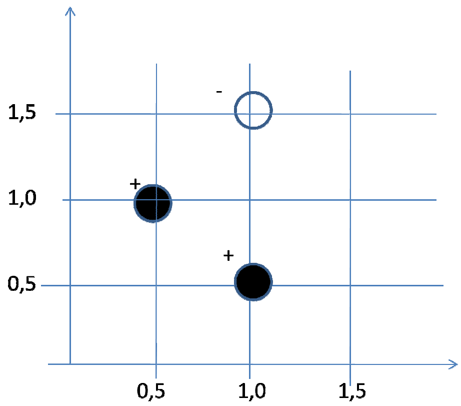
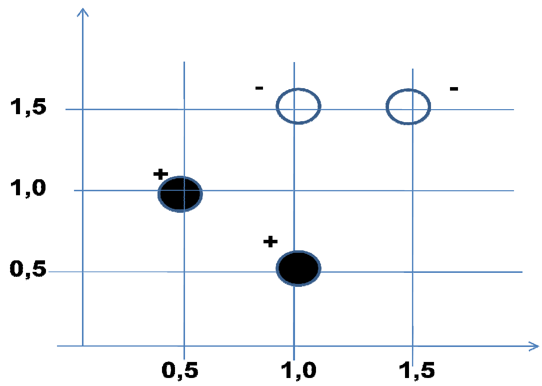
Ejercicios de clasificadores SVM propuestos en exámenes

Ejercicio 1.- Considere los tres vectores bidimensionales linealmente separables de la siguiente figura. Encuentre el SVM lineal que separa de manera óptima las clases al maximizar el margen.



Solución.-
Todos los puntos son vectores soporte el hiperplano de margen H^+ es la línea que pasa por los dos puntos positivos. El hiperplano de margen H^- es la recta que pasa por el punto negativo y es paralela a H^+. La función de decisión es la recta que está entre H^+ y H^-. Esta recta tiene por ecuación $x_2 = -x_1+2$. Habría que hacerlo por el primal y por el dual. En el siguiente ejercicio se obtiene esta recta

Ejercicio 2.- Considere los cuatro vectores bidimensionales linealmente separables de la siguiente figura. Encuentre el SVM lineal que separa de manera óptima las clases al maximizar el margen.



Solución.- Los puntos que son vectores soporte son los positivos (1, 0,5) y el (0,5, 1) y el hiperplano de margen H^+ es la línea que pasa por los dos puntos positivos. El hiperplano de margen H^- es la recta que pasa por el punto negativo (1, 1,5) y es paralela a H^+. La función de decisión es la recta que está entre H^+ y H^-. Esta recta tiene por ecuación $-x_1 + 2 = x_2$.

Tenemos la clase positiva formada por los vectores soporte (0.5 ,1) y (1, 0.5) La clase negativa está formada por el vector soporte (1, 1.5)

La ecuación del hiperplano es $\mathbf{w}^T\mathbf{x} + w_0 = 0$, o lo que es lo mismo $w_1x_1+w_2x_2+w_0=0$

1) Primal

Si sustituimos los vectores soporte en los hiperplanos positivo y negativo

H +: $\langle w \bullet x^{+} \rangle + w_0 = 1$

H ⁻: $\langle w \bullet x^{-} \rangle + w_0 = -1$

tenemos las ecuaciones para los tres vectores soporte

$0.5w_1+w_2+w_0=1$

$w_1+0.5w_2+w_0=1$

$w_1+1.5w_2+w_0=-1$

y operando adecuadamente $w_1=-2$; $w_2=-2$ y $w_0= 4$, por lo que la ecuación del hiperplano separador es $x_1+x_2-2=0$, $x_2=-x_1+2$

2) Dual

$$\max \sum_{i=1} \alpha_i - \frac{1}{2} \sum_{i,j=1} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) = \max \sum_{i=1} \alpha_i - \frac{1}{2} \boldsymbol{\alpha}^{\mathbf{T}} \mathbf{Q} \boldsymbol{\alpha}$$
$$s.a. \ \sum_{i=1} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \text{ i=1,...,n}$$

$$\hat{\mathbf{w}} = \sum_{i=1} \hat{\alpha}_i y_i \mathbf{x_i} = \sum_{i \in Sop} \hat{\alpha}_i y_i \mathbf{x_i}$$

$$\hat{w}_{\mathbf{0}} = \pm 1 - \sum_{j \in Sop} \hat{\alpha}_j y_j (\mathbf{x_j^T x_i}), \quad \text{con } \mathbf{x_i} \in \omega_1 \ y \ \ \hat{\boldsymbol{\alpha}}_{\mathbf{i}} > 0$$

vectores soporte (0.5 ,1) y (1, 0.5) clase positiva y vector soporte (1, 1.5) clase negativa

Ahora la función a maximizar es

$$\text{L}(\alpha_1, \alpha_2, \alpha_3, \lambda;\ \mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}) =$$
$$= \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(\alpha_1^2(\mathbf{x_1^T x_1}) + \alpha_1\alpha_2(\mathbf{x_1^T x_2}) - \alpha_1\alpha_3(\mathbf{x_1^T x_3}) + \alpha_2\alpha_1(\mathbf{x_2^T x_1})$$
$$+ \alpha_2^2(\mathbf{x_2^T x_2}) - \alpha_2\alpha_3(\mathbf{x_2^T x_3}) - \alpha_3\alpha_1(\mathbf{x_3^T x_1}) - \alpha_3\alpha_2(\mathbf{x_3^T x_2})$$
$$+ \alpha_3^2(\mathbf{x_3^T x_3})) - \lambda(\alpha_1 + \alpha_2 - \alpha_3)$$

donde $\mathbf{x_1}^{\mathbf{T}}=(0.5\ ,1)$ $\mathbf{x_2}^{\mathbf{T}}=(1,\ 0.5)$ y $\mathbf{x_3}^{\mathbf{T}}=(1,\ 1.5)$ sustituyendo los valores de los productos escalares tenemos

$$\text{L}(.) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(\alpha_1^2(\frac{5}{4}) + \alpha_1\alpha_2(1) - \alpha_1\alpha_3(2) + \alpha_2\alpha_1(1) + \alpha_2^2(\frac{5}{4}) - \alpha_2\alpha_3(\frac{7}{4}) +$$
$$- \alpha_3\alpha_1(2) - \alpha_3\alpha_2(\frac{7}{4}) + \alpha_3^2(\frac{13}{4})) - \lambda(\alpha_1 + \alpha_2 - \alpha_3)$$

Derivando con respecto a los α_i y λ tenemos

$$\frac{\partial L}{\partial \alpha_1} = 0, \text{ esto es, } 1 - \frac{5}{4}\alpha_1 - \alpha_2 + 2\alpha_3 - \lambda = 0$$

$$\frac{\partial L}{\partial \alpha_2} = 0, \text{ esto es, } 1 - \alpha_1 - \frac{5}{4}\alpha_2 + \frac{7}{4}\alpha_3 - \lambda = 0$$

$$\frac{\partial L}{\partial \alpha_3} = 0, \text{ esto es, } 1 + 2\alpha_1 + \frac{7}{4}\alpha_2 - \frac{13}{4}\alpha_3 + \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = -\alpha_1 - \alpha_2 + \alpha_3 = 0, \text{ luego } \alpha_3 = \alpha_1 + \alpha_2$$

Resolviendo el sistema de 4 ecuaciones con 4 incógnitas ($\alpha_1, \alpha_2, \alpha_3, \lambda$) tenemos

$\alpha_1 = 4,\ \alpha_2 = 0$ y $\alpha_3 = 4$ $\mathbf{x_1}^{\mathbf{T}}$=(0.5 ,1) $\mathbf{x_2}^{\mathbf{T}}$=(1, 0.5) y $\mathbf{x_3}^{\mathbf{T}}$=(1, 1.5)

$$\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i \mathbf{x}_i = 4 \times 1 \times \begin{pmatrix} 0.5 \\ 1 \end{pmatrix} + 0 \times 1 \times \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} + 4 \times (-1) \times \begin{pmatrix} 1 \\ 1.5 \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$$

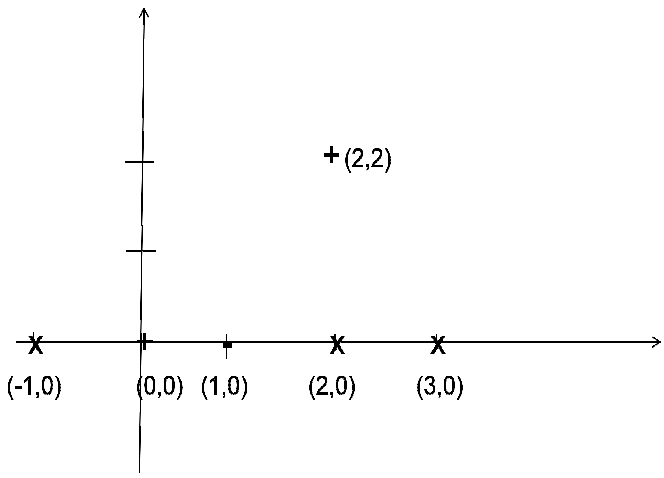
$$\hat{w}_{\mathbf{0}} = 1\text{-} \sum_{j \in Sop} \hat{\alpha}_j y_j (\mathbf{x}_{\mathbf{j}}^{\mathbf{T}} \mathbf{x}_{\mathbf{i}}), \quad \text{con } \mathbf{x}_{\mathbf{i}} \in \omega_1\ y\ \ \hat{\alpha}_i > 0$$

$$\text{o también } \hat{\mathrm{w}}_0 = -1 - ((s_1)^T \hat{\mathbf{w}}) = -1 - (1,1.5)\begin{pmatrix} -2 \\ -2 \end{pmatrix} = -1 - (-5) = 4$$

La ecuación es $-2x_1-2x_2+4=0$, esto es $-x_1-x_2+2=0$, $x_2=-x_1+2$

Ejercicio 3.- Considere los cinco vectores (0,0) y (2,2) de la clase positiva y (-1,0), (2,0) y (3,0) de la clase negativa. Encuentre el modelo SVM lineal que separa de manera óptima las clases al maximizar el margen. Clasifique al nuevo patrón (1,0) de la clase negativa. Utilice el primal y el dual

Solución.- 1)



Los vectores soporte son de la clase positiva el (2,2) y de la negativa el (2,0) y el (3,0) tanto para el primal como para el dual, de esta forma consideramos que el (0,0) está mal clasificado en el conjunto de entrenamiento. $\begin{matrix} H_0 : \mathbf{w.x} + w_0 = 1 \\ H_1 : \mathbf{w.x} + w_0 = -1 \end{matrix}$

$$\begin{matrix} x_1 & x_2 & Clase \\ 2 & 0 & -1 \\ 2 & 2 & +1 \end{matrix}$$

$$2w_1 + 0w_2 + w_0 = -1 \qquad 2w_1 + 2w_2 + w_0 = 1$$

Tenemos dos ecuaciones con tres incógnitas, así que tomamos como vector soporte negativo también el (3,0). Si ponemos las tres ecuaciones tenemos $\begin{matrix} 2w_1 + 2w_2 + w_0 = 1 \\ 2w_1 + 0w_2 + w_0 = -1 \\ 3w_1 + 0w_2 + w_0 = -1 \end{matrix}$

de donde despejando tenemos w_0= -1, w_1=0 y w_2=1, y la ecuación del hiperplano es

$H : 0\,x_1 + x_2 - 1 = 0;\ x_2 = 1$. Esta solución hace que el patrón (1, 0) de la clase negativa tenga un valor de x_2 menor que 1, luego lo clasifica bien.

