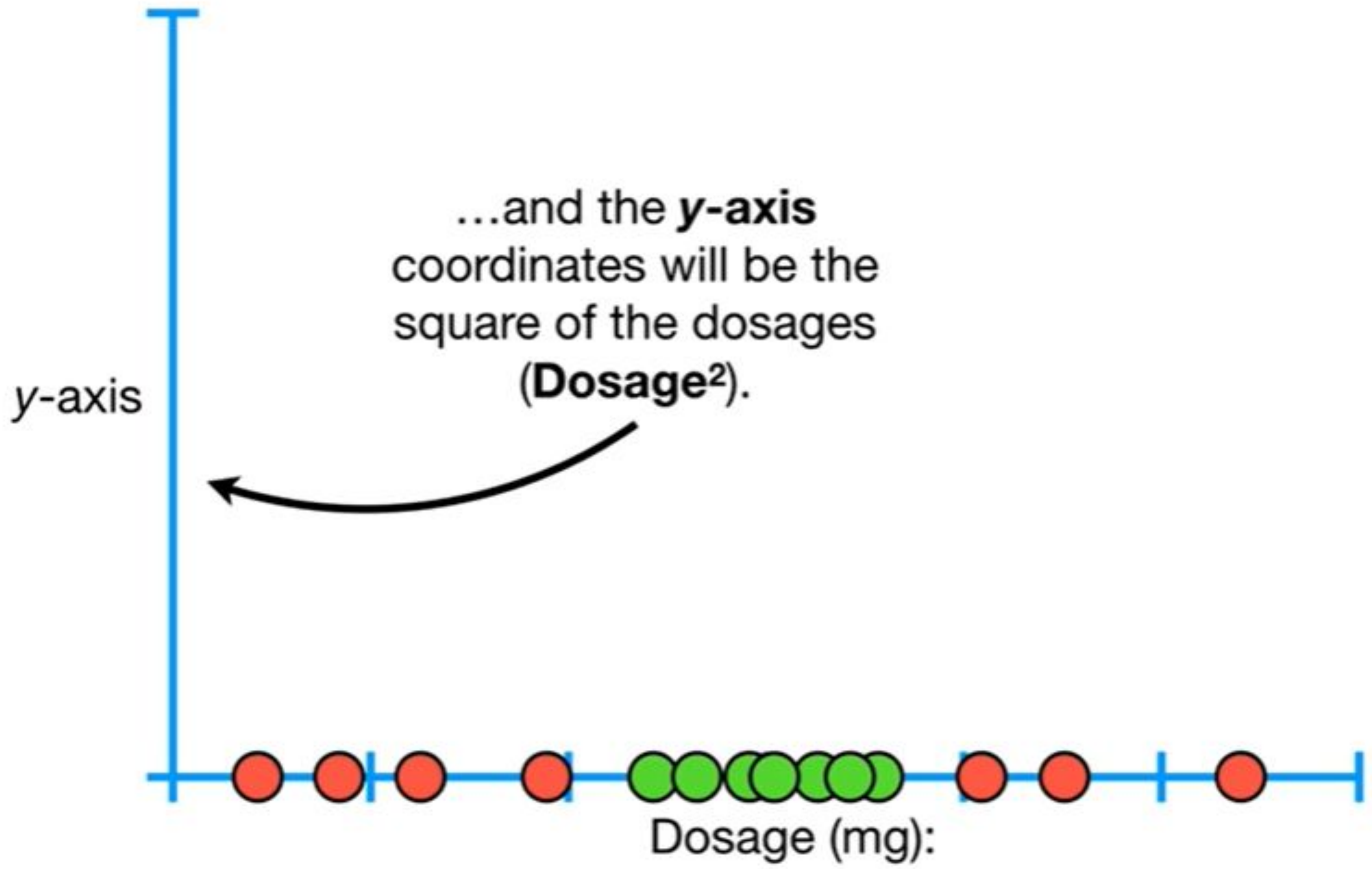
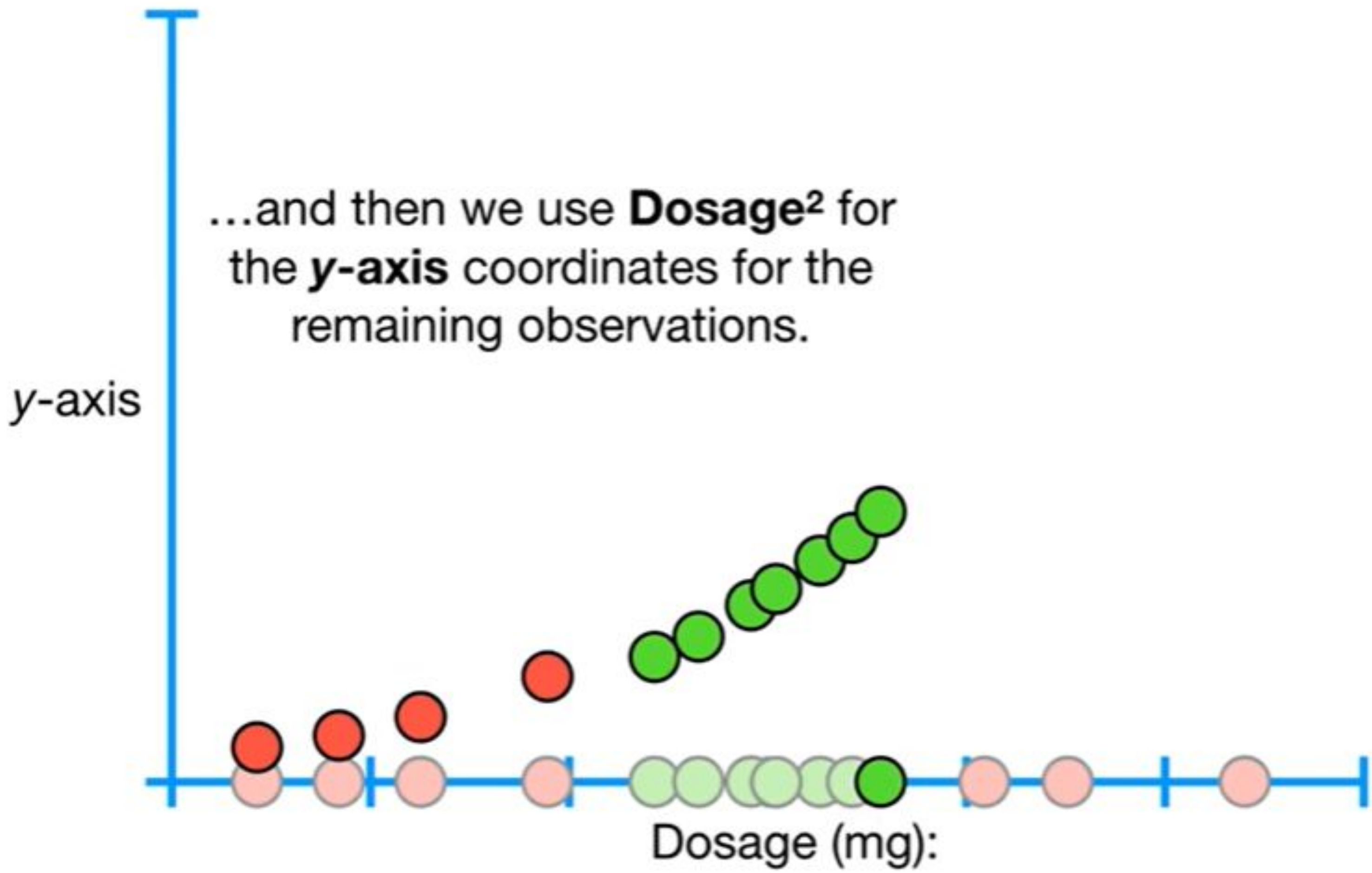
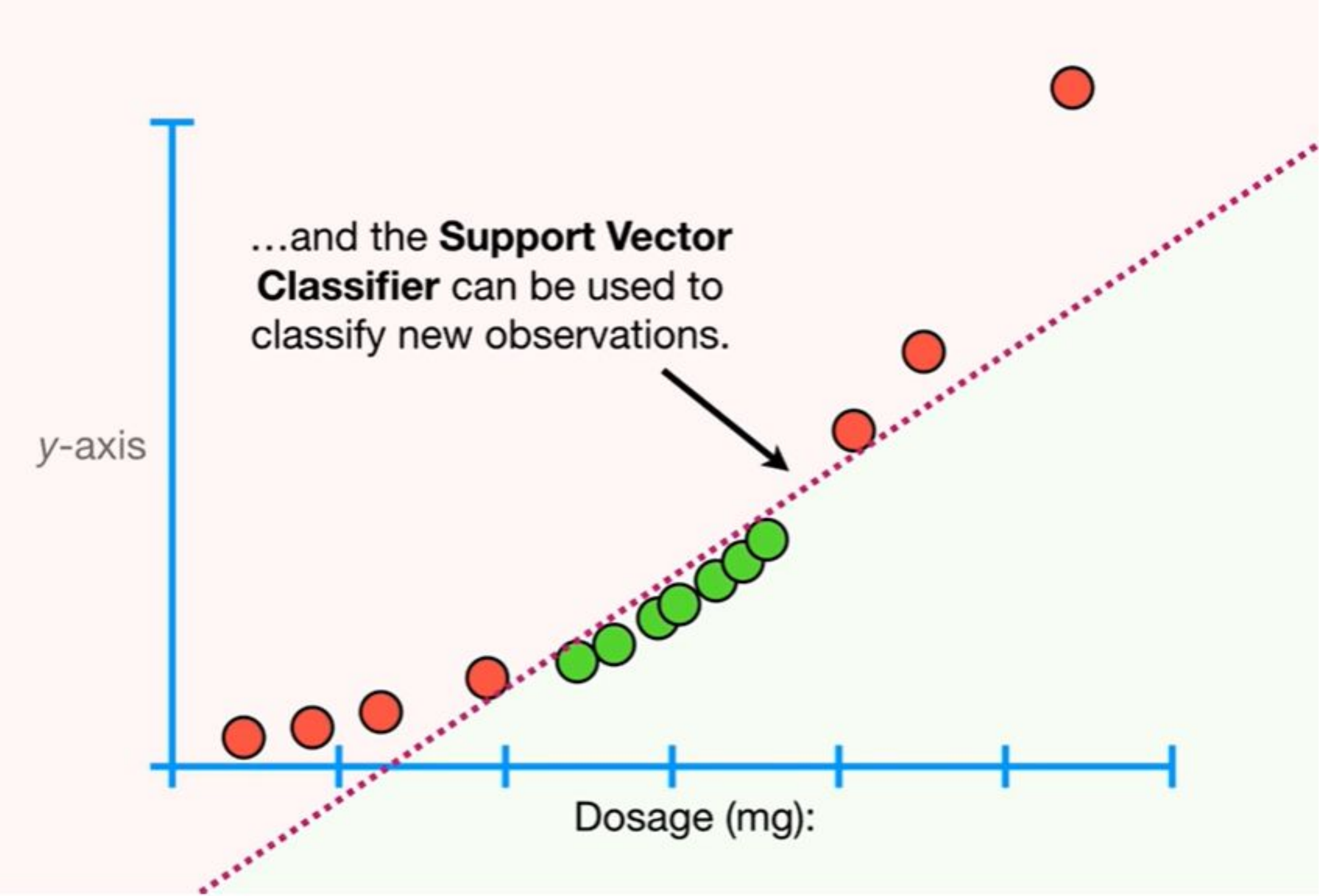
SVM

