Ejercicios y cuestiones de clasificadores SVM

Ejercicio 1.- ¿Cuál es el objetivo del algoritmo SVM? ¿Cuándo se puede aplicar con éxito?

Solución.- SVM son clasificadores lineales que tratan de encontrar el hiperplano separador de dos clases de patrones, los pertenecientes a la clase positiva y los pertenecientes a la clase negativa. Se aplican con éxito cuando los patrones de las dos clases en el conjunto de entrenamiento son linealmente separables. Son apropiados cuando tenemos un espacio de variables independientes de alta dimensionalidad. Los atributos del conjunto de entrenamiento se necesita que sean números reales.

Ejercicio 2.- ¿Qué función lineal es utilizada por SVM para clasificación? ¿Cómo se asigna a una instancia asociada a un vector de entrada la clase positiva o la negativa?

Solución.-SVM encuentra una función lineal de la forma $f(\mathbf{x}) = \langle \mathbf{w}.\mathbf{x} \rangle + b$ donde **x** es el vector de entrada, **w** es el vector de pesos, y b es el sesgo. Para asignar a un vector $\mathbf{x_i}$ la clase positiva (1) o la negativa (-1) se hace mediante la función

$$y_i = \begin{cases} 1 & \text{si } \langle \mathbf{w}.\mathbf{x}_i \rangle + b \geq 0 \\ -1 & \text{si } \langle \mathbf{w}.\mathbf{x}_i \rangle + b < 0 \end{cases}$$

Ejercicio 3.- Si los patrones del conjunto de entrenamiento son linealmente separables, ¿Cuantas bandas de decisión pueden separar los puntos de la clase positiva de los puntos de la negativa? ¿Qué bandas de decisión calcula el algoritmo SVM? ¿Porque?

Solución.- Existen infinitas funciones de decisión que separan ambas clases. El algoritmo SVM encuentra el hiperplano que maximiza el margen. Esta función minimiza la cota superior del error de clasificación.

Ejercicio 4.- ¿Que es el margen? ¿Cuáles son las ecuaciones de los dos hiperplanos que definen el margen H^+ y H^-?.

El margen es la distancia entre los dos hiperplanos que lo delimitan y cuyas ecuaciones son

$$H_+ : \langle \mathbf{w}.\mathbf{x}^+ \rangle + b = 1$$
$$H_- : \langle \mathbf{w}.\mathbf{x}^- \rangle + b = -1$$

donde x+ y x- son los puntos que están más cerca al hiperplano

$$\langle \mathbf{w}.\mathbf{x} \rangle + b = 0$$

Ejercicio 5.- Explique por qué el problema de minimización con restricciones se transforma en un problema dual de maximización. ¿Qué vectores de entrada (patrones de entrenamiento) se utilizan para calcular la solución?

Solución.- La transformación se lleva a cabo para facilitar el hallazgo de la solución óptima (es decir, el hiperplano de decisión con máximo margen).

Sólo los vectores soporte (aquellos que se encuentran en los hiperplanos H^+ y H^- son utilizados para calcular la solución). Esto reduce de forma considerable la complejidad, en tiempo de ejecución, del algoritmo SVM.

Ejercicio 6. Resuma las principales ventajas y desventajas de SVM.

Solución.

Las ventajas son:

• Posee una sólida fundamentación teórica.

• Realiza un procedimiento de clasificación más riguroso que la mayoría de los métodos alternativos, en especial para problemas con una alta dimensionalidad de los datos.

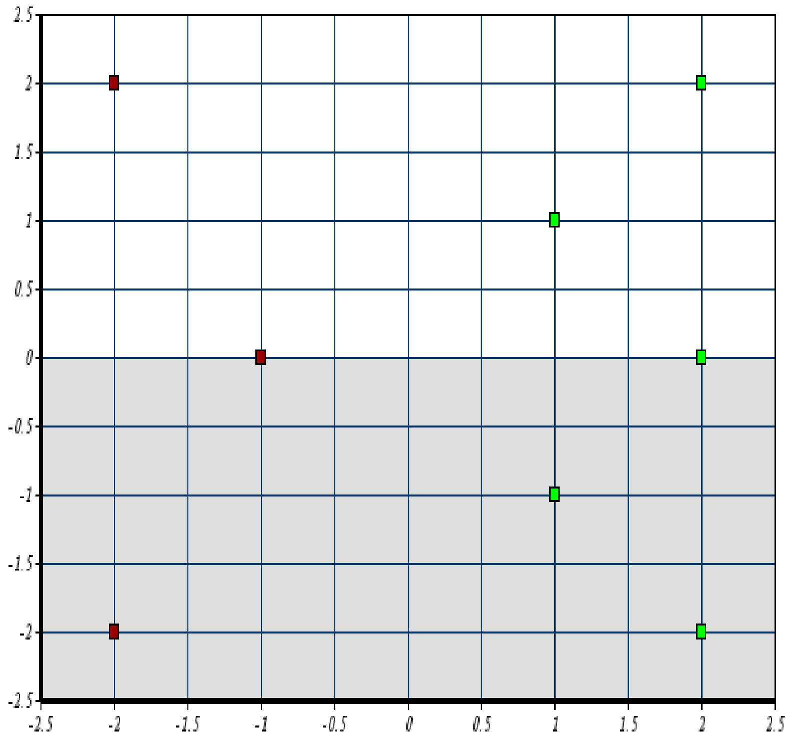
• Se puede aplicar a problemas de clasificación no lineal cuando utilizamos funciones de núcleo "kernel". Utilizando "el truco" del *kernel* se pueden tratar problemas con un coste computacional aceptable.