2) Si consideramos solo los vectores soporte (2,0) como negativo y (2,2) como positivo, tenemos en el dual

$$\max \alpha_1 + \alpha_2 - \frac{1}{2}(\alpha_1^2 (s_1)^T.(s_1) - \alpha_1\alpha_2 (s_1)^T.(s_2) - \alpha_2\alpha_1 (s_2)^T.(s_1) + \alpha_2^2 (s_2)^T.(s_2))$$

$$s.a. \quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

donde $(s_1) = (2,2)^T$ el positivo y $(s_2) = (2,0)^T$ el negativo, por lo que la función de Lagrange es

$$\text{L}(.) = \alpha_1 + \alpha_2 - \frac{1}{2}(8\alpha_1^2 - 4\alpha_1\alpha_2 - 4\alpha_2\alpha_1 + 4\alpha_2^2) - \lambda(-\alpha_1 + \alpha_2)$$

derivando con respecto a $\alpha_1,\ \alpha_2$ y λ, tenemos

$$\begin{cases} 1 - 8\alpha_1 + 4\alpha_2 + \lambda = 0 \\ 1 + 4\alpha_1 - 4\alpha_2 - \lambda = 0 \\ \alpha_1 - \alpha_2 = 0 \end{cases}$$

de donde $\alpha_1 = \alpha_2 = 1/2$ y $\lambda = 1$

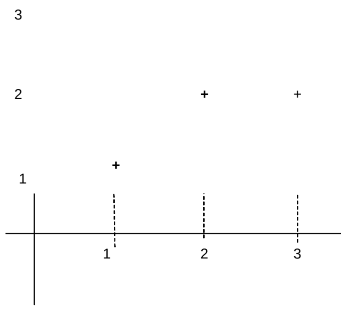
$$\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i (\mathbf{s_i}) = (1/2)(-1)\begin{pmatrix} 2 \\ 0 \end{pmatrix} + (1/2)(1)\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$\hat{\text{w}}_0 = 1 - ((s_1)^T \hat{\mathbf{w}}) = 1 - (2,2)\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 - 2 = -1,$ hemos tomado el VS positivo

y la ecuación del hiperplano separador es $0x_1 + x_2 - 1 = 0$, o lo que es igual $x_2 = 1$

c) El patrón de coordenadas $(1, 0)^T$ al ser el valor de x_2 menor de 1, esto es 0<1, se clasifica en la clase negativa luego se clasifica bien.

Ejercicio 4.- Considere los cinco vectores bidimensionales de la siguiente figura, donde las cruces son patrones de la clase positiva y los círculos patrones de la clase negativa. Encuentre el SVM lineal que separa de manera óptima las clases al maximizar el margen.



Solución.-
Los datos etiquetados como positivos definidos como vectores traspuestos son el $(1,1)^T$ y $(2,2)^T$ y los datos etiquetados como negativos son el $(1,2)^T$, $(2,3)^T$ y $(3,2)^T$.
Sin hacer ninguna trasformación, los vectores soporte pueden ser todos menos el (3,2), si consideramos que será un patrón no correctamente clasificado en entrenamiento. También podemos considerar como vectores soporte el (2,2) como positivo y el (2,3) y el (3,2) como negativos, y considerar ahora que el patrón (1,2) a priori estará mal clasificado. Con estos últimos vectores soporte tenemos.

$$H_0 : \mathbf{w.x} + w_0 = 1 \qquad H_1 : \mathbf{w.x} + w_0 = -1$$

$x1$	$x2$	*Clase*
2	2	1
2	3	−1
3	2	−1

$$w_1x_1 + w_2x_2 + w_0 = 1 \qquad w_1x_1 + w_2x_2 + w_0 = -1$$

$$1)\ 2w_1 + 2w_2 + w_0 = 1 \qquad 2)\ 2w_1 + 3w_2 + w_0 = -1$$

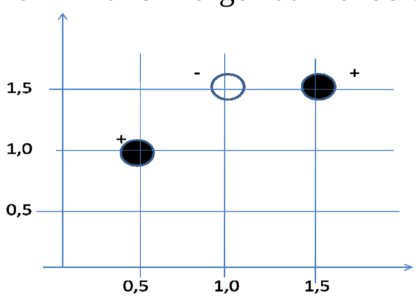
$$3)\ 3w_1 + 2w_2 + w_0 = -1$$

Resolviendo el sistema tenemos que w2=w1=-2 y w0= 9

La ecuación es

$$H: -2x_1 - 2x_2 + 9 = 0;\ x_2 = -x_1 + 4{,}5$$

Ejercicio 5.- Considere los tres vectores bidimensionales linealmente separables de la siguiente figura. Encuentre el SVM lineal que separa de manera óptima las clases al maximizar el margen utilizando el primal y el dual.



Solución.-

1) Los vectores soporte son para la clase negativa, el (1, 1,5), y para la positiva los vectores (0,5 1) y (1,5 1,5)

x_1	x_2	*Clase*
0,5	1	1
1,5	1,5	1
1	1,5	−1

$H_0 : \mathbf{w.x} + w_0 = 1$

$H_1 : \mathbf{w.x} + w_0 = -1$

$w_1x_1 + w_2x_2 + w_0 = 1$

1) $0{,}5w_1 + w_2 + w_0 = 1$

2) $1{,}5w_1 + 1{,}5w_2 + w_0 = 1$

$w_1x_1 + w_2x_2 + w_0 = -1$

3) $w_1 + 1{,}5w_2 + w_0 = -1$

Resolviendo el sistema tenemos que w_1=4, w_2=-8 y w_0= 7

La ecuación es

$$H: 4x_1 - 8x_2 + 7 = 0;\ x_2 = \frac{x_1}{2} + \frac{7}{8}$$

2) Si consideramos los tres vectores soporte $(1{,}5\ 1{,}5)^+$, $(0{,}5\ 1)^+$ $(1,\ 1{,}5)^-$ y tenemos en el dual

$$\max \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(4{,}5\alpha_1^2 + 2{,}25\alpha_1\alpha_2 - 3{,}75\alpha_1\alpha_3 + 1{,}25\alpha_2^2 + 2{,}25\alpha_2\alpha_1$$
$$-2\alpha_2\alpha_3 + 3{,}25\alpha_3^2 - 3{,}75\alpha_3\alpha_1 - 2\alpha_3\alpha_2) - \lambda(\alpha_1 + \alpha_2 - \alpha_3)$$

$$s.a. \quad -\alpha_1 - \alpha_2 + \alpha_3 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0,\ \alpha_3 \geq 0$$

derivando con respecto a α_1, α_2 , α_3 y λ, tenemos

$$\begin{cases} 1 - 4{,}5\alpha_1 - 2{,}25\alpha_2 + 3{,}75\alpha_3 - \lambda = 0 \\ 1 - 2{,}25\alpha_1 - 1{,}25\alpha_2 + 2\alpha_3 - \lambda = 0 \\ 1 - 3{,}75\alpha_1 + 2\alpha_2 - 3{,}25\alpha_3 + \lambda = 0 \\ -\alpha_1 - \alpha_2 + \alpha_3 = 0 \end{cases}$$

de donde $\alpha_1 = 24$, $\alpha_2 = 16$, $\alpha_3 = 40$ y $\lambda = 7$

$$\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i s_i = (24)(1)\begin{pmatrix} 1{,}5 \\ 1{,}5 \end{pmatrix} + (16)(1)\begin{pmatrix} 0{,}5 \\ 1 \end{pmatrix} + (40)(-1)\begin{pmatrix} 1 \\ 1{,}5 \end{pmatrix} = \begin{pmatrix} 4 \\ -8 \end{pmatrix}$$

$$\hat{w}_0 = 1 - (s_1^T \hat{\mathbf{w}}) = 1 - (1{,}5\ \ ,1{,}5)\begin{pmatrix} 4 \\ -8 \end{pmatrix} = 7$$

y la ecuación del hiperplano separador es

$4x_1$-$8x_2$+7=0 , o lo que es igual $x_2 = (1/2)x_1 + (7/8)$

Ejercicio 6.-Dados los datos etiquetados como positivos definidos como vectores traspuestos $(3,0)^T$, $(0,3)^T$, $(-3,0)^T$, $(0, -3)^T$ y los datos etiquetados como negativos $(0,2)^T$, $(1,0)^T$, $(0,-1)^T$ y $(-1,0)^T$ y la función de transformación del espacio de características de entrada.

$$\Phi(x_1,x_2)^T = \begin{cases} (4-x_2+|3x_1-x_2|,\ 4-x_1+|3x_1-x_2|)^T & \text{si } \sqrt{x_1^2+x_2^2} \geq 2 \\ (x_1,x_2)^T & \text{en otro caso} \end{cases}$$

a) Calcule los vectores soporte y la ecuación del hiperplano de separación.
b) Analizar la clasificación de los patrones con etiqueta positiva (4,4), (-4,-4) y (1,3)

1.- La transformación del vector (3,0) es el vector (13,10), la del (0,3) es el vector (4,7), la del vector (-3,0) se transforma en el vector (13,16) y el vector (0,-3) en el vector (10,7); mientras que los negativos todos quedan igual menos el (0,2) que se transforma en el (4,6) . De esta forma, los vectores soporte podrían ser el (4,6) y el (4,7). De esta manera, con estos vectores soporte tenemos.

