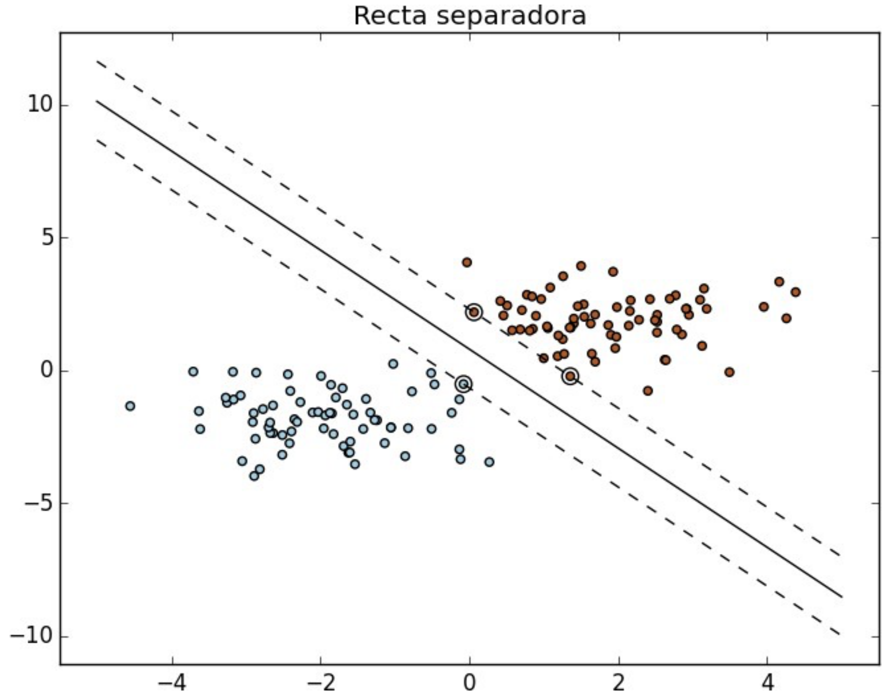


MÁQUINAS DE VECTORES SOPORTE

Aprendizaje automático

Resultados obtenidos

En esta segunda parte del proyecto hemos estudiado las máquinas de vectores de soporte.

En primer lugar, para el conjunto de datos `data/hart/tr.dat` y `data/hart/trlabels.dat`, hemos representado gráficamente los vectores soporte obtenidos mediante la función svmtrain, superponiéndolos a las muestras de entrenamiento.

El resultado se encuentra en el fichero `ex31.ofig`

En segundo lugar, para cada uno de los conjuntos del subdirectorio `data/mini` (`trSep.dat, trSeplabels.dat`) y (`tr.dat, trlabels.dat`) hemos calculado:

- SVM sin kernel. Esto ha sido posible mediante la instrucción:
  `res=svmtrain(xl,X,'-t 0 -c 1000')`
- Los Multiplicadores de Lagrange asociados a cada dato de entrenamiento los hemos obtenido a partir del resultado de la función anterior, mediante la orden `abs(res.sv_coef)`, ya que `res.sv_coef` nos retorna los multiplicadores de Lagrange, pero multiplicados por la clase de la muestra.
  Estos valores están presentes en las dos gráficas en la parte de arriba de cada muestra. Si miramos el valor de los multiplicadores de Lagrange, observaremos que todos valen 0 excepto los multiplicadores de Lagrange de aquellas muestras que son vectores soporte. Esto es porque una muestra es vector soporte si su multiplicador de Lagrange es distinto de 0.

- Vectores soporte. En el caso de la gráfica separable, podemos observar qué muestras son vectores soporte, ya que están marcadas mediante una cruz (+). Además, en la gráfica No Separable están marcados tanto los vectores soporte "correctos" y "erróneos"
- En cuanto al vector de pesos y umbral de la función discriminante lineal se obtienen respectivamente mediante las fórmulas:

$$\boldsymbol{\theta} = \sum_{m \in \mathcal{V}} c_m \; \alpha_m \; \mathbf{x}_m \qquad \theta_0 \; = \; c_m - \boldsymbol{\theta}^t \mathbf{x}_m$$

  En los dos conjuntos de datos hemos obtenido los valores
  `theta=[-0.99955 -1.49978]` y `theta0=7.9987`

