Boletín 5: Máquinas de Soporte Vectorial

Para la realización de las prácticas correspondientes a este boletín se utilizará scikit-learn en el CESGA. Utilizaremos un **SEED_VALUE=1**.

1. Dado el siguiente conjunto de datos de clasificación con 16 observaciones, 2 variables de entrada y una variable de salida, mediante una SVM lineal con C=1 se han obtenido los coeficientes α_i indicados en la última columna:

Observación	X_1	X_2	Y	α_i
1	2	6	1	0
2	4	3	1	1
3	4	4	1	0,3333
4	4	6	1	0
5	6	3	1	1
6	7	7	1	0,1667
7	8	4	1	1
8	9	8	1	1
9	2	1	-1	1
10	6	2	-1	0,5
11	7	4	-1	1
12	8	8	-1	1
13	9	1	-1	0
14	10	3	-1	0
15	10	6	-1	1
16	12	4	-1	0

Indica:

- Cuáles son los vectores de soporte y cuáles de ellos están en el límite del margen.
- Cuáles son los coeficientes del hiperplano (β y β_0) y el valor de M.
- Los valores de ε_i y las observaciones incorrectamente clasificadas.

Nota: este ejercicio debe hacerse sin utilizar ninguna función de scikit-learn.

2. Dado el problema de clasificación Blood Transfusion Service Center:

   a. La clase que implementa las SVM en problemas de clasificación en scikit-learn es *sklearn.svm.SVC* (existen otras dos clases, pero nos centraremos en ésta). Revisa los parámetros y métodos que tiene.

   b. Divide los datos en entrenamiento (80%) y test (20%).

   c. Realiza la experimentación con *SVC* usando los valores por defecto de los parámetros, excepto para los siguientes hiper-parámetros:

      i. *kernel* en donde deberás probar el *'linear', 'poly'* (con gamma=1) y *'rbf'.*

      ii. *C*, parámetro de regularización (para todos los *kernels*). Prueba potencias enteras de 10 (…; 0,01; 0,1; 1; 10; 100; …). Valores muy grandes de C provocan tiempos de cómputo muy elevados. No pruebes en ningún caso valores superiores a 10^{11}.

      iii. *degree*: grado del polinomio en el *kernel* polinómico. Debe ser mayor que 1, si no sería lineal. No pruebes valores superiores a 5. En estos casos debes limitar aún más el valor máximo de C para que el cómputo se haga en un tiempo razonable.

      iv. *gamma* en el caso del *kernel rbf*. Prueba potencias enteras de 10 (…; 0,01; 0,1; 1; 10; 100; …). Para el *kernel* polinómico utiliza *gamma*=1.

   Muestra la gráfica del error de entrenamiento con validación cruzada (5-CV) frente al valor del hiper-parámetro. En el caso del kernel *rbf* muestra la gráfica frente a *C* para algunos valores de *gamma* —los que consideres más representativos. De forma equivalente, para *degree* con el *kernel* polinomial. Justifica la elección del valor más apropiado.

   Para cada tipo de kernel, ¿cuál es el menor error de validación cruzada, su desviación estándar y el valor de los hiper-parámetros para el que se consigue?

   Muestra la gráfica del error de test frente al valor del hiper-parámetro, y valora si la gráfica del error de entrenamiento con validación cruzada ha hecho una buena estimación del error de test.

   Para cada tipo de kernel, ¿cuál es el error de test para el valor de los hiper-parámetros seleccionados por la validación cruzada?

Entregable

Se debe entregar un único fichero comprimido con el nombre *PrimerApellido_SegundoApellido.zip* (también son válidos los formatos .rar y .7z), que contenga dos archivos:

- El primer archivo debe ser de tipo pdf, y contendrá exclusivamente las respuestas a los ejercicios (incluyendo las gráficas necesarias para justificar dichas respuestas). No se incluirá en este archivo ningún otro tipo de texto.

- El segundo archivo será de tipo ipynb, y permitirá reproducir toda la experimentación realizada en el boletín.

