

UNIVERSIDAD CATÓLICA DE SAN PABLO
MAESTRÍA EN CIENCIA DE LA COMPUTACIÓN
SISTEMAS INTELIGENTES

Máquina de Vectores de Soporte (SVM)

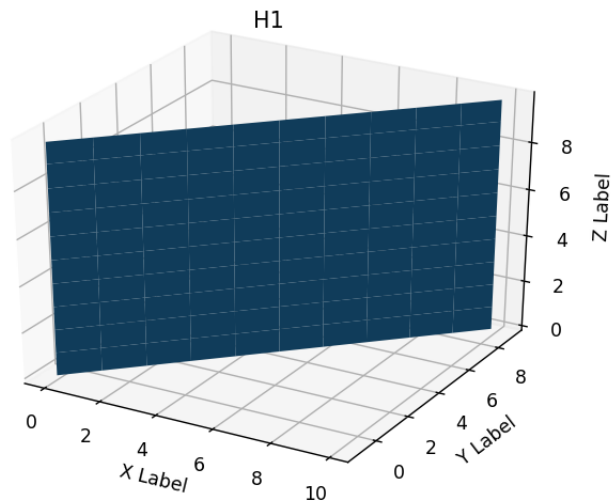
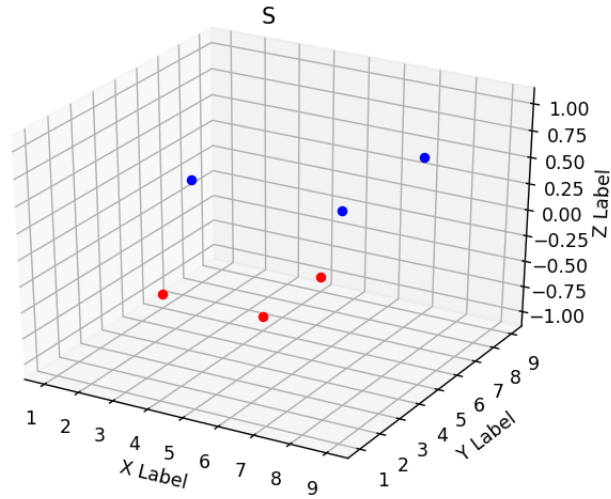
Prof. Graciela Meza Lovón

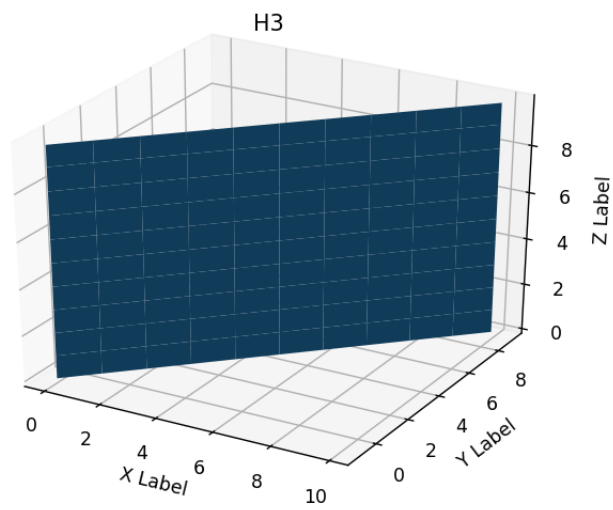
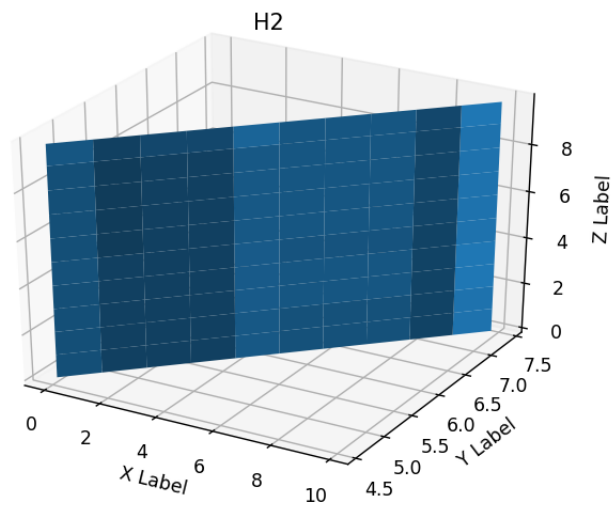
Palomino Paucar, Daniel Alfredo
Noviembre 26, 2018