- Margen: mediante la fórmula $2/|\theta|$. En los dos conjuntos de datos hemos obtenido un margen de 1.11
- Parámetros de la frontera lineal (recta) de separación, calculados mediante la fórmula:

$$\phi(\mathbf{x};\boldsymbol{\theta},\theta_0) = 0 \quad \Rightarrow \quad \theta_1 x_1 + \theta_2 x_2 + \theta_0 = 0 \quad \Rightarrow \quad x_2 = -\frac{\theta_1}{\theta_2} x_1 - \frac{\theta_0}{\theta_2},$$

Esta recta está presente en las dos gráficas, dibujada mediante una línea negra. Además, las dos rectas que definen las fronteras del margen se han representado con los colores magenta y cian para diferenciarlas de la recta de separación, y las hemos obtenido con las fórmulas:

$$\phi(\mathbf{x};\boldsymbol{\theta},\theta_0) = +1 \quad \Rightarrow \quad \theta_1 x_1 + \theta_2 x_2 + \theta_0 = +1 \quad \Rightarrow \quad x_2 = -\frac{\theta_1}{\theta_2}x_1 - \frac{\theta_0 - 1}{\theta_2}$$

$$\phi(\mathbf{x};\boldsymbol{\theta},\theta_0) = -1 \quad \Rightarrow \quad \theta_1 x_1 + \theta_2 x_2 + \theta_0 = -1 \quad \Rightarrow \quad x_2 = -\frac{\theta_1}{\theta_2}x_1 - \frac{\theta_0 + 1}{\theta_2}$$

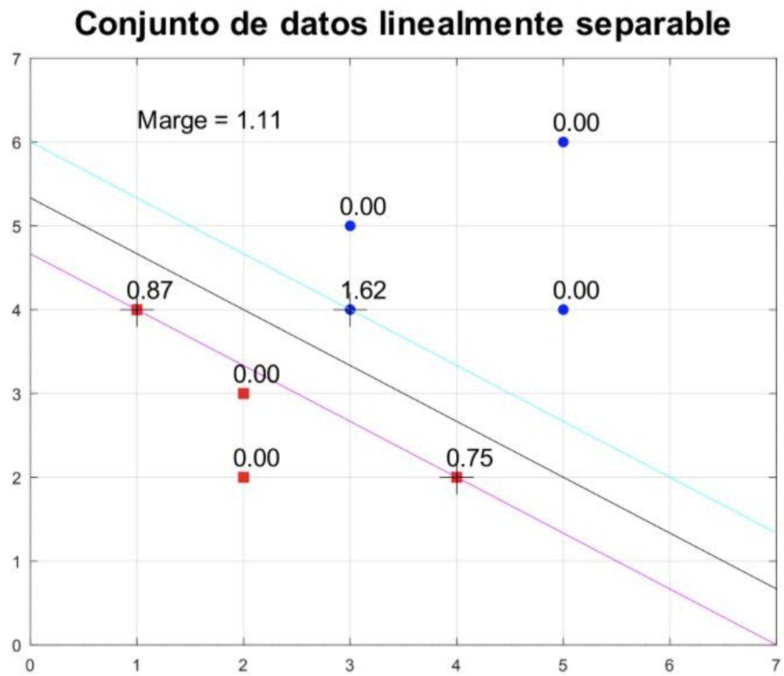


Ilustración 1 Representación gráfica para el caso separable

La gráfica para el conjunto de datos linealmente separable es la Ilustración 1.

A parte de los parámetros calculados anteriormente, para el conjunto no-separable también hemos determinado los valores de tolerancia de margen, asociados a cada dato de entrenamiento. Estos valores están presentes en la Ilustración 2 (gráfica para el conjunto de datos no-separable), ya que es el valor que hay a la derecha de cada muestra.

Esta tolerancia la hemos calculado mediante la fórmula:

$$\zeta_m = 1 - c_m(\boldsymbol{\theta}^t \mathbf{x}_m + \theta_0)$$

Además, los vectores soporte “erróneos” han sido marcados en la representación gráfica con un aspa (x), mientras que los vectores soporte “correctos” están representados mediante una cruz (+).

Por último, hemos variado C, probando los valores 1, 10, 100 y 1000 y hemos creado una gráfica para cada caso concreto.

