Máquinas de Vectores de Soporte (SVM)

E. Wilderd Mamani Condori

November 2018

1 Preguntas de Teoria

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:

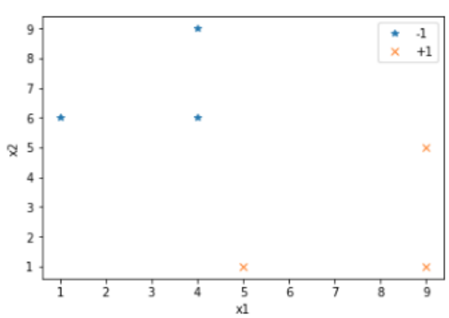
$$H_1 : x_1 - x_2 - 1 = 0,$$
$$H_2 : 2x_1 - 7x_2 + 22 = 0,$$
$$H_3 : \sqrt{\frac{1}{2}}x_1 - \sqrt{\frac{1}{2}}x_2 - \sqrt{\frac{1}{2}} = 0,$$
$$H_4 : 2x_1 - 7x_2 - 32 = 0,$$

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

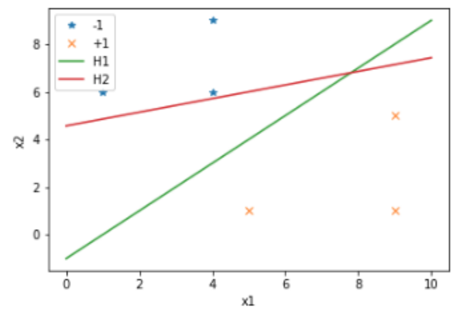
   Solución

   Primero para este ejemplo, vamos hacer uso de lenguage python para poder graficar el conjunto S y su respectiva separación $\{-1, 1\}$, y tambien graficaremos los hiperplanos en 2D.

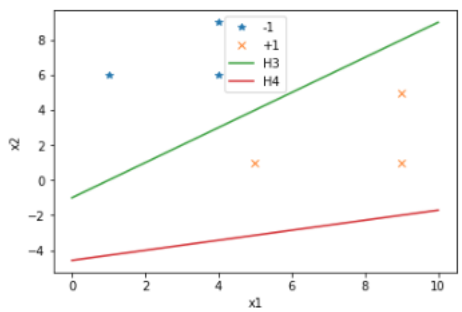
   para los puntos tenemos:



ahora vamos a graficar con H_1 y H_2



para los hiperplanos H_3 y H_4 con los puntos correspondietes tenemos



como podemos observar en las anteriores figuras, lo que tenemos es que dos hiperplanos coenciden asi estos son graficado en el punto medio intuitivamente podemos tambien decir que H_3 factorizando la expresion de este $\sqrt{\frac{1}{2}}$, obtenemos la ecuacion del hiperplano H_1, asi que esta es una razon por la que coenciden.

2. Encuentre los parametros **w** y b que definen los hiper$H_1, H_2, H_3,$ y H_4 y luego determine para $H_1, H_2, H_3,$ y H_4 si son parametros de separación.

   Solución

   En esta pregunta vamos a desarrollar la expresion de acuerdo al enunciado de la forma general del hiperplano para sacar los parametros **W** y b tenemos la expresion:

$$
\begin{aligned}
0 &= W^T X + b \\
0 &= AX_1 + BX_2 + b \\
0 &= \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + b \\
0 &= \begin{bmatrix} A \\ B \end{bmatrix}^T \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + b \\
W^T &= \begin{bmatrix} A \\ B \end{bmatrix}^T \\
X &= \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}
\end{aligned}
$$

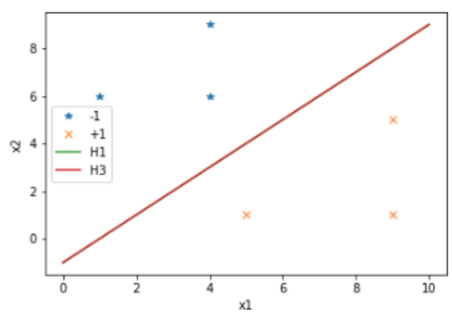
   de lo anterior podemos extraer los parametro correspondientes a W y b asi para los hiperplanos tenemos:

$$H_1: \quad \begin{bmatrix}1 & -1\end{bmatrix}\begin{bmatrix}X_1\\X_2\end{bmatrix} + (-1)$$
$$H_2: \quad \begin{bmatrix}2 & -7\end{bmatrix}\begin{bmatrix}X_1\\X_2\end{bmatrix} + (32)$$
$$H_3: \quad \begin{bmatrix}\sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}X_1\\X_2\end{bmatrix} + (\sqrt{\frac{1}{2}})$$
$$H_1: \quad \begin{bmatrix}2 & -7\end{bmatrix}\begin{bmatrix}X_1\\X_2\end{bmatrix} + (-32)$$