Ejercicio 1

Observación	X_1	X_2	Y	α
1	2	6	1	0
2	4	3	1	1
3	4	4	1	0.3333
4	4	6	1	0
5	6	3	1	1
6	7	7	1	0.1667
7	8	4	1	1
8	9	8	1	1
9	2	1	-1	1
10	6	2	-1	0.5
11	7	4	-1	1
12	8	8	-1	1
13	9	1	-1	0
14	10	3	-1	0
15	10	6	-1	1
16	12	4	-1	0

Tabla 1: Dataset con valores de α

I. **Indica cuáles son los vectores de soporte y cuáles de ellos están en el límite del margen.**

- Vectores de soporte: observaciones 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 15.
- Vectores de soporte en el límite del margen: observaciones 3, 6, 10.

II. **Indica cuáles son los coeficientes del hiperplano (β y β_0) y el valor de M.**

- Coeficientes del hiperplano: $\beta = (-0{,}5, 0{,}5)$, $\beta_0 = 1$
- Valor de M: $\sqrt{2} \approx 1{,}4142$

III. **Indica los valores de ε_i y las observaciones incorrectamente clasificadas.**

Observación	X_1	X_2	Y	α	ε_i
1	2	6	1	0	0
2	4	3	1	1	0.5
3	4	4	1	0.3333	0
4	4	6	1	0	0
5	6	3	1	1	1.5
6	7	7	1	0.1667	0
7	8	4	1	1	2
8	9	8	1	1	0.5
9	2	1	-1	1	1.5
10	6	2	-1	0.5	0
11	7	4	-1	1	0.5
12	8	8	-1	1	2
13	9	1	-1	0	0
14	10	3	-1	0	0
15	10	6	-1	1	0
16	12	4	-1	0	0

Tabla 2: Valores de ε_i para cada observación. En rojo, las observaciones mal clasificadas.

Ejercicio 2

- **Menor error de validación cruzada, su desviación estándar y valor de los hiperparámetros:**
    - Kernel lineal: $\Delta = 0{,}234146$, $\sigma = 0{,}01123$, param_C = 100000.
    - Kernel polinómico: $\Delta = 0{,}220728$, $\sigma = 0{,}037121$, param_C = 10000, param_degree = 3.
    - Kernel radial: $\Delta = 0{,}195616$, $\sigma = 0{,}021161$, param_C = 1, param_gamma = 1.
- **Error de test para los hiperparámetros de validación cruzada:**
    - Kernel lineal: $\Delta = 0{,}22$
    - Kernel polinómico: $\Delta = 0{,}233333$
    - Kernel radial: $\Delta = 0{,}226667$

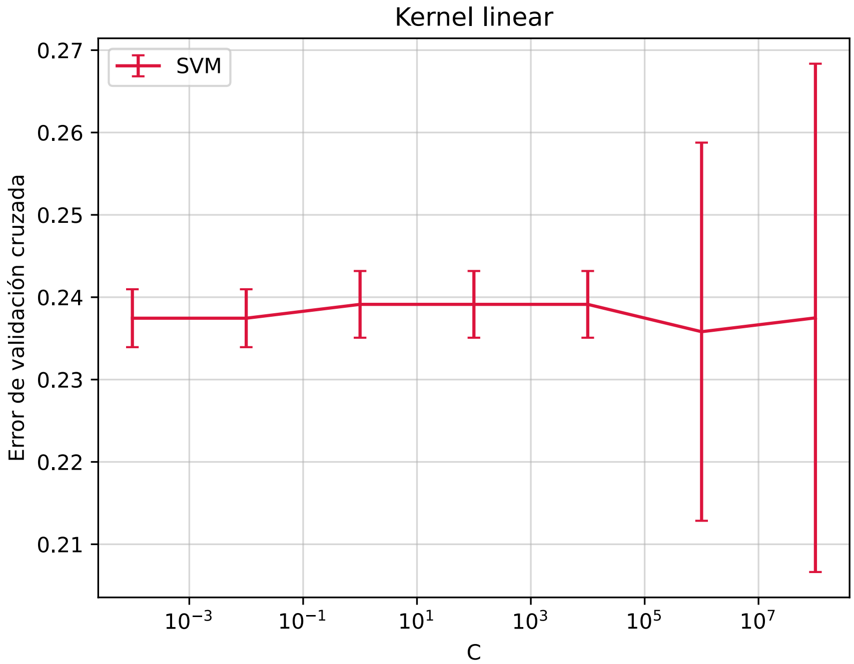

Figura I: Ejercicio 2: error de validación cruzada en datos de validación para el kernel lineal (exploración de grano grueso).

