

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	7
4. Conclusiones	9

Índice de figuras

1. Resultados para el conjunto linealmente separable	3
2. Conjunto separable, $C = 10$	4
3. Conjunto separable, $C = 100$	4
4. Conjunto separable, $C = 1000$	5
5. Resultados para el conjunto no linealmente separable	5
6. Conjunto no separable, $C = 10$	6
7. Conjunto no separable, $C = 100$	6
8. Conjunto no separable, $C = 1000$	7
9. Tabla de errores según el grado del kernel polinómico (MNIST) .	8
10. Gráfica de errores según el grado del kernel polinómico (MNIST)	8

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.

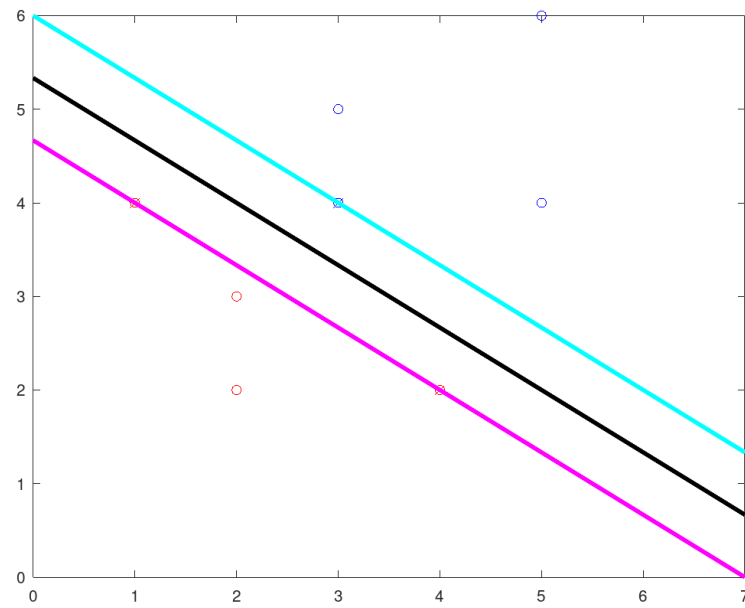
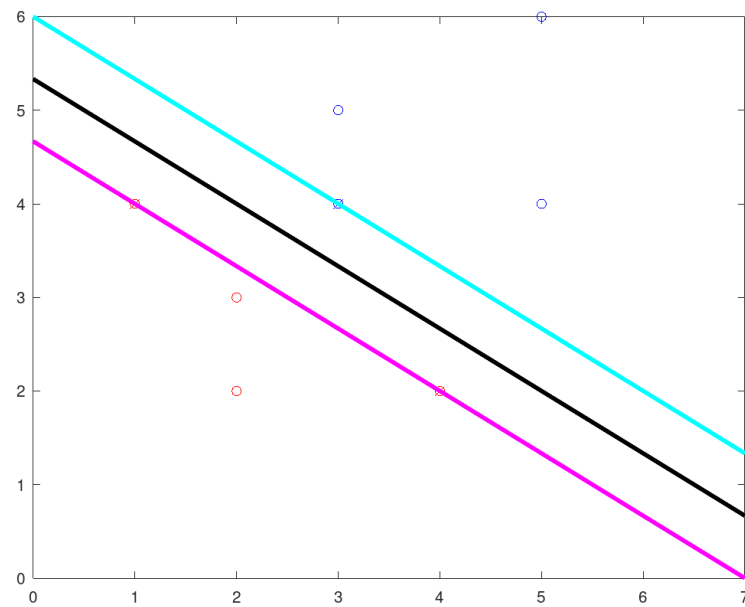
2. Un pequeño ejercicio completo de aprendizaje

