

APR: Práctica sobre SVM

Roselló Beneitez, N. U.; Roselló Oviedo, M.

6 de Enero de 2020

Índice

1. Descripción de la práctica	3
2. Un pequeño ejercicio completo de aprendizaje	3
3. Aplicación de SVM a MNIST	8
4. Conclusiones	9

Índice de figuras

1. Resultados para el conjunto linealmente separable	3
2. Conjunto separable, $C = 0.1$	4
3. Conjunto separable, $C = 1$	4
4. Conjunto separable, $C = 10$	5
5. Conjunto separable, $C = 1000$	5
6. Resultados para el conjunto no linealmente separable	6
7. Conjunto no separable, $C = 0.1$	6
8. Conjunto no separable, $C = 1$	7
9. Conjunto no separable, $C = 10$	7
10. Conjunto no separable, $C = 1000$	8
11. Error según el grado del kernel polinómico (MNIST)	8
12. Error según el grado del kernel polinómico (MNIST)	9

1. Descripción de la práctica

En esta práctica se han trabajado con máquinas de vector soporte implementadas en Octave mediante la librería *libsvm*. Dicha librería admite diversos tipos de parámetros, como pueden ser: la tolerancia del sistema modelada mediante el parámetro C , el tipo del *kernel* y su grado, entre otras opciones. Los experimentos de esta práctica se han centrado mayoritariamente en pruebas con variaciones de los tres parámetros mencionados previamente, puesto que se considera que son los más relevantes dado lo visto en las sesiones de teoría para la correcta realización de la práctica.

Puesto que las fórmulas utilizadas se encuentran al final de la práctica, se procederá directamente a la exposición de los resultados sin entrar en mayor detalle en el desarrollo matemático de estos.

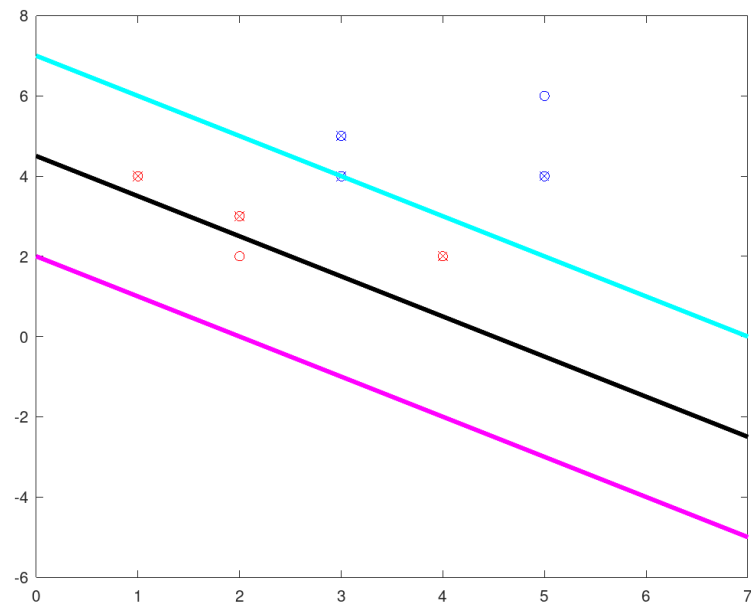
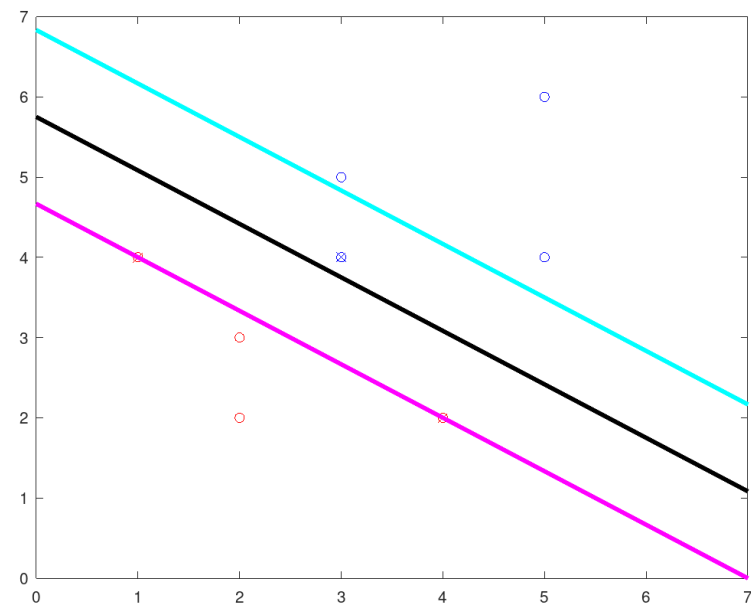
2. Un pequeño ejercicio completo de aprendizaje