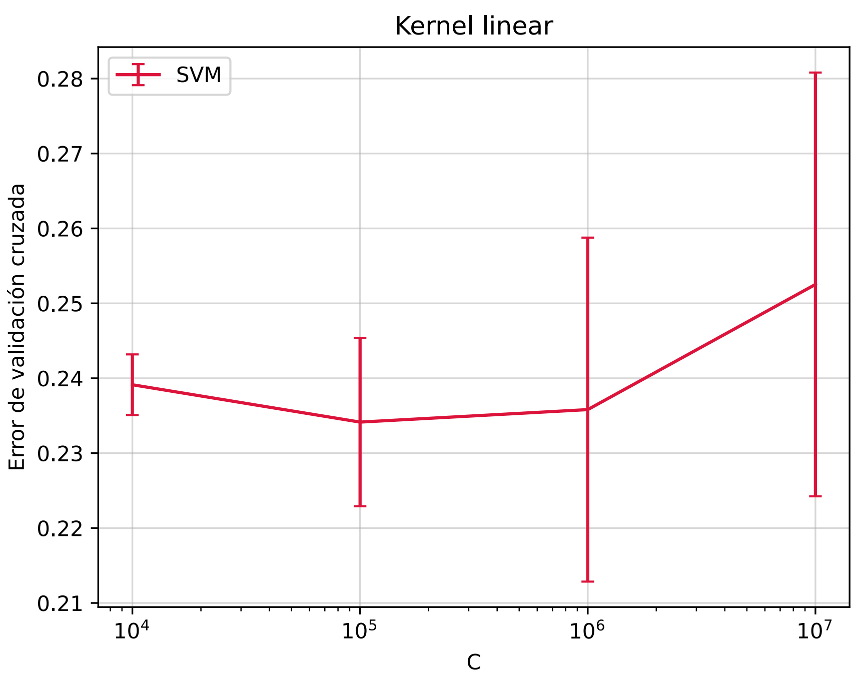

Figura II: Ejercicio 2: error de validación cruzada en datos de validación para el kernel lineal (exploración de grano fino).

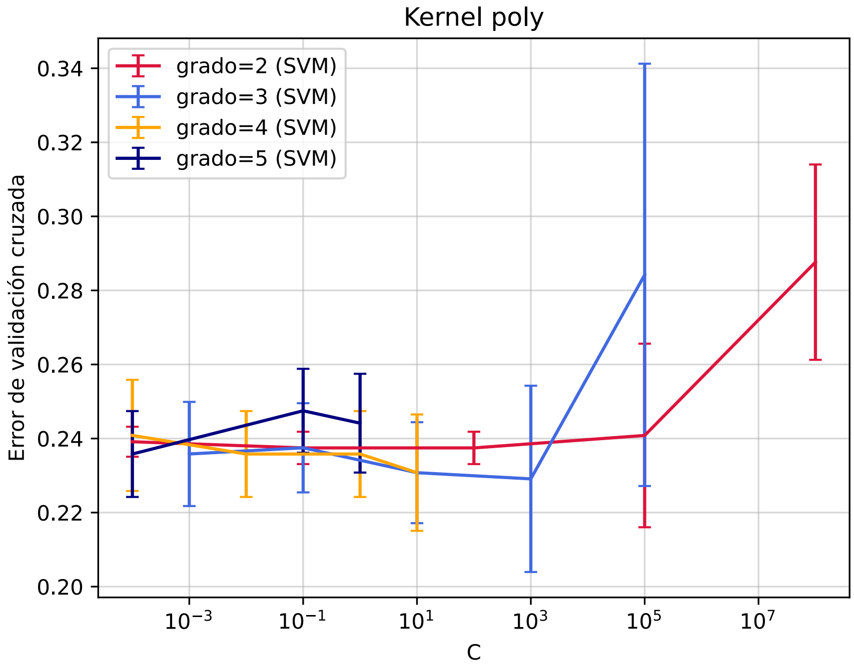

Figura III: Ejercicio 2: error de validación cruzada en datos de validación para el kernel polinómico (exploración de grano grueso).