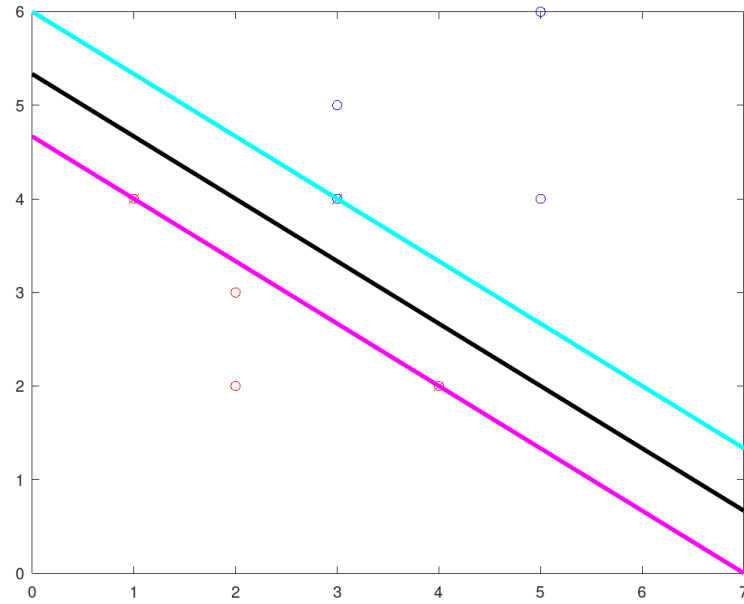
En el primer ejercicio a entregar, se han probado diversas combinaciones para los valores de C con respecto a los datos linealmente separables y a los no linealmente separables dados por el conjunto de datos *mini*. Concretamente, estos valores han seguido una especie de función logarítmica con el fin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

En primer lugar, para el conjunto separable:

Separable			
-c	10	100	1000
Multiplicadores de Lagrange	0.87472	0.87472	0.87472
	0.74989	0.74989	0.74989
	-1.62461	-1.62461	-1.62461
Vectores soporte	(1, 4)	(1, 4)	(1, 4)
	(4, 2)	(4, 2)	(4, 2)
	(3, 4)	(3, 4)	(3, 4)
Vector de pesos	(-0.99955, -1.49978)	(-0.99955, -1.49978)	(-0.99955, -1.49978)
Peso umbral	7.99870	7.99870	7.99870
Margen	1.10970	1.10970	1.10970

Figura 1: Resultados para el conjunto linealmente separable

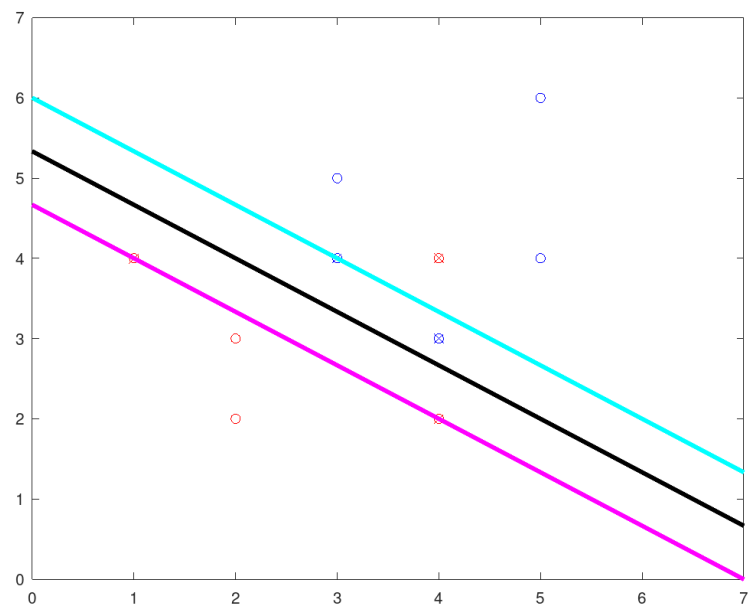
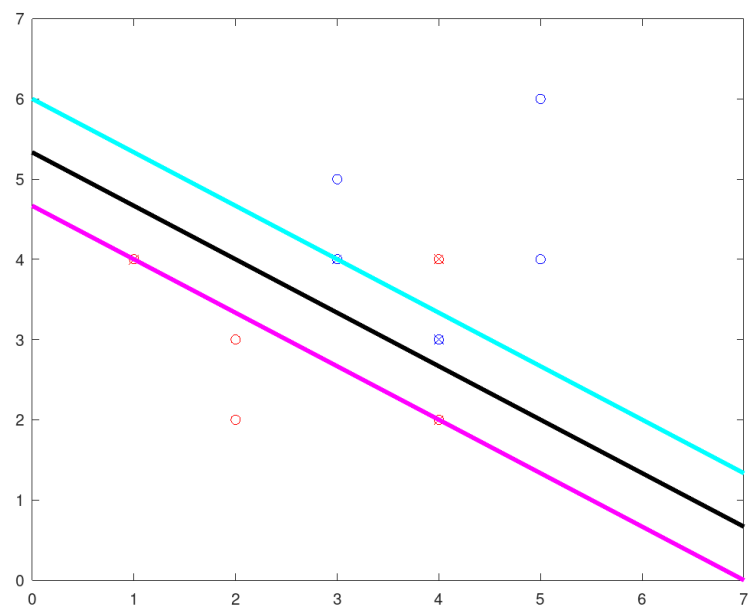
Figura 2: Conjunto separable, $C = 10$ Figura 3: Conjunto separable, $C = 100$