$$H_0 : \mathbf{w.x} + w_0 = 1$$
$$H_1 : \mathbf{w.x} + w_0 = -1$$

x_1	x_2	*Clase*
4	6	-1
4	7	$+1$

$$w_1x_1 + w_2x_2 + w_0 = -1 \qquad w_1x_1 + w_2x_2 + w_0 = 1$$
$$4w_1 + 6w_2 + w_0 = -1 \qquad 4w_1 + 7w_2 + w_0 = 1$$

de esta dos ecuaciones obtenemos que $w_2=2$ y que $4w_1+w_0=-13$, pero tenemos dos ecuaciones con tres incógnitas, así que tomamos como vector soporte positivo también el (10,7). Si ponemos las tres ecuaciones tenemos

$$w_2 = 2$$
$$4w_1 + w_0 = -13$$
$$10w_1 + 7w_2 + w_0 = 1$$

de donde despejando tenemos $w_0= -13$, $w_1=0$ y $w_2=2$, y la ecuación del hiperplano es

$H : 2x_2 - 13 = 0;\ x_2 = 13/2 = 6.5$. Esta solución no es la mejor pero si aproximada

2.- Si consideramos solo los vectores soporte (4,6) como negativo y (4,7) como positivo, tenemos en el dual

$$\max \alpha_1 + \alpha_2 - \frac{1}{2}\begin{pmatrix} \alpha_1^2(\Phi(s_1)^T.\Phi(s_1)) - \alpha_1\alpha_2(\Phi(s_1)^T.\Phi(s_2)) \\ -\alpha_2\alpha_1(\Phi(s_2)^T.\Phi(s_1)) + \alpha_2^2(\Phi(s_2)^T.\Phi(s_2)) \end{pmatrix}$$

$$s.a. \quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

donde $\Phi(s_1) = (4,6)^{\mathrm{T}}$ y $\Phi(s_2) = (4,7)^{\mathrm{T}}$, por lo que

$$\mathrm{L}(\alpha_1,\alpha_2,\lambda) = \alpha_1 + \alpha_2 - \frac{1}{2}\left(52\alpha_1^2 - 58\alpha_1\alpha_2 - 58\alpha_2\alpha_1 + 65\alpha_2^2\right) - \lambda(-\alpha_1 + \alpha_2)$$

derivando con respecto a α_1, α_2 y λ, tenemos

$$\begin{cases} 1 - 52\alpha_1 + 58\alpha_2 + \lambda = 0 \\ 1 + 58\alpha_1 - 65\alpha_2 - \lambda = 0 \\ \alpha_1 - \alpha_2 = 0 \end{cases}$$

de donde $\alpha_1 = \alpha_2 = 2$ y $\lambda = -13$

$$\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i \Phi(s_i) = (2)(-1)\begin{pmatrix}4\\6\end{pmatrix} + (2)(1)\begin{pmatrix}4\\7\end{pmatrix} = \begin{pmatrix}0\\2\end{pmatrix}$$

$$\hat{\mathrm{w}}_0 = -1 - (\Phi(s_1)^T \hat{\mathbf{w}}) = -1 - (4,6)\begin{pmatrix}0\\2\end{pmatrix} = -1 - 12 = -13$$

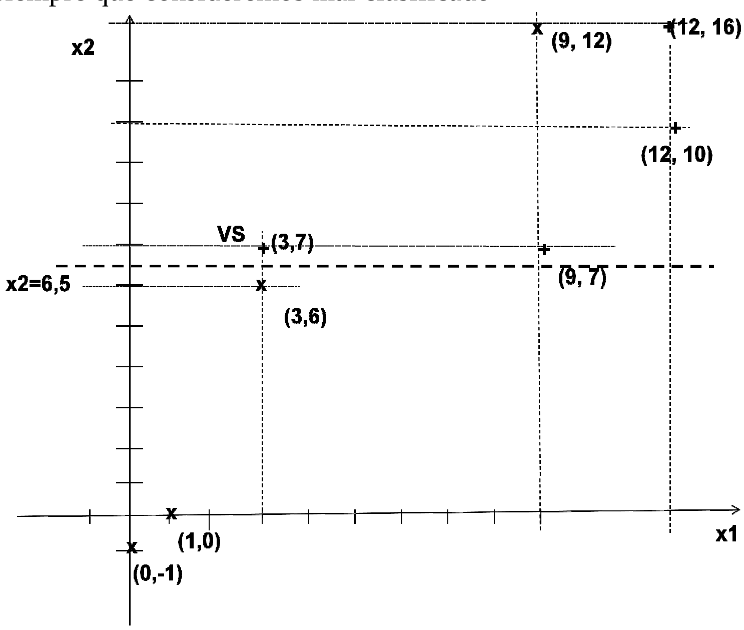
Hemos tomado como vector soporte el negativo
y la ecuación del hiperplano separador es
$0x_1+2x_2-13=0$, o lo que es igual, $x_2 = 13/2 = 6.5$

b) El patrón con etiqueta positiva (4,4) se transforma en el (8,8), luego como 8> 6,5 pertenece a la clase positiva; mientras que el patrón (-4,-4) se transforma en el (16,16) luego como 16> 6,5 pertenece también a la clase positiva. Estarán, por tanto, bien clasificados.
Por último el (1,3) se transforma en el (1,3) y como 3< 6,5 pertenece a la clase negativa, luego está mal clasificado.

Ejercicio 7.- Dados los datos etiquetados como positivos definidos como vectores traspuestos $(3,0)^T$, $(0,3)^T$, $(-3,0)^T$, $(0, -3)^T$ y los datos etiquetados como negativos $(0,2)^T$, $(1,0)^T$, $(0,-1)^T$ y $(-2,0)^T$ y la función de transformación del espacio de características de entrada.

$$\Phi(x_1, x_2)^T = \begin{cases} (3 - x_2 + |3x_1 - x_2|,\ 4 - x_1 + |3x_1 - x_2|)^T & \text{si } \sqrt{x_1^2 + x_2^2} \geq 2 \\ (x_1, x_2)^T & \text{en otro caso} \end{cases}$$

a) Calcule los vectores soporte y la ecuación del hiperplano de separación.
b) Utilice el algoritmo SVM para clasificar el patrón de coordenadas $(-1,0)^T$ ¿Qué conclusiones podemos sacar?
1.- La transformación del vector (3,0) es el vector (12,10), la del (0,3) es el vector (3,7), la del vector (-3,0) se transforma en el vector (12,16) y el vector (0,-3) en el vector (9,7); mientras que los negativos todos quedan igual menos el (0,2) que se transforma en el (3,6) y el (-2,0) que se transforma en el (9,12). De esta forma, los vectores soporte podrían ser el (3,6) y el (3,7); siempre que consideremos mal clasificado



el patrón (-2,0). De esta manera, con estos vectores soporte tenemos.

$$H_0 : \mathbf{w}.\mathbf{x} + w_0 = 1$$
$$H_1 : \mathbf{w}.\mathbf{x} + w_0 = -1$$

x_1	x_2	*Clase*
3	6	-1
3	7	$+1$

$$w_1x_1 + w_2x_2 + w_0 = -1 \qquad w_1x_1 + w_2x_2 + w_0 = 1$$
$$3w_1 + 6w_2 + w_0 = -1 \qquad 3w_1 + 7w_2 + w_0 = 1$$

de esta dos ecuaciones obtenemos que w_2=2 y que $3w_1+w_0$= -13, pero tenemos dos ecuaciones con tres incógnitas, así que tomamos como vector soporte positivo también el (9,7). Si ponemos las tres ecuaciones tenemos

$$w_2 = 2$$
$$3w_1 + w_0 = -13$$
$$9w_1 + 7w_2 + w_0 = 1$$

de donde despejando tenemos w_0= -13, w_1=0 y w_2=2, y la ecuación del hiperplano es

$H : 2x_2 - 13 = 0;\ x_2 = 13/2 = 6.5$. Esta solución hace que el patrón (-2,0) de la clase negativa no esté bien clasificado en el conjunto de entrenamiento.

2.- Si consideramos solo los vectores soporte (3,6) como negativo y (3,7) como positivo, tenemos en el dual

$$\max \alpha_1 + \alpha_2 - \frac{1}{2}\begin{pmatrix} \alpha_1^2(\Phi(s_1)^T.\Phi(s_1)) - \alpha_1\alpha_2(\Phi(s_1)^T.\Phi(s_2)) - \alpha_2\alpha_1(\Phi(s_2)^T.\Phi(s_1)) \\ +\alpha_2^2(\Phi(s_2)^T.\Phi(s_2)) \end{pmatrix}$$

$$s.a. \quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

,

donde $\Phi(s_1) = (3,6)^{\mathrm{T}}$ y $\Phi(s_2) = (3,7)^{\mathrm{T}}$, por lo que

$$L(\alpha_1, \alpha_2, \lambda) = \alpha_1 + \alpha_2 - \frac{1}{2}\left(45\alpha_1^2 - 51\alpha_1\alpha_2 - 51\alpha_2\alpha_1 + 58\alpha_2^2\right) - \lambda(-\alpha_1 + \alpha_2)$$

derivando con respecto a α_1, α_2 y λ, tenemos

$$\begin{cases} 1 - 45\alpha_1 + 51\alpha_2 + \lambda = 0 \\ 1 + 51\alpha_1 - 58\alpha_2 - \lambda = 0 \\ \qquad \alpha_1 - \alpha_2 = 0 \end{cases}$$

de donde $\alpha_1 = \alpha_2 = 2$ y $\lambda = -13$

$$\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i \Phi(s_i) = (2)(-1)\begin{pmatrix} 3 \\ 6 \end{pmatrix} + (2)(1)\begin{pmatrix} 3 \\ 7 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

$$\hat{\mathrm{w}}_0 = -1 - (\Phi(s_1)^T \hat{\mathbf{w}}) = -1 - (3,6)\begin{pmatrix} 0 \\ 2 \end{pmatrix} = -1 - 12 = -13$$