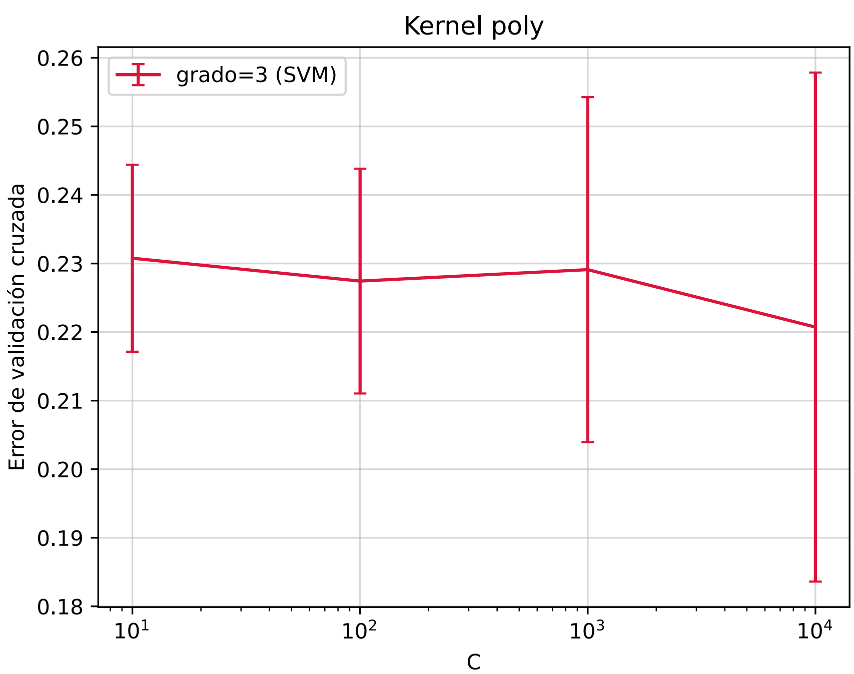

Figura IV: Ejercicio 2: error de validación cruzada en datos de validación para el kernel polinómico (exploración de grano fino).

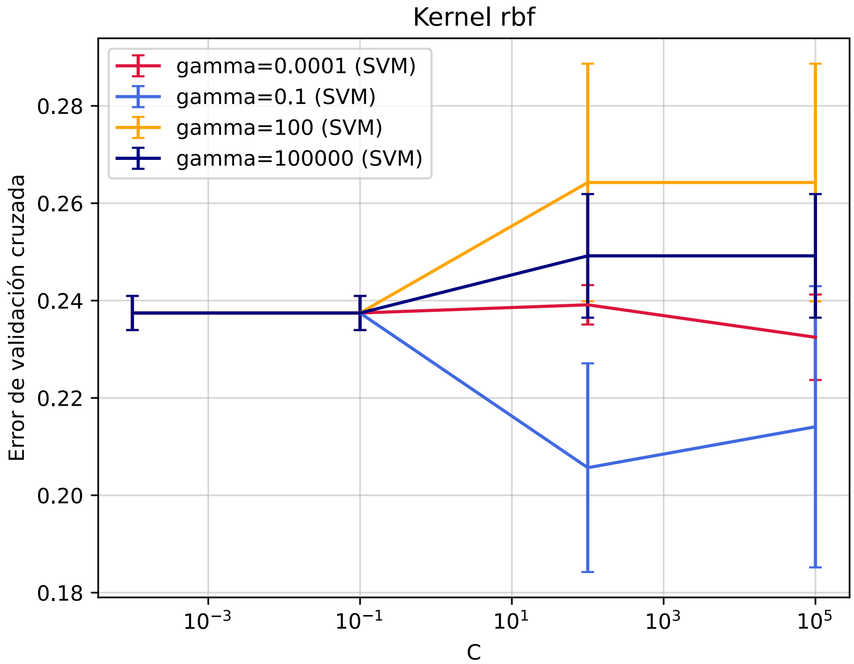

Figura V: Ejercicio 2: error de validación cruzada en datos de validación para el kernel radial (exploración de grano grueso).