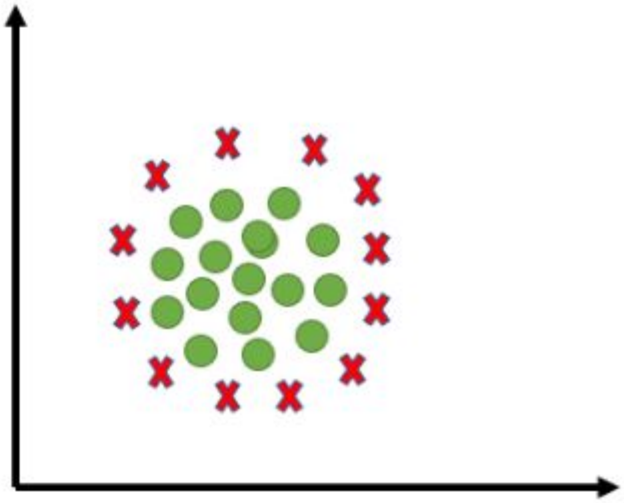
Support Vector Machine

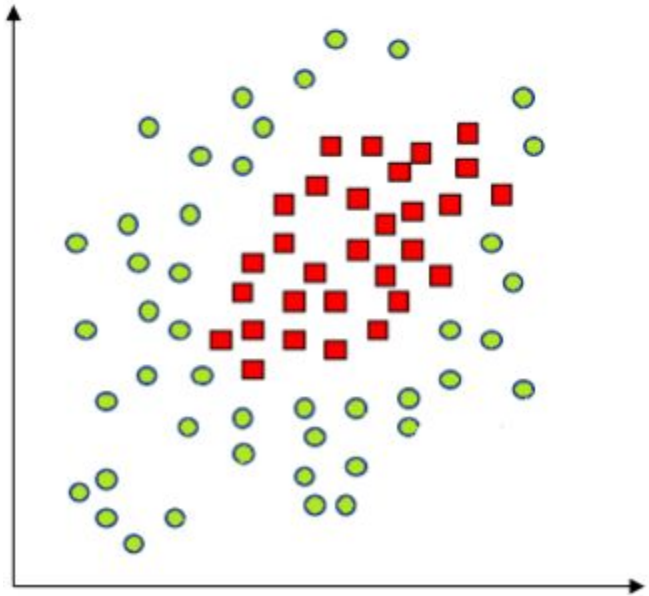

Dosage (mg):

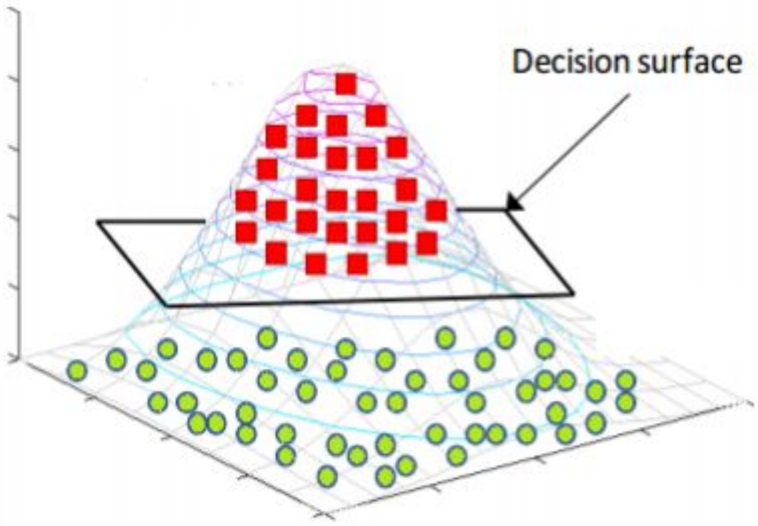

...and the **y-axis** coordinates will be the square of the dosages (**Dosage²**).
y-axis
Dosage (mg):