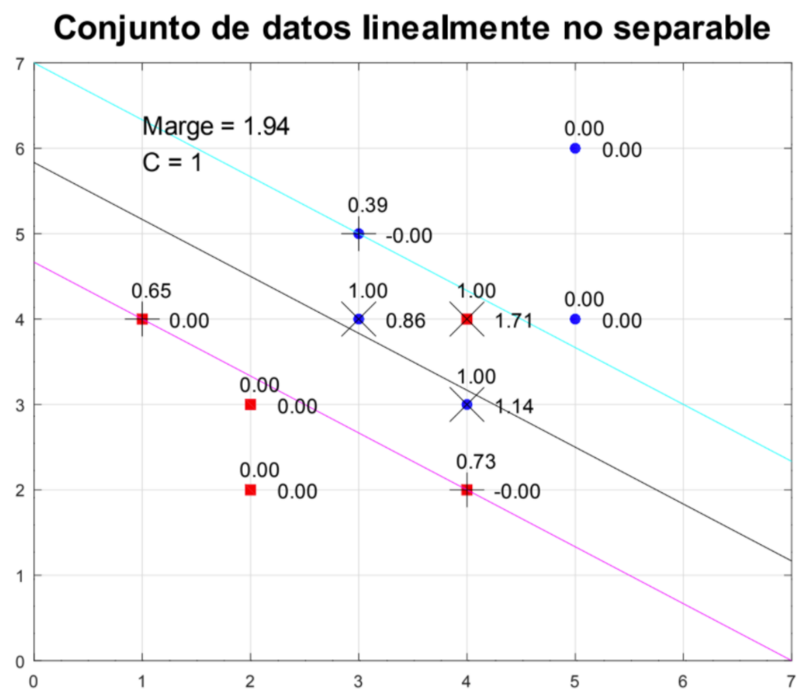


Ilustración 1 Representación gráfica para el caso no-separable con C=1

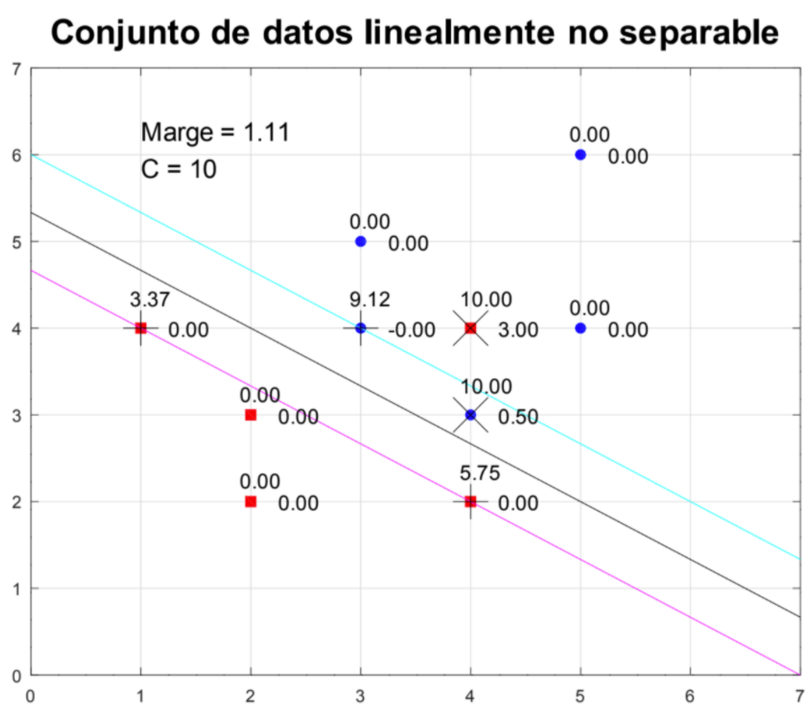


Ilustración 2 Representación gráfica para el caso no-separable con C=10

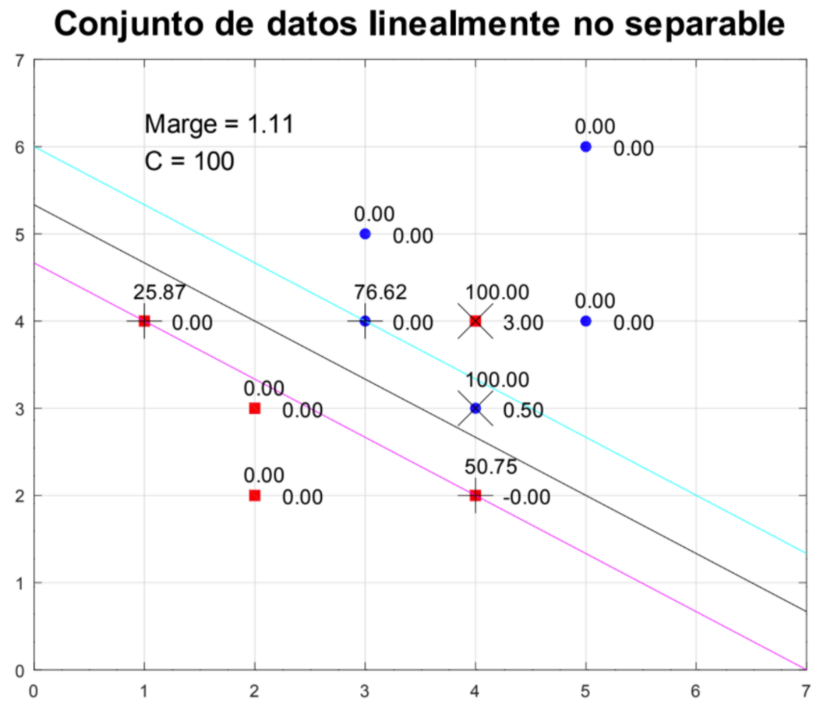


Ilustración 3 Representación gráfica para el caso no-separable con C=100

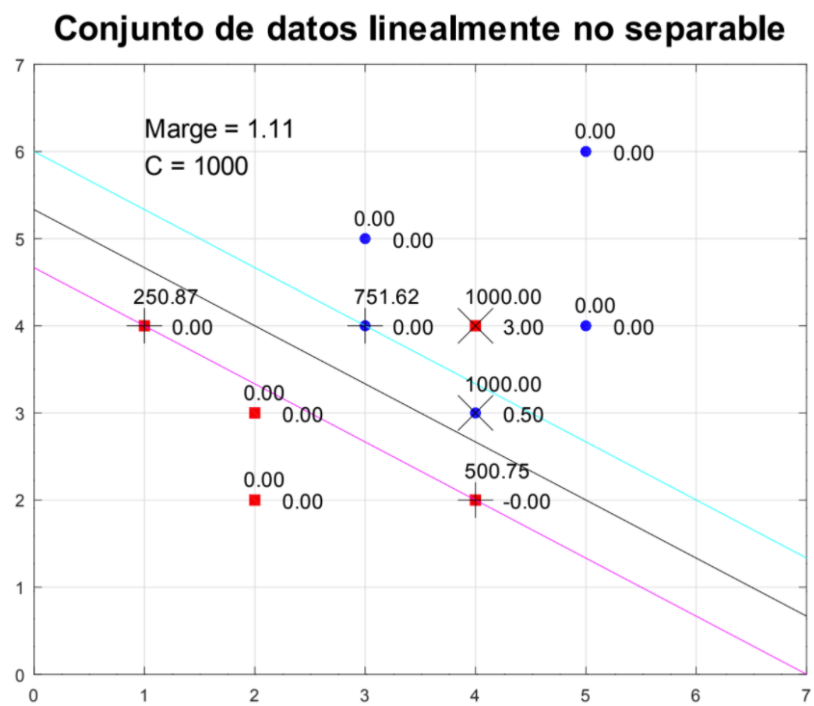


Ilustración 4 Representación gráfica para el caso no-separable con C=1000

Los ficheros `graficaSep.m` y `graficaNoSep.m` son los utilizados para construir las dos gráficas.

