la red es más compleja peor clasifica, esto se debe a que es una base de datos muy sencilla.

Respecto a los valores de tasa de aprendizaje de 0,5 y momento de 0,7, son los más bajos. De hecho, la tasa de aprendizaje que sea pequeña tiene sentido ya que mientras menos entrenemos menos sobre aprendemos ya que al ser tan sencilla sobre aprende muy rápido, a mi parecer, creo que si redujeramos el valor de la tasa de aprendizaje más patrones clasifica correctamente del Test (hasta cierto punto).

En el **segundo experimento** se buscaba comparar las mejores soluciones de la combinación de MSE/Entropía Cruzada con función de salida SoftMax/sigmoide del algoritmo Online frente al algoritmo Offline (las combinaciones son con las mejores estructuras anteriores, validación 0 y $F = 1$, F = decremento de aprendizaje).

Respecto a las pruebas no se van a realizar la combinación de Entropía cruzada con sigmoide ya que los resultados que se pueden obtener son muy malos.

Esto se debe a que SoftMax normaliza los valores de salida entre 0 y 1, mientras que la sigmoide no lo realiza. Por tanto, si utilizamos el Entropía cruzada para obtener los errores de las neuronas de salida y estas no se encuentran normalizadas la probabilidad no va a ser justa para todas las salidas dando resultados bastante malos.

Los resultados obtenidos son los siguientes:

Base de datos XOR:

Offline:

Función de error	Función de activación	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
MSE	Sigmoide	0,0003896	0,0003405	0,0003896	0,0003405	100	0	100	0
MSE	SOFTMAX	0,0001171	1,06E-05	0,0001171	1,06E-05	100	0	100	0
Entropía Cruzada	SOFTMAX	0,0003374	3,40E-05	0,0003374	3,40E-05	100	0	100	0



Online:

Función de error	Función de activación	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
MSE	Sigmoide	0,0250449	0,0499938	0,0250449	0,0499938	95	10	95	10
MSE	SOFTMAX	0,165142	0,209605	0,165142	0,209605	80	24,4949	80	24,4949
Entropía Cruzada	SOFTMAX	8,85E-05	1,17E-05	8,85E-05	1,17E-05	100	0	100	0

Si comparamos entre todas las combinaciones del mismo modo sin duda el mejor es Entropía cruzada con SoftMax ya que tanto Online como offline dan los mejores resultados.

Si comparamos entre los dos modos de ejecución del algoritmo(online u offline) podemos apreciar que el modo offline es mucho mejor en todos los casos que el modo online ya que en el modo online tanto MSE con cualquiera de sus combinaciones no clasifica correctamente todos los patrones de Test mientras que en offline esto sí ocurre.

Por último, si comparamos los dos mejores casos de los dos modos el mejor es el online, lo cual me resulta bastante extraño ya que sale increíblemente pequeño el error de test y las otras dos combinaciones del online eran bastante peor que en el offline. En la base de datos XOR nos interesa sobre entrenar ya que los patrones de test son iguales a los de train por ello el de menor error en general es el mejor y en este caso es online con Entropía Cruzada y SoftMax.

Base de datos noMNIST:

Offline:

Función de error	Función de activación	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
MSE	Sigmoide	0,0208832	0,0013433	0,041327	0,001970	91,9556	0,92322	84,6667	1,09545
MSE	SOFTMAX	0,0204543	0,001741	0,0476643	0,002900	92,2222	0,62459	81	1,41421
Entropía Cruzada	SOFTMAX	0,0254675	0,0038254	0,117271	0,013339	95,1333	1,05503	84,3333	0,86922



Online:

Función de error	Función de activación	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
MSE	Sigmoide	0,133081	0,01646	0,132462	0,0192143	54,5111	5,48245	55,8667	7,58537
MSE	SOFTMAX	0,167946	0,0253355	0,167774	0,0266701	46,6222	8,71173	47,0667	9,30567
Entropía Cruzada	SOFTMAX	1,17658	0,0818415	1,17479	0,0596908	16,7333	0,054433	16,7333	0,13333

Los resultados que nos da esta base de datos son muy interesantes, la más curiosa es que el método offline es mejor, mucho mejor, en todas las combinaciones posibles, de hecho, el método online clasifica correctamente pocos patrones de Test, en concreto la entropía cruzada con SoftMax clasifica una media de 16,7333 patrones de test correctamente.

Si comparamos entre sí, las combinaciones resultantes en el modo offline podemos ver que tanto MSE con sigmoide y Entropía cruzada con SoftMax tiene valores de CCR elevados y parecidos. Por tanto, a la hora de elegir el mejor es más complicado. Si nos fijamos en la media de error de test el MSE con sigmoide es mejor pero a la hora de fijarnos en el CCR para los patrones de test la desviación típica en esta combinación es mayor.

Por consiguiente, elegiremos como mejor configuración Entropía Cruzada con SoftMax aunque prácticamente serviría cualquiera de las dos configuraciones.