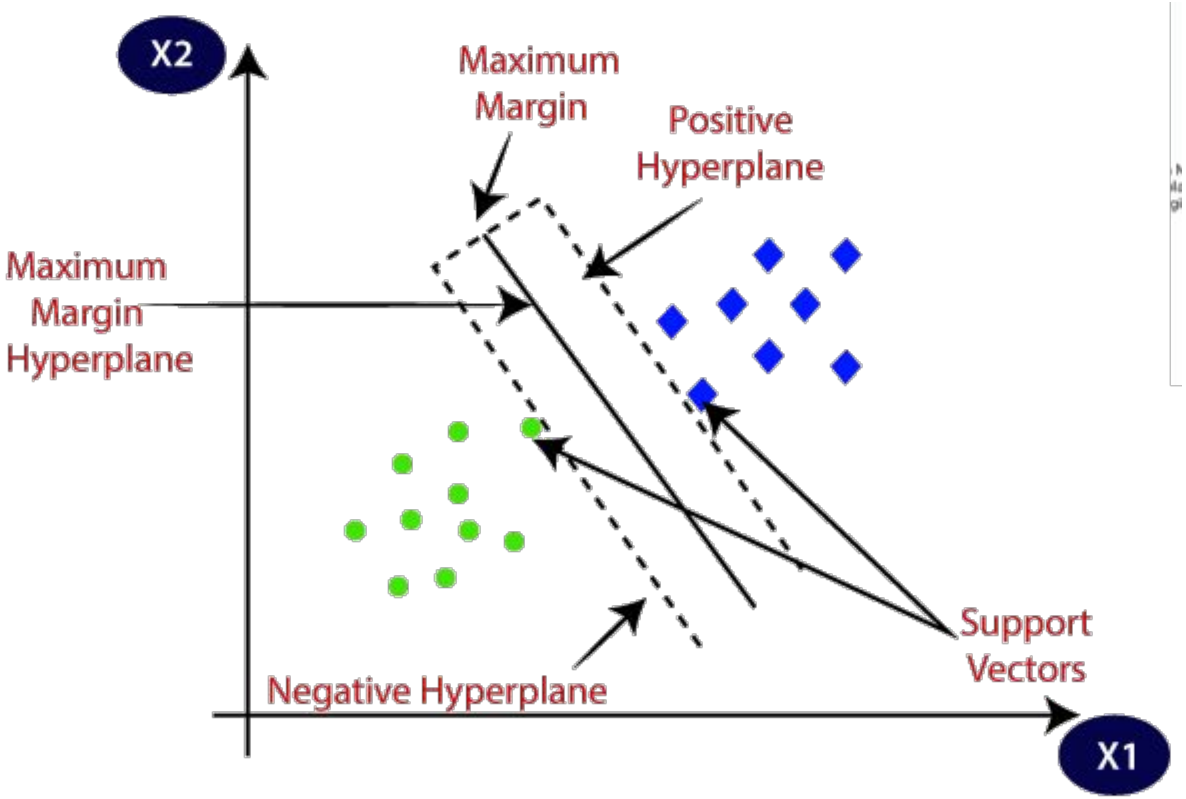

...and then we use **Dosage²** for the ***y*-axis** coordinates for the remaining observations.
y-axis
Dosage (mg):

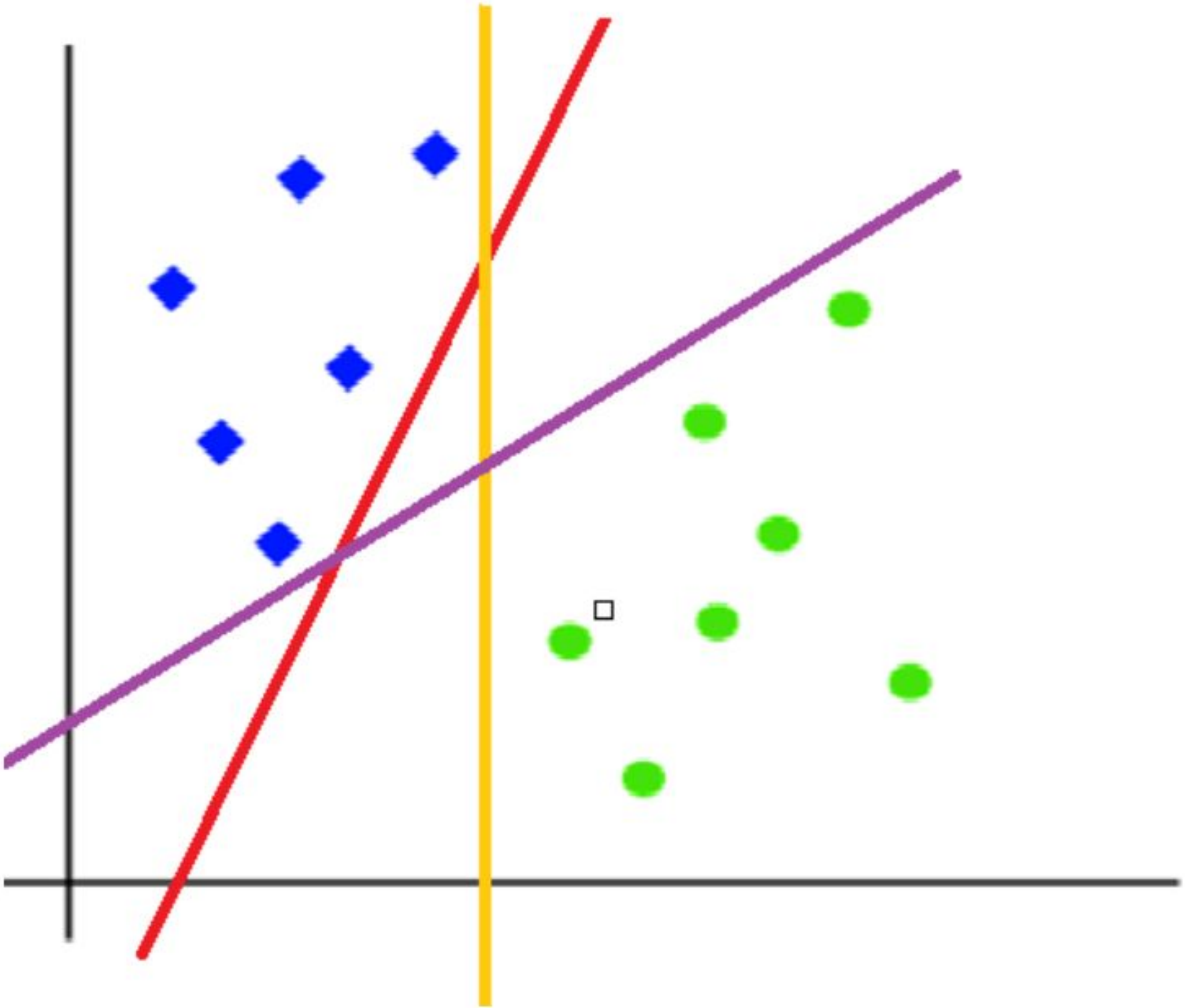

...and the **Support Vector Classifier** can be used to classify new observations.
y-axis
Dosage (mg):