3. En el conjunto H, ¿ cuantos hiperplanos iguales existe?. En el caso que existan, ¿cuales son?, Fundamente

   Solución

   Del la pregunta teorica 1 podemos notar que H_1 y H_3 son dos hiperplanos que tienen sus respectivos graficos que coensiden asi tenemos el siguiente grafico

   ademas de sus ecuaciones podemos factorizar $\sqrt{\frac{1}{2}}$ en H_3 para llegar a obtener H_1 y asi finalmente podemos tener las mismas ecuaciones con lo que concluimos que son iguales.

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

   Solución

   para calcular el margen de separacion segun definicion tenemos que sacar las distancias con respecto a los $H*$ a cada punto para luego sacar los τ_{max} siguiendo :

$$y^{(i)}\frac{|D(x^{(i)})|}{||w||}$$

   para facilitar nuestro calculo procederemos a sacar primero del un cadidato es nuestro punto $(4,6)$ asi que segun el grafico este seria nuestro τ para el hiperplano H_1 y H_3, de esta manera tenemos el τ para cada uno de los hiperplanos asumiento el punto (4,6) el candidato a cumplir:

$$y^{(i)}\frac{|D(x^{(i)})|}{||w||} \geq \tau_{max}, i = 1, \ldots, m$$

entonces decimos sea $x = (4,6)$, hallaremos la distancia con respecto a H_1 y H_3, que son los candidatos a ser optimos, asi tenemos:

$$
\begin{aligned}
H_1 &= x_1 - x_2 - 1 \\
X &= (a,b) = (4,6) \\
&= \frac{|D(x)|}{||w||} = \frac{|AX_1 + BX_2 + C|}{||w||} \\
&= \frac{|1(4) + (-1)(6) - 1|}{\sqrt{2}} \\
&= \frac{3}{\sqrt{2}}
\end{aligned}
$$

de lo anteriormente supuesto decimos que el margen de separación desde un hiperplano H_1 es $\frac{3}{\sqrt{2}}$.

ahora suponiendo que uno de los hiperplanos es esta dentro de $H*$ entonces podemos asumir que estos $\mathrm{H}_1 y H_3 sonsolhiperplanosmasoptimosdebidoaquetienenel \tau_{max}$

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

Solución

para elegir los vectores de soporte primero tenemos que verificar que cumpla la condicion

$$
y^{(i)} \left(\left\langle w, x^{(i)} \right\rangle + b \right) = 1
$$

para esto vamos a elegir un punto representante por el cual supuestamente pasa el vector de soporte entonces sea este punto del grafico anterior p1 = (4,6) y p2 = (9,5),con H_1, tenemos y asi probaremos la condicion anterior tenemos:

$$
\begin{aligned}
\frac{|D(x)|}{||w||} &\geq \tau_{max} \\
\frac{|x_1 - x_2 + b|}{\sqrt{2}} &\geq \frac{3\sqrt{2}}{2} \\
\frac{|x_1 - x_2 + b|}{\sqrt{2}} &= \frac{3\sqrt{2}}{2} \\
x_1 - x_2 + b &= 3 \\
\text{ahora con p1, y } H_1 \text{ tenemos} & \\
X &= (4,6) \\
(4) - (6) + b &= -3 \\
b_1 &= -1 \\
\text{ahora con p2 y } H_1 \text{ tenemos} & \\
X &= (9,5) \\
(9) - (4) + b &= 3 \\
b_2 &= 1
\end{aligned}
$$

de lo anterior podemos notar que los vectores de soporte pasan por los puntos p1, y p2, pero ahora para hallar esos vectores de soporte hallamos la ecuacion del hiperplano $H_1 : x_1 - x_2 - 1 = 0$ entonces sabemos que los vectores de soporte tendran la misma pendiente lo que implica solo hallar la constante por el cual pasa asi tendremos la ecuacion siguiente para el vector de soporte $v_1 : x_1 - x_2 - b_{v1} = 0$ y $v_2 : x_1 - x_2 - b_v = 0$, entonces hallamos para los puntos por donde pasan cada uno de los vectores.