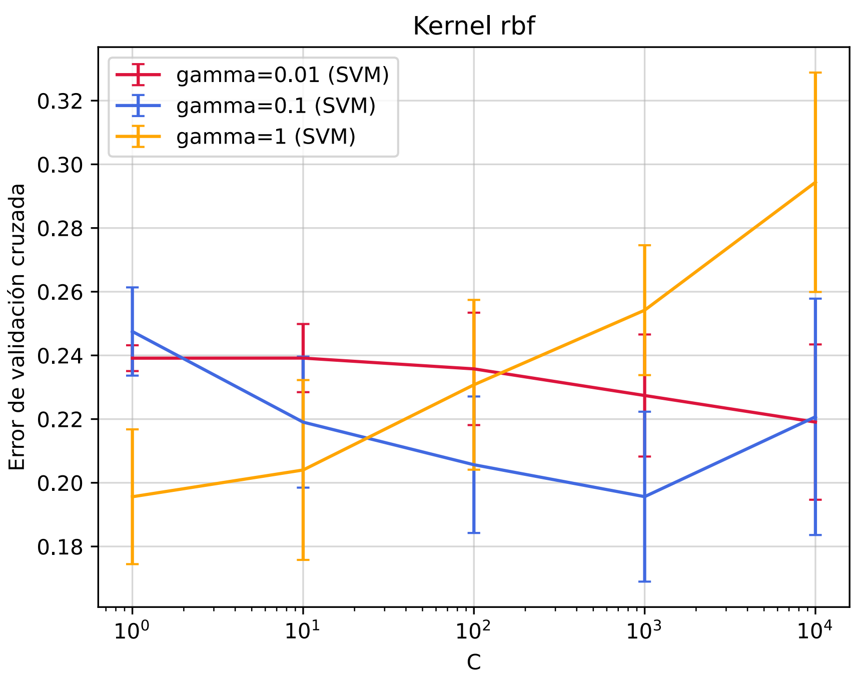

Figura VI: Ejercicio 2: error de validación cruzada en datos de validación para el kernel radial (exploración de grano fino).

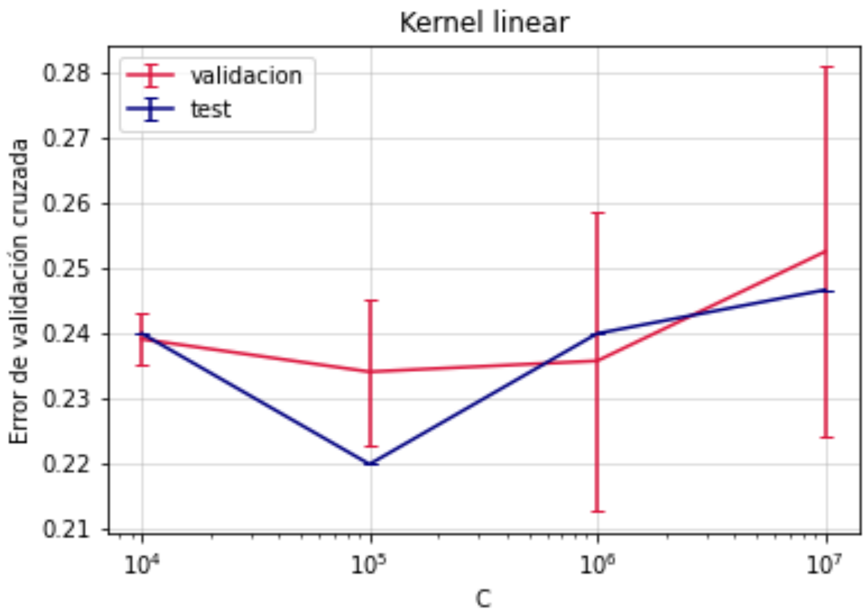

Figura VII: Ejercicio 2: error de validación cruzada en datos de validación frente al mismo error en datos de test para el kernel lineal.