Base de datos divorce:

Offline:

Función de error	Función de activación	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
MSE	Sigmoide	0,0035726	0,0035404	0,0184186	0,0046167	99,685	0,385746	97,6744	0
MSE	SOFTMAX	0,0004869	0,0002279	0,021457	0,0014393	100	0	97,6744	0
Entropía Cruzada	SOFTMAX	0,0011786	0,0005152	0,0596844	0,0096857	100	0	97,6744	0



Online:

Función de error	Función de activación	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
MSE	Sigmoide	3,89E-06	6,64E-07	0,0229496	0,00029647	100	0	97,6744	0
MSE	SOFTMAX	1,97E-06	9,73E-07	0,0231522	4,11E-05	100	0	97,6744	0
Entropía Cruzada	SOFTMAX	1,13E-05	3,72E-06	0,119475	0,00873227	100	0	97,6744	0

En este caso, casi todas las combinaciones dan como resultado una media de 100 de CCR excepto el primer caso, esto se debe a que es un base de datos sencilla.

Por ello, para poder comparar qué combinaciones son las mejores es necesario fijarse en el error de test.

Tanto online como offline, se obtienen resultados similares para cada pareja de combinaciones siendo en este caso la peor pareja de Entropía Cruzada con SoftMax.

Si comparamos MSE con su dos respectivas combinaciones dan buenos resultados pero el mejor lo da al usarse la sigmoide siendo el mejor de los casos cuando se ejecuta el algoritmo en modo Offline.

Como conclusión los mejores casos son hasta el momento:

XOR->2 capas, 100 neuronas, 0,7 tasa de aprendizaje, 0,5 momento, Entropía Cruzada, SoftMax, Online.

noMNIST->1 capas, 8 neuronas, 0,9 tasa de aprendizaje, 1,5 momento, Entropía Cruzada, SoftMax, Offline.

divorce->1 capas, 4 neuronas, 0,7 tasa de aprendizaje, 0,5 momento, MSE, Sigmoide, Offline.

Como podemos observar, para problemas de clasificación, tiende a ser mejor el modo offline al modo online, cabe destacar también que tiende a ser mejor usar Entropía Cruzada con SoftMax para estos tipos de problemas.

Pero lógicamente no todos los problemas se resuelven del mismo método, puede ocurrir como en estos casos que otra configuración nos de mejores resultados.



Como **último experimento** se buscaba establecer los mejores valores para los parámetros de validación y decremento de aprendizaje para las mejores configuraciones conseguidas. (En la base de datos XOR no tiene sentido hacer validación).

Los resultados obtenidos son los siguientes:

Base de datos XOR:

decremento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
1	8,85442E-05	1,1735E-05	8,85442E-05	1,1735E-05	100	0	100	0
2	0,00025602	1,42663E-05	0,00025602	1,42663E-05	100	0	100	0

En este caso el decremento de aprendizaje, al aumentarlo, empeora los resultados (error de test ya que el CCR es el mismo), ya que a nuestra red no le da tiempo a mejorar los pesos de las neuronas más alejadas de la capa de salida y en la XOR se busca ajustar los pesos de XOR hasta tal punto que sobreentrene.

Base de datos noMIST:

Iteraciones	Validación	decremento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
5000	0,0	1	0,0254675	0,0038254	0,117271	0,013339	95,1333	1,05503	84,3333	0,86922
5000	0,0	2	0,0458226	0,0034710	0,116604	0,0087279	91,8444	0,91246	81,2	1,6
2005	0,15	1	0,0455623	0,0086626	0,121298	0,0067511	89,6889	1,80178	79,6	0,67986
2077	0,15	2	0,0840699	0,0054340	0,126549	0,0039355	82,3111	1,01932	77,1333	1,69444
1713	0,25	1	0,0501172	0,0085321	0,112127	0,0064555	86,8	2,19921	80,9333	1,1431
3260	0,25	2	0,0688664	0,0111151	0,124369	0,0065023	83,9556	2,3239	77,2	1,69444

En este caso el aumentar el factor de decremento si aplicar validación mejora la media de error en test pero clasifica peor y de hecho, en el resto de casos donde se



aplica validación, aparte de que la media de error de test es peor, encima clasifica también peor que cuando no se aumenta el factor de decremento.

Si nos fijamos en la validación, está empeora el CCR de test y es normal ya que tiene menos patrones para entrenar, es útil sobre todo para evitar el sobreentrenamiento.

Aun así hay un caso donde el CCR disminuye pero no es una bajada drástica, y el número de iteraciones se reduce en una cantidad importante acelerando el cómputo de la red. Este ha sido el caso de $F = 1$ y una validación de 25% de los patrones.

Por tanto, si queremos agilizar el proceso a costo de empeorar el CCR, nos renta utilizar validación de 25%.

