Advanced Machine Learning

Likhit Nayak

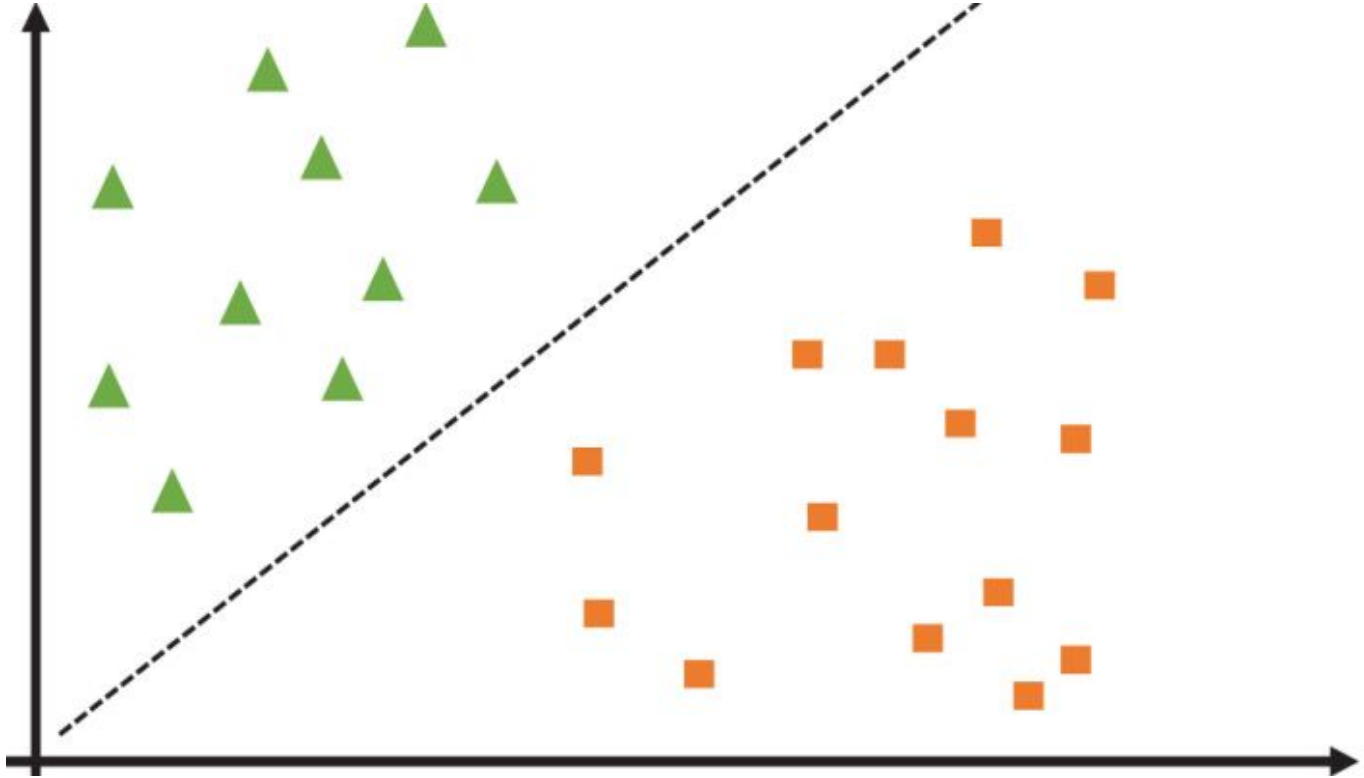
What is a SVM?

Support Vector Machines (SVMs) is a type of machine learning algorithm for:

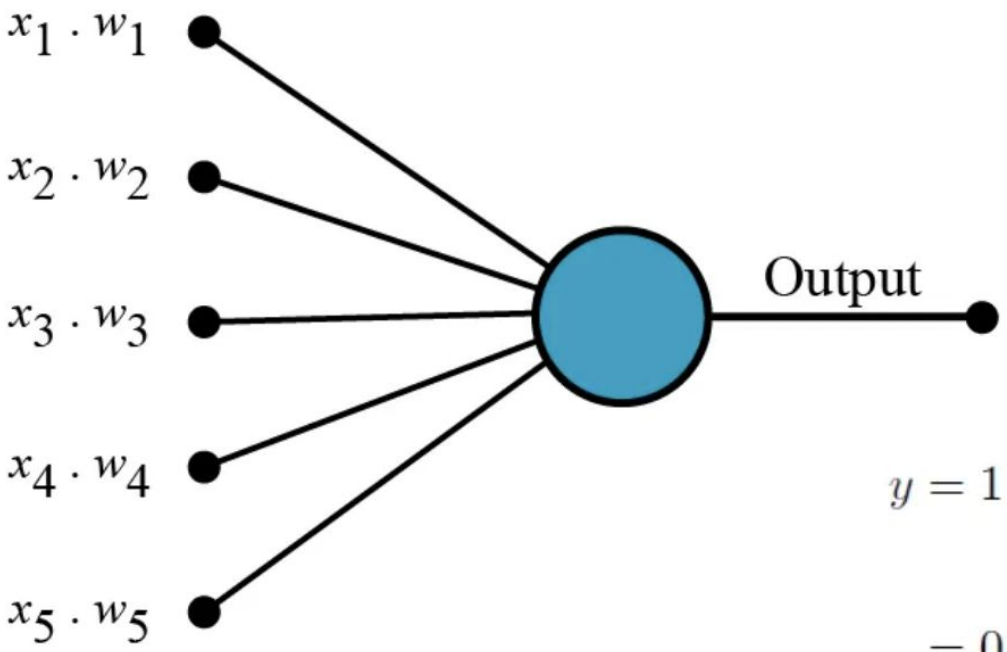
- **Supervised learning**
- **Classification tasks**
- **Regression tasks**

The objective of an SVM algorithm is to find a ***hyperplane*** that separates the data points belong to different classes

What is a SVM?



