

A thick dark blue vertical bar runs down the left side 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, there are several thin, curved, overlapping lines in shades of grey and black, resembling stylized grass or abstract brushstrokes.

Curso 2021-2022

Proyecto de prácticas

Reconocimiento de dígitos
manuscritos: MNIST

Iñaki Diez Lambies y Manuel Diaz Pastor
PERCEPCIÓN

1 CONTENIDO

2	Clasificador Gaussiano.....	2
2.1	Gaussian-exp.py	2
2.2	Gaussian-eva.py	3

Entrega 2

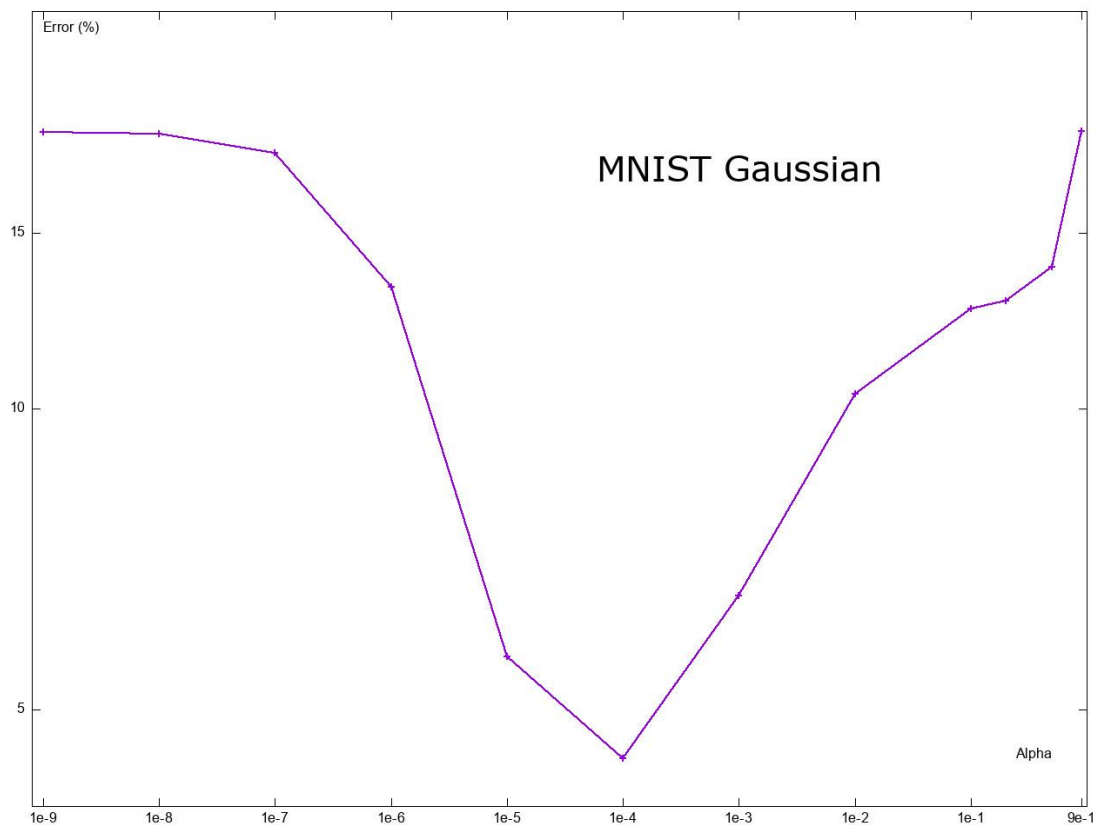
2 CLASIFICADOR GAUSSIANO

Para la segunda pregunta de esta segunda parte del proyecto, hemos construido un clasificador gaussiano en `gaussian.py` que toma como argumentos el conjunto de entrenamiento, sus etiquetas, el conjunto de prueba, sus etiquetas y un grupo de valores de alfa como valores posibles de suavizado para utilizar el algoritmo de “flat-smoothing”.

2.1 GAUSSIAN-EXP.PY

Una vez resueltos todos los problemas encontrados en el desarrollo del código, hemos preparado un experimento en `gaussian-exp.py`, al igual que con el clasificador multinomial donde pasamos los diferentes valores de suavizado $1e-9$, $1e-8$, $1e-7$, $1e-6$, $1e-5$, $1e-4$, $1e-3$, $1e-2$, $1e-1$ y $9e-1$ para la clasificación del 10% de los datos de entrenamiento de MNIS a partir del 90% restante.

El resultado del experimento se expone en la siguiente gráfica, donde el eje “x” representa los valores de suavizado probados y el eje “y” el porcentaje de error resultante.



A la vista de los resultados, podemos comprobar que el valor de suavizado óptimo para la clasificación de estos datos es $1e-4$.

2.2 GAUSSIAN-EVA.PY

Una vez obtenido el valor óptimo de suavizado, hemos de realizar la evaluación mediante el código *gaussian-eva.py* que se encarga de probar la clasificación del conjunto de datos de prueba de MNIST en base al conjunto de entrenamiento para un valor de alfa concreto.

Como hemos mencionado previamente, el valor seleccionado es $1e-4$. Teniendo esto en cuenta, el porcentaje de error en la evaluación del clasificador para MNIST mediante un clasificador gaussiano es del 4'18%. Llama la atención la mejora significativa de este clasificador cuadrático con respecto al clasificador lineal. Además, si volvemos a comprobar los resultados con los existentes en la base de datos de MNIST, podremos observar que si bien la mayoría de ellos siguen superando al clasificador gaussiano, obtenemos mejores resultados que para los clasificadores lineales y el clasificador de k-vecinos más cercanos básico con distancia L2:

Linear Classifiers		
linear classifier (1-layer NN)	none	12.0
linear classifier (1-layer NN)	deskewing	8.4
pairwise linear classifier	deskewing	7.6
K-Nearest Neighbors		
K-nearest-neighbors, Euclidean (L2)	none	5.0

Sin embargo, aún queda por detrás de los clasificadores no lineales que aparecen en la página:

Non-Linear Classifiers		
40 PCA + quadratic classifier	none	3.3
1000 RBF + linear classifier	none	3.6