Hemos tomado como vector soporte el negativo

y la ecuación del hiperplano separador es $0x_1+2x_2-13=0$,

o lo que es igual $x_2 = 13/2 = 6,5$

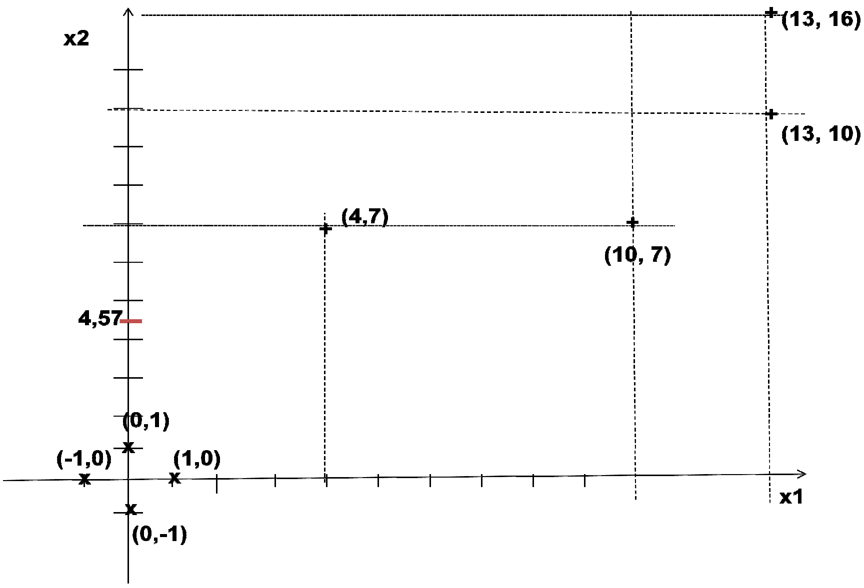
c) El patrón de coordenadas $(-1,0)^T$, se transforma en el $(-1,0)^T$ y al ser el valor de x_2 menor de 6,5 se clasifica en la clase negativa luego se clasifica bien

Ejercicio 8.- Dados los datos etiquetados como positivos definidos como vectores traspuestos $(3,0)^T$, $(0,3)^T$, $(-3,0)^T$, $(0, -3)^T$ y los datos etiquetados como negativos $(0,1)^T$, $(1,0)^T$, $(0,-1)^T$ y $(-1,0)^T$ y la función de transformación del espacio de características de entrada.

$$\Phi(x_1, x_2)^T = \begin{cases} (4 - x_2 + |3x_1 - x_2|,\ 4 - x_1 + |3x_1 - x_2|)^T & \text{si } \sqrt{x_1^2 + x_2^2} > 2 \\ (x_1, x_2)^T & \text{en otro caso} \end{cases}$$

a) Calcule los vectores soporte y la ecuación del hiperplano de separación.
b) Utilice el algoritmo SVM para clasificar el patrón de coordenadas $(1,1)^T$ ¿Qué conclusiones podemos sacar?

1.- La transformación del vector (3,0) es el vector (13,10), la del (0,3) es el vector (4,7), la del vector (-3,0) se transforma en el vector (13, 16) y el vector (0,-3) en el vector (10,7); mientras que los negativos quedan igual. De esta forma, los vectores soporte podrían ser el (1,0) y el (4,7), puesto que la distancia euclídea entre ellos es mayor 7,61, que la distancia entre el (0,1) y el (4,7) que es 7,21. De esta manera, con estos vectores soporte tenemos.



$$H_0 : \mathbf{w.x} + w_0 = 1$$
$$H_1 : \mathbf{w.x} + w_0 = -1$$

x_1	x_2	$Clase$
1	0	−1
4	7	+1

$$w_1 x_1 + w_2 x_2 + w_0 = -1$$
$$1 w_1 + 0 w_2 + w_0 = -1$$
$$w_1 = -1 - w_0$$

$$
\begin{aligned}
&w_1x_1 + w_2x_2 + w_0 = 1\\
&4(-1 - w_0) + 7w_2 + w_0 = 1\\
&-4 - 4w_0 + 7w_2 + w_0 = 1\\
&7w_2 = 3w_0 + 5
\end{aligned}
$$

pero tenemos dos ecuaciones con tres incógnitas, así que tomamos como vector soporte negativo también el (0,1) y aunque no es la mejor solución, nos sirve para tener un hiperplano aproximado

$$
\begin{aligned}
&w_1x_1 + w_2x_2 + w_0 = -1\\
&0w_1 + 1w_2 + w_0 = -1\\
&w_2 = -1 - w_0
\end{aligned}
$$

y si ponemos las tres ecuaciones

$$
\begin{aligned}
&w_1 = -1 - w_0\\
&w_2 = -1 - w_0\\
&7w_2 = 3w_0 + 5
\end{aligned}
$$

de donde despejando tenemos w_0=-1,2, w_1=0,2 y w_2=0,2, y la ecuación del hiperplano es

$H:\ 0{,}2x_1 + 0{,}2x_2 - 1{,}2 = 0;\ x_2 = -x_1 + 6$, esta solución no es la mejor pero si aproximada

El vector (1,1) se transforma en el (1,1) por lo que si le aplicamos la ecuación del hiperplano tenemos que el signo es negativo y estaría en la clase negativa, por lo que el clasificador al menos para este patrón es adecuado

2.- Si consideramos solo los vectores soporte (1,0) y (4,7) tenemos en el dual

$$
\max \alpha_1 + \alpha_2 - \frac{1}{2}\begin{pmatrix} \alpha_1^2(\Phi(s_1)^T.\Phi(s_1)) - \alpha_1\alpha_2(\Phi(s_1)^T.\Phi(s_2)) - \alpha_2\alpha_1(\Phi(s_2)^T.\Phi(s_1)) \\ +\alpha_2^2(\Phi(s_2)^T.\Phi(s_2)) \end{pmatrix}
$$

$$
s.a.\quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0
$$

donde $\Phi(s_1) = (1,0)^{\mathrm{T}}$ y $\Phi(s_2) = (4,7)^{\mathrm{T}}$, por lo que

$$
L(\alpha_1,\alpha_2,\lambda) = \alpha_1 + \alpha_2 - \frac{1}{2}\left(\alpha_1^2 - 4\alpha_1\alpha_2 - 4\alpha_2\alpha_1 + 65\alpha_2^2\right) - \lambda(-\alpha_1 + \alpha_2)
$$

derivando con respecto a α_1, α_2 y λ, tenemos

$$
\begin{cases}
1 - \alpha_1 + 4\alpha_2 + \lambda = 0\\
1 + 4\alpha_1 - 65\alpha_2 - \lambda = 0\\
\alpha_1 - \alpha_2 = 0
\end{cases}
$$

$$
\textit{de donde } \alpha_1 = \alpha_2 = 2/58,\ \text{y } \lambda = -\frac{64}{58} = -\frac{32}{29}
$$

$$
\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i \Phi(s_i) = (2/58)(-1)\begin{pmatrix}1\\0\end{pmatrix} + (2/58)(1)\begin{pmatrix}4\\7\end{pmatrix} = \begin{pmatrix}3/29\\7/29\end{pmatrix}
$$

$$\hat{w}_0 = -1-(\Phi(s_1)^T \hat{\mathbf{w}}) = -1-(1,0)\begin{pmatrix}3/29\\7/29\end{pmatrix} = -32/29$$

y la ecuación del hiperplano separador es

$3x_1+7x_2-32=0$, o lo que es igual

$x_2 = -(3/7)x_1 + (32/7)$, si $x_1=0$, $x_2=32/7=4,57$

el vector de coordenadas (1,1), se trasforma en el (1,1) y tiene valor asociado 3+7-32= -22, negativo luego pertenece a la clase negativa y por tanto el clasificador, al menos para este patrón, es adecuado

Ejercicio 9.-
Dados los vectores de la clase negativa (1,1) (1,-1) (-1,1) y (-1,-1) y los vectores de la clase positiva (3,3) (3,-3) (-3,3) y (-3,-3) encontrar el hiperplano separador que maximiza el margen (utilizar la transformación)

$$\Phi\left(x_1, x_2\right) = \left[x_1, x_2,\ ((x_1^2 + x_2^2) - 6)/4\right]$$

Solución.- Después de la transformación los vectores de la clase negativa tienen de componentes (1,1,-1) (1,-1,-1) (-1,1,-1) (-1,-1,-1) y los vectores de la clase positiva (3,3,3) (3,-3,3) (-3,3,3) y (-3,-3,3).
De esta forma todos son vectores soporte y verifican las ecuaciones, los de la clase negativa

$w_1+ w_2-w_3+w_0= -1$; $w_1- w_2-w_3+w_0 = -1$; $-w_1+ w_2-w_3+w_0= -1$; $-w_1-w_2-w_3+w_0 = -1$

de las anteriores ecuaciones tenemos que $-w_3+w_0 = -1$

Para las positivas

$3w_1+3w_2+3w_3+w_0= 1$; $3w_1-3w_2+3w_3+w_0= 1$; $-3w_1+3w_2+3w_3+w_0= 1$;
$-3w_1-3w_2+3w_3+w_0 = 1$

de las anteriores ecuaciones tenemos que $3w_3+w0 = 1$. Sustituyendo y despejando tenemos que

$w_1=0$; $w_2=0$; $w_3=1/2$ y $w_0=-1/2$

La función discriminante es

$$(0,0,1/2)\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} - 1/2 = 0; \quad 1/2x_3 = 1/2,\ x_3=1$$

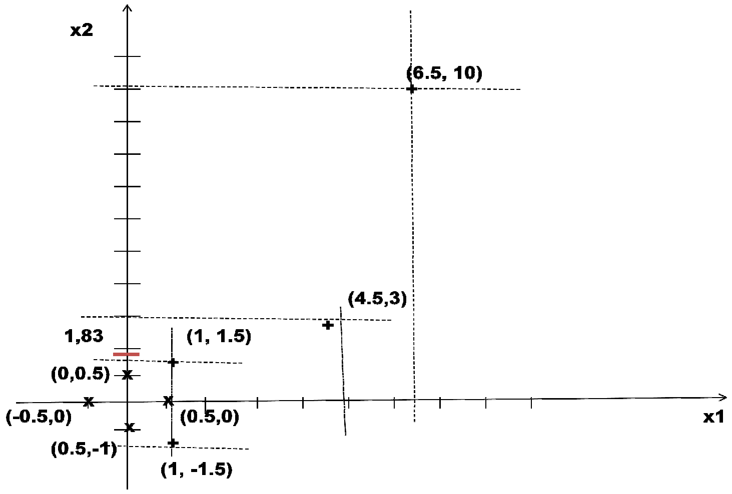
y los hiperplanos separadores $x_3=-1$ y $x_3=3$

Ejercicio 10.- Dados los datos etiquetados como positivos definidos como vectores traspuestos $(2.5, 1)^T$, $(1, 1.5)^T$, $(-2.5, 1)^T$, $(1, -1.5)^T$ y los datos etiquetados como negativos $(0, 0.5)^T$, $(0.5, 0)^T$, $(0, -0.5)^T$ y $(-0.5, 0)^T$ y la función de transformación del espacio de características de entrada.

$$\Phi(x_1, x_2)^T = \begin{cases} (4 - x_2 + |x_1 - x_2|,\ 4 - x_1 + |x_1 - x_2|)^T & \text{si } \sqrt{x_1^2 + x_2^2} > 2 \\ (x_1, x_2)^T & \text{en otro caso} \end{cases}$$

a) Calcule los vectores soporte y la ecuación del hiperplano de separación utilizando el método dual.
b) Utilice el algoritmo SVM para clasificar el patrón de la clase positiva, de coordenadas $(-1,-1)^T$ ¿Qué conclusiones podemos sacar?