• Diferentes funciones de *kernel* pueden ser acopladas al algoritmo utilizando la misma técnica de aprendizaje y estudiar su rendimiento independientemente del algoritmo.

Las desventajas son:

- Trabaja solamente en espacios de entrada asociados a variables reales. Para datos categóricos, se necesita pasarlos a valores numéricos.
- Sólo está diseñado para clasificación binaria. En el caso de problemas multiclase se utilizan técnicas de tipo *ensemble* como son las variantes uno frente a todos o uno frente a uno.
- El hiperplano producido por el algoritmo SVM es difícil de interpretar por los usuarios. La interpretabilidad es aún peor si utilizamos funciones de *kernel*.
- SVM se utiliza en la resolución de problemas en los que no se necesita un entendimiento humano.

Ejercicio 7. Queremos aprender un margen estricto lineal SVM de los puntos de la figura, donde los puntos verdes tienen etiqueta positiva y los rojos etiquetas negativas. Encuéntrelo mediante un algoritmo de máquina de soporte lineal?



Solución: Para resolver el problema mediante el algoritmo de SVM tenemos que encontrar los valores de **w** y b. Hay que hallar dos hiperplanos con el margen más grande

Una solución puede ser utilizar los vectores soporte (-1,0) y (-2,2) para los rojos, mientras que para los verdes el vector soporte seria el (1,-1) De esta manera las ecuaciones son:

$(-1,0).\mathbf{w} + b=-1$ $(-2,2).\mathbf{w} + b=-1$ $(1,-1).\mathbf{w} + b=1$

Que se transforman en

$-w_1+b=-1$ (1); $-2w_1+2w_2+b=-1$ (2)

$w_1-w_2+b=1$ (3)

Cuya solución es $b=1/3$; $w_1=4/3$ y $w_2=2/3$ y la ecuación del hiperplano separador es

$g(\mathbf{x})=4/3x_1+2/3x_2+1/3$ y la frontera es $g(\mathbf{x})=0$; esto es $4x_1+2x_2+1=0$, o lo que es igual

$x_2=-2x_1-1/2$

Otra solución alternativa puede ser considerar los vectores soporte (-1,0) para los rojos y los vectores soporte (1,-1) y (1,1) para los verdes. De esta forma tendríamos las ecuaciones

$-w_1+b=-1$ (1); $w_1-w_2+b=1$ (2)

$w_1+w_2+b=1$ (3)

Cuya solución es $w_1=1$, $w_2=0$, $b=0$, luego la recta es $x_1=0$. Esta solución es mejor que la primera porque la distancia del punto (-1,0) a la recta discriminante es 2; mientras que la distancia del punto (1,-1) a la recta discriminante $4x_1+2x_2+1=0$, es $\frac{3}{\sqrt{20}}$

Ejercicio 8.- Queremos diseñar un clasificador para el problema XOR, en donde los puntos $x_1=(-1, -1)$ y $x_3=(1, 1)$ son de la clase C_1, y $x_2=(-1, 1)$ y $x_4=(1, -1)$ son de la clase C_2.

Considerando la transformación $\varphi : R^2 \rightarrow R^6$

$$y = \phi(x_1, x_2) = \left[1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2 , x_2^2\right]^T ,$$

mostrar que el problema se vuelve separable en el espacio transformado. Diseñar un clasificador de máximo margen para el problema. Especificar los hiperplanos (separador y de márgenes) y el margen.

