

# Memoria

MÁQUINAS DE VECTORES DE SOPORTE

Mireia Pires State | Aprendizaje automático | 13/12/2022

## Contenido

DESCRIPCIÓN DEL TRABAJO RELIZADO .....	2
KERNEL POLINÓMICO .....	3
KERNEL GAUSSIANO .....	6
KERNEL SIGMOIDE .....	7
CONCLUSIÓN .....	7

## DESCRIPCIÓN DEL TRABAJO RELIZADO

Para el desarrollo de esta práctica se han realizado diferentes experimentos para entrenar máquinas de vectores de soporte mediante tres tipos de kernels (polinómico, gaussiano y sigmoide). Para los experimentos se ha cogido un 20% de MNIST para el entrenamiento y un 10% para el test.

Los scripts empleados han sido los proporcionados por el profesorado, pero han sido adaptados para los diferentes kernels y ajustes de parámetros. En general, los parámetros explorados han sido:

- **C:** Parámetro para determinar la penalización de una muestra mal clasificada. Es decir, representa como de tolerante es el modelo. Un valor elevado de C significaría que el modelo evita que existan muestras mal clasificadas (bias bajo, varianza elevada). Un valor pequeño quiere decir que pueden existir muestras mal clasificadas (bias alto, varianza baja).

Los valores con los que se ha experimentado han sido  $C = (0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000)$ .

Ha sido explorado en todos los kernels.

- **Gamma:** Este valor define como de lejos llega la influencia de una sola muestra. Un numero elevado representa que el radio de influencia es pequeño, por lo que el modelo tenderá a estar sobre ajustado. Lo contrario ocurre con valores pequeños, pero hace que el modelo tenga una menor precisión. En conclusión, es preferible valores intermedios.

Los valores con los que se ha experimentado han sido  $\gamma = (0.1, 1, 10)$ .

Ha sido explorado en todos los kernels.

- **Termino independiente:** Ha sido explorado en el kernel polinómico y sigmoide.

Los valores con los que se ha experimentado han sido  $\text{coefo} = (0.01, 0.1, 1)$ .

- **Dimensión:** Ha sido explorado en el kernel polinómico.

Los valores con los que se ha experimentado han sido  $d = (1, 2, 3, 4)$ .

Finalmente, una vez obtenidos los resultados del experimento con cada kernel, escogemos el que de un error mínimo para entrenar un modelo con todos los datos y guardarlo.

## KERNEL POLINÓMICO

En la siguiente tabla, podemos observar los resultados obtenidos para los diferentes valores que han sido asignados a la dimensionalidad.

Dimensión = 1		
C	EDT	EDV
0.001	60.73	60.68
0.01	15.26	14.80
0.1	11.06	10.80
1	9.82	10.25
10	9.26	10.17
100	9.15	10.20
1000	9.08	10.15
10000	9.05	10.17
Dimensión = 2		
C	EDT	EDV
0.001	84.47	84.90
0.01	16.76	16.92
0.1	7.12	7.33
1	3.85	4.95
10	1.39	4.78
100	0.15	5.82
Dimensión = 3		
C	EDT	EDV
0.001	84.54	84.92
0.01	21.34	22.28
0.1	6.03	6.65
1	2.26	3.78
10	0.27	3.83
Dimensión = 4		
C	EDT	EDV
0.001	84.99	85.23
0.01	39.82	40.55
0.1	11.37	12.73
1	3.02	5.75
10	0.33	5.10

Una vez obtenido el mejor valor ( $C = 1$ ,  $EDT = 2.26$ ,  $EDV = 3.78$ ,  $D = 3$  y  $\gamma = \text{Default}$ ), procedemos a probar diferentes  $\gamma$ .

$\gamma = 0.1$		
C	EDT	EDV
0.001	88.71	88.88
0.01	88.71	88.88
0.1	88.71	88.88
1	88.71	88.88
10	85.97	86.2
100	31.35	31.83
1000	7.29	7.6
10000	2.75	4.18
$\gamma = 1$		
C	EDT	EDV
0.001	88.71	88.88
0.01	85.97	86.2
0.1	31.36	31.83
1	7.29	7.6
10	2.75	4.18
100	0.41	3.57
1000	0.01	4.08
$\gamma = 10$		
C	EDT	EDV
0.001	7.29	7.6
0.01	2.75	4.18
0.1	0.41	3.57
1	0.01	4.08

Observamos que la mejor gamma es  $\gamma = 1$  y  $C = 100$  y  $\gamma = 10$  y  $C = 0.1$ . Por último, volvemos a realizar el experimento (con  $\gamma = 1$  y  $D = 3$  y  $\gamma = 10$  y  $D = 3$ ) para diferentes términos independientes.