Solución
1) Los vectores de la clase positiva se transforman sólo dos en la forma:
El $(2.5, 1)^T$ pasa al $(4{,}5 , 3)^T$; el $(-2.5, 1)^T$ pasa a $(6{,}5 , 10)^T$, mientras que los de la clase negativa no cambian ninguno.
Los vectores soporte son el $(0{,}5, 0)^T$ y el $(0, 0{,}5)^T$ como negativos y $(1, 1{,}5)^T$ como positivo, siempre que admitamos que el patrón (1,-1.5) estaría mal clasificado



a) Si lo resolvemos por el primal tenemos

$$H_0 : \mathbf{w.x} + w_0 = 1 \qquad H_1 : \mathbf{w.x} + w_0 = -1$$

x_1	x_2	*Clase*
0,5	0	−1
0	0,5	-1
1	1,5	+1

$$\frac{1}{2}w_1 + 0w_2 + w_0 = -1$$

$$0w_1 + \frac{1}{2}w_2 + w_0 = -1$$

de la ecuaciones anteriores $w_1 = w_2$

$$w_1 + \frac{3}{2}w_2 + w_0 = 1$$

$$2w_1 + 3w_2 + 2w_0 = 2$$

$$w_0 = 1 - \frac{5}{2}w_1$$

Teniendo en cuenta estos resultados
$4w_1 = 3$, *luego* $w_1 = w_2 = 3/4$ y sustituyendo $w_0 = -11/8$, luego la ecuación del hiperplano es
$\frac{3}{4}x_1 + \frac{3}{4}x_2 - \frac{11}{8} = 0$, o lo que es igual $6x_1 + 6x_2 - 11 = 0$
de donde $x_2 = -x_1 + \frac{11}{6}$, si $x_1 = 0$, $x_2 = 1{,}83$

b) El patrón de coordenadas (-1,-1) se transforma en el mismo por lo que si sustituimos sus coordenadas en la ecuación de la recta tenemos que -6-6-11=-23 por lo que se le asigna la clase negativa y estaría mal clasificado.

Dual

Si lo hacemos por el dual y consideramos los vectores (0,0.5) y (1,1.5) como vectores soporte por ser los más cercanos, tenemos

$$\max \alpha_1+\alpha_2-\ \frac{1}{2}\begin{pmatrix}\alpha_1^2(\Phi(s_1)^T.\Phi(s_1))-\alpha_1\alpha_2(\Phi(s_1)^T.\Phi(s_2))-\alpha_2\alpha_1(\Phi(s_2)^T.\Phi(s_1))\\ +\alpha_2^2(\Phi(s_2)^T.\Phi(s_2))\end{pmatrix}$$

$$s.a.\quad -\alpha_1+\alpha_2=0,\ \alpha_1\geq 0,\ \alpha_2\geq 0$$

donde $\Phi(s_1)=(0,0.5)^{\mathrm{T}}$ y $\Phi(s_2)=(1,\ 1.5)^{\mathrm{T}}$, por lo que

$$\max\ \alpha_1+\alpha_2-\ \frac{1}{2}\left(0,25\alpha_1^2-0,75\alpha_1\alpha_2-0,75\alpha_2\alpha_1+3,25\alpha_2^2\right)$$

$$s.a.\quad -\alpha_1+\alpha_2=0,\ \alpha_1\geq 0,\ \alpha_2\geq 0$$

derivando $L=\alpha_1+\alpha_2-\ \frac{1}{2}\left(0,25\alpha_1^2-0,75\alpha_1\alpha_2-0,75\alpha_2\alpha_1+3,25\alpha_2^2\right)-\lambda(-\alpha_1+\alpha_2)$

con respecto a α_1, α_2 y λ, tenemos

$$\begin{cases}1-0,25\alpha_1+0,75\alpha_2+\lambda=0\\ 1+0,75\alpha_1-3,25\alpha_2-\lambda=0\\ \qquad\alpha_1-\alpha_2=0\end{cases}$$

de donde $\alpha_1=\alpha_2=1$, y $\lambda=-\frac{3}{2}$

$$\hat{\mathbf{w}}=\sum_{i\in Sop}\hat{\alpha}_i y_i\Phi(s_i)=(1)(-1)\begin{pmatrix}0\\0,5\end{pmatrix}+(1)(1)\begin{pmatrix}1\\1,5\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}$$

$$\hat{\mathrm{w}}_0=1-(\Phi(s_1)^T\hat{\mathbf{w}})=1-(1,1.5)\begin{pmatrix}1\\1\end{pmatrix}=-1.5$$

y la ecuación del hiperplano separador es $x_1+x_2-1.5=0$,

o lo que es igual $x_2=-x_1+1,5$, si $x_1=0$, $x_2=1,5$

b) El patrón de coordenadas (-1,-1) se transforma en el mismo por lo que si sustituimos sus coordenadas en la ecuación de la recta tenemos que -1-1-1.5=-3.5 por lo que se le asigna la clase negativa y estaría mal clasificado.

Ejercicio 11.- Dados los datos etiquetados como positivos definidos como vectores traspuestos $(2,5,\ 0)^T$, $(0,\ 2,5)^T$, $(-2,5,0)^T$, $(0,\ -2,5)^T$ y los datos etiquetados como negativos $(0,1)^T$, $(1,0)^T$, $(0,-1)^T$ y $(-1,0)^T$ y la función de transformación del espacio de características de entrada.

$$\Phi(x_1,x_2)^T=\begin{cases}(4-x_2+|3x_1-x_2|,\ 4-x_1+|3x_1-x_2|)^T & \text{si } \sqrt{x_1^2+x_2^2}>2\\ (x_1,x_2)^T & \text{en otro caso}\end{cases}$$

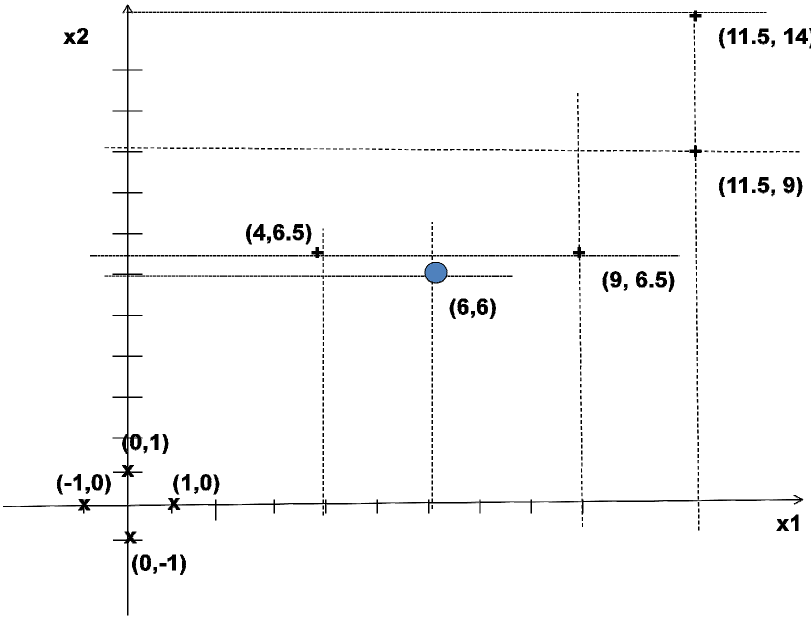
a) Calcule los vectores soporte y la ecuación del hiperplano de separación.

b) Utilice el algoritmo SVM para clasificar el patrón de coordenadas $(2,2)^T$ ¿Qué conclusiones podemos sacar?

Solución.- La transformación del vector (2,5, 0) es el vector (11,5, 9), la del (0, 2,5) es el vector (4, 6,5), la del vector (-2,5, 0) se transforma en el vector (11,5, 14) y el vector (0, -2,5) en el vector (9, 6,5); mientras que los negativos quedan igual. De esta forma, los vectores soporte podrían ser el (0,1) y el (4, 6,5), puesto que la distancia euclídea entre ellos **es menor** que la distancia entre (1,0) y (4, 6,5) dado que

$d((1,0),(4,\ 6,5)) = \sqrt{51,25};$ mientras que d((0,1),(4, 6,5))=$\sqrt{46,25}$

Esta elección no maximiza el margen, pero si cogemos como vectores soporte el (1,0) y el (4,6.5), no tenemos la seguridad de que el punto (0,1) no esté en el margen.



De esta manera, con dos vectores soporte tenemos que trabajar en el espacio dual, dado que en el primal tendremos dos ecuaciones con tres incógnitas, a no ser que elijamos también como vector soporte el (1,0).

$$H_0 : \mathbf{w.x} + w_0 = 1$$
$$H_1 : \mathbf{w.x} + w_0 = -1$$

los vectores soporte son

x_1	x_2	$Clase$
0	1	-1
4	6,5	$+1$

$$w_1x_1 + w_2x_2 + w_0 = -1$$
$$0w_1 + 1w_2 + w_0 = -1$$
$$w_2 = -1 - w_0$$

para el primer vector soporte y para el segundo

$$w_1x_1 + w_2x_2 + w_0 = 1$$
$$4w_1 + 6,5(-1 - w_0) + w_0 = 1$$
$$4w_1 = 5,5w_0 + 7,5$$

pero tenemos dos ecuaciones con tres incógnitas, así que tomamos como vector soporte negativo también el (1,0) y aunque no es la mejor solución, nos sirve para tener un hiperplano aproximado

$w_1x_1 + w_2x_2 + w_0 = -1$ $\quad$ $w_1 = -1 - w_0$

$1w_1 + 0w_2 + w_0 = -1$ $\quad$ $w_2 = -1 - w_0$

$w_1 = -1 - w_0$ y si ponemos las tres ecuaciones $\quad$ $4w_1 = 5,5w_0 + 7,5$

de donde despejando tenemos w_0= -1,21, w_1= 0,21 y w_2= 0,21, y la ecuación del hiperplano es $H:\ (0,21,\ 0,21)\cdot \mathbf{x}^T -1,21 = 0,21x_1 + 0,21x_2 - 1,21 = 0;\ x_2 = -x_1 + 5,76$, esta solución no es la mejor pero si aproximada.

