

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	Ejercicio Opcional: PCA+Gaussian.py	2
2.1	PCA+Gaussian-exp.py.....	2

Entrega 2

2 EJERCICIO OPCIONAL: PCA+GAUSSIAN.PY

Tras obtener los resultados de los mejores valores de suavizado de alfa para el suavizado “flat smoothing” con los datos en bruto, hemos de realizar un análisis de este mismo proceso para los datos reducidos mediante la técnica de Principal Component Analysis desarrollado previamente. Es de esta forma que hemos probado todos los valores de dimensiones para PCA probados para el clasificador de vecinos más cercanos.

2.1 PCA+GAUSSIAN-EXP.PY

De esta forma obtenemos los resultados en una gráfica donde cada curva representa los valores de PCA utilizados, el eje “x” el valor de suavizado empleado y el eje “y” porcentaje de error resultante.

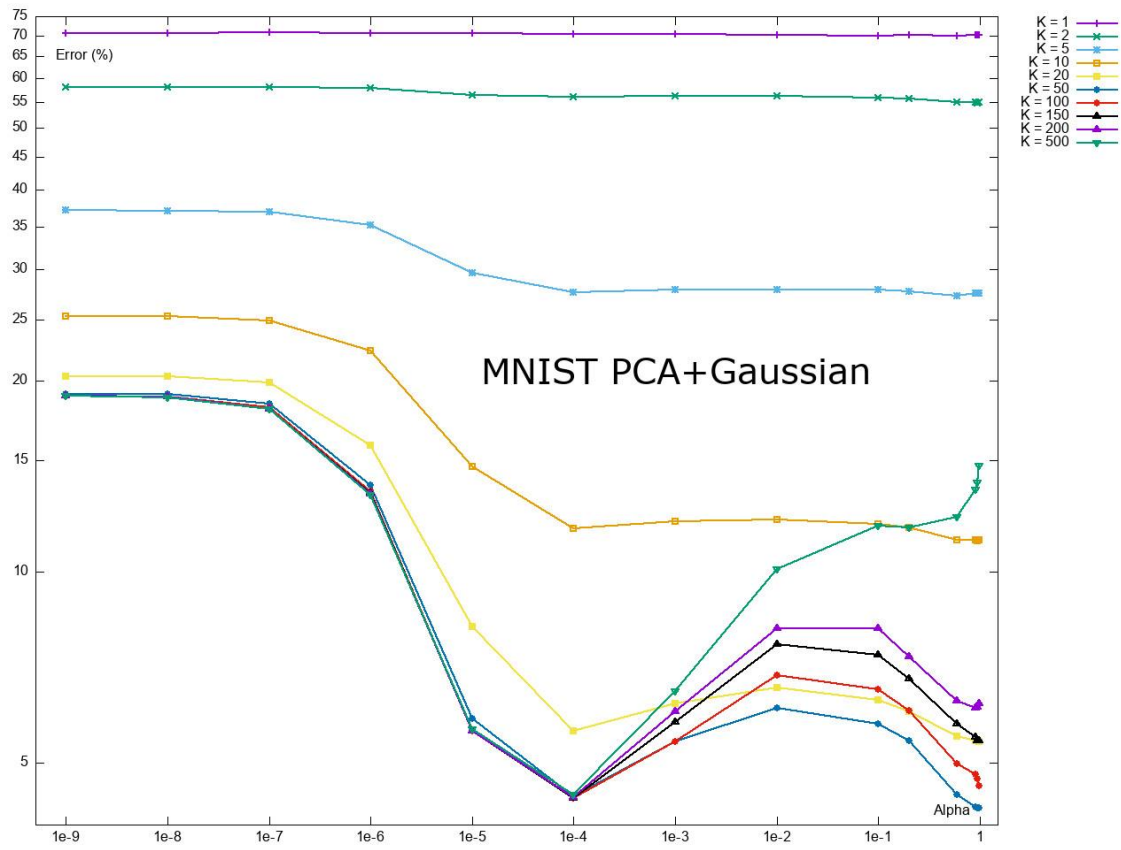
El código genera la matriz de proyección de datos para cada valor de “k” (dimensiones en PCA) y llama a la función de clasificador gaussiano con los datos proyectados y el vector de valores de suavizado.