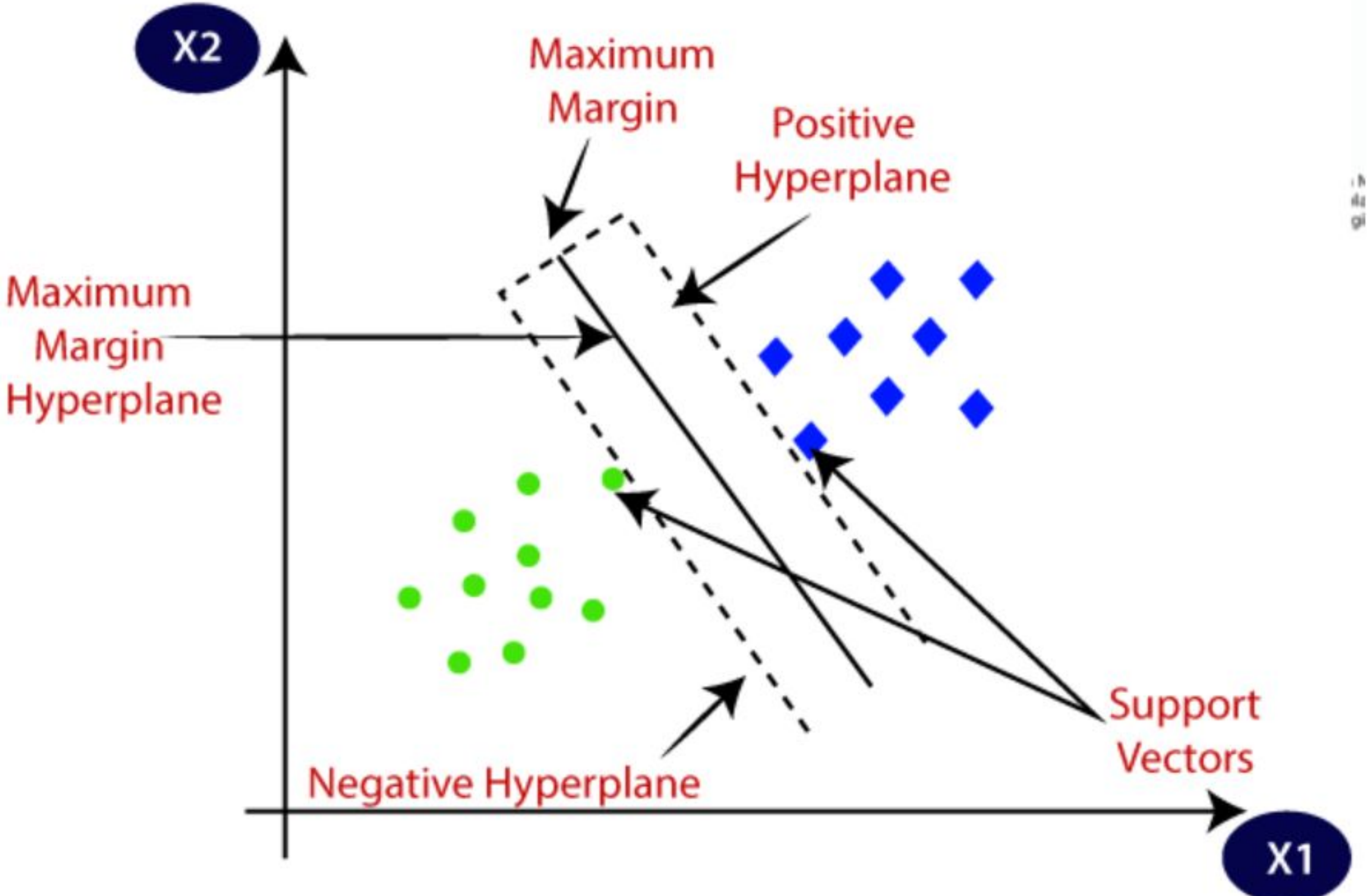




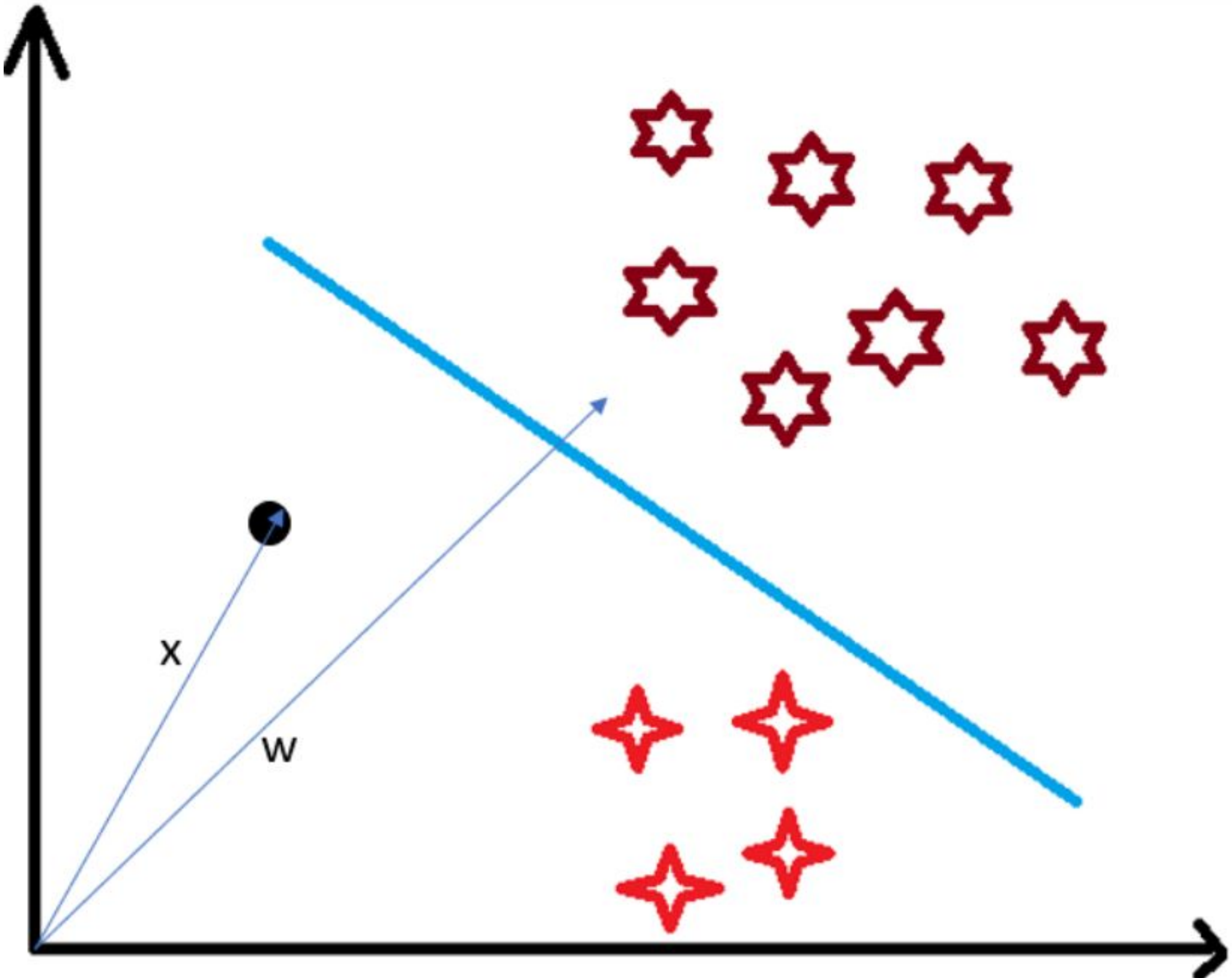

Decision surface

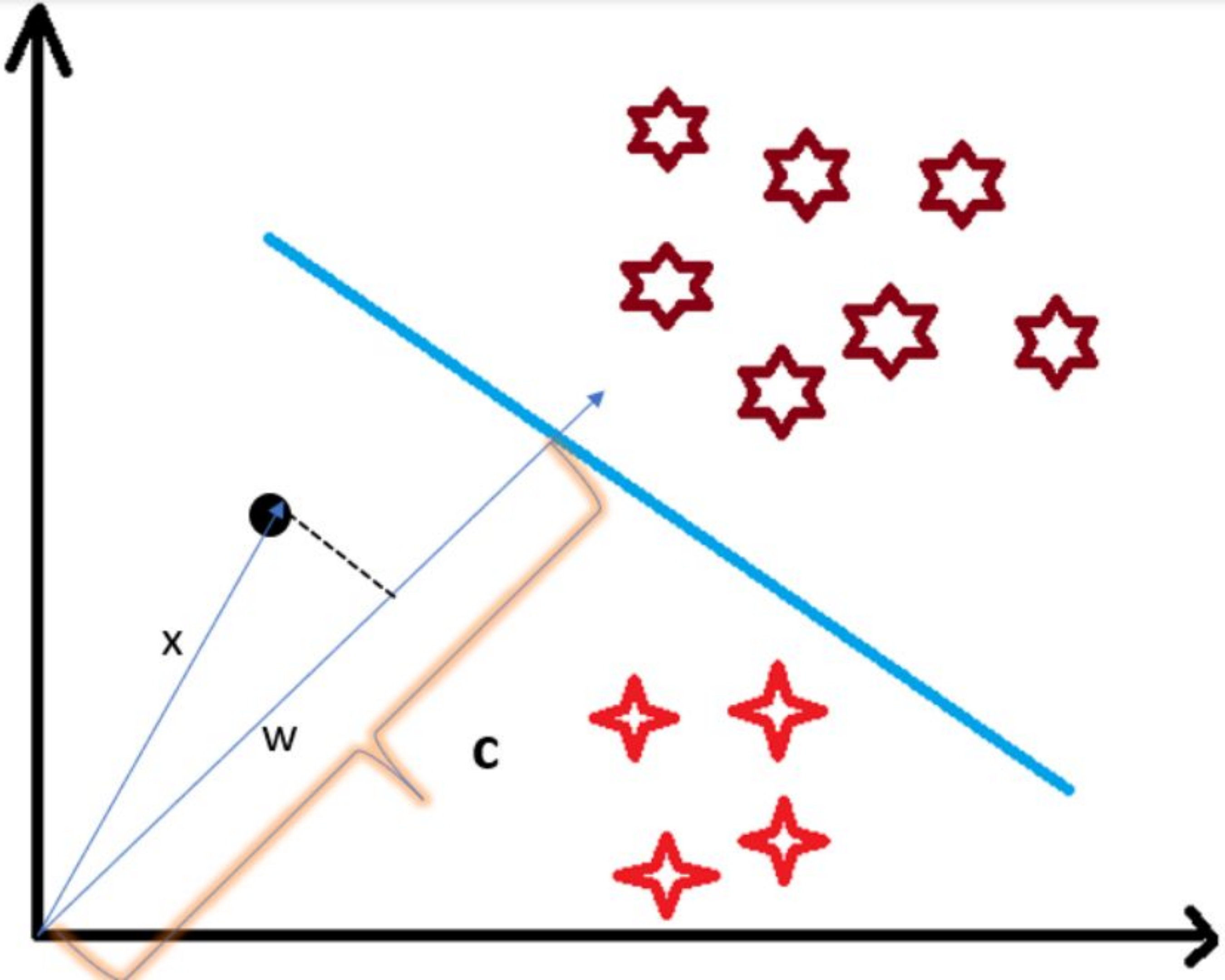

X2
Maximum
Margin
Positive
Hyperplane
Maximum
Margin
Hyperplane
Support
Vectors
Negative Hyperplane
X1




X2
Maximum
Margin
Positive
Hyperplane
Maximum
Margin
Hyperplane
Support
Vectors
Negative Hyperplane
X1

Mathematical Intuition behind SVM


x
w


X
W
C

We already know that projection of any vector or another vector is called dot-product. Hence, we take the dot product of x and w vectors. If the dot product is greater than ‘c’ then we can say that the point lies