

# Máy học véc-tơ hỗ trợ (SVM)

## Cài đặt với thư viện tối ưu

27/2/2021

# Đối tượng

- Sinh viên Khoa học máy tính

# Mục tiêu

- Hiểu được bài toán SVM và đưa nó về bài toán quy hoạch toàn phương
- Áp dụng gói thư viện để huấn luyện mô hình máy học SVM

# Tài liệu

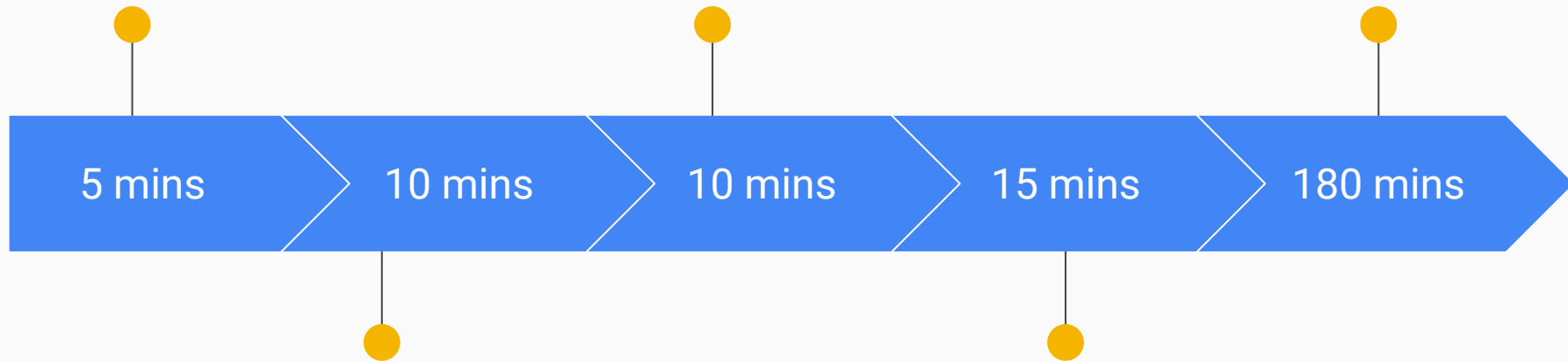
- Slide này
- Slides 2 - SVM
- Giáo trình nguyên lý máy học

# Kế hoạch

Nhắc lại bài toán SVM

Tìm hiểu thư viện tối  
ưu

Bài tập



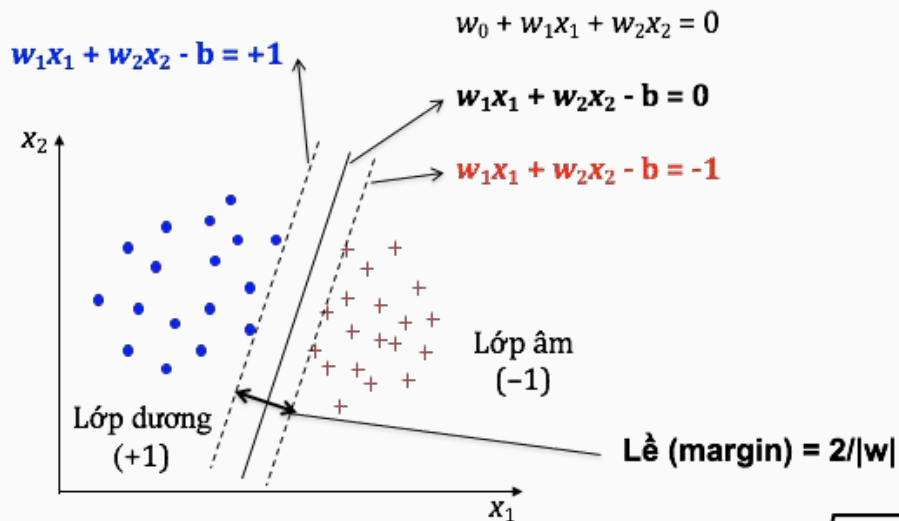
Đưa bài toán SVM về  
bài toán Quy hoạch  
toàn phương

Cài đặt SVM bằng thư  
viện tối ưu

# Bài toán SVM

Tìm **w** và **b** sao cho:

- Mỗi lớp nằm về một phía của 2 đường thẳng hỗ trợ
- Lệ lớn nhất



Các phần tử càng cách xa đường biên càng tốt ~ Lệ càng lớn càng tốt.

$$|w| = \sqrt{w_1^2 + w_2^2}$$

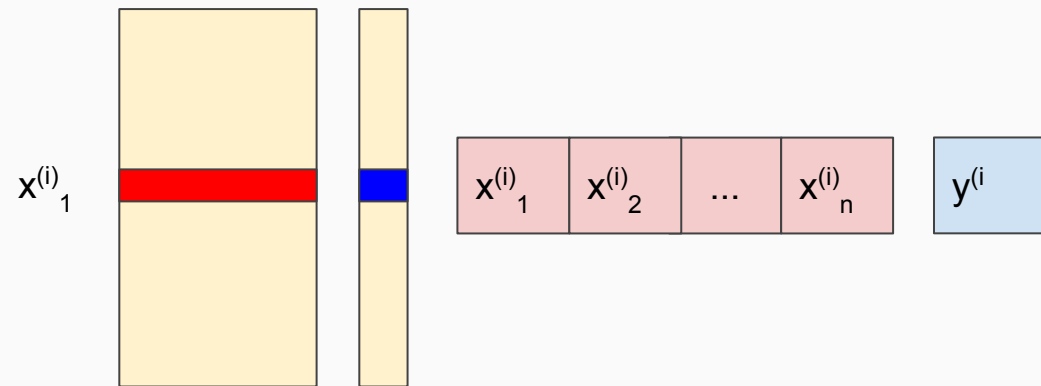
# Bài toán SVM

- Với mỗi phần tử thuộc lớp dương:  $y^{(i)} = 1$   
 $-w_1x_1^{(i)} + w_2x_2^{(i)} - b \geq +1$
- Lớp âm:  $y^{(i)} = -1$   
 $-w_1x_1 + w_2x_2 - b \leq -1$
- Nhân 2 vế cho  $y^{(i)}$ :  
 $-y^{(i)}(w_1x_1^{(i)} + w_2x_2^{(i)} - b) \geq +1$
- Bài toán trở thành tìm  $\mathbf{w} = (w_1, w_2)$  và  $b$  sao cho:  
 $2/\|\mathbf{w}\| \rightarrow \max$  (hay  $\|\mathbf{w}\|^2 \rightarrow \min$ )  
với ràng buộc:  $y^{(i)}(w_1x_1^{(i)} + w_2x_2^{(i)} - b) \geq +1$



# Bài toán SVM

Biểu diễn dưới dạng ma trận



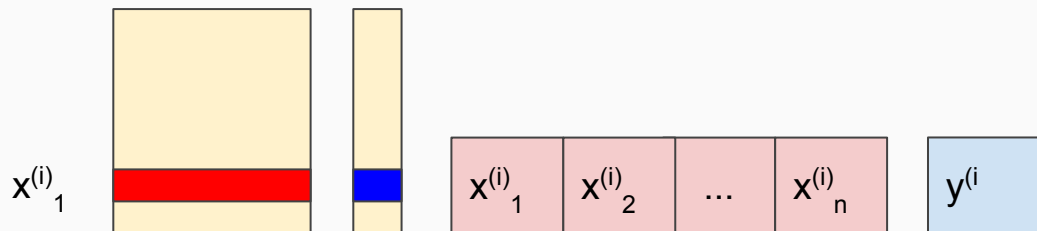
**Dữ liệu huấn luyện**

Dữ liệu huấn luyện là một ma trận có  $m$  hàng và  $n$  cột

Mỗi hàng là một phần tử huấn luyện, mỗi phần tử huấn luyện có  $n$  thuộc tính ( $n$  chiều)

# Bài toán SVM

Biểu diễn dưới dạng ma trận



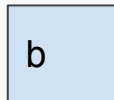
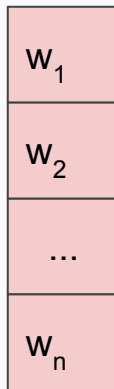
Dữ liệu huấn luyện

Véc-tơ pháp tuyến  $w$ :

- Là một véc-tơ có  $n$  phần tử tương ứng với  $n$  thuộc tính của dữ liệu huấn luyện.