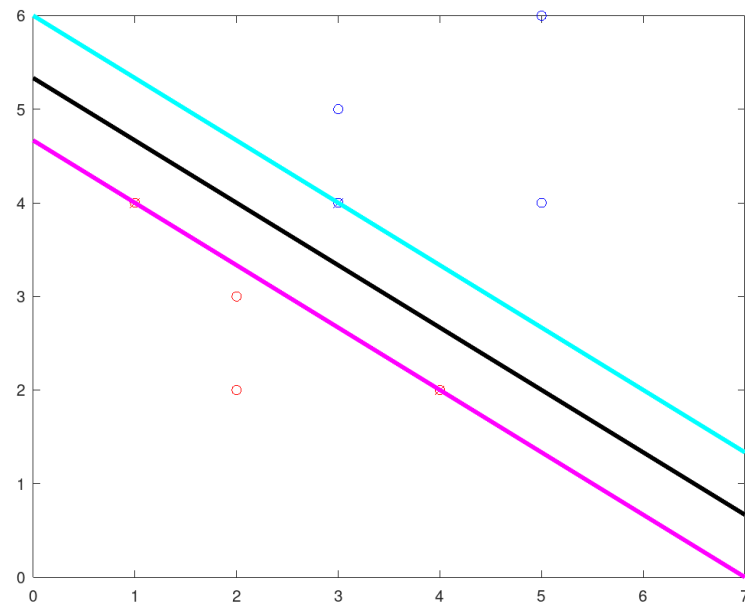
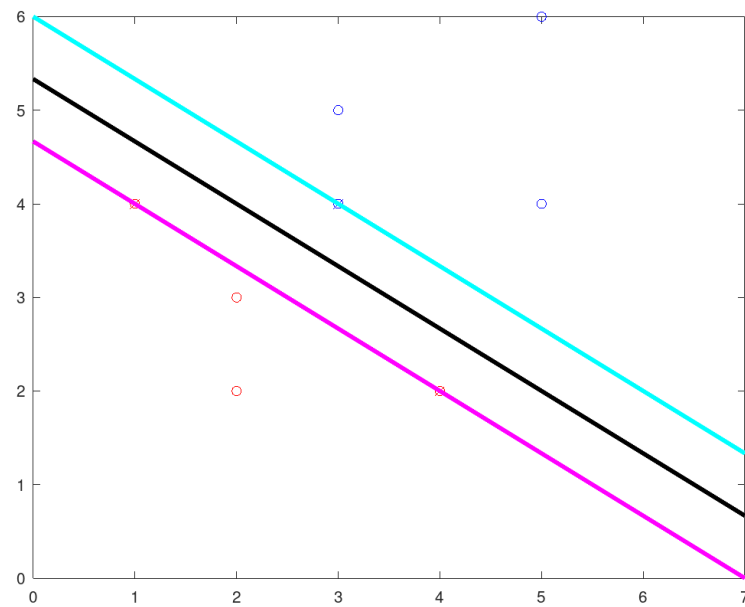
En el primer ejercicio a entregar, se han probado diversas combinaciones para los valores de C con respecto a los datos linealmente separables y a los no linealmente separables dados por el conjunto de datos *mini*. Concretamente, estos valores han seguido una especie de función logarítmica con el fin de ver el orden de magnitud en el cual los modelos mostraban sus mejores comportamientos.