Finalmente, los mejores resultados los obtenemos con los coeficientes  $\text{coefo} = 0.1$  con  $C = 100$ ,  $EDT = 0.25$  y  $EDV = 3.08$  (para  $\gamma = 1$ ); y  $\text{coefo} = 1$  con  $C = 0.1$ ,  $EDT = 0.25$  y  $EDV = 3.08$  (para  $\gamma = 10$ ). Dependiendo de con que características se quiere el modelo se elegirá un parámetro u otro.

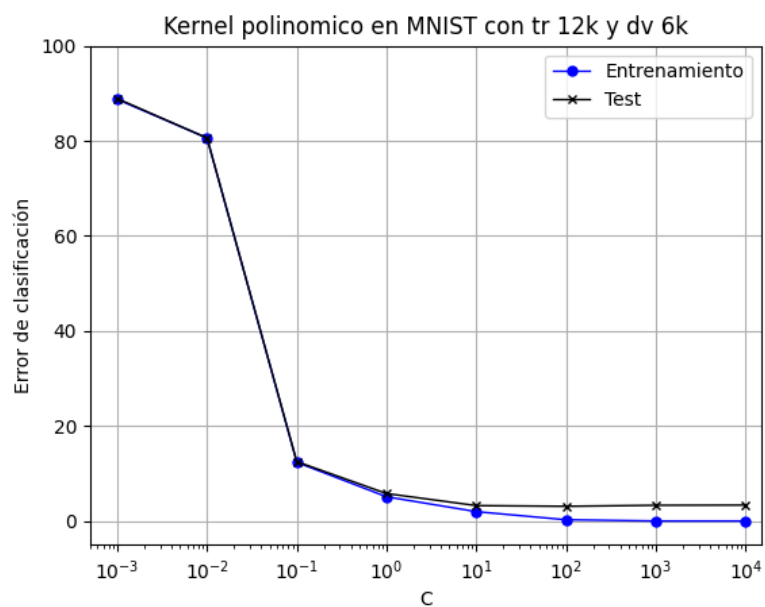


Fig 1. Gráfica que representa el error del experimento

con  $d = 3$ , coefo = 0.1 y  $\gamma = 1$

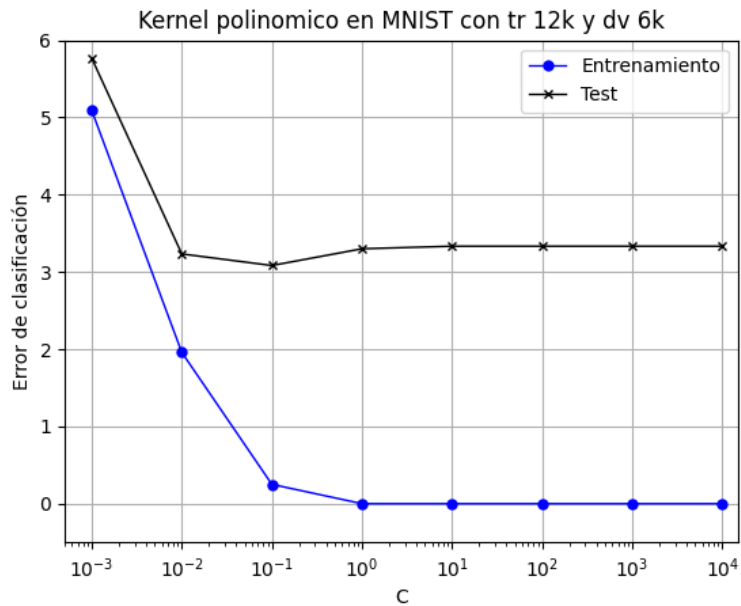


Fig 2. Gráfica que representa el error del experimento

con  $d = 3$ , coefo = 1 y  $\gamma = 10$

Por último, entrenamos el modelo con todos los datos y obtenemos un error del 0,42%.

## KERNEL GAUSSIANO

En cuanto a este kernel, se ha realizado el experimento con diferentes valores de  $\gamma$ . Comparando los resultados, se ha concluido que los mejores parámetros son  $C = 10$  y  $\gamma = \text{Default}$  porque  $\text{EDT} = 0.26$  y  $\text{EDV} = 2.85$ .