Base de datos divorce:

Iteraciones	Validación	decremento	Media (Train)	Desviación típica (Train)	Media (Test)	Desviación típica (Test)	Media (Train CCR)	Desviación típica (Train CCR)	Media (Test CCR)	Desviación típica (Test CCR)
5000	0	1	0,0041385	0,003750	0,0189444	0,0035012	99,685	0,385746	97,6744	0
5000	0	2	0,0095012	0,003295	0,0214826	0,0022137	99,0551	0,589237	97,6744	0
3674	0,15	1	0,0057394	0,005664	0,0234062	0,0045765	99,2126	0,995993	97,6744	0
4204	0,15	2	0,0109215	0,003354	0,0259675	0,0052324	98,8976	0,802995	97,6744	0
5000	0,25	1	0,0052438	0,007463	0,0222801	0,0010754	98,8976	1,06808	97,6744	0
5000	0,25	2	0,0111907	0,005971	0,0246561	0,0015950	98,2677	0,589237	97,6744	0

Para la base de datos divorce ocurre lo mismo que en la base de datos noMIST, tanto el factor de decremento al aumentarlo empeora como que, al realizar validación, se reduce el número de iteraciones pero en este caso al aplicar una validación del 25% no se puede reducir el número de iteraciones.

Por eso, en este caso, si se quiere reducir el número de iteraciones, es necesario realizar una validación del 15%.



4. Resumen:

Tras analizar toda la información vista en las tablas, llegamos a la conclusión de que, para problemas de clasificación, tiende a ser mejor el modo offline al modo online, cabe destacar también que tiende a ser mejor usar Entropía Cruzada con SoftMax para estos tipos de problemas.

Pero para cada base de datos esto no puede tenerse por que ser así, sino siempre se utilizaría el mismo método y ese no es el caso.

Respecto al uso de la validación como el decremento, llegamos a las mismas conclusiones que en la práctica pasada.

Por último, si quisiéramos mejorar el CCR resultante, tendríamos que realizar pruebas de ensayo y error modificando la configuración de nuestra red ya sea modificando las capas, las neuronas, el factor de aprendizaje, el momento, etc, ya que puede ser que la base de datos noMNIST con 2 capas ocultas y 64 neuronas si le aplicamos ciertos parámetros y aplicamos validación, probablemente obtengamos mejores resultados que la obtenida por mis experimentos.

Para afianzar todos estos conceptos se va a mostrar la matriz de confusión de test del mejor modelo de red neuronal obtenido para la base de datos noMNIST:

También, se analizará los errores cometidos, incluyendo las imágenes de algunas de las letras en las que el modelo de red se equivoca, para comprobar visualmente si son confusos.

Matriz de confusión de test para la mejor arquitectura (Mejor SEED 5):

42	4	0	3	1	0
3	41	2	1	2	1
3	0	46	0	0	1
1	3	2	43	0	1
1	3	2	0	42	2
1	1	1	0	3	44

Como podemos observar en la matriz, la diagonal son los casos donde se acierta, son aquellos que se utilizan para calcular el CCR.

Por ejemplo, para la clase letra A se aciertan 42 y se confunde con 4 B, ninguna C, 3 letras D, una letra E y no se confunde con ninguna F.



Para este caso los patrones que han fallado son los siguientes:

```
Error encontrado en -> 0
Error encontrado en -> 17
Error encontrado en -> 28
Error encontrado en -> 30
Error encontrado en -> 37
Error encontrado en -> 39
Error encontrado en -> 41
Error encontrado en -> 48
Error encontrado en -> 52
Error encontrado en -> 69
Error encontrado en -> 75
Error encontrado en -> 80
Error encontrado en -> 84
Error encontrado en -> 85
Error encontrado en -> 91
Error encontrado en -> 96
Error encontrado en -> 99
Error encontrado en -> 108
Error encontrado en -> 109
Error encontrado en -> 118
Error encontrado en -> 143
Error encontrado en -> 157
Error encontrado en -> 160
Error encontrado en -> 161
Error encontrado en -> 172
Error encontrado en -> 173
Error encontrado en -> 182
Error encontrado en -> 188
Error encontrado en -> 207
Error encontrado en -> 209
Error encontrado en -> 210
Error encontrado en -> 216
Error encontrado en -> 219
Error encontrado en -> 223
Error encontrado en -> 237
Error encontrado en -> 238
Error encontrado en -> 260
Error encontrado en -> 261
Error encontrado en -> 275
Error encontrado en -> 285
Error encontrado en -> 289
Error encontrado en -> 299
```

Ahora comprobaremos como son algunos de esos patrones y con qué y si tiene sentido el por qué han fallado.

Por ejemplo, el patrón 237 es una E y la ha confundido con una A. El patrón es el siguiente:





En este caso el patrón era complicado, incluso a una persona le costaría identificar la E en este caso.

Otro ejemplo es el patrón 143, que es una letra C y la ha confundido con una A. El patrón es el siguiente:



En este caso, para una persona, claramente es una c pero si no ha entrenado con patrones donde la c se encontraba sobreada con negro, puede ser y de hecho es, que no la clasifica correctamente.

No todos los patrones donde se ha equivocado son difíciles, también hay algunos patrones simples donde se ha equivocado como por ejemplo el patrón 0 que es el siguiente:



Claramente es una A pero nuestra red la clasifica como una E e incluso casi como una F.