on the right side. If the dot product is less than ‘c’ then the point is on the left side and if the dot product is equal to ‘c’ then the point lies on the decision

$$\overrightarrow{X}.\overrightarrow{w} = c \quad (the\ point\ lies\ on\ the\ decision\ boundary)$$

$$\overrightarrow{X}.\overrightarrow{w} > c \ (positive\ samples)$$

$$\overrightarrow{X}.\overrightarrow{w} < c \ (negative\ samples)$$

To classify a point as negative or positive we need to define a decision rule. We can define decision rule as:

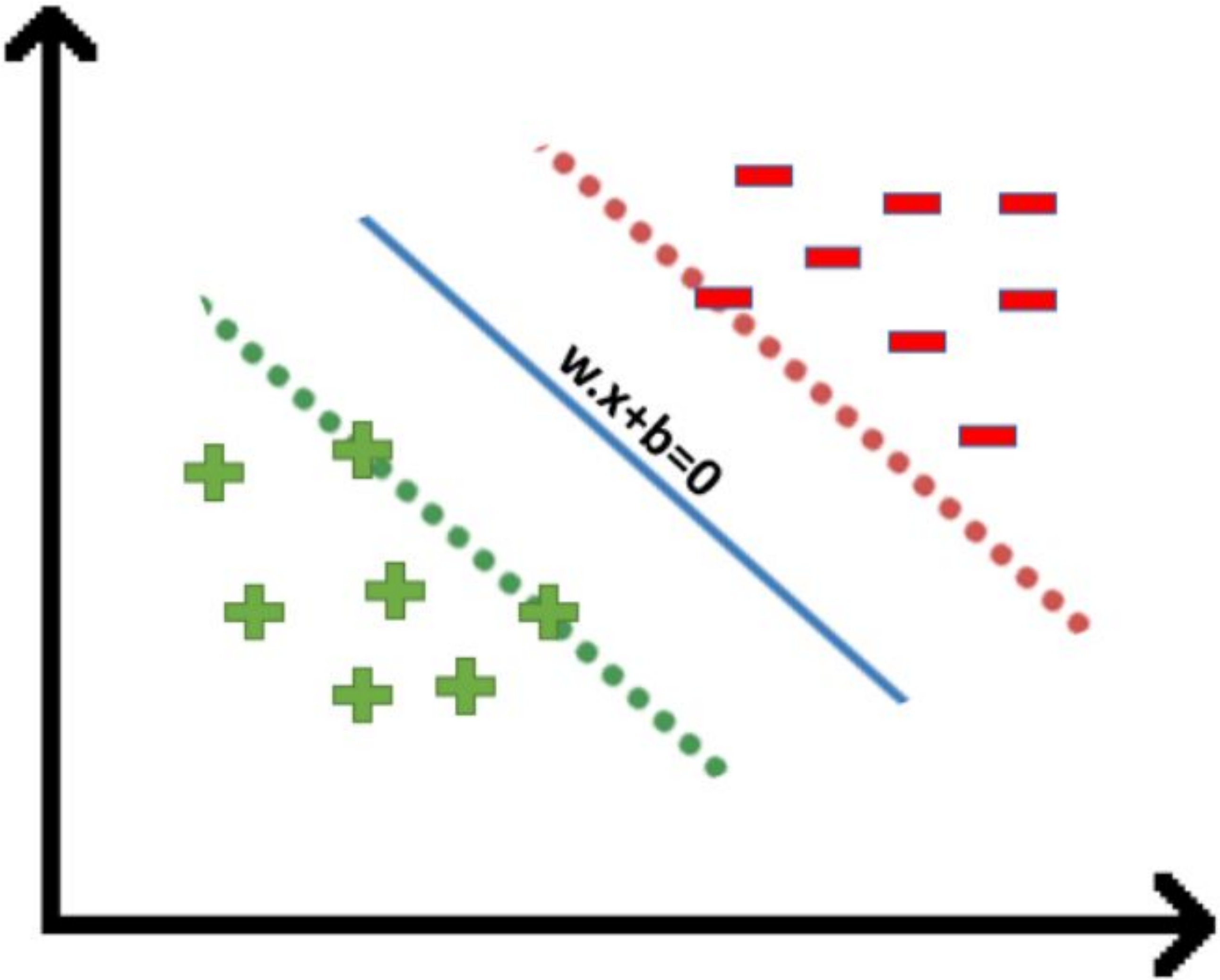
$$\overrightarrow{X}.\overrightarrow{w} - \mathrm{c} \geq 0$$