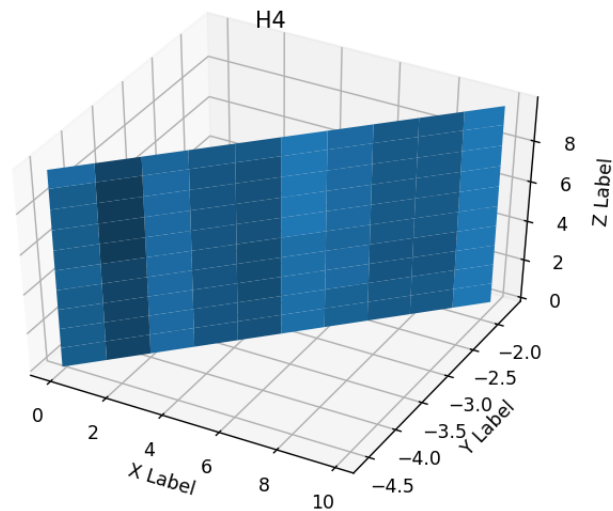
1. PREGUNTAS DE TEORÍA

Sea el conjunto $S = ((1, 6), -1), ((4, 9), -1), ((4, 6), -1), ((5, 1), 1), ((9, 5), 1), ((9, 1), 1)$ y un conjunto de cuatro hiperplanos $H = H_1, H_2, H_3, H_4$ definidos como: $H_1 : x_1 - x_2 - 1 = 0$, $H_2 : 2x_1 - 7x_2 + 32 = 0$, $H_3 : \sqrt{\frac{1}{2}}x_1 - \sqrt{\frac{1}{2}}x_2 - \sqrt{\frac{1}{2}} = 0$, $H_4 : 2x_1 - 7x_2 - 32 = 0$

1. Usando cualquier lenguaje de programación grafique S , H_1 , H_2 , H_3 y H_4 .







2. Encuentre los parámetros w y b que definen los hiperplanos H_1 , H_2 , H_3 y H_4 , y luego determine para H_1 , H_2 , H_3 y H_4 si son hiperplanos de separación. Fundamente.

Dadas la ecuación vectorial de un hiperplano: $W.X + b = 0$, se tiene al extender la ecuación: $w_1.x_1 + w_2.x_2 + w_3.w_3 + b = 0$.

Dando la forma a la ecuaciones dadas:

a) H_1 :

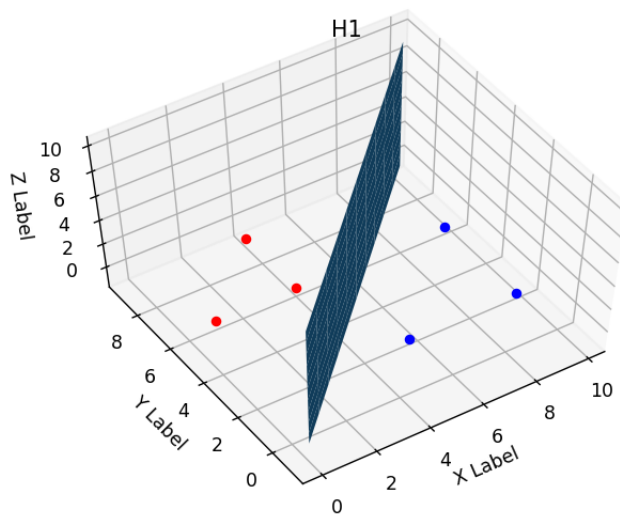
$$x_1 - x_2 - 1 = 0$$

$$(1, -1, 0).X - 1 = 0$$

Entonces:

$$W = (1, -1, 0) \wedge b = -1$$

Graficando:



Por lo tanto H_1 : Sí es un hiperplano de separación.

b) H_2 :