para el punto p1 = (4,6) = (x_1, x_2)

$$\begin{aligned} v_1 : x_1 - x_2 - b_{v1} &= 0 \\ \text{con} \quad b_1 &= -1 \\ (4) - (6) - b_{v1} &= 0 \\ b_{v1} &= 2 \end{aligned}$$

para el punto p2 = (9,5) = (x_1, x_2)

$$\begin{aligned} v_2 : x_1 - x_2 - b_{v2} &= 0 \\ \text{con} \quad b_2 &= 1 \\ (9) - (5) - b_{v1} &= 0 \\ b_{v1} &= 4 \end{aligned}$$

por consiguiente tenemos los vectores de soporte $v_1 : x_1 - x_2 - 2 = 0$ y $v_2 : x_1 - x_2 - 4 = 0$

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

Solución

para demostrar esta condicion partimos de la ecuación al cual se le llama problema de optimización dual, para lo cual calculamos el lagrangiano:

$$L(w, \beta, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{m} \alpha_i \left[y^{(i)}(W^T X^{(i)} + b) - 1 \right]$$

primero vamos a derivar con respecto a W, para obtener una expresion como el siguiente: para la derivada las constantes son iugal a cero asi, tenemos

$$\frac{\partial L}{w}(w, \beta, \alpha) = (1)w - \sum_{i=1}^{m} \alpha_i \left[y^{(i)}((1)X^{(i)} + (0)) - (0) \right]$$

aplicando la ecuacion (1) que dice

$$\frac{\partial L}{w}(w*, \beta*, \alpha*) = 0, i = 1, ..., I$$

tenemos:

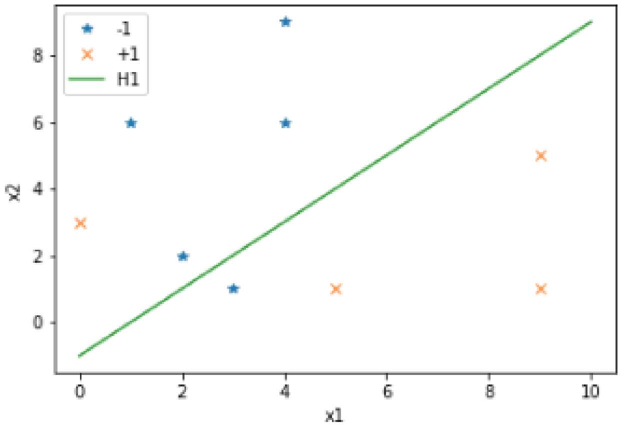
$$\frac{\partial L}{w}(w*, \beta*, \alpha*) = w * - \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} = 0$$

de lo anterior queda demostrado la ecuacion 7, con la primera condicion solamente aplicacando la primera condicion.

7. * Sea el conjunto $N = ((1,6),1),((4,9),1),((4,6),1),((5,1),1),((9,5),1),((9,1),1),((0,3),1),((2,2),1),((3,1),1)$y el hiperplano H_1 definido anteriormente. Usando cualquier lenguaje de programación, grafique N y H_5 .

Solución

acontinación mostramos los puntos correspondientes a la clasificacion con sus etiquetas y tambien el



hiperplano H_1 07.png

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.

   Solución

   para este ejemplo y determinar los ejemplos no separables primero los puntos tienen que cumplir la siguiente condición:

   $$y^{(i)}(W^T x^{(i)}) \geq 1$$

   de lo anterior tenemos dos caso no separables : primero cuando no se cumple la condicion data anteriormente , para este ejemplo, del grafico son para los puntos $p_1 : (2,2)$ y $p_2 : (3,1)$, asi por la condición:

   $$\begin{aligned} H_1 &: x_1 - x_2 - 1 = 0 \\ W &= (1,-1) \quad x = (x_1, x_2) \\ &\text{reemplazando p1 en H1 tenemos:} \\ (1)(2) - (1)(2) &\geq 1 \\ 0 &\ngeq 1 \text{ ahora reemplazando p2 en H1 tenemos:} \\ (1)(1) - (1)(3) &\geq 1 \\ -2 &\ngeq 1 \end{aligned}$$

   por lo tanto decimos que p_1 y p_2 son no separables, por que caen dentro del margen asociado a la clase correcta.