En primer lugar, para el conjunto separable:

Separable				
-c	0.1	1	10	1000
Multiplicadores de Lagrange	0.10000	0.53846	0.87472	0.87472
	0.10000	0.46154	0.74989	0.74989
	0.10000	-1.00000	-1.62461	-1.62461
	-0.10000	-	-	-
	-0.10000	-	-	-
	-0.10000	-	-	-
Vectores soporte	(1, 4)	(1, 4)	(1, 4)	(1, 4)
	(2, 3)	(4, 2)	(4, 2)	(4, 2)
	(3, 4)	(3, 4)	(3, 4)	(3, 4)
	(3, 5)	-	-	-
	(5, 4)	-	-	-
Vector de pesos	(-0.40000, -0.40000)	(-0.61538, -0.92308)	(-0.99955, -1.49978)	(-0.99955, -1.49978)
Peso umbral	1.80000	5.30770	7.99870	7.99870
Margen	3.53550	1.80280	1.10970	1.10970

Figura 1: Resultados para el conjunto linealmente separable

Figura 2: Conjunto separable, $C = 0.1$ Figura 3: Conjunto separable, $C = 1$

Figura 4: Conjunto separable, $C = 10$ Figura 5: Conjunto separable, $C = 1000$

En cuanto al conjunto no separable, se ha obtenido:

No Separable				
-c	0.1	1	10	1000
Multiplicadores de Lagrange	250.87	250.87	250.87	250.87
	500.75	500.75	500.75	500.75
	1000.00	1000.00	1000.00	1000
	-751.62	-751.62	-751.62	-751.62
	-1000.00	-1000.00	-1000.00	-1000
Vectores soporte	(1, 4)	(1, 4)	(1, 4)	(1, 4)
	(4, 2)	(4, 2)	(4, 2)	(4, 2)
	(4, 4)	(4, 4)	(4, 4)	(4, 4)
	(3, 4)	(3, 4)	(3, 4)	(3, 4)
	(4, 3)	(4, 3)	(4, 3)	(4, 3)
Vector de pesos	(-0.99955, -1.49977)	(-0.99955, -1.49986)	(-0.99955, -1.49977)	(-0.99955, -1.49986)
Peso umbral	7.99770	7.99860	7.99770	7.99860
Margen	1.10970	1.10970	1.10970	1.10970
Márgenes de tolerancia	0.00090	0.00090	0.00090	0.00000
	0.00000	0.00000	0.00000	0.00000
	2.99865	2.99865	2.99955	2.99865
	0.00000	0.00000	0.00000	0.00090
	0.50023	0.50023	0.50023	0.50113

