Ejercicios 2018-2019

Para comenzar, usaremos siempre el mismo conjunto de datos:

$$\mathbf{X}^T = \begin{pmatrix} 0 & 2 & 3 & 2 \\ 0 & 2 & 0 & 0 \end{pmatrix}, \quad \mathbf{y}^T = (-1,-1,1,1)$$

Ejercicio 1: Plano de separación

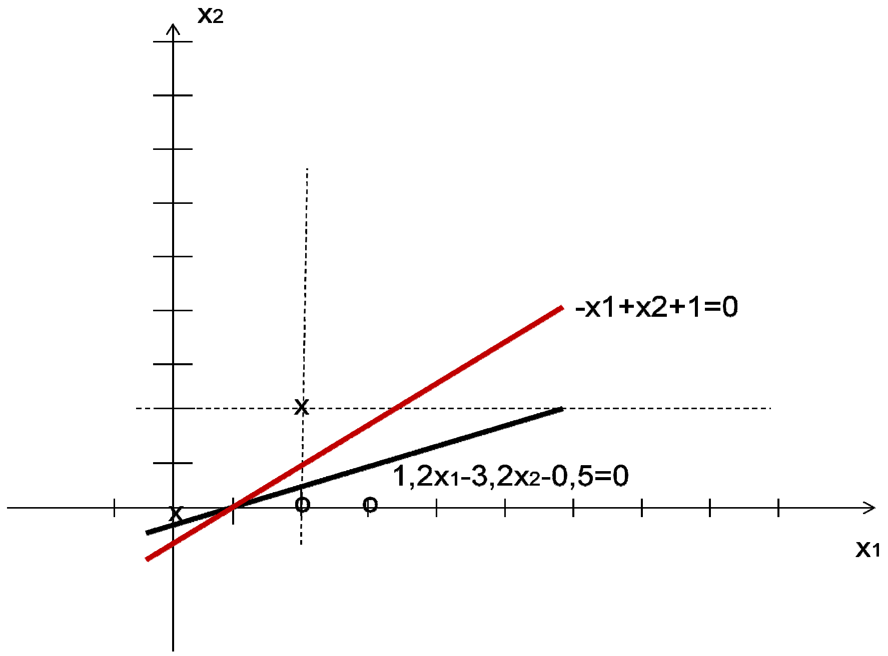
Sea $\mathbf{w}^T$= (1,2, -3,2) y b= -0.5 un hiperplano que separa los datos.

1.1) Mostrar los datos de aprendizaje y el plano de separación. ¿es óptimo en el sentido de la metodología de SVM?

1.2) Calcule el margen funcional ρ de este hiperplano

1.3) Calcule los parámetros de un hiperplano equivalente cuyo margen es igual a 1.

Solución.- Hay dos puntos de cada clase, cuya representación gráfica es



1.1) No es óptimo porque no tiene un margen máximo.

El margen máximo con restricciones es:

$$\min_{\mathbf{w}} \ \frac{1}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w},$$
$$s.a. \ \ y_i(\mathbf{w}^{\mathbf{T}}\mathbf{x}_i + \mathrm{w}_0) \ \geq 1, \quad i = 1...n$$

Por una parte el margen es

$$\frac{1}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w} = \frac{1}{2}\left((1,2)^2 + (-3,2)^2\right) = 5,84$$

por la otra, los cuatro vectores deben de cumplir las restricciones, esto es

-[(1,2, -3,2)$(0,0)^T$-0,5]=0,5 NO es ≥1

-[(1,2, -3,2)$(2,2)^T$-0,5]=4,5 SI es ≥1

[(1,2, -3,2)$(3,0)^T$-0,5]=3,1 SI es ≥1

[(1,2, -3,2)$(2,0)^T$-0,5]=1,9 SI es ≥1

De esta forma uno de los vectores no cumple las restricciones

1.2) El margen funcional es la mínima distancia de los puntos (0,0) y (2,0) al hiperplano. Esto es

$$d(P,r)=\frac{|Ax_1+Bx_2+C|}{\sqrt{A^2+B}} \qquad d((0,0),r)=\frac{|-0.5|}{\sqrt{(1,2)^2+(3,2)^2}}=0,146$$

$$d((2,0),r)=\frac{|1,2\times 2-0.5|}{\sqrt{(1,2)^2+(3,2)^2}}=0,556$$

Por lo tanto el margen es 0,146

1.3) Para ello contamos con tres vectores soporte, a saber (0,0), (2,2) y (2,0)
por lo tanto las ecuaciones son