Độ lệch  $b$ :

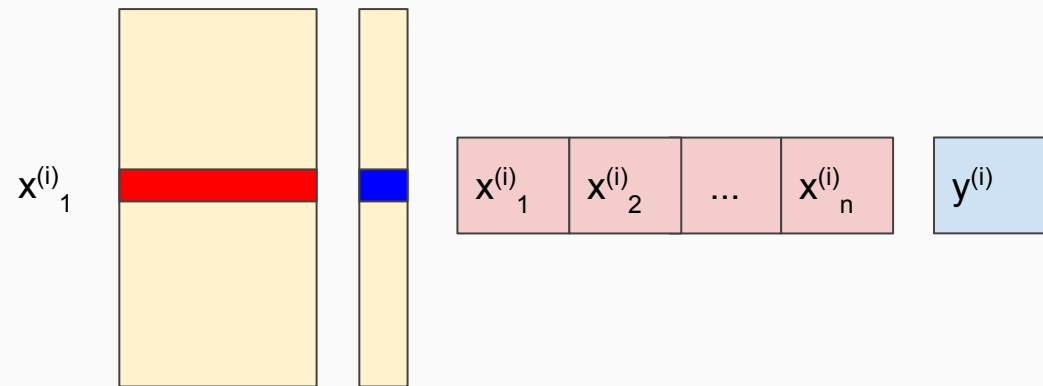
- Là một số thực.



$w$  và  $b$

# Bài toán SVM

Biểu diễn dưới dạng ma trận



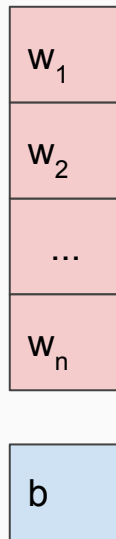
Dữ liệu huấn luyện

Các ràng buộc

$$y^{(i)}(w_1 x^{(i)}_1 + w_2 x^{(i)}_2 - b) \geq +1$$

Lấy từng hàng của dữ liệu huấn luyện đem nhân với  $w$ .

Có  $m$  phần tử  $\Rightarrow$  có  $m$  ràng buộc.



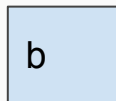
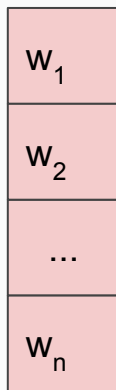
$w$  và  $b$

# Bài toán SVM



Giả sử ta có tập huấn luyện:

x1	x2	y
2	2	+1
3	1	+1
1	1	-1



**Các ràng buộc:**

$$y^{(i)}(w_1 x^{(i)}_1 + w_2 x^{(i)}_2 - b) \geq +1$$

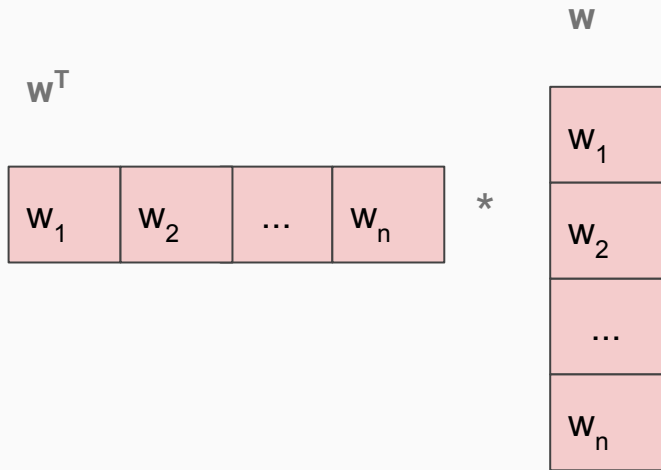
$$+1 \cdot (2w_1 + 2w_2 - b) \geq 1$$

$$+1 \cdot (3w_1 + 1w_2 - b) \geq 1$$

$$-1 \cdot (1w_1 + 1w_2 - b) \geq 1$$

**Chú ý các hệ số trong hệ bất phương trình này, nó từ đâu tới ?**

# Bài toán SVM



**Hàm mục tiêu:**

$$|w|^2 \rightarrow \min$$

$$|w|^2 = w_1 \cdot w_1 + w_2 \cdot w_2 + \dots = w^T \cdot w$$

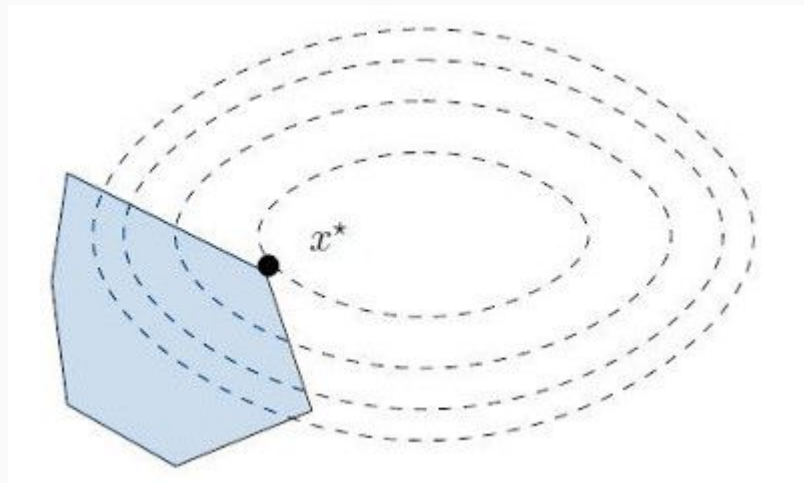
# Bài toán quy hoạch toàn phương

Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

x: véc-tơ các biến ta cần phải tìm sao cho hàm mục tiêu tối ưu.



# Đưa bài toán SVM về dạng chuẩn toàn phương

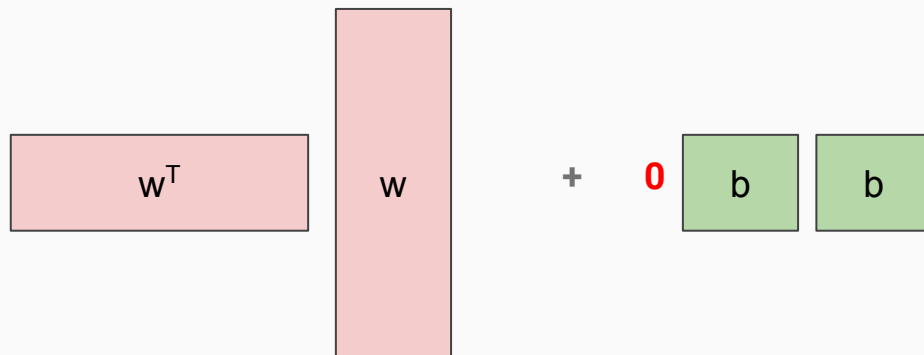
Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

**Hàm mục tiêu:**  $|w|^2 \rightarrow \min$

$$|w|^2 = w^T \cdot w = w^T \cdot w + 0 \cdot b \cdot b$$



# Đưa bài toán SVM về dạng chuẩn toàn phương

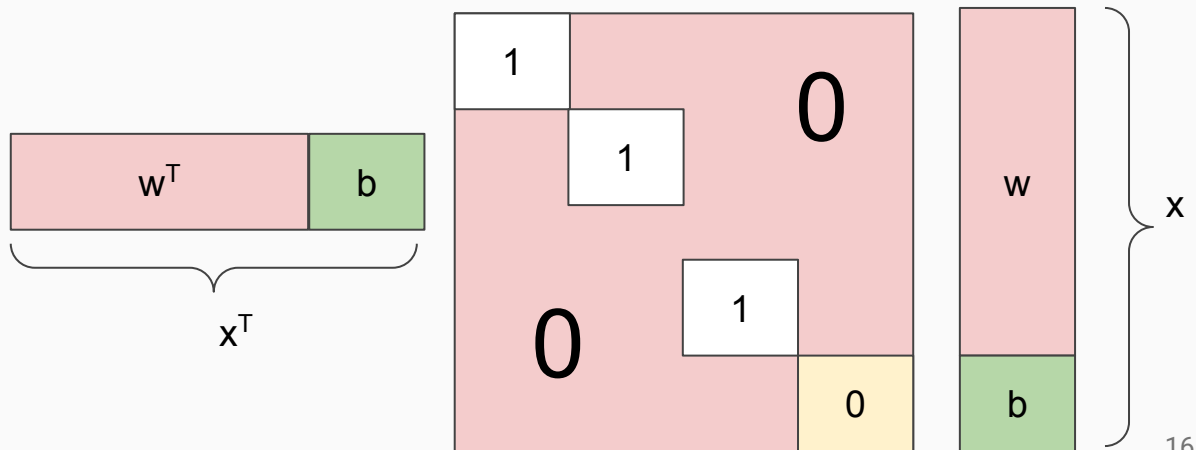
Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T P x + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