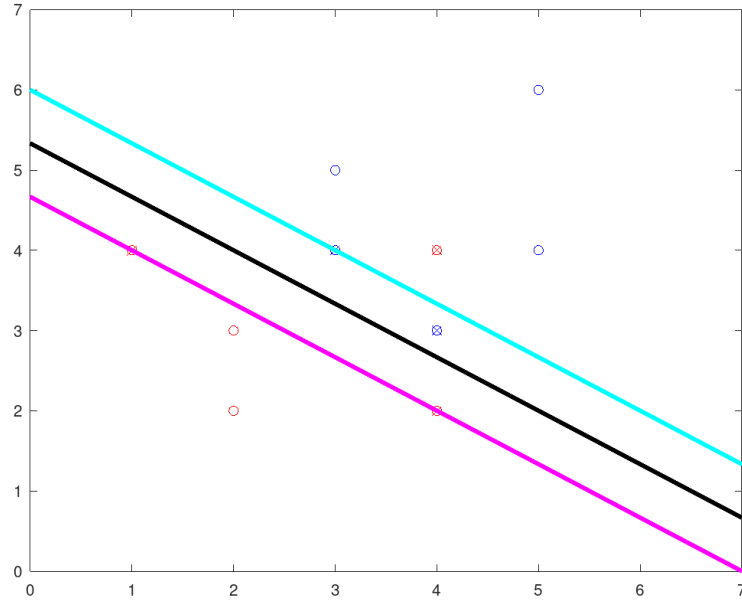
Figura 4: Conjunto separable, $C = 1000$

En cuanto al conjunto no separable, se ha obtenido:

No Separable			
-c	10	100	1000
Multiplicadores de Lagrange	3.37490	25.87500	250.87000
	5.74980	50.75000	500.75000
	10.00000	100.00000	1000.00000
	-9.12470	-76.00000	-751.62000
	-10.00000	-100.00000	-1000.00000
Vectores soporte	(1, 4)	(1, 4)	(1, 4)
	(4, 2)	(4, 2)	(4, 2)
	(4, 4)	(4, 4)	(4, 4)
	(3, 4)	(3, 4)	(3, 4)
	(4, 3)	(4, 3)	(4, 3)
Vector de pesos	(-1.00000, -1.49960)	(-0.99955, -1.49978)	(-0.99955, -1.49986)
Peso umbral	7.99860	7.99870	7.99860
Margen	1.10960	1.10970	1.10970
Márgenes de tolerancia	0.00000	0.00000	0.00000
	0.00070	-0.00090	0.00000
	3.00000	2.99866	2.99865
	0.00000	0.00090	0.00090
	0.49965	0.50112	0.50113

Figura 5: Resultados para el conjunto no linealmente separable

Figura 6: Conjunto no separable, $C = 10$ Figura 7: Conjunto no separable, $C = 100$