(0,0).**w** + b=1 (2,2).**w** + b=1 (2,0).**w** + b=-1

La primera se convierte en b=1

La segunda en $2w_1+2w_2+b=1$ y la tercera en $2w_1+b=-1$

Sustituyendo b en la tercera ecuación, tenemos que $w_1=-1$, y sustituyendo los valores de b y w_1 en la tercera tenemos que $w_2=1$. La ecuación de la recta discriminante es
$-x_1+x_2+1=0$, o lo que es igual $x_2= x_1-1$

Ejercicio 2.- Formulación del dual en la metodología SVM

2.1. Formule el problema de optimización de la SVM (versión dual) bajo la forma de un problema de optimización cuadrática, es decir, determinar las matrices A y Q así como el vector $\boldsymbol{p}$ en la siguiente expresión:

$$\begin{cases}\max\limits_{\alpha}\dfrac{1}{2}\boldsymbol{\alpha}^{\mathbf{T}}\mathbf{Q}\boldsymbol{\alpha}+\mathbf{p}^{\mathbf{T}}\boldsymbol{\alpha}\\ s.a.\quad \mathbf{A}\boldsymbol{\alpha}\geq\ \mathbf{0}_{(n+2)}\end{cases}$$

2.2. Visualice los valores del vector $\boldsymbol{\alpha}$. ¿Qué observamos? Por qué ?

2.3. Calcula los parámetros del hiperplano (w, b). Compare los resultados con los obtenidos en el primal.

2.4. Implementar la función de decisión (expresión dual). seleccionados los ejemplos (2,3,+) y (4,0,-) compruebe que están correctamente clasificados.

Solución

2.1 Contamos con tres vectores soporte, a los cuales por comodidad les cambiamos las etiquetas de clase, a saber (0,0,+1), (2,2,+1) y (2,0,-1)

Dual

$$\begin{cases}\max\limits_{\alpha}-\dfrac{1}{2}\boldsymbol{\alpha}^{\mathbf{T}}\mathbf{Q}\boldsymbol{\alpha}+\mathbf{p}^{\mathbf{T}}\boldsymbol{\alpha}\\ s.a.\quad \mathbf{A}\boldsymbol{\alpha}\geq\ \mathbf{0}_{(n+2)}\end{cases}$$

$$\max-\frac{1}{2}\sum_{i,j=1}\alpha_i\alpha_j y_i y_j(\mathbf{x}_i^T\mathbf{x}_j)+\sum_{i=1}\alpha_i=\max-\frac{1}{2}\boldsymbol{\alpha}^{\mathbf{T}}\mathbf{Q}\boldsymbol{\alpha}+\sum_{i=1}\alpha_i$$

$$s.a.\ \sum_{i=1}\alpha_i y_i=0,\ \ \alpha_i\geq 0,\ \text{i=1,...,n}$$

En nuestro caso sujeto a $\alpha_1+\alpha_2-\alpha_3$=0. Los coeficientes del hiperplano se calculan como

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

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

Ahora la función a maximizar es

$$\mathrm{L}(\alpha_1,\alpha_2,\alpha_3,\lambda;\ \mathbf{x_1},\mathbf{x_2},\mathbf{x_3}) =$$

$$= \alpha_1+\alpha_2+\alpha_3 - \frac{1}{2}\begin{pmatrix}\alpha_1^2(\mathbf{x_1^T}.\mathbf{x_1})+\alpha_1\alpha_2(\mathbf{x_1^T}.\mathbf{x_2})-\alpha_1\alpha_3(\mathbf{x_1^T}.\mathbf{x_3})+\alpha_2\alpha_1(\mathbf{x_2^T}.\mathbf{x_1})+\alpha_2^2(\mathbf{x_2^T}.\mathbf{x_2})-\alpha_2\alpha_3(\mathbf{x_2^T}.\mathbf{x_3})+ \\ -\alpha_3\alpha_1(\mathbf{x_3^T}.\mathbf{x_1})-\alpha_3\alpha_2(\mathbf{x_3^T}.\mathbf{x_2})+\alpha_3^2(\mathbf{x_3^T}.\mathbf{x_3})\end{pmatrix}$$