Figura 6: Resultados para el conjunto no linealmente separable

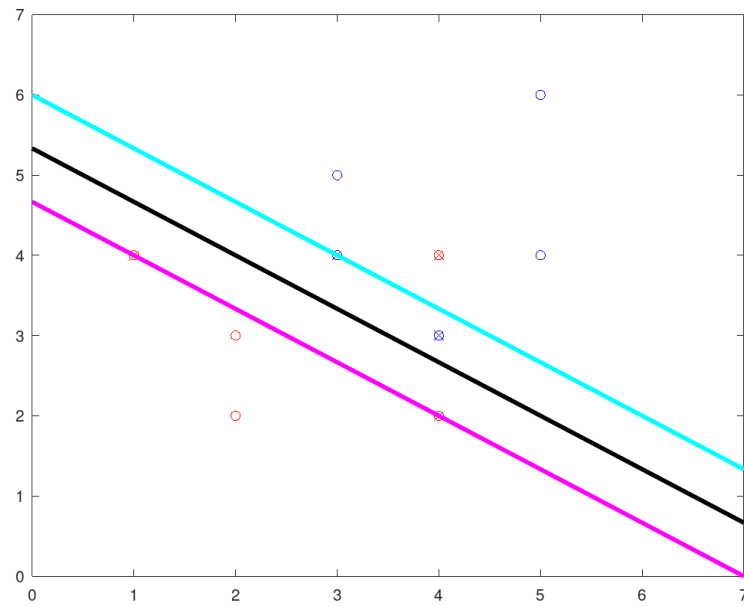
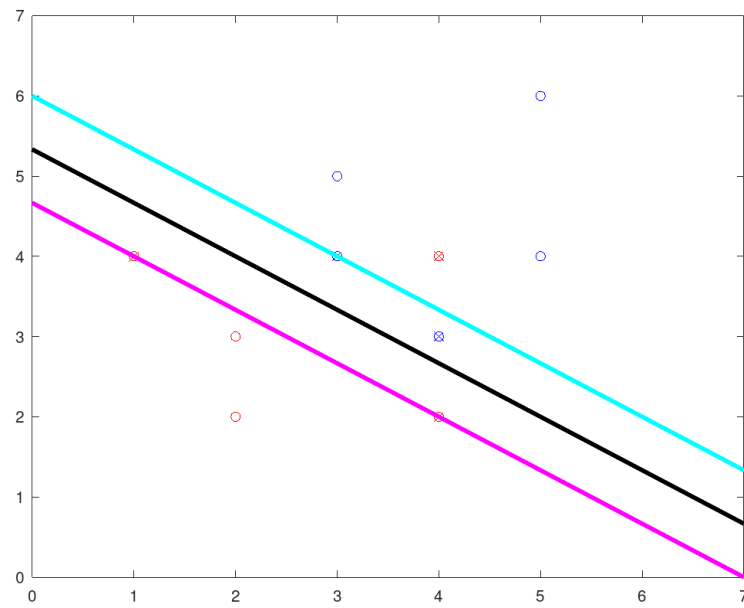
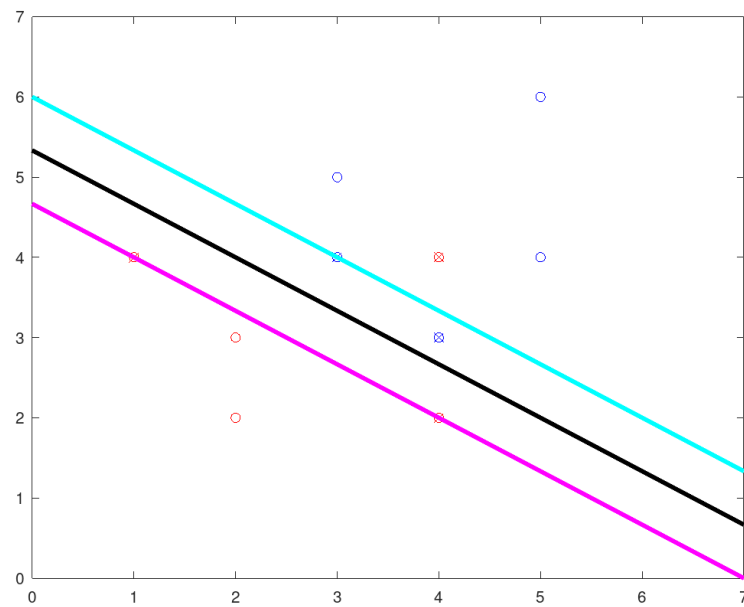
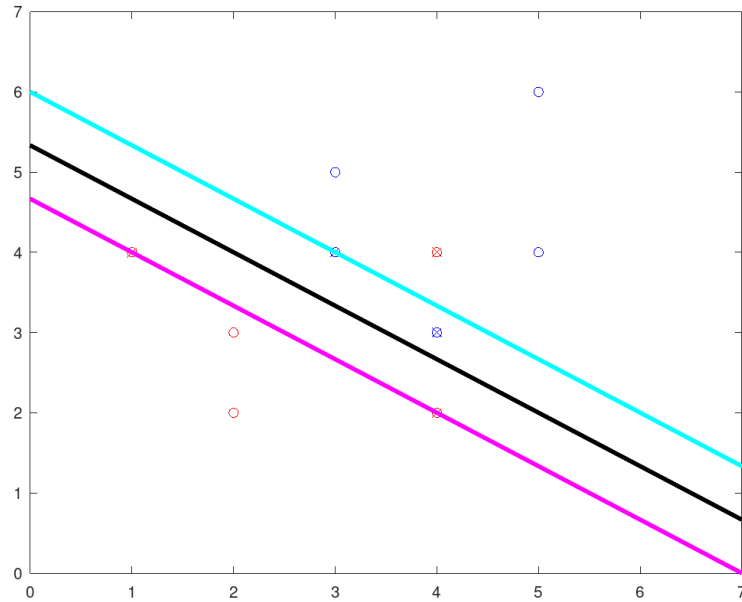


Figura 7: Conjunto no separable, $C = 0.1$

Figura 8: Conjunto no separable, $C = 1$ Figura 9: Conjunto no separable, $C = 10$

Figura 10: Conjunto no separable, $C = 1000$

3. Aplicación de SVM a MNIST

En el segundo ejercicio, correspondiente a la base de datos MNIST, se han probado tanto diversos valores de C como diversos grados del polinomio correspondiente al *kernel* con el que se ha ejecutado el entrenamiento. Cabe destacar que únicamente se ha probado este *kernel* polinómico debido a que los otros tipos de *kernel* disponibles (lineal, gaussiano o de función radial y sigmoide) ya han sido testeados en el entregable de teoría correspondiente al tema 4 y se ha observado como su tasa de acierto es muy reducida para este caso concreto. Además, se conoce que el *kernel* polinómico es el de mayor acierto gracias a la página web del MNIST. Tras diversas pruebas, se ha determinado que el valor de C no es relevante para este caso de estudio, ya que se obtienen los mismos resultados para tolerancias en el rango $[0,001 - 1000]$.