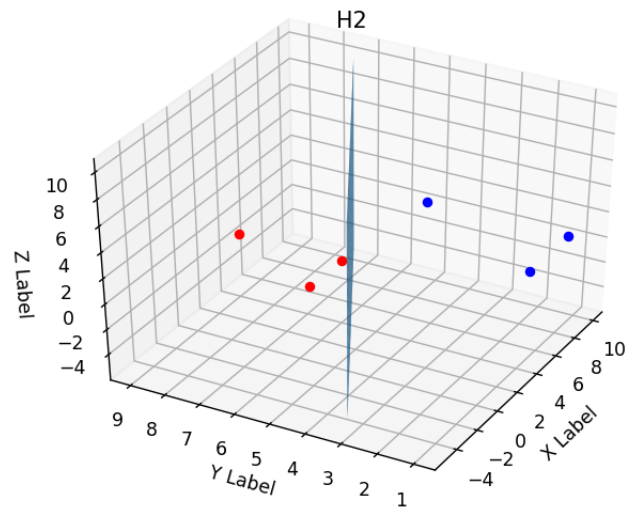
$$2x_1 - 7x_2 + 32 = 0$$

$$(2, -7, 0) \cdot X + 32 = 0$$

Entonces:

$$W = (2, -7, 0) \wedge b = 32$$

Graficando:



Por lo tanto H_2 : Sí es un hiperplano de separación.

c) H_3 :

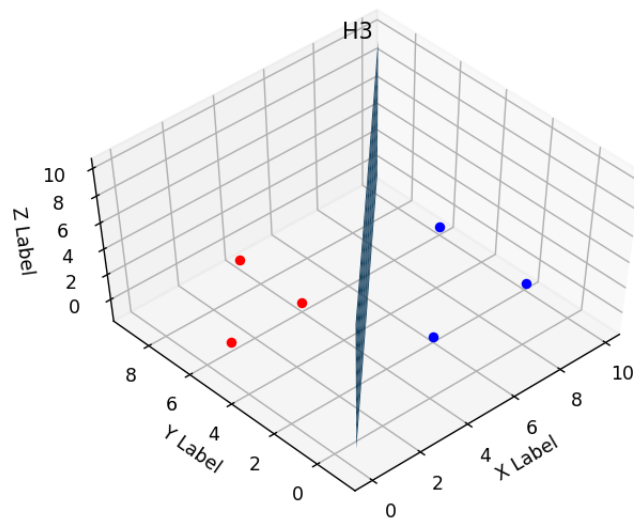
$$\sqrt{\frac{1}{2}}x_1 - \sqrt{\frac{1}{2}}x_2 - \sqrt{\frac{1}{2}} = 0$$

$$(1, -1, 0) \cdot X - 1 = 0$$

Entonces:

$$W = \left(\sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{2}}, 0\right) \wedge b = -\sqrt{\frac{1}{2}}$$

Graficando:



Por lo tanto H_3 : Sí es un hiperplano de separación.

d) H_4 :

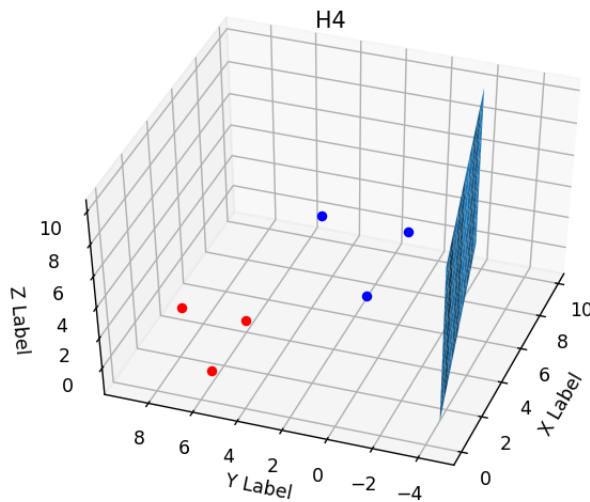
$$2x_1 - 7x_2 - 32 = 0$$

$$(2, -7, 0) \cdot X - 32 = 0$$

Entonces:

$$W = (2, -7, 0) \wedge b = -32$$

Graficando:



Por lo tanto H_4 : No es un hiperplano de separación.

3. En el conjunto H , ¿cuántos hiperplanos iguales existen?. En el caso de que existan ¿cuáles son éstos? Fundamente.

Existen 2 hiperplanos iguales H_1 y H_3 , dado que si a la ecuación de H_3 la dividimos entre $\sqrt{\frac{1}{2}}$ tendríamos la misma ecuación que H_1 .