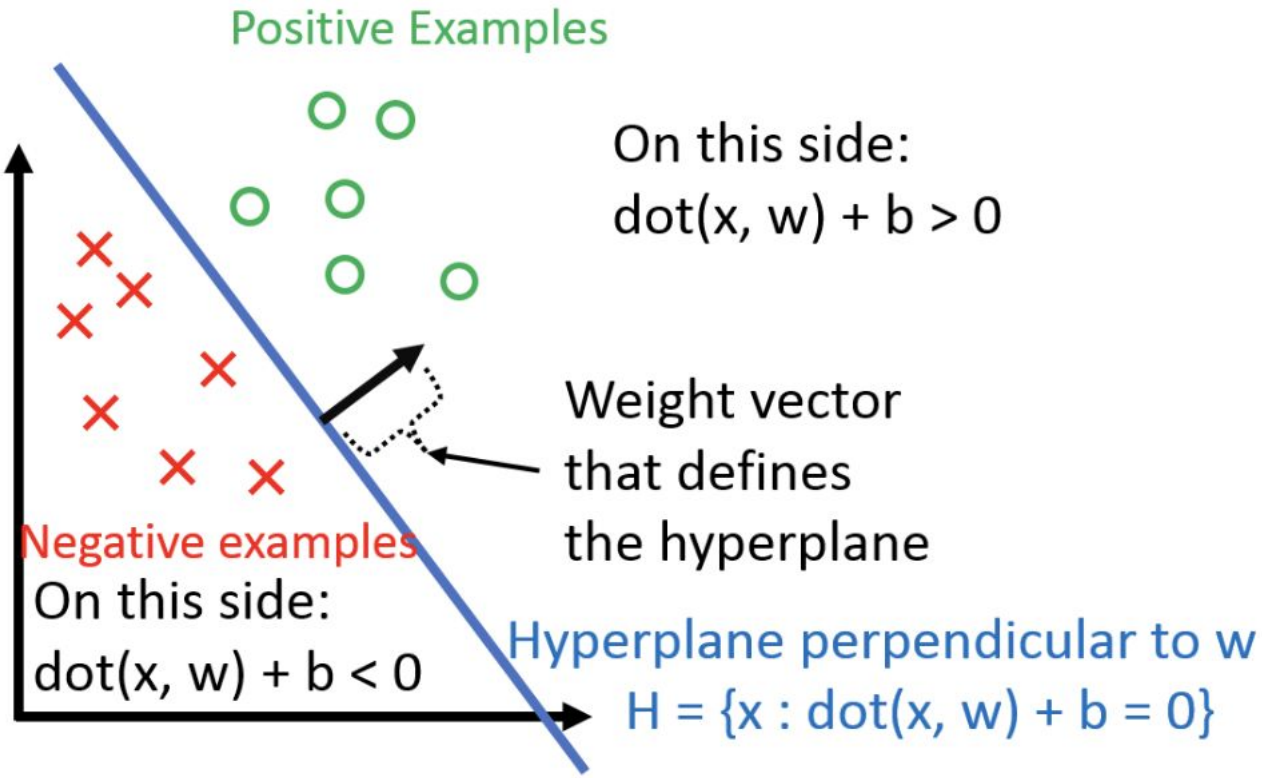
Perceptron



$$y = 1 \quad if \sum_{i=0}^{n} w_i * x_i \geq 0$$

$$= 0 \quad if \sum_{i=0}^{n} w_i * x_i < 0$$

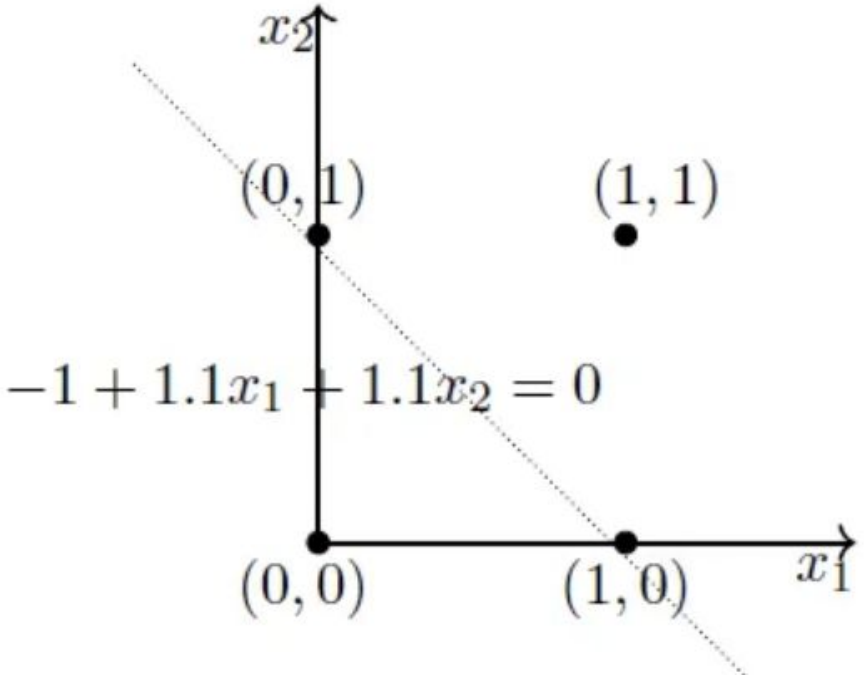
Perceptron


Positive Examples
On this side:
dot(x, w) + b > 0
Weight vector
that defines
the hyperplane
Negative examples
On this side:
dot(x, w) + b < 0
Hyperplane perpendicular to w
H = {x : dot(x, w) + b = 0}

Perceptron - Learning OR function

x_1	x_2	OR	
0	0	0	$w_0 + \sum_{i=1}^{2} w_i x_i < 0$
1	0	1	$w_0 + \sum_{i=1}^{2} w_i x_i \geq 0$
0	1	1	$w_0 + \sum_{i=1}^{2} w_i x_i \geq 0$
1	1	1	$w_0 + \sum_{i=1}^{2} w_i x_i \geq 0$