El vector (2,2) se transforma en el (6,6) por lo que si le aplicamos la ecuación del hiperplano tenemos que el signo es positivo, 1,31, y estaría en la clase positiva, por lo que el clasificador al menos para este patrón es adecuado; aunque esta en el margen.

Sol 2.- Si consideramos solo los vectores soporte (0,1) y (4, 6,5) en el espacio de características tenemos en el dual

$$\max \alpha_1 + \alpha_2 - \frac{1}{2}\begin{pmatrix} \alpha_1^2(\Phi(s_1)^T.\Phi(s_1)) - \alpha_1\alpha_2(\Phi(s_1)^T.\Phi(s_2)) - \alpha_2\alpha_1(\Phi(s_2)^T.\Phi(s_1)) \\ +\alpha_2^2(\Phi(s_2)^T.\Phi(s_2)) \end{pmatrix}$$

$$s.a.\quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

,

donde $\Phi(s_1) = (0,1)^T$ y $\Phi(s_2) = (4,\ 6,5)^T$, por lo que

$$\max\ \alpha_1 + \alpha_2 - \frac{1}{2}\left(\alpha_1^2 - 6,5\alpha_1\alpha_2 - 6,5\alpha_2\alpha_1 + 58,25\alpha_2^2\right)$$

$$s.a.\quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

derivando $L = \alpha_1 + \alpha_2 - \frac{1}{2}\left(\alpha_1^2 - 6,5\alpha_1\alpha_2 - 6,5\alpha_2\alpha_1 + 58,25\alpha_2^2\right) - \lambda(-\alpha_1 + \alpha_2)$

con respecto a α_1, α_2 y λ, tenemos

$$\begin{cases} 1 - \alpha_1 + 6,5\alpha_2 + \lambda = 0 \\ 1 + 6,5\alpha_1 - 58,25\alpha_2 - \lambda = 0 \\ \alpha_1 - \alpha_2 = 0 \end{cases}$$

$$de\ donde\ \alpha_1 = \frac{2}{46,25},\ \alpha_2 = \frac{2}{46,25},\ \text{y}\ \lambda = -\frac{103,5 - 46,25}{46,25}$$

$$\hat{\mathbf{w}} = \sum_{i\in Sop} \hat{\alpha}_i y_i \Phi(s_i) = (2/46,25)(-1)\begin{pmatrix}0\\1\end{pmatrix} + (2/46,25)(1)\begin{pmatrix}4\\6,5\end{pmatrix} = \begin{pmatrix}8/46,25\\11/46,25\end{pmatrix}$$

$$\hat{\mathrm{w}}_0 = -1 - (\Phi(s_1)^T\hat{\mathbf{w}}) = -1 - (0,1)\begin{pmatrix}8/46,25\\11/46,25\end{pmatrix} = -57,25/46,25$$

y la ecuación del hiperplano separador es

$(8/46,25)x_1 + (11/46,25)x_2 - 57,25/46,25 = 0$, o lo que es igual

$8x_1 + 11x_2 - 57,25 = 0$, de donde x_2= -0,73 x_1+5,2, si x_1=0, x_2=5,2

El vector (2,2) se transforma en el (6,6) por lo que si le aplicamos la ecuación del hiperplano tenemos que el signo es positivo, 56,75, y estaría en la clase positiva, por lo que el clasificador al menos para este patrón es adecuado.

Ejercicio 12.- Dados los datos etiquetados como positivos definidos como vectores traspuestos $(2, 0)^T$, $(0, 2)^T$, $(-2, 0)^T$, $(0, -2)^T$ y los datos etiquetados como negativos $(0, 0.5)^T$, $(0.5, 0)^T$, $(0, -0.5)^T$ y $(-0.5, 0)^T$ y la función de transformación del espacio de características de entrada.

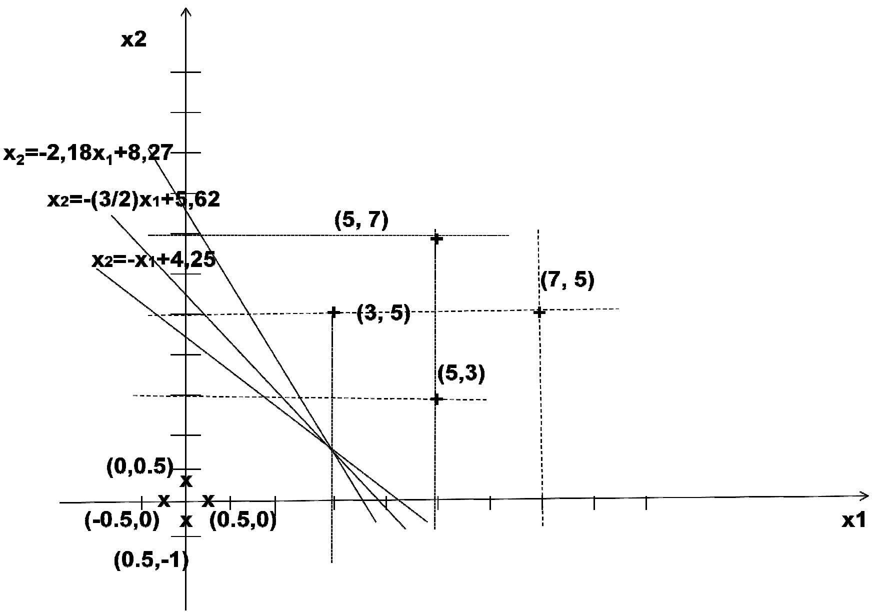
$$\Phi(x_1, x_2)^T = \begin{cases} (3 - x_2 + |x_1 - x_2|,\ 3 - x_1 + |x_1 - x_2|)^T & \text{si } \sqrt{x_1^2 + x_2^2} > 1 \\ (x_1, x_2)^T & \text{en otro caso} \end{cases}$$

a) Calcule los vectores soporte y la ecuación del hiperplano de separación utilizando el método dual.

b) Utilice el algoritmo SVM para clasificar el patrón de la clase positiva de coordenadas $(-1,-1)^T$ ¿Qué conclusiones podemos sacar?

Solución.-

La transformación del vector (2, 0) es el vector (5, 3), la del (0, 2) es el vector (3, 5), la del vector (-2, 0) se transforma en el vector (5, 7) y el vector (0, -2) en el vector (7, 5); mientras que los negativos quedan igual. De esta forma, los vectores soporte son los dos vectores $(0, 0.5)^T$, $(0.5, 0)^T$ de la clase negativa y los vectores transformado de la positiva $(3,5)^T$ y $(5,3)^T$, puesto que las recta que pasa por los dos primeros es la $x_2=-x_1+0,5$ mientras que la que pasa por los dos últimos es la $x_2=-x_1+8$



los vectores soporte son

x_1	x_2	*Clase*
0	0,5	−1
0,5	0	-1
3	5	1
5	3	1

La ecuaciones para los negativos son

$$0w_1 + 0,5w_2 + w_0 = -1$$

$$0,5w_1 + 0w_2 + w_0 = -1$$

de donde 1) $w_1 = w_2$ y 2) $0,5w_1 + w_0 = -1$

para los dos vectores soporte negativos y para los positivos tenemos

$3w_1 + 5w_2 + w_0 = 1$

$5w_1 + 3w_2 + w_0 = 1$

de donde sustituyendo 1) tenemos

3) $8w_1 + w_0 = 1$

Resolviendo el sistema de ecuaciones 2) y 3) tenemos $w_1=2/(7,5)$, $w_2=2/(7,5)$ y w0=-8,5/(7,5)

Luego la ecuación de la recta es $2x_1+2x_2-8,5=0$, lo que es igual $\mathbf{x_2=-x_1+(8,5/2)}$

b) El vector (-1,-1) se transforma en el (4,4) por lo que si le aplicamos la ecuación del hiperplano tenemos que el signo es positivo, 8+8-8,5=17,5, y estaría en la clase positiva, por lo que el clasificador al menos para este patrón es adecuado.

Solución 2.- Si consideramos solo los vectores soporte (0.5,0) y (5, 3) en el espacio de características tenemos en el dual

$$\max \alpha_1 + \alpha_2 - \frac{1}{2}\begin{pmatrix} \alpha_1^2(\Phi(s_1)^T.\Phi(s_1)) - \alpha_1\alpha_2(\Phi(s_1)^T.\Phi(s_2)) - \alpha_2\alpha_1(\Phi(s_2)^T.\Phi(s_1)) \\ +\alpha_2^2(\Phi(s_2)^T.\Phi(s_2)) \end{pmatrix}$$

$$s.a. \quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

,

donde $\Phi(s_1) = (0.5, 0)^{\mathrm{T}}$ y $\Phi(s_2) = (5,\ 3)^{\mathrm{T}}$, por lo que

$$\max\ \alpha_1 + \alpha_2 - \frac{1}{2}\left(0,25\alpha_1^2 - 2,5\alpha_1\alpha_2 - 2,5\alpha_2\alpha_1 + 34\alpha_2^2\right)$$

$$s.a. \quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

derivando $L=\alpha_1 + \alpha_2 - \frac{1}{2}\left(0,25\alpha_1^2 - 2,5\alpha_1\alpha_2 - 2,5\alpha_2\alpha_1 + 34\alpha_2^2\right) - \lambda(-\alpha_1 + \alpha_2)$

con respecto a α_1, α_2 y λ, tenemos

$$\begin{cases} 1 - 0,25\alpha_1 + 2,5\alpha_2 + \lambda = 0 \\ 1 + 2,5\alpha_1 - 34\alpha_2 - \lambda = 0 \\ \alpha_1 - \alpha_2 = 0 \end{cases}$$

de donde $\alpha_1 = \dfrac{2}{29,25}$, $\alpha_2 = \dfrac{2}{29,25}$, y $\lambda = -\dfrac{33,75}{29,25}$

$$\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i \Phi(s_i) = (2/29,25)(-1)\begin{pmatrix} 0,5 \\ 0 \end{pmatrix} + (2/29,25)(1)\begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 9/29,25 \\ 6/29,25 \end{pmatrix}$$

$$\hat{\mathrm{w}}_0 = 1 - (\Phi(s_1)^T \hat{\mathbf{w}}) = 1 - (5,3)\begin{pmatrix} 9/29,25 \\ 6/29,25 \end{pmatrix} = -\frac{33,75}{29,25}$$

y la ecuación del hiperplano separador es

$(9/29,25)x_1+(6/29,25)x_2-33,75/29,25=0$, o lo que es igual

$9x_1 + 6x_2 - 33,75 = 0$, de donde $x_2 = -3/2x_1+33,75/6$, si $x_1=0$, $x_2=33,75/6=5,62$

b) El vector (-1,-1) se transforma en el (4,4) por lo que si le aplicamos la ecuación del hiperplano tenemos que el signo es positivo, 36+24-33,75=26,25, y estaría en la clase positiva, por lo que el clasificador al menos para este patrón es adecuado.