4. Calcule el margen τ para cada hiperplano de separación. Luego, suponga que el conjunto H contiene al hiperplano óptimo H^* , ¿cuál sería H^* ? Fundamente.

a) H_1 :

$$\tau_1 = 2,1213$$

b) H_2 :

$$\tau_2 = 0,2742$$

c) H_3 :

$$\tau_3 = 2,1213$$

De acuerdo a los datos calculados, si el hiperplano H^* existiera en el conjunto H , este sería H_1 o H_3 (pues representan el mismo plano), dado que poseen el τ con mayor valor.

5. ¿Cuáles son los vectores de soporte del hiperplano H^* escogido en la pregunta anterior?. Fundamente. (No necesita encontrar los valores α)

Los vectores de soporte del hiperplano óptimo elegido (H_1 o H_3) serían:

$$(4, 6, -1), (5, 1, 1), (9, 5, 1)$$

pues se encuentran en el margen óptimo, teniendo cada vector el τ mínimo de 2,1313.

6. Demuestre la primera condición KKT, i.e. (Ec. 7 de las diapositivas) $\frac{\partial L}{\partial w}(w^*, b^*, \alpha) = w^* - \sum_{i=1}^m \alpha_i y^{(i)} x^i$

Sea:

$$L = \frac{1}{2} W \cdot W^T - \sum_{i=1}^m \alpha_i [y^{(i)} (W^T X^{(i)} + b) - 1]$$

Expandiendo los términos:

$$L = \frac{1}{2} W \cdot W^T - \sum_{i=1}^m \alpha_i [y^{(i)} W^T X^{(i)}] - \sum_{i=1}^m \alpha_i [y^{(i)} b] + \sum_{i=1}^m \alpha_i$$

Para encontrar los W^*, b^*, α^* óptimo, necesitamos derivar respecto de los parámetros, lo cual se describe en la ecuación 1:

$$\frac{\partial L(W^*, b^*, \alpha^*)}{\partial w_i} = 0, i = 1, \dots, n$$

Entonces, derivando L respecto de w_i tenemos:

$$\frac{\partial L(W^*, b^*, \alpha^*)}{\partial w_i} = \frac{\partial(\frac{1}{2} W \cdot W^T - \sum_{i=1}^m \alpha_i [y^{(i)} W^T X^{(i)}] - \sum_{i=1}^m \alpha_i [y^{(i)} b] + \sum_{i=1}^m \alpha_i)}{\partial w_i}$$

Separando los términos de la derecha:

$$\frac{\partial L(W^*, b^*, \alpha^*)}{\partial w_i} = \frac{\partial(\frac{1}{2} W \cdot W^T)}{\partial w_i} - \frac{\partial(\sum_{i=1}^m \alpha_i [y^{(i)} W^T X^{(i)}])}{\partial w_i} - \frac{\partial(\sum_{i=1}^m \alpha_i [y^{(i)} b])}{\partial w_i} + \frac{\partial(\sum_{i=1}^m \alpha_i)}{\partial w_i}$$

El tercer y cuarto término de la derecha no depende de w_i , por lo tanto al derivarse desaparecerán:

$$\frac{\partial L(W^*, b^*, \alpha^*)}{\partial w_i} = \frac{\partial(\frac{1}{2} W \cdot W^T)}{\partial w_i} - \frac{\partial(\sum_{i=1}^m \alpha_i [y^{(i)} W^T X^{(i)}])}{\partial w_i}$$

Finalmente calculando la derivada parcial:

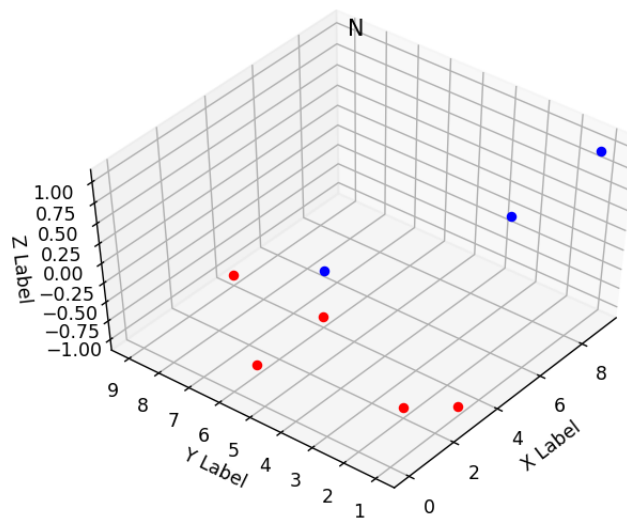
$$\frac{\partial L}{\partial w}(w^*, b^*, \alpha) = w^* - \sum_{i=1}^m \alpha_i y(i) x^i = 0$$