MNIST kernel polinómico		
Grado	% error	% IC ($\alpha = 0.05$)
2	1.95	$\pm 0.1106\%$
3	2.09	$\pm 0.1145\%$
4	2.63	$\pm 0.1280\%$
5	3.35	$\pm 0.1440\%$
6	4.12	$\pm 0.1590\%$
7	5.27	$\pm 0.1788\%$
8	6.36	$\pm 0.1953\%$
9	7.42	$\pm 0.2097\%$

Figura 11: Error según el grado del kernel polinómico (MNIST)

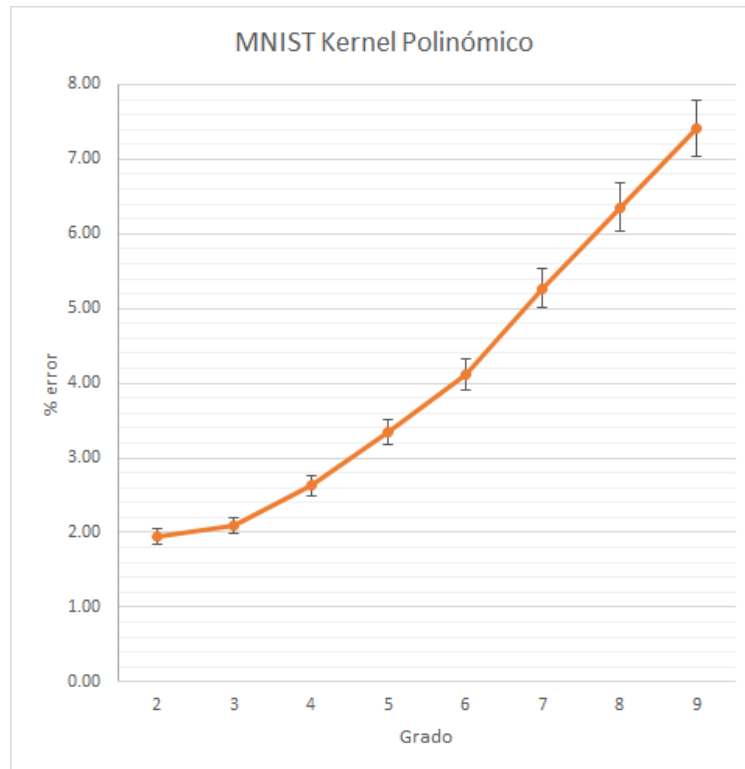


Figura 12: Error según el grado del kernel polinómico (MNIST)

4. Conclusiones

Resulta interesante observar el efecto de la tolerancia (C) en las muestras linealmente separables y no linealmente separables (ejercicio 3). Este parámetro es similar al parámetro b (margen) del *perceptron*: cuanto más C , más tienden los parámetros a ser los “correctos”, es decir, los más generalizables. Esto es porque con un valor alto de C se premia la minimización de la tolerancia de margen, pero siempre respetando la restricción genérica de las máquinas de vector soporte de que las muestras cuyos multiplicadores de Lagrange no sean 0 estén a una distancia mínima del hiperplano separador.

Cabe mencionar también que, cuando C tiende a infinito, se puede observar que la tolerancia de margen de los datos no linealmente separables tiende a ser idéntica a la de los datos linealmente separables. Sin embargo, un valor excesivamente alto de C acaba siendo de poca utilidad debido a que ya no es posible minimizar más la tolerancia de margen; para $C = 10$ y $C = 1000$ se ha comprobado que los resultados son prácticamente idénticos (excepto los multiplicadores de Lagrange, que se tienen que adaptar al cambio en la C).

En cuanto a la base de datos de dígitos MNIST, como se ha podido comprobar en la sección de resultados, el valor de C es irrelevante en la medición del error. Esto puede ser debido a que, dado cualquier kernel polinómico de grado mayor o igual a dos, las muestras son polinómicamente separables.