De esta forma obtenemos los siguientes resultados para todas las combinaciones contempladas:

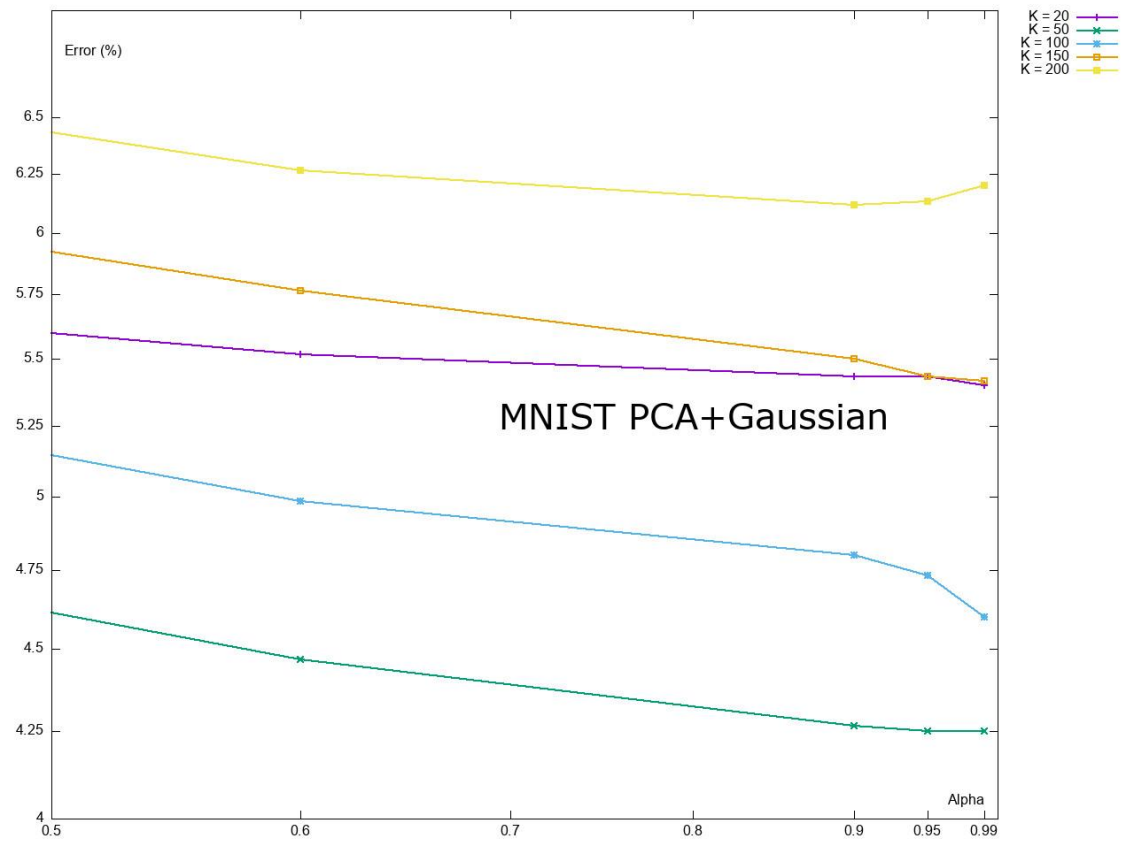
alphas\Ks	K=1	K=2	K=5	K=10	K=20	K=50	K=100	K=150	K=200	K=500
1,00E-09	70,65	58,08	37,12	25,30	20,35	19,10	18,93	18,95	18,95	18,93
1,00E-08	70,67	58,08	37,08	25,30	20,32	19,07	18,88	18,92	18,90	18,85
1,00E-07	70,68	58,08	36,97	24,87	19,88	18,38	18,18	18,08	18,03	18,03
1,00E-06	70,62	57,83	35,20	22,30	15,83	13,68	13,38	13,30	13,27	13,25
1,00E-05	70,55	56,37	29,60	14,67	8,20	5,88	5,62	5,62	5,62	5,65
0,0001	70,37	56,08	27,60	11,70	5,62	4,43	4,40	4,40	4,42	4,47
0,001	70,27	56,25	27,83	12,00	6,22	5,40	5,40	5,80	6,03	6,48
0,01	70,03	56,20	27,87	12,10	6,57	6,10	6,87	7,70	8,15	10,12
0,1	69,97	55,88	27,80	11,92	6,28	5,77	6,53	7,40	8,15	11,82
0,2	70,20	55,58	27,68	11,75	6,03	5,43	6,05	6,78	7,35	11,77
0,6	69,92	54,88	27,22	11,25	5,52	4,47	4,98	5,77	6,27	12,22
0,9	70,05	54,90	27,50	11,22	5,43	4,27	4,80	5,50	6,12	13,50
0,95	70,08	54,85	27,52	11,20	5,43	4,25	4,73	5,43	6,13	13,85
0,99	70,02	54,95	27,52	11,22	5,40	4,25	4,60	5,42	6,20	14,70