$$\text{putting } -\mathrm{c} \text{ as } \mathrm{b}, \text{we get}$$

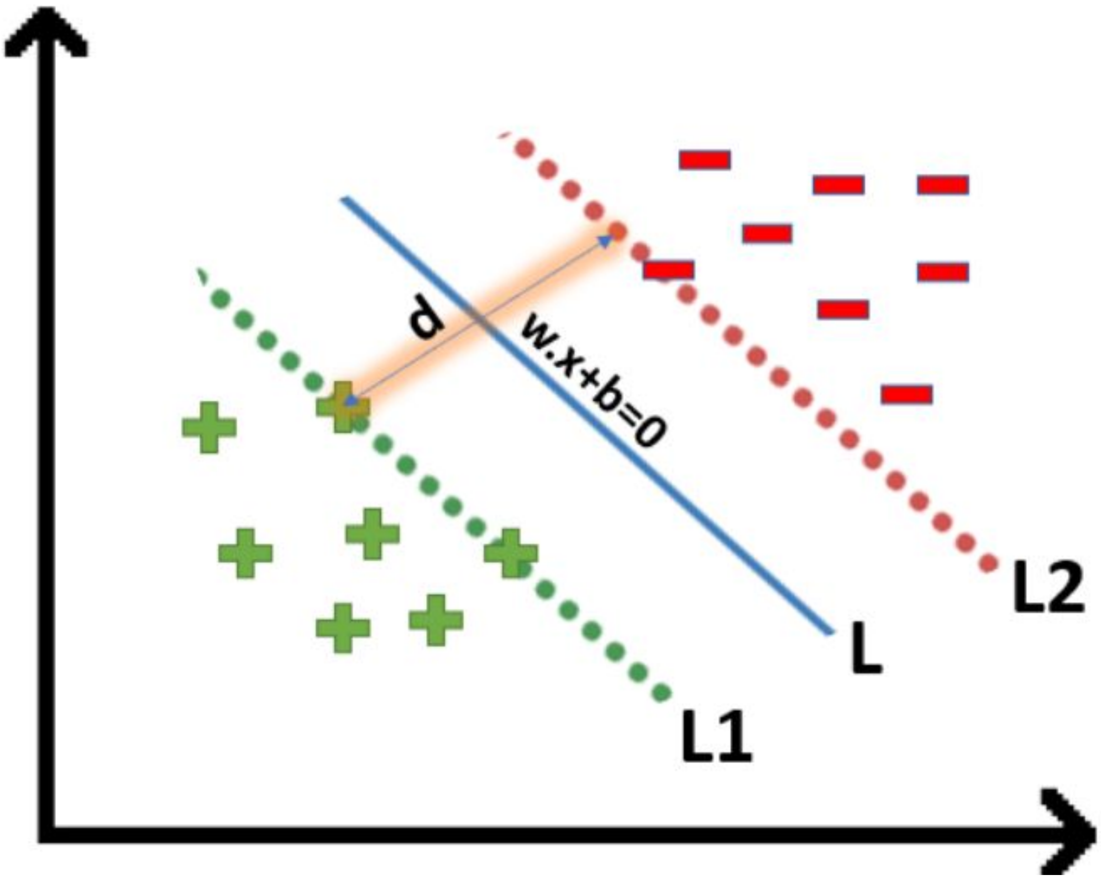
$$\overrightarrow{X}.\overrightarrow{w} + \mathrm{b} \geq 0$$

$$\text{hence}$$

$$\mathrm{y} = \begin{cases} +1 & \text{if } \overrightarrow{X}.\overrightarrow{w} + \mathrm{b} \geq 0 \\ -1 & \text{if } \overrightarrow{X}.\overrightarrow{w} + \mathrm{b} < 0 \end{cases}$$


w.x+b=0

If the value of w.x+b>0 then we can say it is a positive point otherwise it is a negative point. Now we need (w,b) such that the margin has a maximum distance. Let's say this distance is 'd'.



https://www.desmos.com/calculator/15dbwehq9g

We’ll calculate the distance (d) in such a way that no positive or negative point can cross the margin line

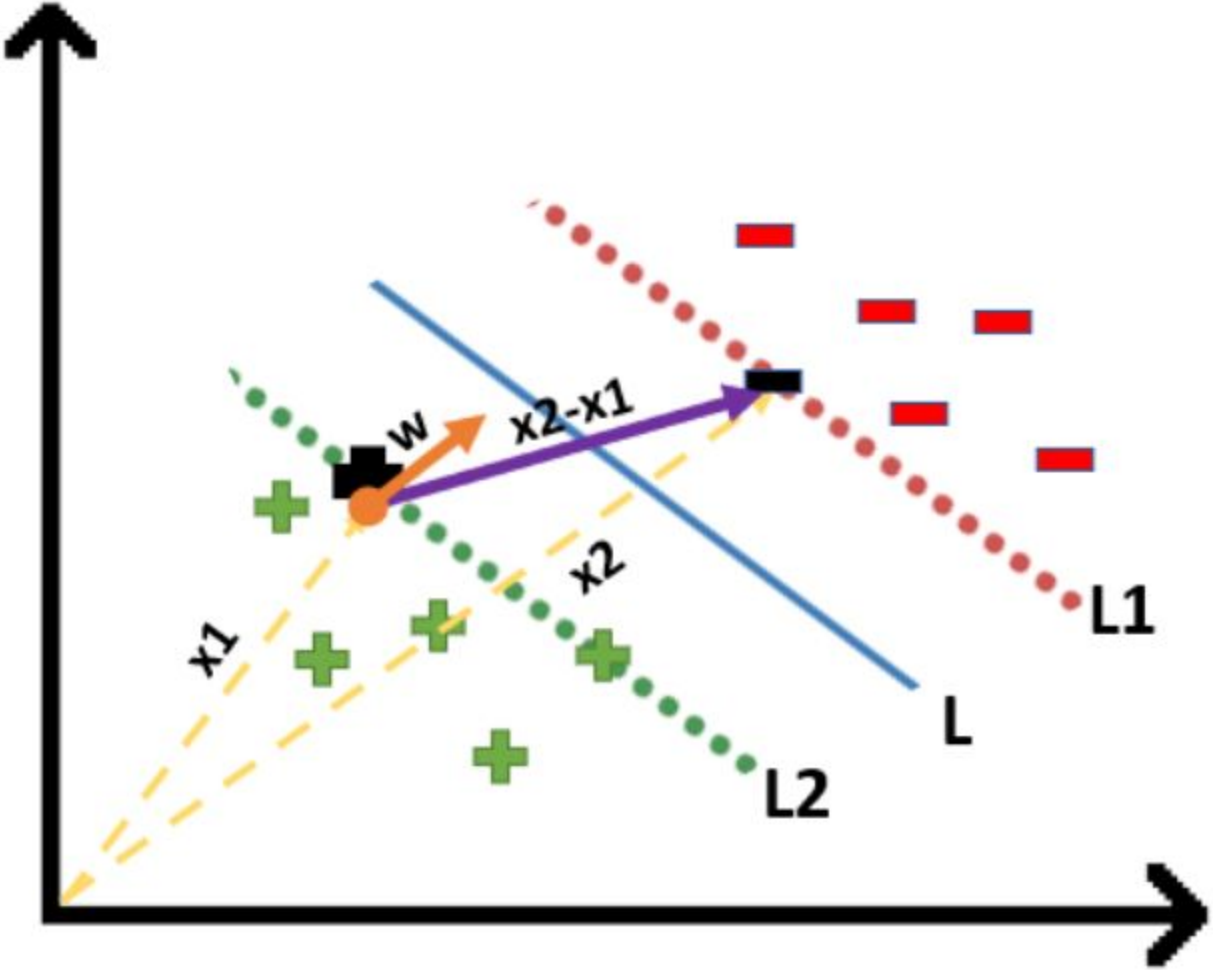
$$For\ all\ the\ Red\ points\ \vec{w}.\overrightarrow{X}+b \leq -1$$

$$For\ all\ the\ Green\ points\ \vec{w}.\overrightarrow{X}+b \geq 1$$

Rather than taking 2 constraints forward, we’ll now try to simplify these two constraints into 1. We assume that negative classes have ***y=-1*** and positive classes have ***y=1.***