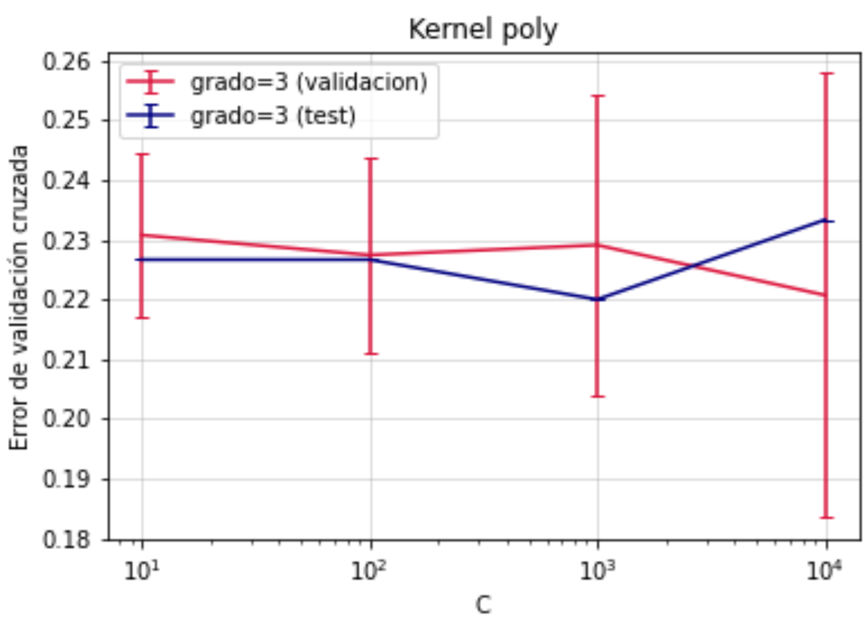

Figura VIII: Ejercicio 2: error de validación cruzada en datos de validación frente al mismo error en datos de test para el kernel polinómico de grado 3.

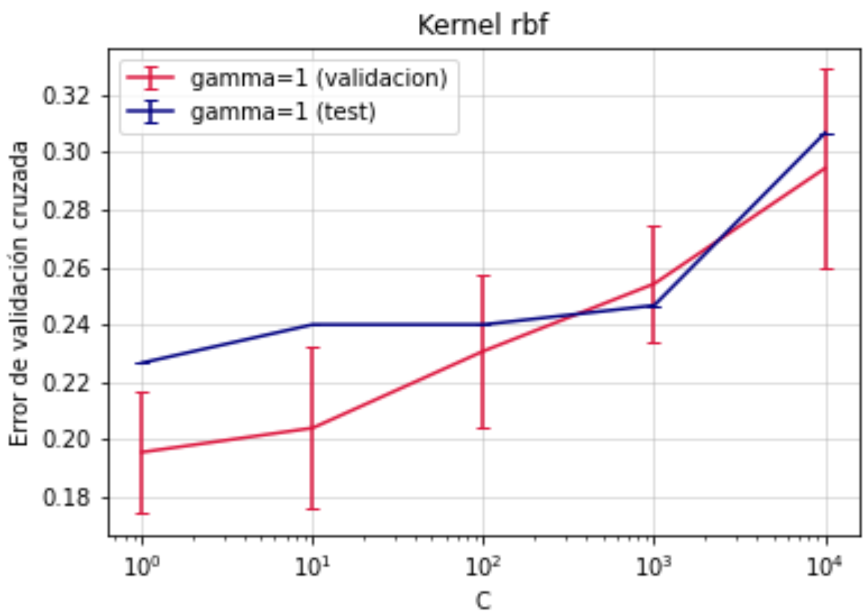

Figura IX: Ejercicio 2: error de validación cruzada en datos de validación frente al mismo error en datos de test para el kernel radial.

Luis Ardévol Mesa

Boletín 5. Ejercicio 1

Observación	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_1	2	4	4	4	6	7	8	9	2	6	7	8	9	10	10	12
x_2	6	3	4	6	3	7	4	8	1	2	4	8	1	3	6	4
y	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1
α_i	0	1	0.3333 ↳ 1/3	0	1	0.1667 ↳ 1/6	1	1	1	0.5 ↳ 1/2	1	1	0	0	1	0

- Vectores de soporte $\longrightarrow \alpha_i > 0$

  - En el límite del margen $\longrightarrow \xi_i = 0$, $0 < \alpha_i < C$
  - Resto $\longrightarrow \xi_i > 0$, $\alpha_i = C$

  $\Big\}$ $\underline{C = 1}$

1.) Vectores de soporte $\longrightarrow$ $\boxed{2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 15}$

En el límite del margen $\longrightarrow$ $\boxed{3, 6, 10}$

2.) Para obtener β_0, necesitamos primero $\beta = \sum_{i=1}^{N} \alpha_i y_i x_i$

$$(x_1): \beta_1 = \cancel{4} + \frac{4}{3} + \cancel{6} + \frac{7}{6} + \cancel{8} + \cancel{9} + \cancel{(-2)} + (-3) + \cancel{(-7)} + \cancel{(-8)} + \cancel{(-10)} = \frac{4}{3} + \frac{7}{6} - 3 = -0.5$$

$$(x_2): \beta_2 = \cancel{3} + \frac{4}{3} + \cancel{3} + \frac{7}{6} + \cancel{4} + \cancel{8} + (-1) + (-1) + \cancel{(-4)} + \cancel{(-8)} + \cancel{(-6)} = \frac{4}{3} + \frac{7}{6} - 2 = 0.5$$

$$\boxed{\beta = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}}$$

Obtenemos β_0 usando los vectores en el límite $(\xi_i = 0) \longrightarrow \alpha_i[y_i(x_i^T\beta + \beta_0) - (1 - \cancel{\xi_i})] = 0 \Rightarrow \beta_0 = \frac{1}{y_i} - x_i^T\beta$