Hàm mục tiêu:  $|w|^2 \rightarrow \min$

$$|w|^2 = w^T \cdot w = w^T \cdot w + 0 \cdot b \cdot b$$





# Đưa bài toán SVM về dạng chuẩn toàn phương

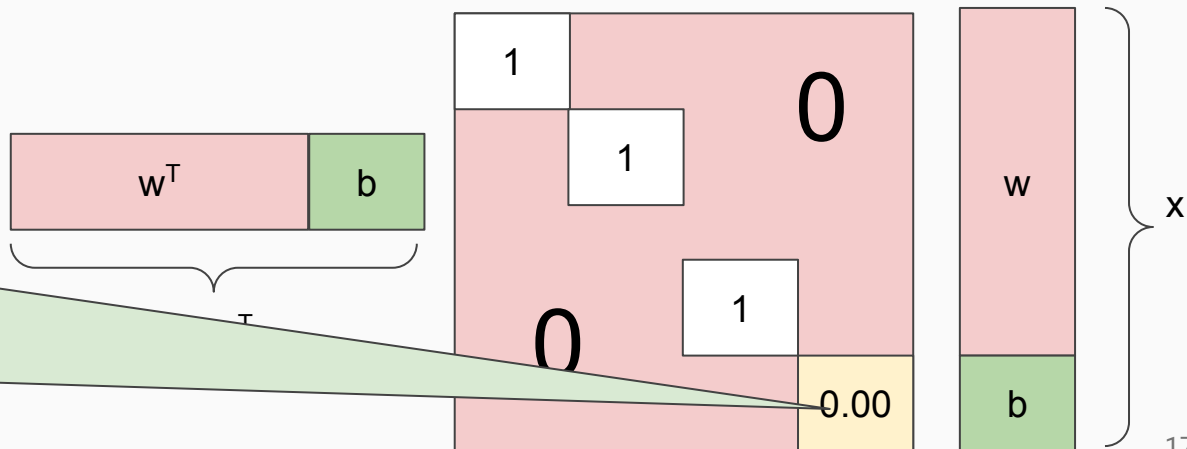
Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\text{minimize } (1/2)x^T Px + q^T x$$

Trong cài đặt  
KHÔNG NÊN đặt giá  
trị này = **0** mà NÊN  
gán cho nó 1 giá trị  
đủ vừa nhỏ, vd:  
**0.000001**

Hàm mục tiêu:  $|w|^2 \rightarrow \min$

$$|w|^2 = w^T \cdot w = w^T \cdot w + 0.b.b$$



# Đưa bài toán SVM về dạng chuẩn toàn phương

Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

Như thế, ta có

**P =**

1			
	1		
		1	
			0.00

# Đưa bài toán SVM về dạng chuẩn toàn phương

Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

Như thế, ta có

q =

0
0
0

# Đưa bài toán SVM về dạng chuẩn toàn phương

Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

## Các ràng buộc:

$$y^{(i)}(w_1 x^{(i)}_1 + w_2 x^{(i)}_2 - b) \geq +1$$

$$+1 \cdot (2w_1 + 2w_2 - b) \geq 1$$

$$+1 \cdot (3w_1 + 1w_2 - b) \geq 1$$

$$-1 \cdot (1w_1 + 1w_2 - b) \geq 1$$

Nhân 2 vế bất phương trình cho -1, ta được

$$-y^{(i)}(w_1 x^{(i)}_1 + w_2 x^{(i)}_2 - b) \leq -1$$

# Đưa bài toán SVM về dạng chuẩn toàn phương

Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

## Các ràng buộc:

$$-y^{(i)}(w_1 x^{(i)}_1 + w_2 x^{(i)}_2 - b) \leq -1$$

$$-2w_1 - 2w_2 + b \leq -1$$

$$-3w_1 - 1w_2 + b \leq -1$$

$$1w_1 + 1w_2 - b \leq -1$$

Viết dưới dạng trận:

$$Gx \leq h$$

# Đưa bài toán SVM về dạng chuẩn toàn phương

Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

minimize  
subject to

**Tập huấn luyện:**

x1	x2	y
2	2	+1
3	1	+1
1	1	-1

- P: ma tr
- q: véc-tơ
- G, A: ma
- h, b: véc

**Các ràng buộc:**

$$-y^{(i)}(w_1x^{(i)}_1 + w_2x^{(i)}_2 - b) \leq -1$$

$$-2w_1 - 2w_2 + b \leq -1$$

$$-3w_1 - 1w_2 + b \leq -1$$

$$1w_1 + 1w_2 - b \leq -1$$

Viết dưới dạng trận:

$$Gx \leq h$$

# Đưa bài toán SVM về dạng chuẩn toàn phương

Tập huấn luyện:

x1	x2	y
2	2	+1
3	1	+1
1	1	-1

Các ràng buộc:

$$-y^{(i)}(w_1x^{(i)}_1 + w_2x^{(i)}_2 - b) \leq -1 \quad \mathbf{G} = ?$$

$$-2w_1 - 2w_2 + b \leq -1$$

$$-3w_1 - 1w_2 + b \leq -1$$

$$1w_1 + 1w_2 - b \leq -1$$

Viết dưới dạng trận:

$$\mathbf{Gx} \leq \mathbf{h}$$

Các ràng buộc:

- Thêm 1 cột cuối vào tập huấn luyện, chứa toàn giá trị -1
- Nhân từng phần tử mỗi hàng với  $-y^{(i)}$

$$\mathbf{h} = -1 \text{ (vector)}$$

# Đưa bài toán SVM về dạng chuẩn toàn phương

Tập huấn luyện:

x1	x2	y
2	2	+1
3	1	+1
1	1	-1

Tập huấn luyện:

x1	x2	Mới	y
2	2	-1	+1
3	1	-1	+1
1	1	-1	-1

Tập huấn luyện:

x1	x2	Mới	y
-2	-2	+1	+1
-3	-1	+1	+1
1	1	-1	-1

- Thêm 1 cột cuối vào tập huấn luyện, chứa toàn giá trị -1
- Nhân từng phần tử mỗi hàng với  $-y^{(i)}$

G



# Đưa bài toán SVM về dạng chuẩn toàn phương

Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

**Các ràng buộc:**

$$\mathbf{A} = \mathbf{0}$$

$$\mathbf{b} = \mathbf{0}$$

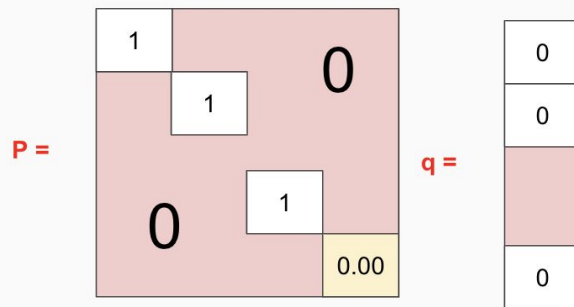
# Đưa bài toán SVM về dạng chuẩn toàn phương

Dạng chuẩn của bài toán Quy hoạch toàn phương có dạng như sau:

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- P: ma trận vuông
- q: véc-tơ
- G, A: ma trận ràng buộc
- h, b: véc-tơ

**Tóm lại:**



- **G = Ma trận huấn luyện, thêm cột, nhân với -y**
- **h = -1 (vector)**
- **A = 0, b = 0**

# Các thư viện

Bài toán Quy hoạch toàn phương có dạng chuẩn

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

Có thể giải bằng nhiều thư viện khác nhau. Mỗi thư viện lại có cách biểu diễn khác 1 chút, bạn cần phải chuyển bài toán dạng chuẩn về đúng dạng của thư viện.

Các thư viện giải bài toán này: cvxopt, quadprog, ...