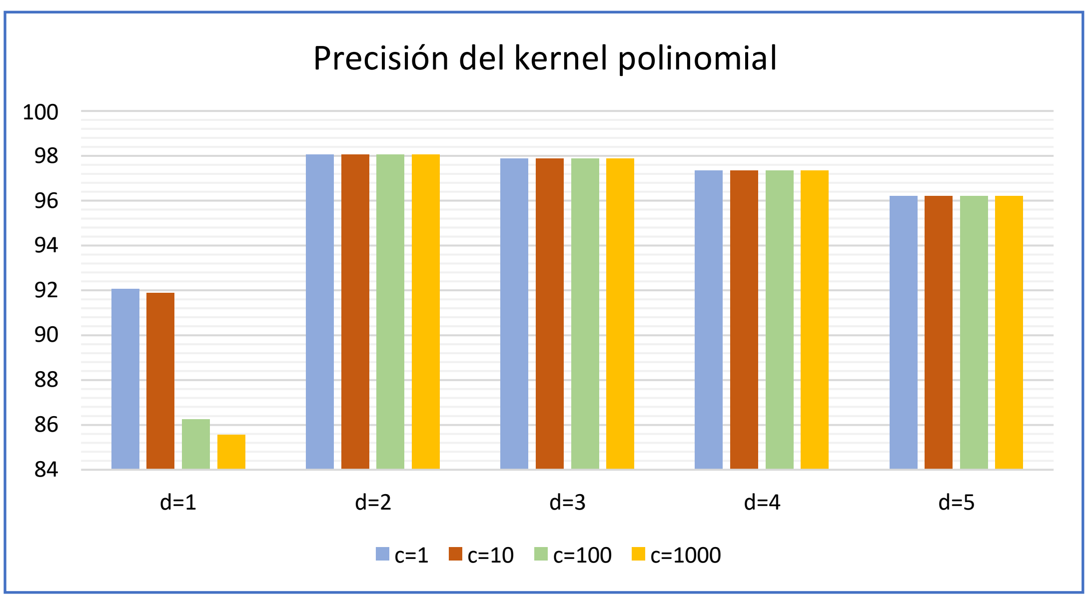
A continuación, hemos realizado un experimento para evaluar el error de clasificación en función de los parámetros del clasificador basado en SVM. Para ello hemos considerado los siguientes valores del parámetro C (-c 1, 10, 100, 1000, 10000) y los siguientes tipos de kernel (-t 0, 1, 2, 3) utilizando la partición entrenamiento-evaluación 90-10. Los valores que se muestran a continuación son para los valores 0, 2 y 3 del kernel:

T	C	PRECISIÓN (%)	INTERVALO (%-%)
0	1 – 10000	84.7167	[84.428800 – 85.004600]
2	1	11.71667	[11.459370 – 11.973970]
2	10 – 10000	11.63333	[11.376673 – 11.889930]
3	1	11.63333	[11.376730 – 11.889930]
3	10 – 10000	11.61667	[11.360270 – 11.873070]

Además, en el caso del kernel polinomial (t = 1), hemos explorado el grado del polinomio (-d 1 2 3 4 5) para averiguar como varia en función a él. Los resultados obtenidos se encuentran en la siguiente tabla:

T	C	D	PRECISIÓN(%)	INTERVALO (%-%)
1	1	1	92.06667	[91.850415 – 92.282918]
1	**1**	**2**	**98.05000**	**[97.939358 – 98.160642]**
1	1	3	97.88333	[97.768158 – 97.998509]
1	1	4	97.35000	[97.221480 – 97.478520]
1	1	5	96.21666	[96.064000 – 96.369333]
1	10	1	91.88333	[91.664815 – 92.101852]
1	**10**	**2**	**98.05000**	**[97.939358 – 98.160642]**
1	10	3	97.88333	[97.768158 – 97.998509]
1	10	4	97.35000	[97.221480 – 97.478520]
1	10	5	96.21666	[96.064000 – 96.369333]
1	100	1	86.25000	[85.974443 – 86.525557]
1	**100**	**2**	**98.05000**	**[97.939358 – 98.160642]**
1	100	3	97.88333	[97.768158 – 97.998509]
1	100	4	97.35000	[97.221480 – 97.478520]
1	100	5	96.21667	[96.064000 – 96.369333]
1	1000	1	85.55000	[84.660300 – 86.439700]
1	**1000**	**2**	**98.05000**	**[97.700100 – 98.399900]**
1	1000	3	97.88333	[97.519130 – 98.247530]
1	1000	4	97.35000	[96.943600 – 97.756400]
1	1000	5	96.21667	[95.733870 – 96.699470]

Como se puede observar en la tabla, el valor de la precisión es contante para los casos en los que se usa el grado del polinomio 2, independientemente del valor del parámetro C. Esto también queda reflejado en la siguiente gráfica:



A partir de los resultados obtenidos con `svm-exp.m` hemos extraído los parámetros para las mejores tasas de error: t = 1, d = 2 y c = {1, 10, 100}. Al ejecutar el programa `svm-eva.m` con esos parámetros obtenemos una precisión de 98.05% en el intervalo [97.77983 - 98.321017], es decir, un error aproximado de 1.95%.

Para comparar los resultados de este clasificador con los de la página de MNIST, hemos utilizado los errores correspondientes a “SVM, Gaussian Kernel” (1.4%) y “SVM deg 4 polynomial” (1.1%). Como se puede comprobar nuestro error es superior a los extraídos de la página de MNIST, lo cual se debe parcialmente al preprocesamiento de los datos mediante “deskewing”.

Según los datos de la página de MNIST, este clasificador también es mejor que el utilizado tanto en la práctica anterior (mixtura de gausianas) como los utilizados en la asignatura de PER (el clasificador lineal y el KNN).