Observación 3: $\beta_0^{(3)} = 1 - (4\ 4)\begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} = 1$

Observación 6: $\beta_0^{(6)} = 1 - (7\ 7)\begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} = 1$

Observación 10: $\beta_0^{(10)} = -1 - (6\ 2)\begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} = 1$

$$\boxed{\beta_0 = \frac{1}{3}\left(\beta_0^{(3)} + \beta_0^{(6)} + \beta_0^{(10)}\right) = 1}$$

El valor de M lo obtenemos como $\boxed{M = \|\beta\|^{-1} = \left[\left(1/4 + 1/4\right)\right]^{-1/2} = \sqrt{2} \approx 1.4142}$

3.) Para las observaciones 1, 4, 13, 14, 16, $\xi_i = 0$ ya que $\alpha_i = 0$; veamos por qué

$$\left.\begin{array}{l} \left.\begin{array}{l} \alpha_i = 0 \\ \alpha_i = C - \mu_i \end{array}\right\} \mu_i = C = 1 \neq 0 \\ \qquad\qquad\quad \uparrow \\ \qquad\qquad\ \text{dato} \\ \mu_i \xi_i = 0 \end{array}\right\} \Rightarrow \xi_i = 0 \quad \forall i \ / \ \underline{\alpha_i = 0}$$

Por tanto, $\xi_1=\xi_4=\xi_{13}=\xi_{14}=\xi_{16}=0$. Además, también sabemos que para los vectores de soporte en el margen $\xi_i=0$, por tanto $\xi_3=\xi_6=\xi_{10}=0$. Para el resto de observaciones, que son vectores de soporte.

$$\alpha_i\left[y_i(x_i^T\beta+\beta_0)-(1-\xi_i)\right]=0 \overset{\alpha_i\neq 0}{\Longrightarrow} \xi_i=1-y_i(x_i^T\beta+\beta_0)$$

$$\xi_2=1-\left[(4\ \ 3)\begin{pmatrix}-0.5\\0.5\end{pmatrix}+1\right]=1-\left[-2+\frac{3}{2}+1\right]=\frac{1}{2}<1$$

$$\xi_5=1-\left[(6\ \ 3)\begin{pmatrix}-0.5\\0.5\end{pmatrix}+1\right]=1-\left[-3+3/2+1\right]=\frac{3}{2}>1$$

$$\xi_7=1-\left[(8\ \ 4)\begin{pmatrix}-0.5\\0.5\end{pmatrix}+1\right]=1-\left[-4+2+1\right]=2>1$$

$$\xi_8=1-\left[(9\ \ 8)\begin{pmatrix}-0.5\\0.5\end{pmatrix}+1\right]=1-\left[-\frac{9}{2}+4+1\right]=\frac{1}{2}<1$$

$$\xi_9=1+\left[(2\ \ 1)\begin{pmatrix}-0.5\\0.5\end{pmatrix}+1\right]=1+\left[-1+\frac{1}{2}+1\right]=\frac{3}{2}>1$$

$$\xi_{11}=1+\left[(7\ \ 4)\begin{pmatrix}-0.5\\0.5\end{pmatrix}+1\right]=1+\left[-\frac{7}{2}+2+1\right]=\frac{1}{2}<1$$

$$\xi_{12}=1+\left[(8\ \ 8)\begin{pmatrix}-0.5\\0.5\end{pmatrix}+1\right]=1+1=2>1$$

$$\xi_{15}=1+\left[(10\ \ 6)\begin{pmatrix}-0.5\\0.5\end{pmatrix}+1\right]=1+\left[-5+3+1\right]=0<1$$

Hagamos una tabla con las observaciones y sus respectivas ξ_i. Indicaremos las observaciones mal clasificadas ($\xi_i>1$) en rojo.

Observación	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ξ_i	0	0.5	0	0	1.5	0	2	0.5	1.5	0	0.5	2	0	0	0	0