Để cài đặt, từ dòng lệnh gõ: **pip install <tên thư viện>**

# Thư viện Quadprog

Thư viện quadprog biểu diễn bài toán hơi khác tí so với bài toán chuẩn:

$$\frac{1}{2}x^T Gx - a^T x$$

Ràng buộc:

$$C^T x \geq b$$

$$\begin{array}{ll} \text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Gx \leq h \\ & Ax = b \end{array}$$

- $G = P$
- $a = -q$
- $C^T = -G$  hay  $C = -G^T$
- $b = -h$  (vector **1**)

# Thư viện Quadprog

$$\frac{1}{2}x^T Gx - a^T x$$

$$C^T x \geq b$$

**Xây dựng trực tiếp từ dữ liệu huấn luyện:**

$$y^{(i)}(w_1 x^{(i)}_1 + w_2 x^{(i)}_2 - b) \geq +1$$

$G =$

1			0
	1		
		1	
0			0.00

- $a = 0$  (vector)
- Xây dựng  $C^T$  từ ma trận dữ liệu huấn luyện
- $b = y$  (vector)

**Tập huấn luyện:**

x1	x2	y
2	2	+1
3	1	+1
1	1	-1

$C^T$

2	2	-1
3	1	-1
-1	-1	1

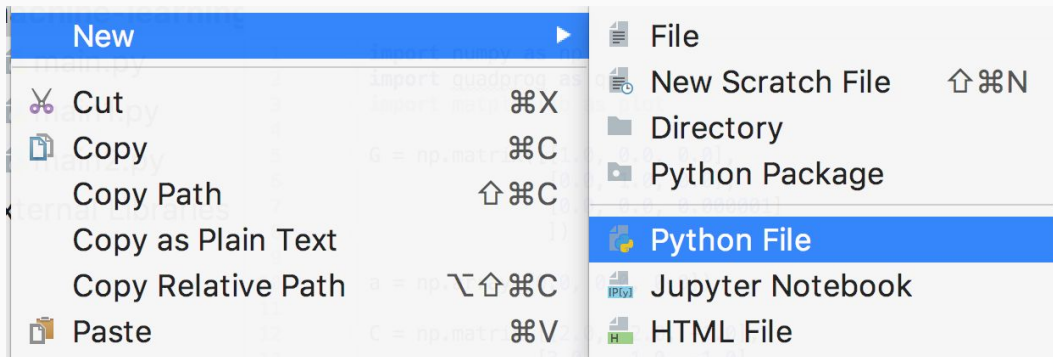
# Cài đặt

1. Cài đặt python 3.x
2. Cài đặt các thư viện: numpy, quadprog, matplotlib
3. Cài đặt PyCharm

# Cài đặt

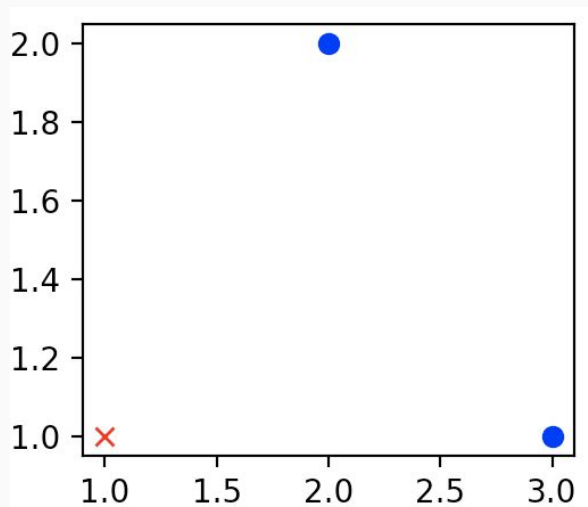
Tạo project mới, đặt tên

Thêm 1 file python



# Cài đặt

Dữ liệu: dương (xanh), âm (đỏ)



```
1 import numpy as np
2 import quadprog as qp
3 import matplotlib.pyplot as plot
4
5 #Dữ liệu
6 X = np.matrix([[2, 2],
7                [3, 1],
8                [1, 1]])
9
10 y = np.array([1, 1, -1])
11
12
13 # Vẽ dữ liệu lên hệ trục tọa độ 0x1x2
14 plot.plot(X[0:2, 0], X[0:2, 1], 'bo')
15 plot.plot(X[2:3, 0], X[2:3, 1], 'rx')
16 plot.show()
```