We can say that for every point to be correctly classified this condition should always be true:

$$y_i(\vec{w}.\overrightarrow{X}+b) \geq 1$$


w
x2-x1
x1
x2
L1
L
L2

We will take 2 support vectors, 1 from the negative class and 2nd from the positive class. The distance between these two vectors x1 and x2 will be *(x2-x1) vector*. What we need is, the shortest distance between these two points which can be found using a trick we used in the dot product. We take a vector 'w' perpendicular to the hyperplane and then find the projection of (x2-x1) vector on 'w'.

$$\Rightarrow (x2 - x1) . \frac{\overrightarrow{w}}{\|w\|}$$ where $\|\vec{w}\| = \sqrt{\sum_i w_i^2}$

$$\Rightarrow \frac{x2 . \overrightarrow{w} - x1 . \overrightarrow{w}}{\|w\|} \qquad ----(1)$$

Since x2 and x1 are support vectors and they lie on the hyperplane, hence they will follow **y_i* (2.x+b)=1** so we can write it as:

$$\text{for positive point } y = 1$$

$$\Rightarrow 1 \times (\overrightarrow{w} . x1 + b) = 1$$

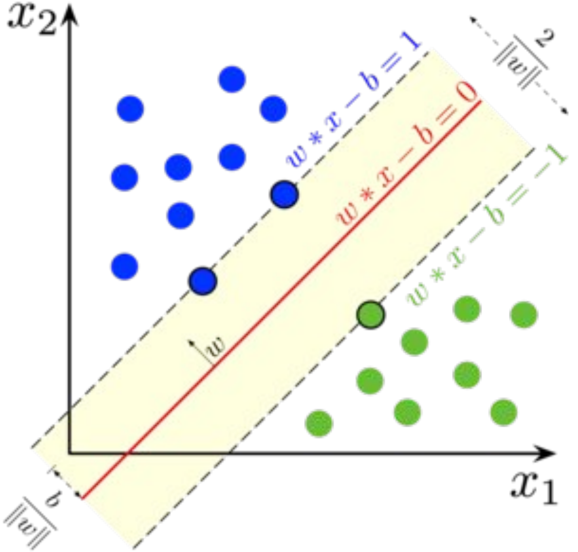
$$\Rightarrow \overrightarrow{w} . x1 = 1 - b \qquad -----(2)$$

$$\text{Similarly for negative point } y = -1$$

$$\Rightarrow -1 \times \left(\overrightarrow{w} . x2 + b\right) = 1$$

$$\Rightarrow \overrightarrow{w} . x2 = -b - 1 \qquad -----(3)$$

Putting equations (2) and (3) in equation (1) we get:



$$\Rightarrow \frac{(1-\mathrm{b})-(-\mathrm{b}-1)}{\|\mathrm{w}\|}$$

$$\Rightarrow \frac{1-\mathrm{b}+\mathrm{b}+1}{\|\mathrm{w}\|} = \frac{2}{\|\mathrm{w}\|} = \mathrm{d}$$

Hence the equation which we have to maximize is:

$$\operatorname{argmax}(\mathrm{w}^*, \mathrm{b}^*)\ \frac{2}{\|\mathrm{w}\|} \text{ such that } \mathrm{y_i}\left(\vec{w}.\vec{X}+b\right) \geq 1$$

The resulting Lagrange multiplier equation we try to optimize is:

$$L = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i (y_i(\vec{w} \cdot \vec{x_i} + b) - 1)$$

α is Lagrange multiplier

Lagrange Multiplier Theorem for Single Constraint

In this case, we consider the functions of two variables. That means the optimization problem is given by:

Max $f(x, Y)$

Subject to:

$g(x, y) = 0$

(or)

We can write this constraint by adding an additive constant such as $g(x, y) = k$.

Let us assume that the functions f and g (defined above) contain first-order partial derivatives. Thus, we can write the Lagrange function as:

$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda\, g(x, y)$

To maximize this distance, we can minimize the square of the denominator to give us a quadratic programming problem given by:

$$\min \frac{1}{2}||w||^2 \ \text{subject to} \ y_i(w^T x_i + b) \geq +1, \forall i$$

Solution Via The Method Of Lagrange Multipliers

To solve the above quadratic programming problem with inequality constraints, we can use the method of Lagrange multipliers. The Lagrange function is therefore:

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 + \sum_i \alpha_i (y_i(w^T x_i + b) - 1)$$

To solve the above, we set the following:

$$\frac{\partial L}{\partial w} = 0, \frac{\partial L}{\partial \alpha} = 0, \frac{\partial L}{\partial b} = 0$$

Plugging above in the Lagrange function gives us the following conditions

$$w = \sum_i \alpha_i y_i x_i \quad \text{and} \quad 0 = \sum_i \alpha_i y_i$$

Deciding The Classification of a Test Point

The classification of any test point x can be determined using this expression:

$$y(x) = \sum_i \alpha_i y_i x^T x_i + b$$

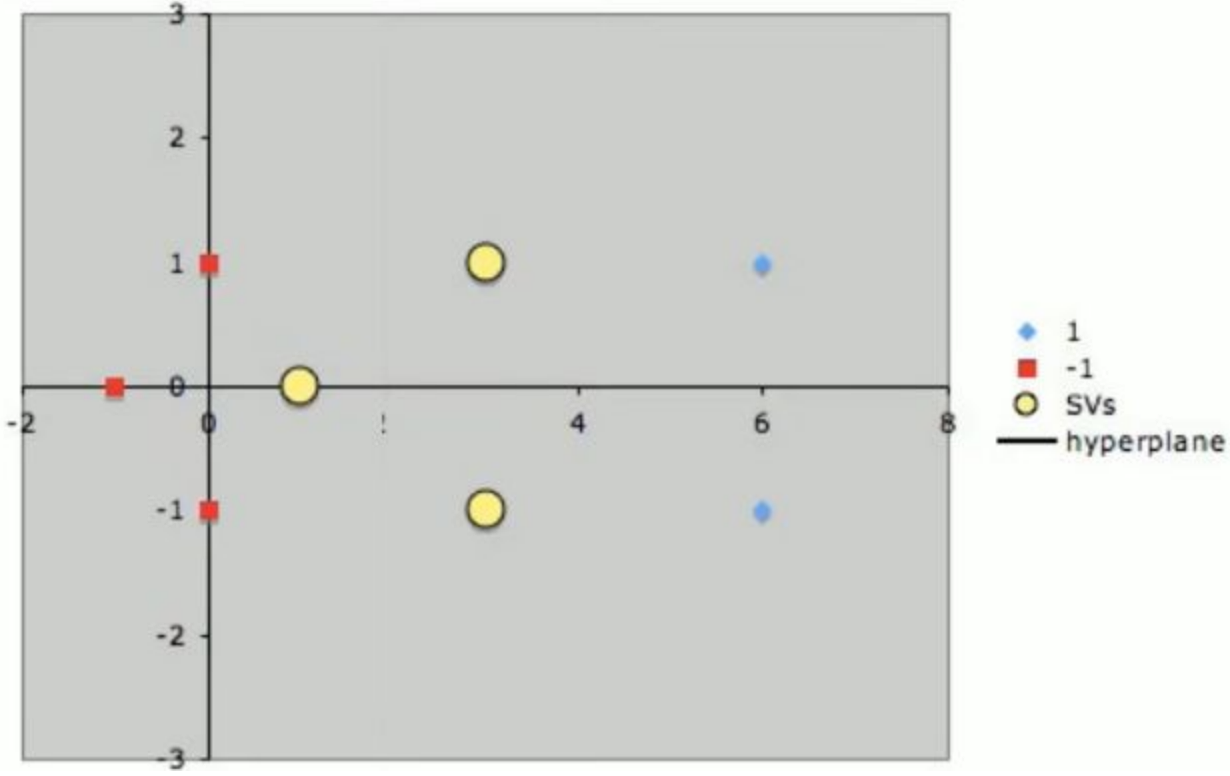
A positive value of $y(x)$ implies $x \in +1$ and a negative value means $x \in -1$