Solución.-

Si $\Phi(\mathbf{x}) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2)$, se tiene que

$$\Phi(\mathbf{x})^T.\Phi(\mathbf{x}') = (1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2x_1^2,, x_2^2)\begin{pmatrix} 1 \\ \sqrt{2}x_1^{'} \\ \sqrt{2}x_2^{'} \\ \sqrt{2}x_1^{'}x_2^{'} \\ x_1^{'2} \\ x_2^{'2} \end{pmatrix} =$$

$$= \left(1 + 2x_1x_1^{'} + 2x_2x_2^{'} + 2x_1x_2x_1^{'}x_2^{'} + x_1^2x_1^{'2} + x_2^2x_2^{'2}\right) =$$

$$= (1 + x_1x_1^{'} + x_2x_2^{'})^2 = (1 + \mathbf{x}^T.\mathbf{x}')^2 = k(\mathbf{x}, \mathbf{x}')$$

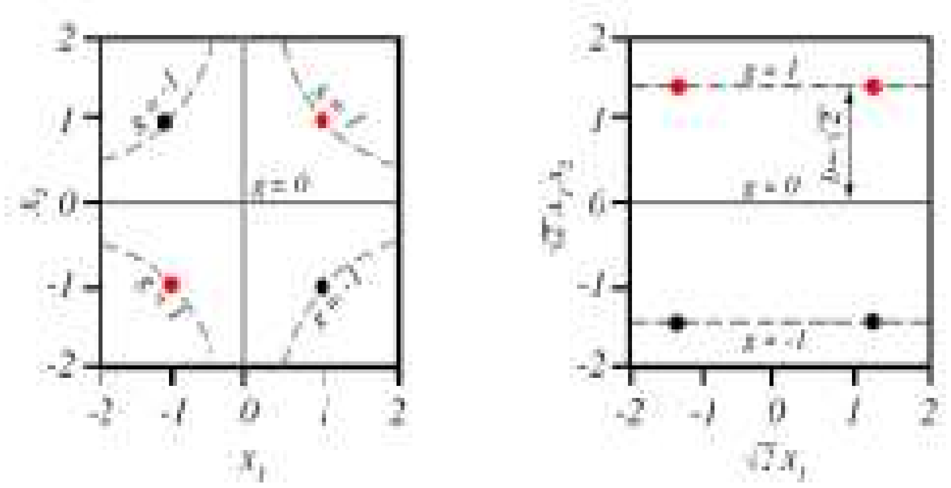
Luego el problema de optimización dual utilizando el truco del kernel es

$$\text{Maximizar } \tilde{L}(\mathbf{a}) = \sum_{n=1}^{4} a_n - \frac{1}{2}\sum_{n=1}^{4}\sum_{m=1}^{4} a_n a_m t_n t_m (\mathbf{x_n^T x_m} + 1)^2$$

$$\text{sujeto a: } a_1\text{-}a_2\text{+}a_3\text{-}a_4\text{=}0, \quad a_n \geq 0, \ n \in \{1, 2, 3, 4\}$$

La solución es $a_1=a_2=a_3=a_4=1/8$ y $w_0=\sqrt{2}$

La función discriminante es $g(x)=x_1x_2$



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

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

$$\mathrm{g}(\mathbf{x}_j) = \hat{\mathbf{w}}^{\mathbf{T}}\mathbf{x}_j + \hat{\mathrm{w}}_0 = \left(\sum_{i\in Sop} \hat{a}_i y_i \mathbf{x}_i\right)^T \mathbf{x}_j + \hat{w}_0 = \sum_{i\in Sop} \hat{a}_i y_i (\mathbf{x_i^T}\mathbf{x}_j) + \hat{w}_0$$

Ejercicio 9. Dados dos patrones de entrenamiento en R^1 $x_1=0, y_1=1$ $x_2=1, y_2=-1$. Determinar el hiperplano separador.

Solución.-
Ahora los datos son $x_1= 1$, $x_2= 0$ con $y = [+1,-1]^T$. Tenemos que encontrar los coeficientes w y $b\in R^1$, de forma tal que se verifique

$$\min_{\mathbf{w}} \quad \frac{1}{2}w^2$$
$$s.a. \ \ 1(w\times 1+\mathrm{b}) \ \geq 1 \ \ (1)$$
$$-1(\mathrm{w}\times 0+\mathrm{b}) \geq 1, \ (2)$$

De la ecuación (1) tenemos $-b \geq 1$. y sustituyendo en (2), $w \geq 2$. Así, para todo par (w, b) que satisfaga las ecuaciones (1) y (2), $w \geq 2$. Como queremos minimizar $(1/2)w^2$ tenemos que el valor más pequeño es w=2, por lo que la solución óptima es (w, b) = (2,−1). El hiperplano separador es entonces $g(x)=w\times x+b=0$, y sustituyendo tenemos $2x - 1 = 0$, que es el punto medio entre x_1 y x_2