Solución 3.-

Si consideramos además de los vectores soporte (0.5,0) de la clase negativa y el (5, 3) en la positiva, el (3,5) también de la positiva tenemos en el dual

$$\max \alpha_1+\alpha_2+\alpha_3-\frac{1}{2}(\alpha_1^2(\Phi(s_1)^T.\Phi(s_1))-\alpha_1\alpha_2(\Phi(s_1)^T.\Phi(s_2))-\alpha_1\alpha_3(\Phi(s_1)^T.\Phi(s_3))$$
$$-\alpha_2\alpha_1(\Phi(s_2)^T.\Phi(s_1))+\alpha_2^2(\Phi(s_2)^T.\Phi(s_2))+\alpha_2\alpha_3(\Phi(s_2)^T.\Phi(s_3))-\alpha_3\alpha_1(\Phi(s_3)^T.\Phi(s_1))$$
$$+\alpha_3\alpha_2(\Phi(s_3)^T.\Phi(s_2))+\alpha_3^2(\Phi(s_3)^T.\Phi(s_3)))$$
$$s.a.\quad -\alpha_1+\alpha_2+\alpha_3=0,\ \alpha_1\geq 0,\ \alpha_2\geq 0,\ \alpha_3\geq 0$$

donde $\Phi(s_1)=(0.5,0)^{\mathrm{T}}$, $\Phi(s_2)=(5,\ 3)^{\mathrm{T}}$ y $,\Phi(s_3)=(3,\ 5)^{\mathrm{T}}$ por lo que

$$\max\ \alpha_1+\alpha_2+\alpha_3-\frac{1}{2}(0,25\alpha_1^2-2,5\alpha_1\alpha_2-1,5\alpha_1\alpha_3-2,5\alpha_2\alpha_1+34\alpha_2^2+30\alpha_2\alpha_3-1,5\alpha_3\alpha_1$$
$$+30\alpha_3\alpha_2+34\alpha_3^2)$$
$$s.a.\quad -\alpha_1+\alpha_2+\alpha_3=0,\ \alpha_1\geq 0,\ \alpha_2\geq 0,\ \alpha_3\geq 0$$

derivando $L=\alpha_1+\alpha_2+\alpha_3-\frac{1}{2}(0,25\alpha_1^2-2,5\alpha_1\alpha_2-1,5\alpha_1\alpha_3-2,5\alpha_2\alpha_1+34\alpha_2^2+30\alpha_2\alpha_3-1,5\alpha_3\alpha_1$

$$+30\alpha_3\alpha_2+34\alpha_3^2)-\lambda(-\alpha_1+\alpha_2+\alpha_3)$$

con respecto a α_1, α_2, α_3 y λ, tenemos

$$\begin{cases} 1-0,25\alpha_1+2,5\alpha_2+1,5\alpha_3+\lambda=0 \\ 1+2,5\alpha_1-34\alpha_2-30\alpha_3-\lambda=0 \\ 1+1,5\alpha_1-30\alpha_2-34\alpha_3-\lambda=0 \\ \alpha_1=\alpha_2+\alpha_3 \end{cases}$$

de donde $\alpha_1=0,0666,\ \alpha_2=0,0774,$ y $\alpha_3=-0,0108$

$$\hat{\mathbf{w}}=\sum_{i\in Sop}\hat{\alpha}_i y_i\Phi(s_i)=(0,0666)(-1)\begin{pmatrix}0,5\\0\end{pmatrix}+(0,0774)(1)\begin{pmatrix}5\\3\end{pmatrix}+(-0,0108)(1)\begin{pmatrix}3\\5\end{pmatrix}=\begin{pmatrix}0,3879\\0,1782\end{pmatrix}$$

$$\hat{\mathrm{w}}_0=1-(\Phi(s_1)^T\hat{\mathbf{w}})=1-(5,3)\begin{pmatrix}0,3879\\0,1782\end{pmatrix}=-1,4741$$

y la ecuación del hiperplano separador es

$0,3879x_1+0,1782x_2-1,4741=0$, o lo que es igual

$x_2=-2,18x_1+8,27$, si $x_1=0$, $x_2=8,27$

Ejercicio 13.-
Dados los datos etiquetados como positivos definidos como vectores traspuestos $(1.5, 0)^T$, $(0, 1.5)^T$, $(-1.5,0)^T$, $(0, -1.5)^T$, los datos etiquetados como negativos $(0.5,0)^T$, $(0,0.5)^T$, $(-0.5,0)^T$ y $(0,-0.5)^T$ y la función de transformación del espacio de características de entrada.

$$\Phi(x_1,x_2)^T = \begin{cases} (2-x_2+|x_1-2x_2|,\ 2-x_1+|x_1-3x_2|)^T, & \text{si } \sqrt{x_1^2+x_2^2}<1, \\ (x_1,x_2)^T, & \text{en otro caso.} \end{cases}$$

(a) Calcule los vectores soporte y la ecuación del hiperplano de separación utilizando la formulación primal y dual del problema.

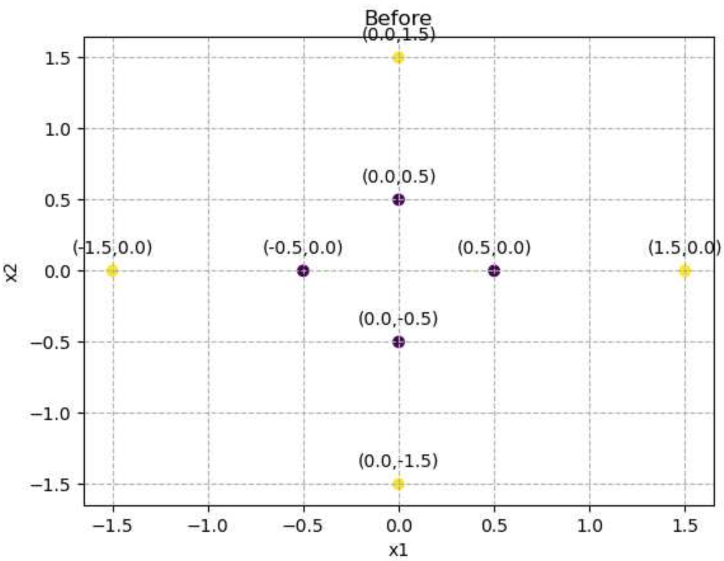
(b) Utilice el algoritmo SVM para clasificar el patrón de coordenadas $(0,0)^T$ de la clase negativa ¿Qué conclusiones podemos sacar?

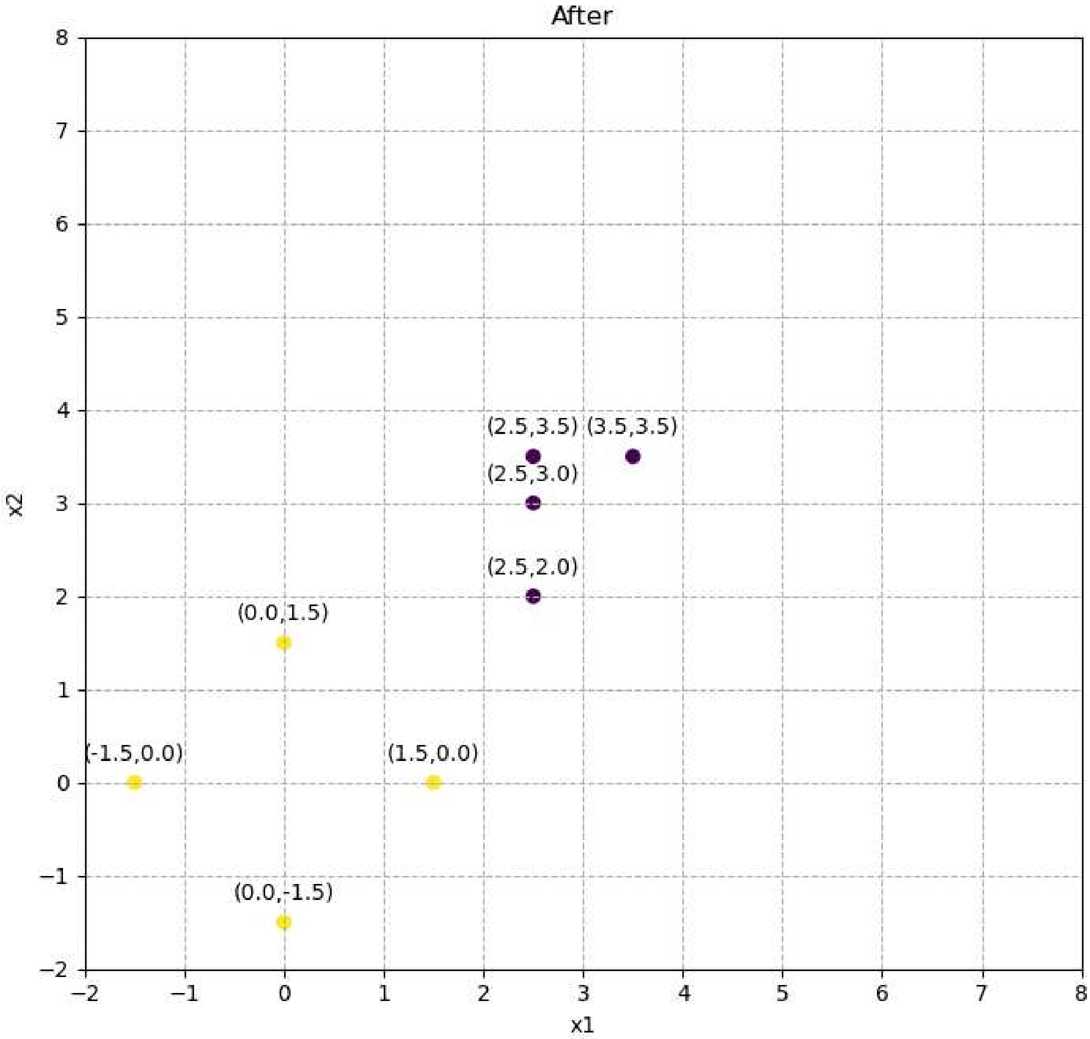
Solución

a) En primer lugar, transformamos todos los puntos de acuerdo a la transformación:

$\mathbf{x}$	$\Phi(\mathbf{x})$
$(1.5,0)$	$(1.5,0)$
$(0,1.5)$	$(0,1.5)$
$(-1.5,0)$	$(-1.5,0)$
$(0,-1.5)$	$(0,-1.5)$
$(0,0.5)$	$(2.5,2)$
$(0.5,0)$	$(2.5,3.5)$
$(0,-0.5)$	$(2.5,3)$
$(-0.5,0)$	$(3.5,3.5)$

Los puntos antes y después de la transformación son:





Queda claro que los vectores soporte serán (2.5,2), (0,1.5) y (1.5,0).