Figura 8: 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 (95%)
2	1.95	$\pm 0.2710\%$
3	2.09	$\pm 0.2804\%$
4	2.63	$\pm 0.3137\%$
5	3.35	$\pm 0.3527\%$
6	4.12	$\pm 0.3896\%$
7	5.27	$\pm 0.4379\%$
8	6.36	$\pm 0.4783\%$
9	7.42	$\pm 0.5137\%$

Figura 9: Tabla de errores según el grado del kernel polinómico (MNIST)

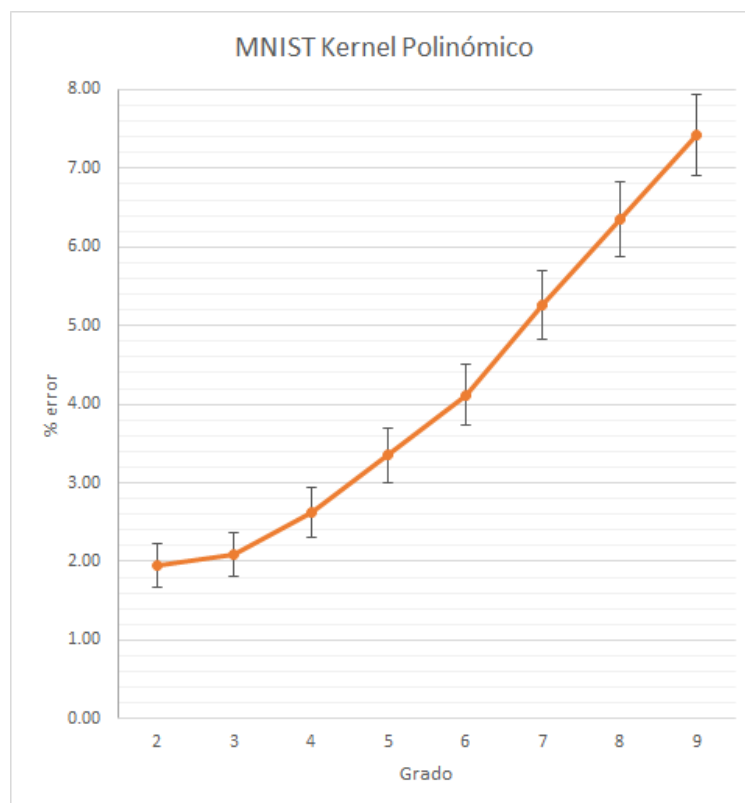


Figura 10: Gráfica de errores según el grado del kernel polinómico (MNIST)

4. Conclusiones

Resulta interesante observar el efecto de la tolerancia (C) en las muestras linealmente separables y no linealmente separables (ejercicio 3). Los valores altos de C tienden a clasificar mejor las muestras, a costa de minimizar la tolerancia de margen. Sin embargo, al experimentar con valores bajos de C (por ejemplo, $C = 0.1$, que no se ha incluido en la memoria ya que se pedían valores altos), se ha observado que el sistema tiende a proporcionar una peor clasificación pero produce un margen mayor. En este caso práctico —un modelo muy reducido de solo dos dimensiones— la distribución de las muestras provoca que se premie a valores altos de C , puesto que hay poca separación espacial entre ellas. Es probable que, con un volumen de datos mayor de otra colección distinta, la decisión sobre el valor de C venga dada por la distribución de dichos datos, pudiendo ser beneficioso en algunos casos un valor más pequeño y precisando de experimentos para determinar empíricamente el mejor valor.

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. Por ejemplo, en el conjunto de datos *mini*, para $C \in \{10, 100, 1000\}$ se ha comprobado que los parámetros de los hiperplanos separadores son prácticamente idénticos.

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. El error mínimo, encontrado en el kernel polinómico de grado 2 y con un valor de 1.95 %, se acerca bastante a los errores de SVM reportados en la página web del MNIST, valores cuyo máximo es 1.4 %. El resultado es bastante satisfactorio, dado que los valores de la página web tienen algún preproceso extra, como aplicación de distorsiones o torsiones del número en cuestión, lo cual puede ayudar al reconocimiento tal y como se vio en PER.