Como podemos observar en la tabla, el mejor resultado obtenido para el clasificador nos indica que la mejor combinación para el clasificador gaussiano es con suavizado con alfa 0’95 y 50 dimensiones de PCA. Esto resulta extremadamente interesante, puesto que un alto valor de alfa para el suavizado en el clasificador cuadrático sin PCA significaba un mayor porcentaje de error. Sin embargo, como se puede observar también en las siguientes tablas, los resultados obtenidos indican lo contrario con la disminución de dimensionalidad. Es decir, al contrario que ocurría previamente, el mejor resultado se obtiene con un alto valor de alfa. Esto probablemente

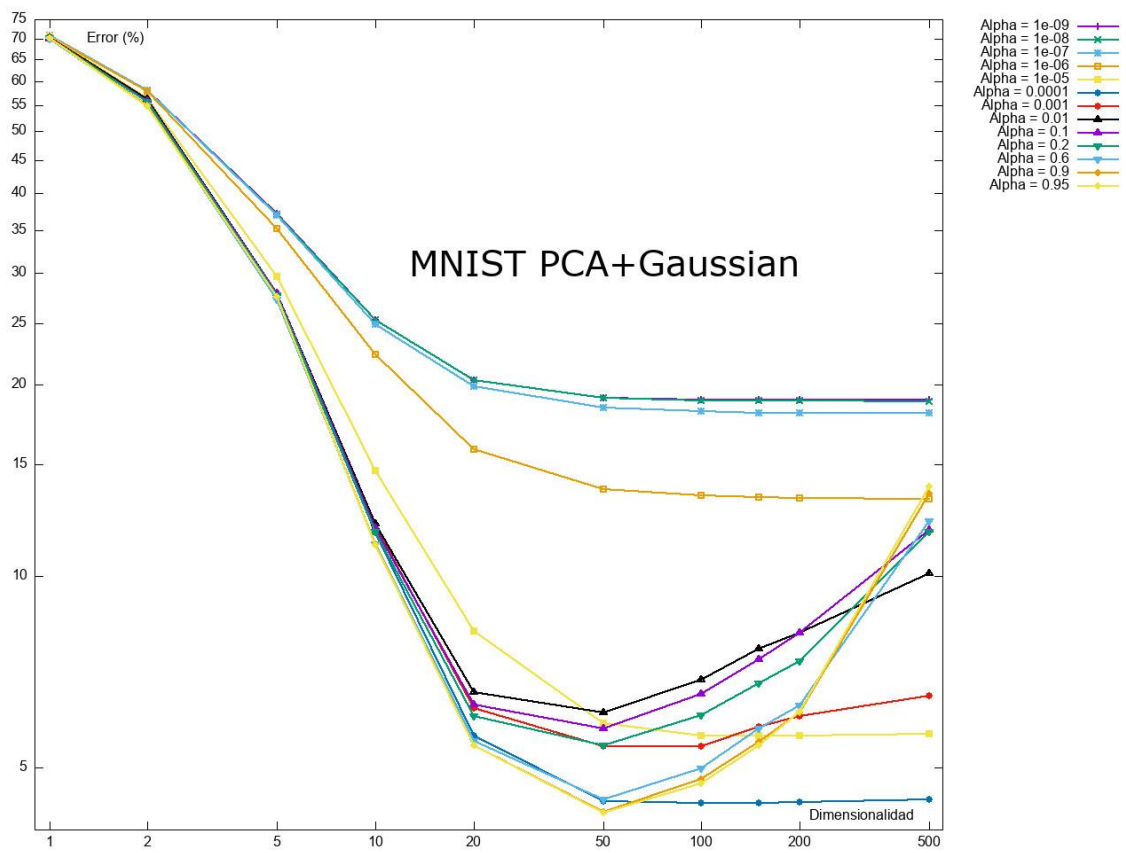
se debe a la influencia de la reducción de datos para mantener la máxima dispersión en los datos.



Gráfica con todas las combinaciones.



Gráfica centrada en los valores más relevantes.



Gráfica respecto al suavizado

De esta forma, con el valor de alfa 0'95 y 50 dimensiones, nos disponemos a preparar la evaluación de los datos de prueba de MNIST. El resultado obtenido nos indica que el porcentaje de error para un clasificador gaussiano es del 3'7%. Esto implica una mejoría con respecto a la versión sin reducción de dimensionalidad, acercando más este resultado a los obtenidos por clasificadores cuadráticos registrados en la página de MNIST:

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

Siendo estos muy próximos al segundo clasificador y algo menos al cuadrático con 40 dimensiones por PCA.