

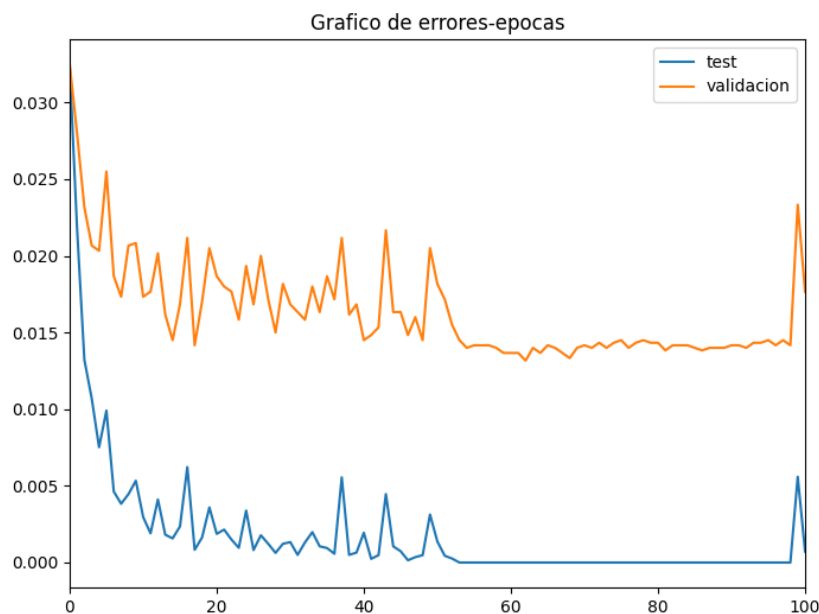
APRENDIZAJE AUTOMÁTICO

Memoria de Redes Neuronales

- Fabián Scherle Carboneres
- Lishuang Sun (María)

1) A partir del código base proporcionado, representa gráficamente la evolución del error en entrenamiento y validación en función del número de épocas.

Para poder observar la tasa de error del clasificador a medida que aumenta el número de épocas decidimos crear un script que fuese capaz de generar una gráfica haciendo uso de la librería *matplotlib.pyplot*. Obteniendo como resultado:



Se puede apreciar que a partir de una época cercana a 60 el clasificador comienza a sufrir de sobre-entrenamiento, debido a que su tasa de error sobre el conjunto de entrenamiento se mantiene en 0 y sobre el conjunto de validación incrementa o fluctúa sobre 0.015 pero no decrementa a medida que aumentan las épocas.

2) Para ajustar los parámetros de la red MLP, estudia el comportamiento de la tasa de error en el conjunto de validación.

Para realizar dicho experimento creamos un script el cual prueba los diferentes algoritmos de optimización (SGD, Adadelta, Adagrad, Adam), funciones de activación (ReLU, Sigmoid, Tanh), números de capas ocultas (1, 2, 3) y número de neuronas por capa (misma cantidad de neuronas para las capas ocultas: 100, 200, 500, 700, 800), teniendo en cuenta un máximo de 20 épocas. Obtenemos los siguientes resultados:

Optimización	Activación	Capas ocultas	Neuronas	Error
SGD	ReLU	1	100	13.35
SGD	ReLU	1	200	13.083
SGD	ReLU	1	500	12.883
SGD	ReLU	1	700	12.85
SGD	ReLU	1	800	12.817
SGD	ReLU	2	100	16.45
SGD	ReLU	2	200	16.333
SGD	ReLU	2	500	14.583
SGD	ReLU	2	700	14.283
SGD	ReLU	2	800	13.467
SGD	ReLU	3	100	49.883
SGD	ReLU	3	200	37.917
SGD	ReLU	3	500	25.767
SGD	ReLU	3	700	19.567
SGD	ReLU	3	800	21.3
SGD	Sigmoid	1	100	34.05
SGD	Sigmoid	1	200	32.167
SGD	Sigmoid	1	500	27.45
SGD	Sigmoid	1	700	27.633
SGD	Sigmoid	1	800	27.45
SGD	Sigmoid	2	100	88.883
SGD	Sigmoid	2	200	88.883
SGD	Sigmoid	2	500	88.883
SGD	Sigmoid	2	700	88.883
SGD	Sigmoid	2	800	88.883
SGD	Sigmoid	3	100	88.883
SGD	Sigmoid	3	200	88.883
SGD	Sigmoid	3	500	88.883
SGD	Sigmoid	3	700	88.883
SGD	Sigmoid	3	800	88.883
SGD	Tanh	1	100	13.7
SGD	Tanh	1	200	13.233
SGD	Tanh	1	500	12.783
SGD	Tanh	1	700	12.667
SGD	Tanh	1	800	12.617
SGD	Tanh	2	100	15.383
SGD	Tanh	2	200	13.367
SGD	Tanh	2	500	13.117
SGD	Tanh	2	700	12.383
SGD	Tanh	2	800	12.4
SGD	Tanh	3	100	16.683
SGD	Tanh	3	200	15.833
SGD	Tanh	3	500	13.6
SGD	Tanh	3	700	13.217
SGD	Tanh	3	800	12.817
Adadelata	ReLU	1	100	18.717
Adadelata	ReLU	1	200	16.55
Adadelata	ReLU	1	500	15.683
Adadelata	ReLU	1	700	14.65