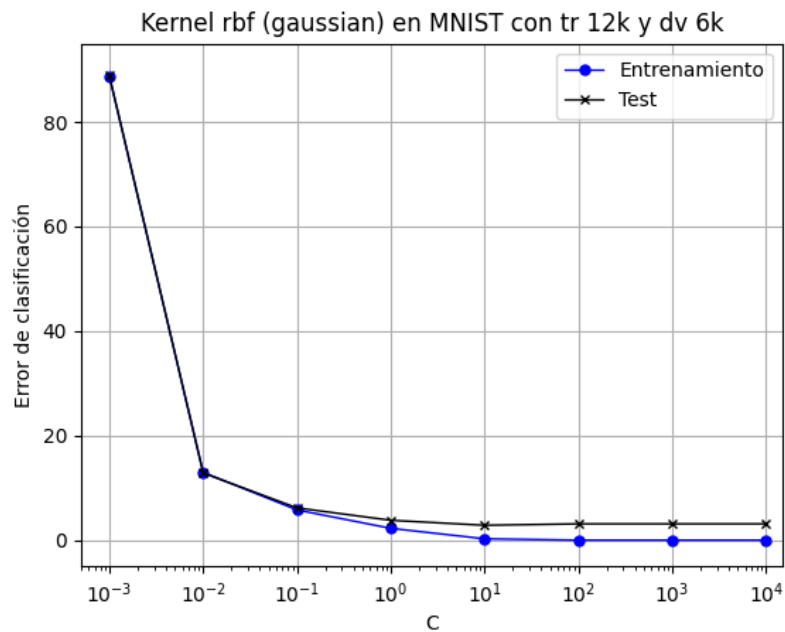


Fig 3. Gráfica que representa el error del experimento con  $\gamma = \text{Default}$

Finalmente, entrenamos el modelo con todos los datos y obtenemos un error del 0.35%. Como podemos observar, este modelo ha reducido el error en comparación al modelo del kernel polinómico.

## KERNEL SIGMOIDE

En cuanto a este kernel, se ha realizado el experimento con diferentes valores de  $\gamma$ . Comparando los resultados, se ha concluido que los mejores parámetros son  $C = 100$  y  $\gamma = 0.1$  porque EDT = 9.42 y EDV = 10.12. Una vez obtenidos estos resultados, también se han explorado diferentes valores para el termino independiente, pero con  $\text{coefo} = 0$  es cuando se obtienen mejores resultados.

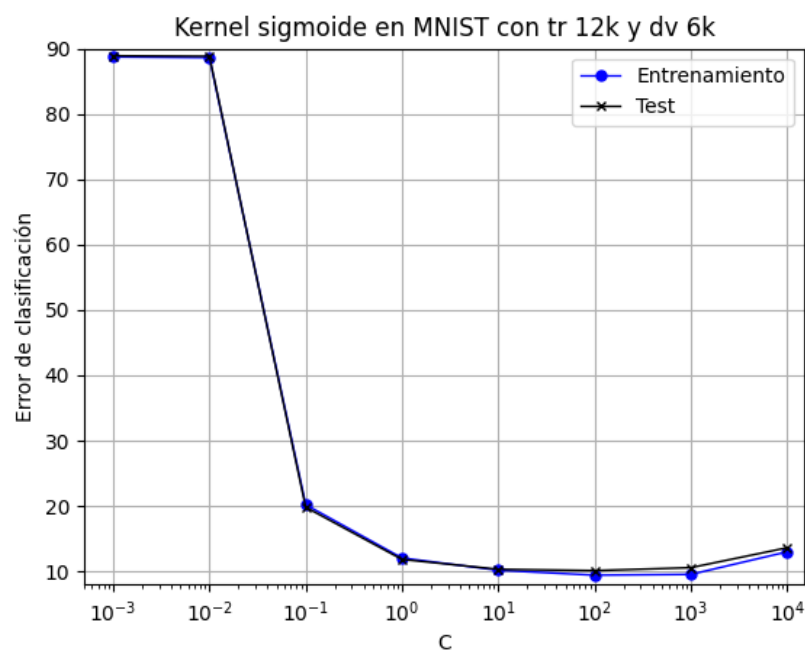


Fig 4. Gráfica que representa el error del experimento con  $\gamma = 0.1$  y  $\text{coefo} = 0$

Finalmente, entrenamos el modelo con todos los datos y obtenemos un error del 9.82%.

## CONCLUSIÓN

Como se ha podido observar a lo largo de esta memoria, el kernel que mejores resultados ha dado es el gaussiano puesto que el modelo entrenado con todos los datos ha sido el que menor error ha generado. Además, es un modelo con un valor C que hace que este no sea excesivamente restrictivo ni excesivamente tolerante.