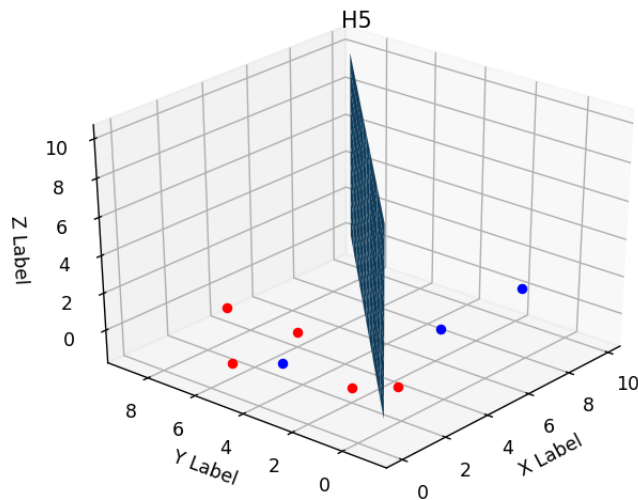
Quedando demostrado la ecuación 7 de las diapositivas.

Sea el conjunto $N = ((1, 6), -1), ((4, 9), -1), ((4, 6), -1), ((5, 1), 1), ((9, 1), 1), ((0, 3), 1), ((2, 2), -1), ((3, 1), -1)$ y el hiperplano H_1 definido anteriormente.

7. Usando cualquier lenguaje de programación grafique N y H_5 .

Graficando:





8. Identifique los ejemplos que son separables y los que no lo son. Luego, determine los ejemplos que son clasificados correctamente y los que no.

Para que un vector sea separable debe cumplir:

$$y_i(W^T X) \geq 1$$

Por lo tanto haciendo los cálculos para los ejemplos del espacio N y el hiperplano H_5 :

- a) $((1, 6), -1): 5$
- b) $((4, 9), -1): 5$
- c) $((4, 6), -1): 2$
- d) $((5, 1), 1): 4$
- e) $((9, 1), 1): 8$
- f) $((0, 3), 1): -3$
- g) $((2, 2), -1): 0$
- h) $((3, 1), -1): -2$

Por lo tanto: Los ejemplos separables son:

$((1, 6), -1), ((4, 9), -1), ((4, 6), -1), ((5, 1), 1), ((9, 1), 1)$

y los no separables son:

$((0, 3), 1), ((2, 2), -1), ((3, 1), -1)$

Así mismo, los ejemplos clasificados correctamente son:

$((1, 6), -1), ((4, 9), -1), ((4, 6), -1), ((5, 1), 1), ((9, 1), 1), ((2, 2), -1)$

y los no clasificados correctamente son:
 $((0, 3), 1), ((3, 1), -1)$

9. Calcule la ... de los ejemplos no separables.
Enunciado NO definido.

2. PREGUNTAS DE INVESTIGACIÓN

1. Explique el significado de la constante C en el término $C \sum_{i=1} \epsilon_i$ que se agrega a la función objetivo en el caso de ejemplos casi linealmente separables. Luego, explique la influencia de C en la capacidad de generalización de una SVM.
2. Describa el significado del parámetro γ en el kernel gaussiano. Luego, explique la influencia de γ en la capacidad de generalización de una SVM.

3. IMPLEMENTACIÓN

1. Usando Scikit-learn de Python, implemente (comente su código) una svm que clasifique el conjunto de datos (por definir).
2. Experimente y muestre resultados usando diferentes valores para los parámetros de los kernels: lineal, polinomial, gaussiano, y el parámetro C . Los resultados deben ser mostrados en el documento pdf.
3. Dentro de la sección de Implementación incluya una subsección donde indique las instrucciones para ejecutar el código.