Adadelata	ReLU	1	800	14.783
Adadelata	ReLU	2	100	30.867
Adadelata	ReLU	2	200	21.933
Adadelata	ReLU	2	500	17.417
Adadelata	ReLU	2	700	16.233
Adadelata	ReLU	2	800	16.117
Adadelata	ReLU	3	100	49.217
Adadelata	ReLU	3	200	47.833
Adadelata	ReLU	3	500	29.517
Adadelata	ReLU	3	700	24.05
Adadelata	ReLU	3	800	24.783
Adadelata	Sigmoid	1	100	36.333
Adadelata	Sigmoid	1	200	35.5
Adadelata	Sigmoid	1	500	37.033
Adadelata	Sigmoid	1	700	33.733
Adadelata	Sigmoid	1	800	33.217
Adadelata	Sigmoid	2	100	88.883
Adadelata	Sigmoid	2	200	88.883
Adadelata	Sigmoid	2	500	88.883
Adadelata	Sigmoid	2	700	88.883
Adadelata	Sigmoid	2	800	88.883
Adadelata	Sigmoid	3	100	88.883
Adadelata	Sigmoid	3	200	88.883
Adadelata	Sigmoid	3	500	88.883
Adadelata	Sigmoid	3	700	88.883
Adadelata	Sigmoid	3	800	88.883
Adadelata	Tanh	1	100	17.2
Adadelata	Tanh	1	200	15.783
Adadelata	Tanh	1	500	14.567
Adadelata	Tanh	1	700	14.167
Adadelata	Tanh	1	800	14.1
Adadelata	Tanh	2	100	21.667
Adadelata	Tanh	2	200	17.15
Adadelata	Tanh	2	500	14.533
Adadelata	Tanh	2	700	13.917
Adadelata	Tanh	2	800	13.067
Adadelata	Tanh	3	100	23.85
Adadelata	Tanh	3	200	20.033
Adadelata	Tanh	3	500	14.833
Adadelata	Tanh	3	700	14.65
Adadelata	Tanh	3	800	13.667
Adagrad	ReLU	1	100	8.517
Adagrad	ReLU	1	200	7.45
Adagrad	ReLU	1	500	6.567
Adagrad	ReLU	1	700	6.133
Adagrad	ReLU	1	800	6.067
Adagrad	ReLU	2	100	7.783
Adagrad	ReLU	2	200	6.85
Adagrad	ReLU	2	500	4.883
Adagrad	ReLU	2	700	4.533