$$-\lambda(\alpha_1+\alpha_2-\alpha_3)$$

Sustituyendo los valores de los vectores soporte $x_1^T=(0,0)$ $y_1=+1$, $x_2^T=(2,2)$ $y_2=+1$ y $x_3^T=(2,0)$ $y_3=-1$, tenemos

$$\mathrm{L}(.) = \alpha_1+\alpha_2+\alpha_3 - \frac{1}{2}\begin{pmatrix}\alpha_1^2(0)+\alpha_1\alpha_2(0)-\alpha_1\alpha_3(0)+\alpha_2\alpha_1(0)+\alpha_2^2(8)-\alpha_2\alpha_3(4)+ \\ -\alpha_3\alpha_1(0)-\alpha_3\alpha_2(4)+\alpha_3^2(4)\end{pmatrix}$$

$$-\lambda(\alpha_1+\alpha_2-\alpha_3)$$

Derivando con respecto a los α_i y λ tenemos

$$\frac{\partial L}{\partial \alpha_1} = 0, \text{ esto es } 1-\lambda = 0$$

$$\frac{\partial L}{\partial \alpha_2} = 0, \text{ esto es, } 1-8\alpha_2+4\alpha_3-\lambda = 0$$

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

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

2.2) Resolviendo el sistema de 4 ecuaciones con 4 incognitas tenemos

$\alpha_1 = 1/2;\ \alpha_2 = 1/2 \text{ y } \alpha_3 = 1$ luego el vector $\boldsymbol{\alpha}$=(1/2,1/2,1)

2.3) Los coeficientes de la recta son

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

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

$$\hat{w}_{\mathbf{0}} = 1$$

La ecuación es $-x_1+x_2+1=0$, esto es $x_2=x_1-1$, que coincide con la solución del primal

2.4) Sea el patrón (2,3) que supuestamente pertenece a la clase positiva, la función discriminante le asigna el valor

-2+3+1= 2 luego está bien clasificado en la clase positiva

Por otra parte si queremos clasificar el patrón de componentes (4,0) tenemos que

-4+0+1= -3, que queda clasificado en la clase negativa, luego está bien clasificado

Ejercicio 3. Examen 7/07/2015

¿En que se basa el algoritmo de las Maquinas de Vectores Soporte? Dados los datos etiquetados como positivos definidos como vectores traspuestos $(2,1.5)^T$, $(2,-1.5)^T$, $(-1.5,-2)^T$, $(-1.5, 2)^T$ y los datos etiquetados como negativos $(1,1)^T$, $(1,-1)^T$, $(-1,-1)^T$ y $(-1,1)^T$ y la función de transformación del espacio de características de entrada

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

Calcular los vectores soporte, la ecuación del hiperplano de separación y utilizar el algoritmo SVM para clasificar el patrón de coordenadas $(-2,4)^T$

Solución

El algoritmo SVM se basa en crear un clasificador biclase lineal, en una dimensión mayor a la dada en el espacio de variables independientes iniciales del problema, Conforme aumentamos la dimensionalidad, la probabilidad de que las clases en este nuevo espacio sean linealmente separables aumenta.

La idea por tanto es minimizar el margen, por una parte, y por la otra minimizar el número de errores de clasificación en una versión soft del algoritmo.

Dados $p_1=(2,1.5)^T$, $p_2=(2,-1.5)^T$, $p_3=(-1.5,-2)^T$, $p_4=(-1.5, 2)^T$ que son patrones de la clase C+ y $p_5=(1,1)^T$, $p_6=(1,-1)^T$, $p_7=(-1,-1)^T$ y $p_8=(-1,1)^T$ que son patrones de la clase C-

Transformamos los patrones según la función no lineal del enunciado. De esta forma.