   Por otro lado tenemos un punto $P_3 : (0,3)$ que es un conjunto fuera de su clasificacion pues aparece como x en otro lado donde no esta su grupo asi, entonces decimos que $(0,3)$ cae al otro lado del hiperplano de clasificación en un grupo donde no esta correctamente clasificado.

9. Calcule las variables de holgura de los ejemplos no separables.

   Solución

   para calcular la variable de holgura para los ejemplos no separables debemos ver la relajacion que existe y el grado de separabilidad de los conjuntos con respecto al borde de su clasificación correspondiente.De lo anterior ahora vamos a seguir para ejemplo no separables se cumple.

$$y^{(i)}\left(\left\langle W, X^{(i)}\right\rangle + b\right) \geq 1-\varepsilon, \quad \varepsilon 0, \quad \varepsilon \in R \quad i=1,\ldots,m$$

entonces ahora vamos a calcular para los puntos $(2,2),(3,1)$ y $(0,3)$ entonces aplicando para cada uno teneos que obtener casa variable de holgura. sin embargo para este ejemplo en particular τ es $b=\frac{3}{2}$ asi que probamos en nuestro borde $x_1 - x_2 + b$, como sigue:

$$\begin{aligned} & p_1 : (2,2) \\ & H_1 : x_1 - x_1 + b = 0 \\ \text{reemplazando} \\ & (1)(2) - (1)(2) - \frac{3}{\sqrt{2}} \geq 1-\varepsilon \\ & -\frac{3}{\sqrt{2}} \geq 1-\varepsilon \\ & {}_1 = 3.11 \end{aligned}$$

asi este punto tiene una desviacion y grado de separabilidad de 3.11 , ahora para el siguiente punto .

$$\begin{aligned} & p_1 : (3,1) \\ & H_1 : x_1 - x_1 + b = 0 \\ \text{reemplazando} \\ & (1)(3) - (1)(1) - \frac{3}{\sqrt{2}} \geq 1-\varepsilon \\ & -\frac{3}{\sqrt{2}} \geq 1-\varepsilon \\ & {}_1 = 2.11 \end{aligned}$$

de igual modo para todos los el otro caso :

$$\begin{aligned} & p_1 : (0,3) \\ & H_1 : x_1 - x_1 + b = 0 \\ \text{reemplazando} \\ & (1)(0) - (1)(3) + \frac{3}{\sqrt{2}} \geq 1-\varepsilon \\ & -\frac{3}{\sqrt{2}} \geq 1-\varepsilon \\ & {}_1 = 1.89 \end{aligned}$$

asi tenemos las variables de holgura 3.11,2.11 y 1.89

2 Preguntas de Investigación

1. Explique el significado de la constante C en el término $C\sum_{i=1} i$ 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.

Solución,

la constante C es denominado una constante que controla el grado en que el costo de ejemplos no separables incluye en la minimización de la norma.

esta constante es muy importante en el hallasgo de los vectores de soporte por que ayuda en minimizar la variable de holgura por lo tanto tambien ayuda en encontrar la mínima distancia, con lo que podemos hallar de mejor forma los vectores de soporte incluyendo esta constante dentro de nuesta ecuación.

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.

   Solución

   a γ se le denomina parametro de generalizacion que nos indica que tan cercano o distante esta xi de xj, por lo que podemos saber la similiradidad entre estos mas formalmente podemos decir que un gamma pequeño significa un gaussiano con una gran variación, por lo que la influencia de x$_j esmayor, esdecir, six_j esunvectordesoporte, ungammapequeñoimplicaquelaclasedeestevectordesoportetendráinflue$

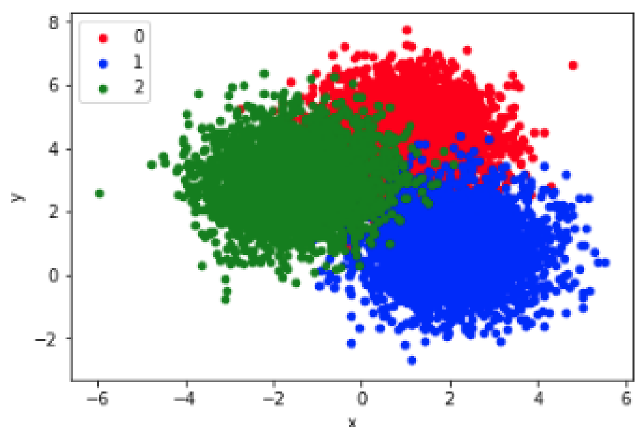
3 Implementación

La implementación será enviado a su correo en jupyter notebook para mas detalles de la implementacion y los paso seguidos, asi mismo se muestra algunos alcances en este documento.