Perceptron - Learning OR function



Perceptron Learning Algorithm

Algorithm: Perceptron Learning Algorithm

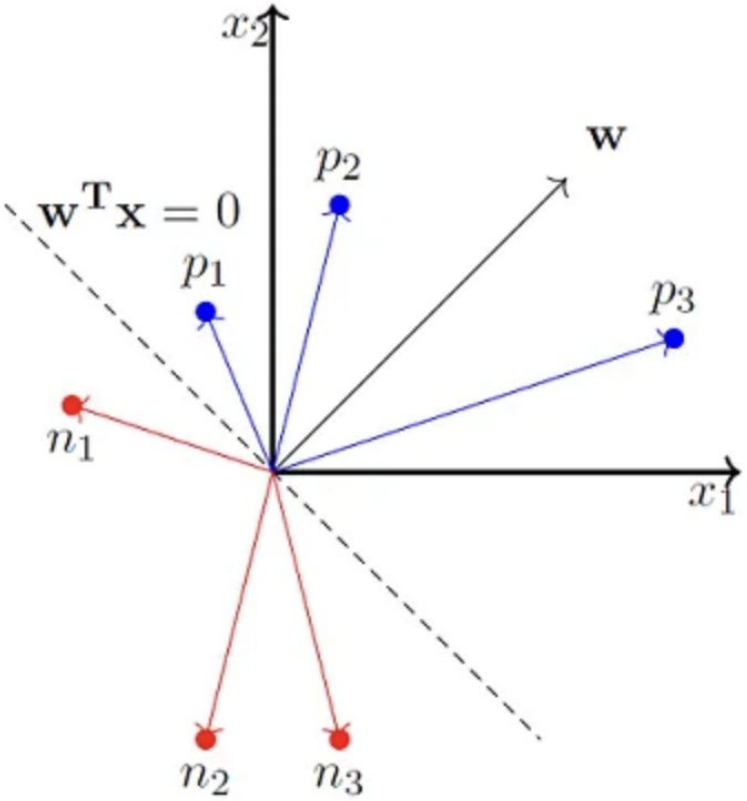
$P \leftarrow$ *inputs with label* 1;
$N \leftarrow$ *inputs with label* 0;
Initialize $\mathbf{w}$ randomly;
while !*convergence* **do**
  Pick random $\mathbf{x} \in P \cup N$;
  if $\mathbf{x} \in P$ *and* $\mathbf{w}.\mathbf{x} < 0$ **then**
    $\mathbf{w} = \mathbf{w} + \mathbf{x}$;
  end
  if $\mathbf{x} \in N$ *and* $\mathbf{w}.\mathbf{x} \geq 0$ **then**
    $\mathbf{w} = \mathbf{w} - \mathbf{x}$;
  end
end
//the algorithm converges when all the inputs are classified correctly

$$\mathbf{w} = [w_0, w_1, w_2, ..., w_n]$$

$$\mathbf{x} = [1, x_1, x_2, ..., x_n]$$

$$\mathbf{w} \cdot \mathbf{x} = \mathbf{w}^{\mathrm{T}}\mathbf{x} = \sum_{i=0}^{n} w_i * x_i$$