Adagrad	ReLU	2	800	3.883
Adagrad	ReLU	3	100	8.017
Adagrad	ReLU	3	200	6.417
Adagrad	ReLU	3	500	4.317
Adagrad	ReLU	3	700	2.9
Adagrad	ReLU	3	800	2.817
Adagrad	Sigmoid	1	100	11.767
Adagrad	Sigmoid	1	200	10.5
Adagrad	Sigmoid	1	500	9.4
Adagrad	Sigmoid	1	700	9.133
Adagrad	Sigmoid	1	800	8.833
Adagrad	Sigmoid	2	100	14.333
Adagrad	Sigmoid	2	200	11.333
Adagrad	Sigmoid	2	500	8.983
Adagrad	Sigmoid	2	700	8.583
Adagrad	Sigmoid	2	800	8.433
Adagrad	Sigmoid	3	100	26.183
Adagrad	Sigmoid	3	200	15.55
Adagrad	Sigmoid	3	500	11.017
Adagrad	Sigmoid	3	700	10.283
Adagrad	Sigmoid	3	800	9.683
Adagrad	Tanh	1	100	8.65
Adagrad	Tanh	1	200	7.883
Adagrad	Tanh	1	500	7.4
Adagrad	Tanh	1	700	7.433
Adagrad	Tanh	1	800	7.367
Adagrad	Tanh	2	100	7.717
Adagrad	Tanh	2	200	7.083
Adagrad	Tanh	2	500	6.433
Adagrad	Tanh	2	700	6.117
Adagrad	Tanh	2	800	6.083
Adagrad	Tanh	3	100	7.4
Adagrad	Tanh	3	200	6.8
Adagrad	Tanh	3	500	5.7
Adagrad	Tanh	3	700	5.35
Adagrad	Tanh	3	800	5.333
Adam	ReLU	1	100	2.317
Adam	ReLU	1	200	1.9
Adam	ReLU	1	500	1.683
Adam	ReLU	1	700	1.883
Adam	ReLU	1	800	1.817
Adam	ReLU	2	100	2.317
Adam	ReLU	2	200	2.25
Adam	ReLU	2	500	2.033
Adam	ReLU	2	700	1.8
Adam	ReLU	2	800	1.767
Adam	ReLU	3	100	2.35
Adam	ReLU	3	200	2.183
Adam	ReLU	3	500	1.667
Adam	ReLU	3	700	1.5

Adam	ReLU	3	800	1.917
Adam	Sigmoid	1	100	2.467
Adam	Sigmoid	1	200	1.933
Adam	Sigmoid	1	500	1.967
Adam	Sigmoid	1	700	1.75
Adam	Sigmoid	1	800	1.883
Adam	Sigmoid	2	100	2.567
Adam	Sigmoid	2	200	2.05
Adam	Sigmoid	2	500	2.05
Adam	Sigmoid	2	700	2.05
Adam	Sigmoid	2	800	2.017
Adam	Sigmoid	3	100	2.8
Adam	Sigmoid	3	200	2.317
Adam	Sigmoid	3	500	2.033
Adam	Sigmoid	3	700	1.967
Adam	Sigmoid	3	800	2.05
Adam	Tanh	1	100	2.5
Adam	Tanh	1	200	2.0
Adam	Tanh	1	500	2.15
Adam	Tanh	1	700	1.567
Adam	Tanh	1	800	1.767
Adam	Tanh	2	100	2.467
Adam	Tanh	2	200	1.983
Adam	Tanh	2	500	2.133
Adam	Tanh	2	700	2.233
Adam	Tanh	2	800	2.383
Adam	Tanh	3	100	2.583
Adam	Tanh	3	200	2.267
Adam	Tanh	3	500	2.667
Adam	Tanh	3	700	2.133
Adam	Tanh	3	800	2.6

Se puede apreciar que la mínima tasa de error lo obtiene el algoritmo de optimización Adam con una función de activación ReLU, 3 capas ocultas de 700 neuronas cada una. Sin embargo, usando el mismo algoritmo con la función de activación Tanh y 1 capa oculta de 700 neuronas se obtiene una tasa de error bastante similar a la mínima.

3) Utiliza los valores óptimos de los parámetros del clasificador para entrenar y evaluar un clasificador final en los conjuntos oficiales MNIST.

Tomando como base los parámetros óptimos obtenidos en el apartado anterior evaluamos el conjunto de muestra de prueba de MNIST:

Optimización	Activación	Capas ocultas	Neuronas	Error	Intervalo
Adam	ReLU	3	700	1.79	[1.53,2.05]
Adam	Tanh	1	700	1.67	[1.419,1.921]

Podemos concluir que el porcentaje de error mínimo para el experimento realizado lo obtiene el algoritmo de optimización Adam con la función de activación Tanh y 1 capa oculta de 700 neuronas.

En relación a las tasas de error de la MNIST web (<http://yann.lecun.com/exdb/mnist/>) comprobamos que se acerca a la tasa de error del clasificador de “2-layer NN, 800 HU, Cross-Entropy Loss” y al de “2-layer NN, 300 HU” con preprocesado deskewing.