1. Usando el Scikit-learn de Python, implemente (comente su código) una SVM que clasifique el conjunto de datos ¡por definir¿

   Para la clasificación se genero datos aleatorios con caractaristicas de tres clases diferentes asi tenemos una imagen de la primera parte de la clasificacion acontinuacion:

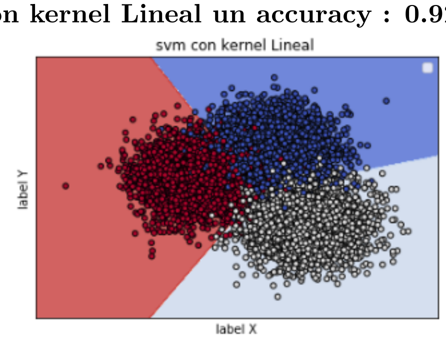
```python
#ahora vamos a generar nuestros datos con blobs de sklearn
from matplotlib import pyplot
from pandas import DataFrame
from sklearn.datasets.samples_generator import make_blobs

# Generamos clasificador 2D del dataset
X, y = make_blobs(n_samples=10000, centers=3, n_features=2, random_state=0)

# plotear puntos con su valor de color
df = DataFrame(dict(x=X[:,0], y=X[:,1], label=y))

colors = {0:'red', 1:'blue', 2:'green'}
fig, ax = pyplot.subplots()
grouped = df.groupby('label')
for key, group in grouped:
    group.plot(ax=ax, kind='scatter', x='x', y='y', label=key, color=colors[key])
pyplot.show()
```



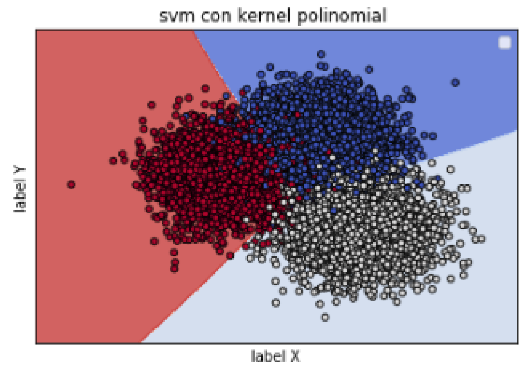
2. Experimente y muestre resultados usando diferentes valores para los parámetros de los kernels: lineal, polinomial, gaussiano, y el parámetro C.

   al implementar el desarrollo de los diferentes kernels se obtuvo los siguientes resultados de accuracy

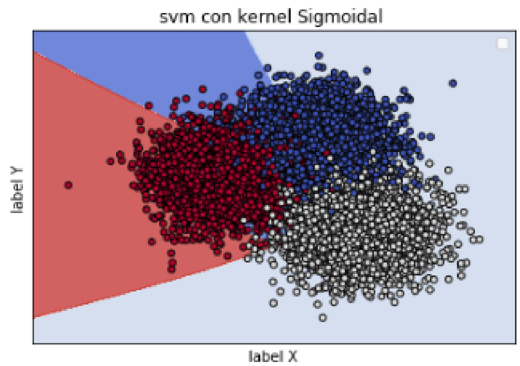
para SVM con kernel Lineal un accuracy : 0.92166666666



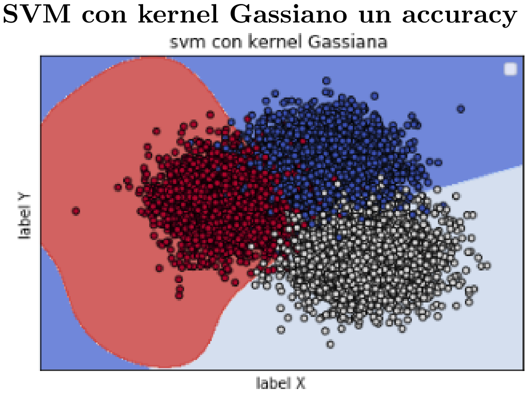
para SVM con kernel Polinomial un accuracy : 0.9183333333333333



para SVM con kernel sigmoide un accuracy : 0.472



para SVM con kernel Gassiano un accuracy : 0.924



3. Dentro de la sección de Implementación incluya una subsección donde indique las instrucciones para ejecutar el código.