Perceptron Learning Algorithm

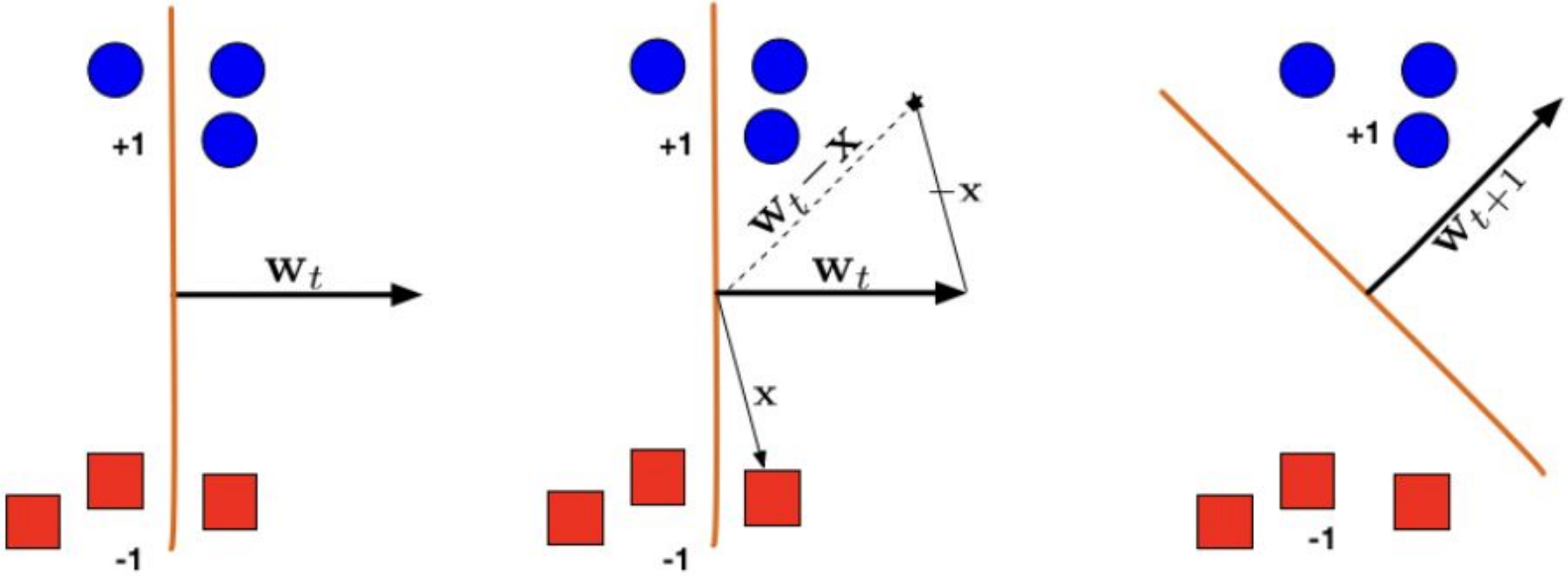


$$cos\alpha = \frac{\mathbf{w}^T\mathbf{x}}{||\mathbf{w}||||\mathbf{x}||}$$

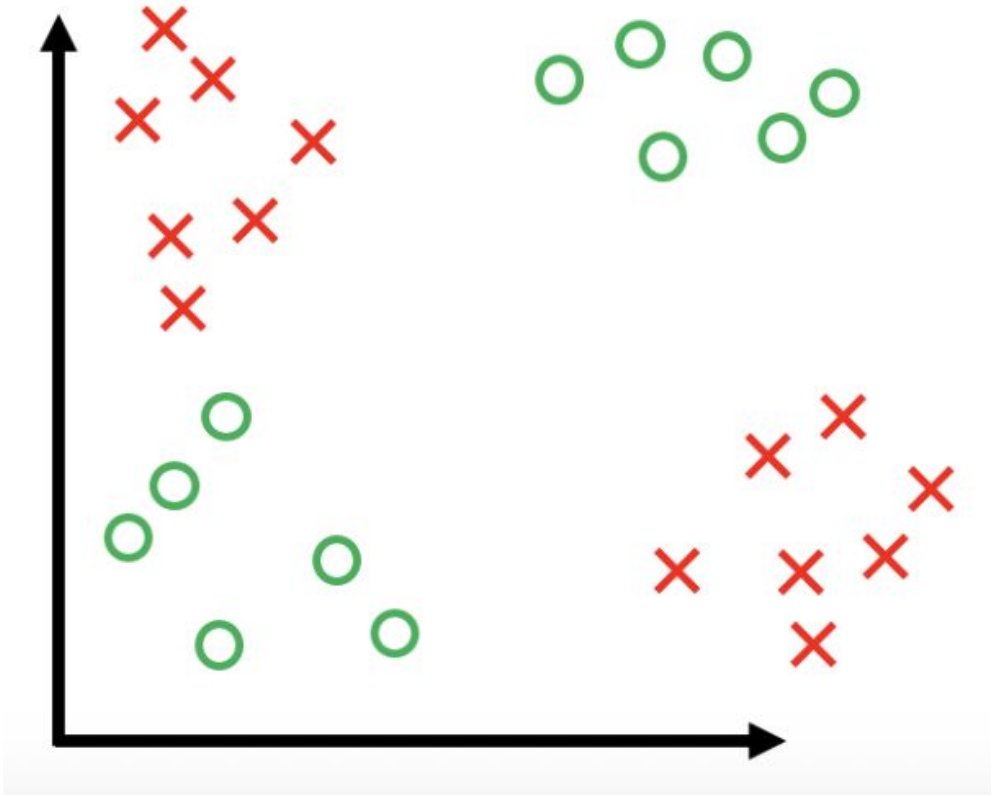
So if $\mathbf{w}^T\mathbf{x} > 0 \Rightarrow cos\alpha > 0 \Rightarrow \alpha < 90$

Similarly, if $\mathbf{w}^T\mathbf{x} < 0 \Rightarrow cos\alpha < 0 \Rightarrow \alpha > 90$

Perceptron Learning Algorithm



Where does a Perceptron fail?