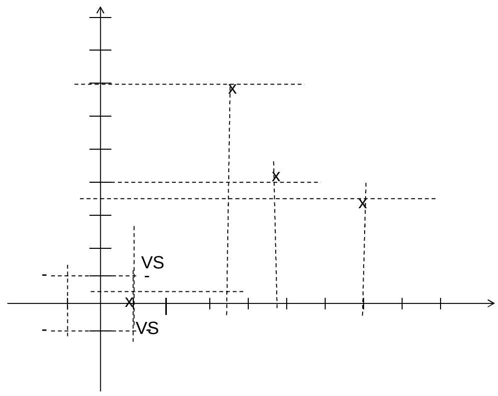
$$p_1' = \phi(p_1) = \begin{pmatrix} 2-1,5+|2-1,5| \\ 2-2+|2-1,5| \end{pmatrix} = \begin{pmatrix} 1 \\ 0,5 \end{pmatrix};$$

puesto que para p_1 $\sqrt{(2)^2+(1,5)^2} > 2;$

de forma similar $p_2' = \phi(p_2) = (7 \quad 3,5)^T$

$p_3' = \phi(p_3) = (4,5 \quad 4)^T$; $p_4' = \phi(p_4) = (3,5 \quad 7)^T$

Mientras que dado que p_5 es tal que $\sqrt{1^2+1^2} = \sqrt{2} = 1,41 < 2$. Entonces no cambia de valor $p'_5=(1,1)^T$, y de forma similar $p'_6=(1,-1)^T$, $p'_7=(-1,-1)^T$ y $p'_8=(-1,1)^T$. Los 8 nuevos patrones se muestran en la figura



Los puntos más cercanos de las dos clases son el $(1\ \ 0{,}5)^T$ y el $(1,\ 1)^T$

De esta forma $\overline{S}_1 = (1\ 0{,}5\ 1)^T$ para C^+; mientras que $\overline{S}_2 = (1\ 1\ 1)^T$ para C^-, de esta forma hemos ampliado el número de componentes de los vectores al añadirles un 1 para contemplar el cálculo del sesgo

Las ecuaciones del dual (antes hay que construir el primal con la función a optimizar y con las restricciones) son ahora

$$\begin{cases} \alpha_1.\overline{S}_1.\overline{S}_1^T + \alpha_2.\overline{S}_2.\overline{S}_1^T = +1 \\ \alpha_1.\overline{S}_1.\overline{S}_2^T + \alpha_2.\overline{S}_2.\overline{S}_2^T = -1 \end{cases}$$

pero

$$\overline{S}_1.\overline{S}_1^T = 2,25 \ \ ; \overline{S}_2.\overline{S}_1^T = 2,5 \ ; \overline{S}_1.\overline{S}_2^T = 2,5 \ y\ \overline{S}_2.\overline{S}_2^T = 3$$

Sustituyendo y resolviendo la ecuación tenemos que

α_1= 11 y α_2= -9,5; por lo que el vector de pesos es de la forma

$$\mathbf{w} = 11\begin{pmatrix} 1 \\ 0,5 \\ 1 \end{pmatrix} - 9,5\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1,5 \\ -4 \\ 1,5 \end{pmatrix}$$

La ecuación de la recta es **w.x**+b,

$1{,}5x_2 = 4x_1 - 1{,}5$ o lo que es igual $x_2 = 2{,}6x_1 - 1$

Para clasificar el punto $(-2\ 4)^T$ tenemos que como

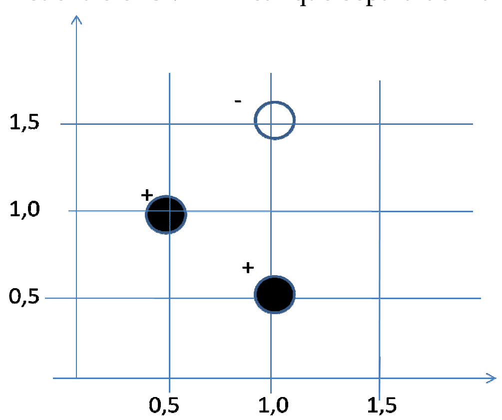
$$\phi(-2\ \ 4) = \begin{pmatrix} 2-4+|-2-4| \\ 2+2+|-2-4| \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \end{pmatrix}$$, si le añadimos la componente del sesgo, entonces

$$f(\begin{pmatrix} 4 \\ 10 \end{pmatrix}) = \sigma\left[11.\begin{pmatrix} 1 \\ 0,5 \\ 1 \end{pmatrix}.(4 \quad 10 \quad 1) - 9,5\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.(4 \quad 10 \quad 1)\right] = \sigma(-32,5)$$

por lo que el patrón pertenece a la clase C-, dado que la salida es negativa

Ejercicio 4.-Examen 25/01/2016

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.



Haga las comprobaciones pertinentes utilizando el método primal y el dual

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 $-x + 2 = 0$.