

A thick dark blue vertical bar runs along the left edge of the page. A purple arrow-shaped banner points to the right from this bar, containing the text 'Curso 2021-2022'. In the bottom-left corner, several thin, curved lines in dark blue and light grey sweep upwards and to the right.

Curso 2021-2022

Proyecto de prácticas

Reconocimiento de dígitos
manuscritos: MNIST

Iñaki Diez Lambies y Manuel Diaz Pastor
PERCEPCIÓN

1 CONTENIDO

2	Ejercicio obligatorio: Clasificador multinomial.....	2
2.1	Multinomial-exp.py	2
2.2	Multinomial-eva.py	3

Entrega 2

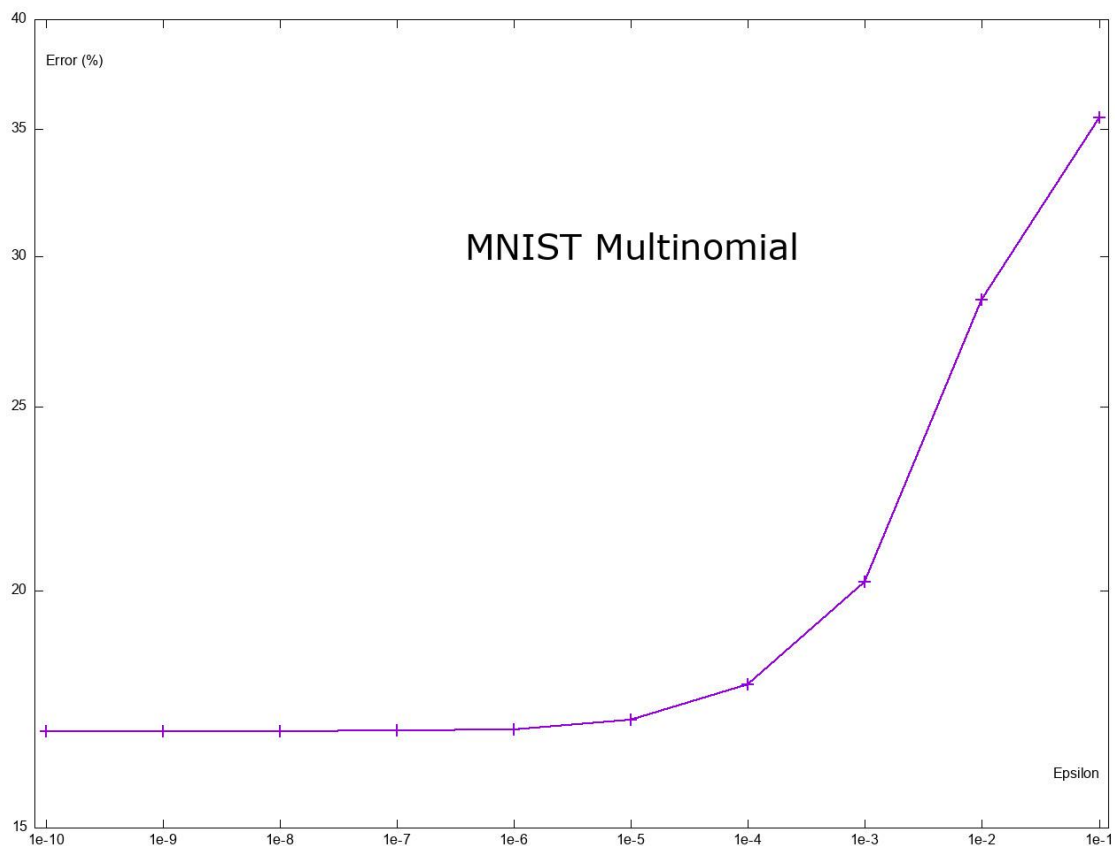
2 EJERCICIO OBLIGATORIO: CLASIFICADOR MULTINOMIAL

Para la realización de esta entrega se nos ha pedido preparar un algoritmo de clasificación multinomial que recibe como argumento un conjunto de datos y etiquetas de entrenamiento, un conjunto de datos y etiquetas de validación y un vector con una serie de valores de suavizado para el algoritmo de suavizado de Laplace.

Para el proceso de la función multinomial de `multinomial.py` hemos dividido el procedimiento en dos partes: cálculo de parámetros y clasificación con suavizado.

2.1 MULTINOMIAL-EXP.PY

Además de esto, hemos realizado un experimento en `multinomial-exp.py` para comprobar el impacto de los diferentes valores de suavizado, siendo estos $1e-10$, $1e-9$, $1e-8$, $1e-7$, $1e-6$, $1e-5$, $1e-4$, $1e-3$, $1e-2$ y $1e-1$, en la clasificación del 10 % de los datos que hemos usado para test. A la vista de los resultados hemos obtenido la siguiente gráfica;



Como podemos observar, el resultado nos indica que a mayor valor de suavizado peores porcentajes de error genera. Esto se debe principalmente a que la mayor funcionalidad del suavizado en este caso es evitar fallos en la ejecución por la abundancia de ceros en la matriz.

Por lo que, en conclusión, extraemos que el mejor valor de suavizado para este conjunto de datos con la distribución multinomial es cualquiera entre $1e-10$ y $1e-8$.

2.2 MULTINOMIAL-EVA.PY

Una vez obtenido el valor de suavizado $1e-10$, hemos realizado la evaluación del clasificador para el conjunto de datos de validación con este número mediante el código *multinomial-eva.py*. El porcentaje de error en la clasificación del clasificador multinomial para el conjunto de datos MNIST es del 16'32%.

Si comparamos el resultado obtenido con los documentados en la web MNIST podremos comprobar que resulta ser el peor analizado, siendo superado incluso por los tres clasificadores lineales registrados (por pares y de una capa). También podemos concluir que resulta una pérdida de eficacia considerable con respecto a los clasificadores de vecinos más cercanos vistos en la primera parte de este proyecto.

Linear Classifiers		
linear classifier (1-layer NN)	none	12.0
linear classifier (1-layer NN)	deskewing	8.4
pairwise linear classifier	deskewing	7.6