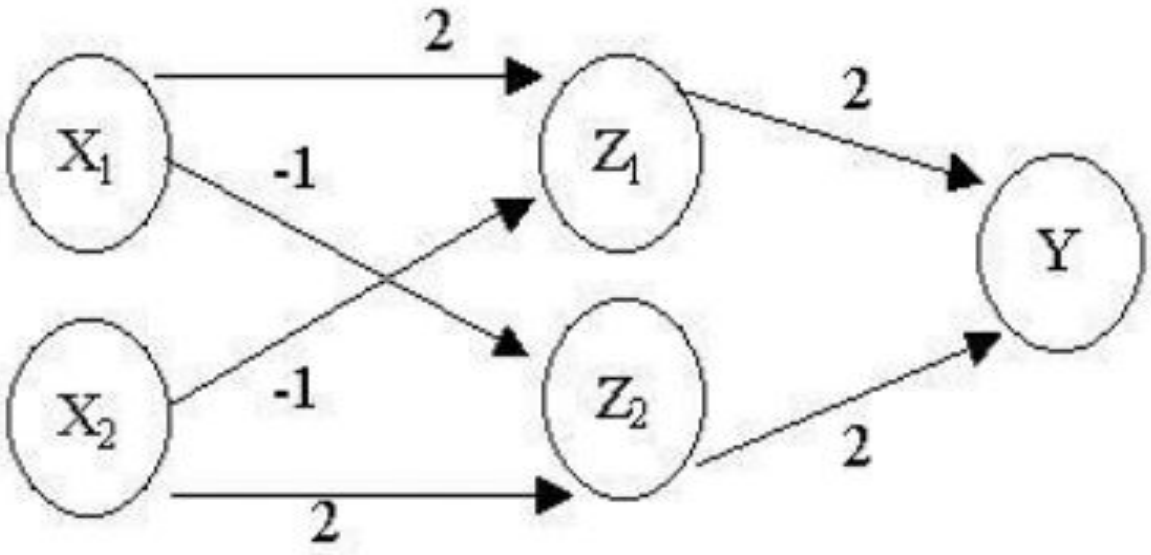
Where does a Perceptron fail?

Learning the XOR function:

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

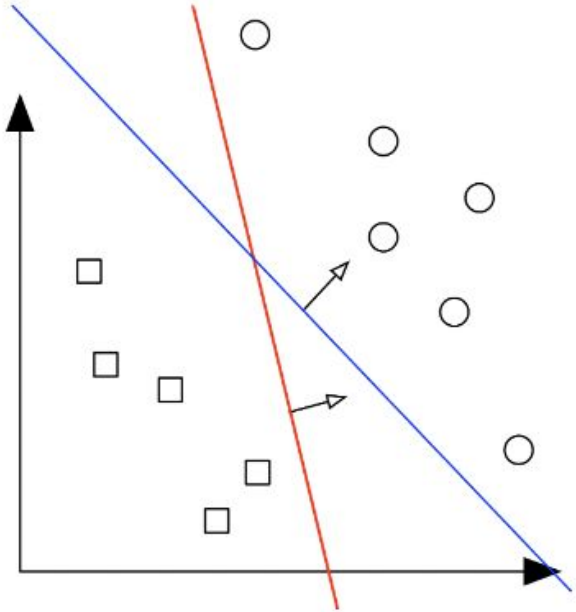
Solution to XOR function

Learning the XOR function: Using MLP (Multilayer Perceptron)



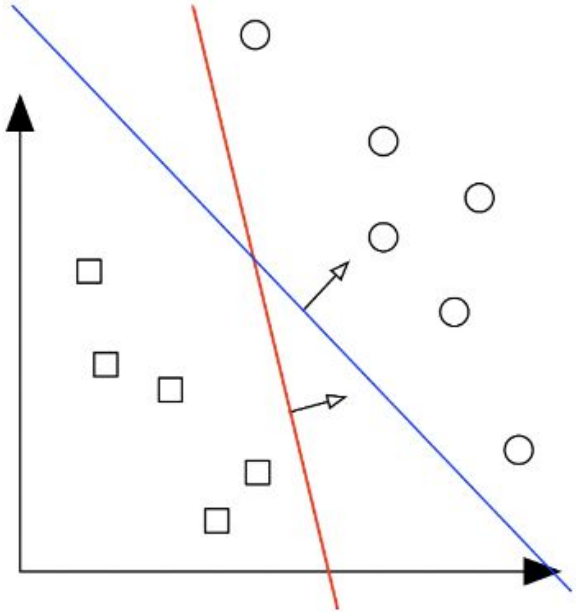
Support Vector Machines (SVM)