Primal)

Definimos el problema en el primal:

$$\min \ m = \frac{\langle \mathbf{w}, \mathbf{w} \rangle}{2}$$

$$\text{s.t. } z_i \left(\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b \right) \geq 1, \ i \in \{1, 2, \ldots, n\}$$

Consideramos las ecuaciones de los hiperplanos en el margen para los vectores soporte:

$$w_1 0 + w_2 1.5 + b = 1$$
$$w_1 1.5 + w_2 0 + b = 1$$
$$w_1 2.5 + w_2 2 + b = -1$$

Lo que nos lleva a:

$w_1 = -2/3$
$w_2 = -2/3$
$b = 2$

Así, el hiperplano queda definido como:

$$-\frac{2}{3}x_1^* - \frac{2}{3}x_2^* + 2 = 0; \quad \frac{2}{3}x_2^* = -\frac{2}{3}x_1^* + 2; \quad x_2^* = -x_1^* + 3$$

donde x_i^* es la i-ésima coordenada del vector transformado.

Dual)

$x_1 = (0,1.5); z_1 = 1$
$x_2 = (1.5,0); z_1 = 1$
$x_3 = (2.5,2); z_1 = -1$

$$\max \lambda_1 + \lambda_2 + \lambda_3 - \frac{1}{2}\begin{pmatrix} 2.25\lambda_1^2 + 2.25\lambda_2^2 + 10.25\lambda_3^2 + 2\cdot 0\lambda_1\lambda_2 - 2\cdot 3\lambda_1\lambda_3 \\ -2\cdot 3.75\lambda_2\lambda_3 \end{pmatrix} - \alpha(\ \lambda_1 + \lambda_2 - \lambda_3)$$

s.t. $\lambda_i \geq 0, \alpha \geq 0$

Ahora obtenemos las derivadas:

$$\begin{cases} 1 - 2.25\lambda_1 + 3\lambda_3 - \alpha = 0 \\ 1 - 2.25\lambda_2 + 3.75\lambda_3 - \alpha = 0 \\ 1 - 10.25\lambda_3 + 3\lambda_1 + 3.75\lambda_2 + \alpha = 0 \\ -\lambda_1 - \lambda_2 + \lambda_3 = 0 \end{cases} \begin{cases} -2.25\lambda_1 + 0\lambda_2 + 3\lambda_3 - \alpha = -1 \\ 0\lambda_1 - 2.25\lambda_2 + 3.75\lambda_3 - \alpha = -1 \\ 3\lambda_1 + 3.75\lambda_2 - 10.25\lambda_3 + \alpha = -1 \\ -\lambda_1 - \lambda_2 + \lambda_3 + 0\alpha = 0 \end{cases}$$

de lo que se puede obtener $\lambda_1 = 0.15$, $\lambda_2 = 0.3$, $\lambda_3 = 0.44$ y $\alpha = 2$

Obtenemos ahora el vector de proyección:

$$\mathbf{w} = \sum_{i\in S}\lambda_i z_i \mathbf{x}_i = (0.15)(1)\begin{pmatrix} 0 \\ 1.5 \end{pmatrix} + (0.3)(1)\begin{pmatrix} 1.5 \\ 0 \end{pmatrix} + (0.44)(-1)\begin{pmatrix} 2.5 \\ 2 \end{pmatrix} = \begin{pmatrix} -0.6666 \\ -0.6666 \end{pmatrix} = \begin{pmatrix} -2/3 \\ -2/3 \end{pmatrix}$$

Y el sesgo:

$$b = 1 - \left\langle \mathbf{w}, \mathbf{x}_{i|z_i=1,\lambda_i\neq 0} \right\rangle = 1 - \left\langle (-2/3,-2/3),(0,1.5) \right\rangle = 1 - (0-1) = 2$$

Así, el hiperplano queda definido como:

$$-\frac{2}{3}x_1^* - \frac{2}{3}x_2^* + 2 = 0; \quad \frac{2}{3}x_2^* = -\frac{2}{3}x_1^* + 2; \quad x_2^* = -x_1^* + 3$$

donde x_i^* es la i-ésima coordenada del vector transformado.

b) El patrón (0,0) se transforma como (2,2).

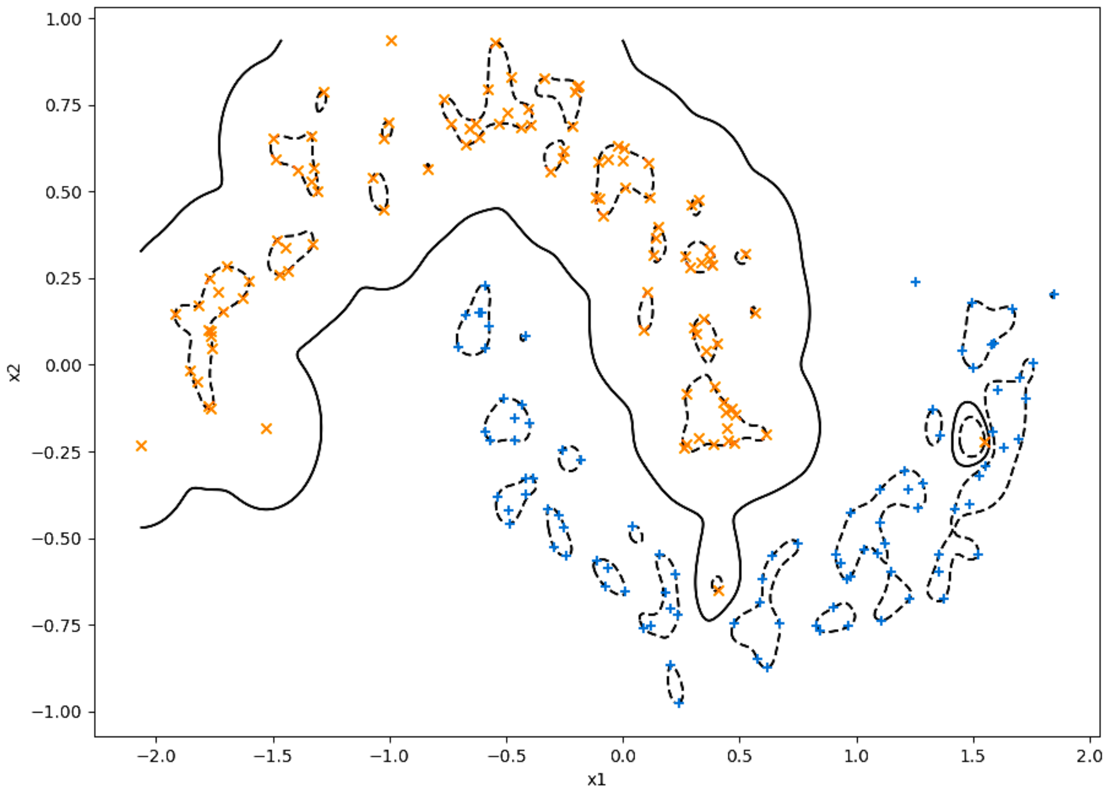
Ahora obtenemos el valor de la función de puntuación de la SVM:

$f(\mathbf{x}) = (-2/3)2 + (-2/3)2 + 2 = -2/3$ → Por lo que el patrón es de la clase negativa, lo cual tiene sentido de acuerdo a la disposición geométrica del mismo.

Cuestión.- La siguiente figura representa el resultado de entrenar una SVM no lineal en un conjunto de datos con dos variables de entrada (x1 y x2), siendo los puntos marcados con "+" de la clase positiva y los puntos marcados con "x" de la clase negativa.

(a) Razone si la SVM ha obtenido un resultado correcto o si, por el contrario, se ha producido sobre-entrenamiento o infra-entrenamiento. Justifique su respuesta.

(b) Indique si habría que aumentar, disminuir o dejar igual los valores de los parámetros *C* y *gamma* de la SVM. Justifique su respuesta.



Solución

(a) El resultado es incorrecto ya que se está produciendo un sobre-entrenamiento. El hiperplano separador es muy rugoso y llega a clasificar en la clase negativa un punto que claramente es un *outlier*, incluido dentro de la clase positiva (abajo a la derecha). Además se observa como los hiperplanos margen son demasiado específicos, formando "islas" muy locales, lo que demuestra una falta de capacidad de generalización y un sobre-aprendizaje de los patrones de entrenamiento.

(b) Como el objetivo es

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w} + C\sum_{i=1}^{n}\xi_i$$

$$s.a. \;\; y_i(\mathbf{w}^{\mathbf{T}}\mathbf{x}_i + \mathrm{w}_0) \geq 1\text{-}\xi_i, \quad i = 1...n, \text{ con } \xi_i \geq 0$$

Podríamos evitar esta situación bajando el valor del parámetro *C*, ya que esto forzaría a aumentar el margen y penalizaría menos los errores de entrenamiento. También podríamos disminuir el parámetro *gamma*, lo que supondría aumentar el radio de la función de *kernel*, haciendo así que la SVM considere como vecinos patrones que se encuentran más alejados, lo que redunda en una mayor suavidad en los hiperplanos obtenidos.