# Cài đặt

Xây dựng các ma trận cần thiết

```
19      # Ma trận G
20      G = np.matrix([[1.0, 0.0, 0.0],
21                      [0.0, 1.0, 0.0],
22                      [0.0, 0.0, 0.000001]
23                      ])
24
25      a = np.array([0.0, 0.0, 0.0])
26
27      # Xây dựng ma trận C
28      C = np.matrix([[2.0, 2.0, -1.0],
29                      [3.0, 1.0, -1.0],
30                      [-1.0, -1.0, 1.0]]).T
31
32      b = np.array([1.0, 1.0, 1.0])
```

# Cài đặt

Giải bài toán QP

Nếu làm đúng, kết quả sẽ là:

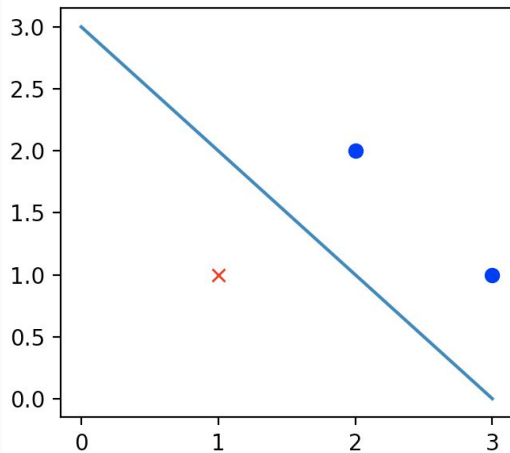
```
[1. 1. 3.]
```

Có nghĩa là  $w_1 = 1$ ,  $w_2 = 1$  và  $b = 3$

```
34 # Giải bài toán quy hoạch toàn phương  
35 sol = qp.solve_qp(G, a, C, b)  
36  
37 # Kết quả (w, b) được lưu trong sol[0]  
38 wb = sol[0]  
39  
40 print(wb)
```

# Cài đặt

## Vẽ kết quả



```
44 # Vẽ dữ liệu lên hệ trục tọa độ 0x1x2
45 plot.plot(X[0:2, 0], X[0:2, 1], 'bo')
46 plot.plot(X[2:3, 0], X[2:3, 1], 'rx')
47
48 x1 = np.array([0, 3])          #x1 từ 1 đến 3
49 x2 = (wb[2] - wb[0]*x1)/wb[1] # x2 tính theo phương trình
50 plot.plot(x1, x2)
51
52 plot.show()
```

Dừng lại chút và  
lập trình chạy thử đi !

# Homework

## Bài tập 1 (**hạn cuối: 11h00 thứ 7 ngày 6/3/2021**)

Báo cáo lại những thứ bạn đã học được qua slide này (có thể đọc thêm tài liệu khác) bằng cách soạn một tài liệu Word/Google Docs mô tả lại những thứ bạn đã học được bằng **ngôn ngữ của bạn (Không copy bài giảng của thầy nộp lại cho thầy)**.

Trong bài báo cáo có:

- Bổ sung các lệnh cần thiết để vẽ thêm hai đường thẳng hỗ trợ.
- Bổ sung các lệnh cần thiết để dự đoán lớp cho dữ liệu mới (0.5, 2)

# Homework

## Bài tập 2 (**hạn cuối: 11h00 thứ 7 ngày 13/3/2021**)

Viết một chương trình Python đọc dữ liệu đầu vào là dữ liệu huấn luyện được cho dưới dạng, tập tin csv; huấn luyện một mô hình SVM cho tập dữ liệu này; xuất kết quả là các tham số  $w_1, w_2, w_3, \dots, w_n$  và  $b$ .

Nộp:

- 01 báo cáo mô tả những thứ đã làm và kết quả tương ứng
- Chương trình Python

# Tham khảo

Nộp bài vào bài tập tương ứng trên classroom (nhớ đúng deadline).

Vẽ đồ thị với Mathplot: <https://matplotlib.org/tutorials/introductory/pyplot.html>

Thư viện Quadprog: <https://pypi.org/project/quadprog/>