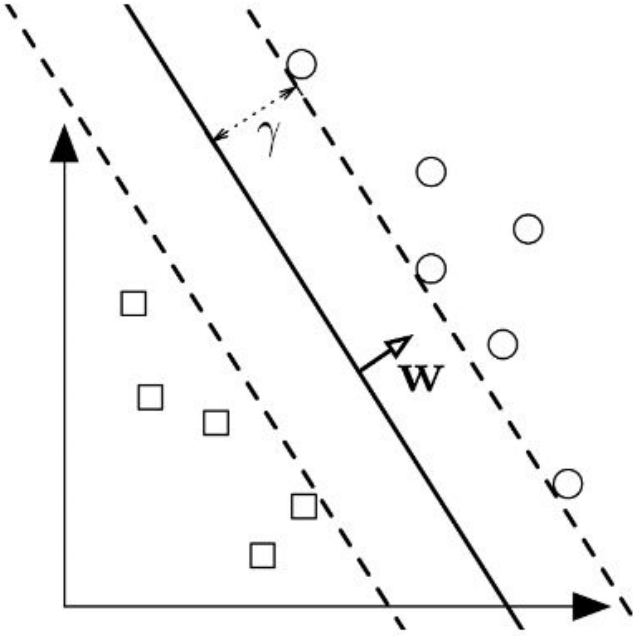
Question: For a linearly separable dataset, what is the best separating hyperplane?



Support Vector Machines (SVM)

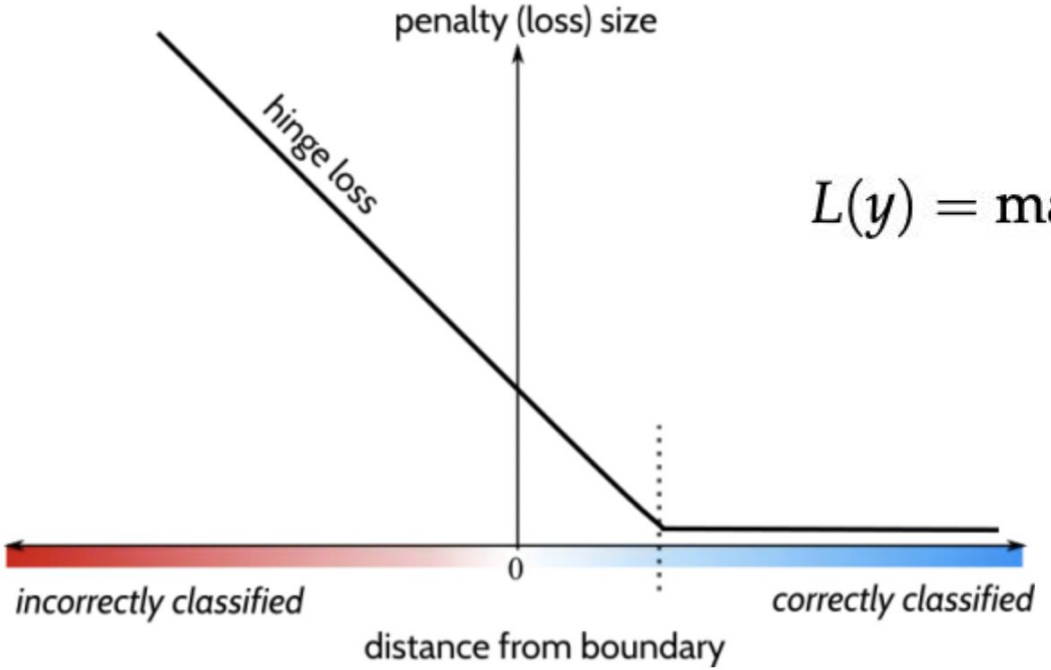
Answer: It is the hyperplane that maximizes the margin, i.e., the distance from the hyperplane to the closest point across both classes





Support Vector Machine (SVM)

It works by minimizing **hinge loss**



$$L(y) = \max(0, 1 - t \cdot y)$$