Suppose we are given the following positively labeled data points,

$$\left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 6 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 \\ -1 \end{pmatrix} \right\}$$

and the following negatively labeled data points,

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\}$$



3
2
1
0
-1
-2
-3
-2
0
4
6
8
1
-1
SVs
hyperplane

- Each vector is augmented with a 1 as a bias input

- So, $s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, then $\widetilde{s_1} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

- Similarly,

- $s_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$, then $\widetilde{s_2} = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$ and $s_3 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, then $\widetilde{s_3} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$

$$
\begin{aligned}
\alpha_1 \tilde{s_1} \cdot \tilde{s_1} + \alpha_2 \tilde{s_2} \cdot \tilde{s_1} + \alpha_3 \tilde{s_3} \cdot \tilde{s_1} &= -1 \\
\alpha_1 \tilde{s_1} \cdot \tilde{s_2} + \alpha_2 \tilde{s_2} \cdot \tilde{s_2} + \alpha_3 \tilde{s_3} \cdot \tilde{s_2} &= +1 \\
\alpha_1 \tilde{s_1} \cdot \tilde{s_3} + \alpha_2 \tilde{s_2} \cdot \tilde{s_3} + \alpha_3 \tilde{s_3} \cdot \tilde{s_3} &= +1
\end{aligned}
$$

$$\alpha_1(1+0+1)+\alpha_2(3+0+1)+\alpha_3(3+0+1) = -1$$

$$\alpha_1(3+0+1)+\alpha_2(9+1+1)+\alpha_3(9-1+1) = 1$$

$$\alpha_1(3+0+1)+\alpha_2(9-1+1)+\alpha_3(9+1+1) = 1$$

$$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix}+\alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix}+\alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} = -1$$

$$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix}+\alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix}+\alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} = 1$$

$$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix}+\alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix}+\alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} = 1$$

$$2\alpha_1 + 4\alpha_2 + 4\alpha_3 = -1$$

$$4\alpha_1 + 11\alpha_2 + 9\alpha_3 = 1$$

$$4\alpha_1 + 9\alpha_2 + 11\alpha_3 = 1$$

$$\alpha_1 = -3.5$$

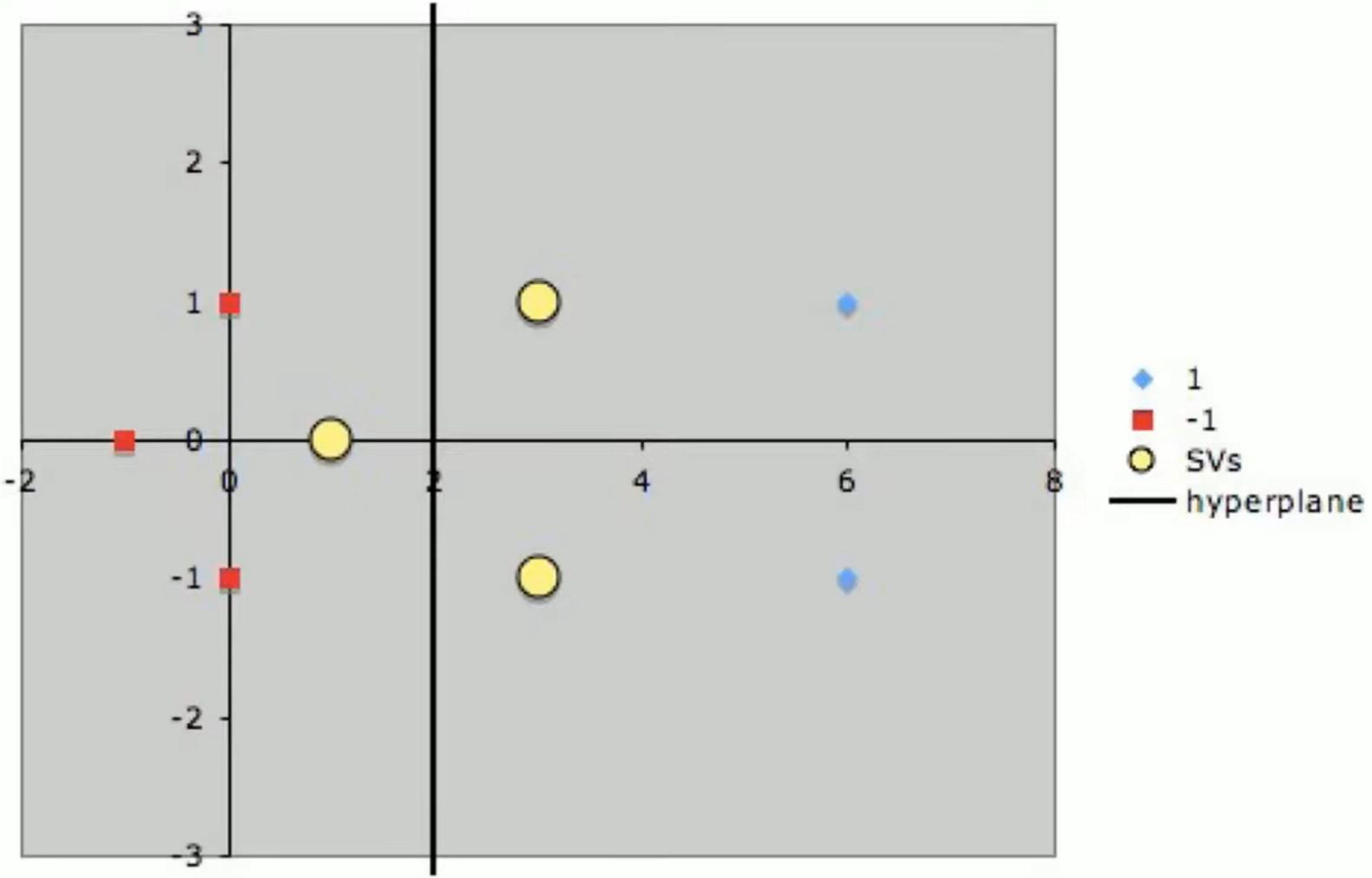
$$\alpha_2 = 0.75$$

$$\alpha_3 = 0.75$$

Support Vector Machine - Linear Example Solved

$$\begin{aligned}
\tilde{w} &= \sum_i \alpha_i \tilde{s}_i \\
&= -3.5 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 0.75 \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} + 0.75 \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}
\end{aligned}$$

- Finally, remembering that our vectors are augmented with a bias.
- We can equate the last entry in $\widetilde{w}$ as the hyperplane offset b and write the separating
- Hyperplane equation $\boldsymbol{y} = \boldsymbol{w}\boldsymbol{x} + \boldsymbol{b}$
- with $\boldsymbol{w} = \binom{1}{0}$ and $\boldsymbol{b} = -\mathbf{2}$.


3
2
1
0
-1
-2
-3
-2
0
2
4
6
8
1
-1
SVs
hyperplane