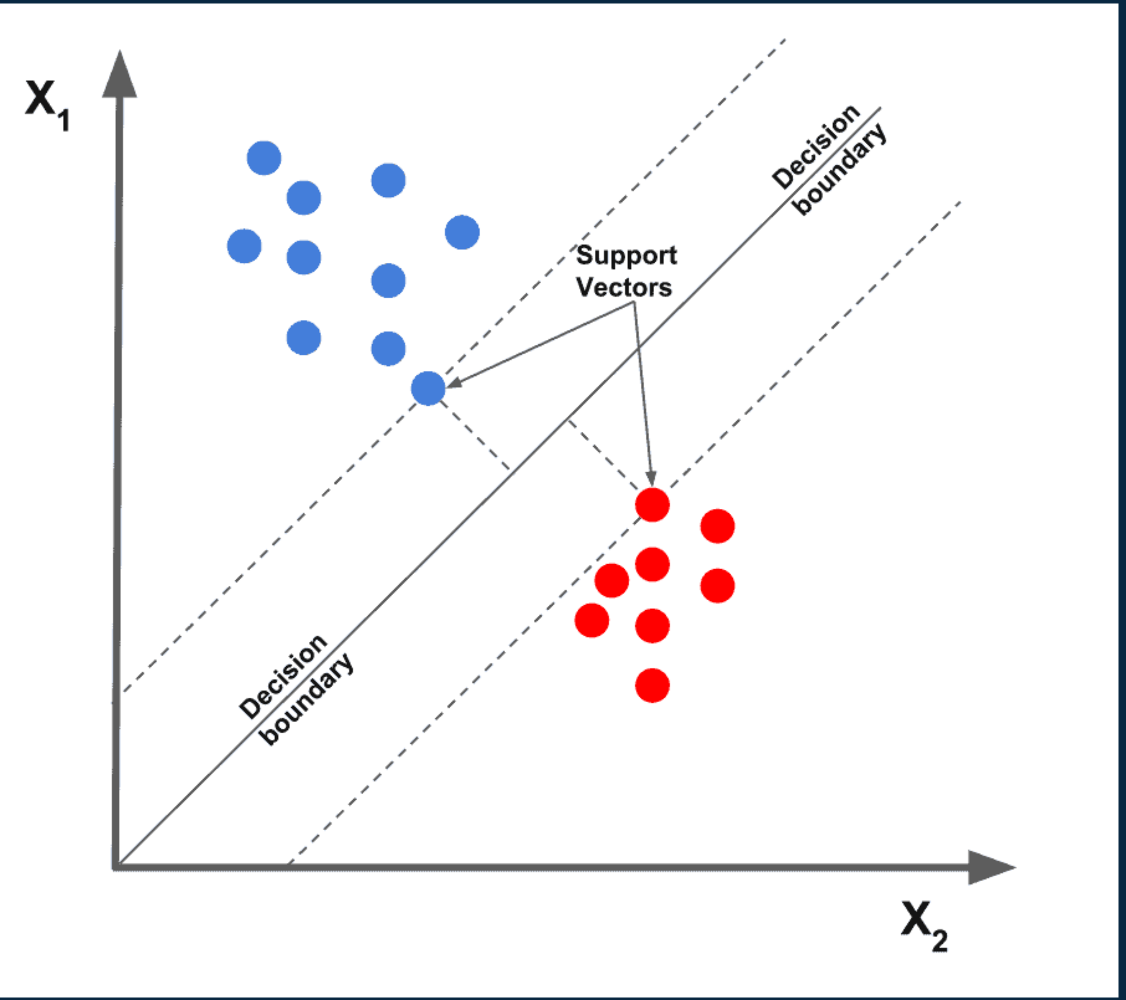
SUPPORT VECTOR MACHINES

WWW.TANTROSTAV.CO.IN

- Support Vector Machines (SVMs) are a type of supervised machine learning algorithm used for classification and regression analysis.

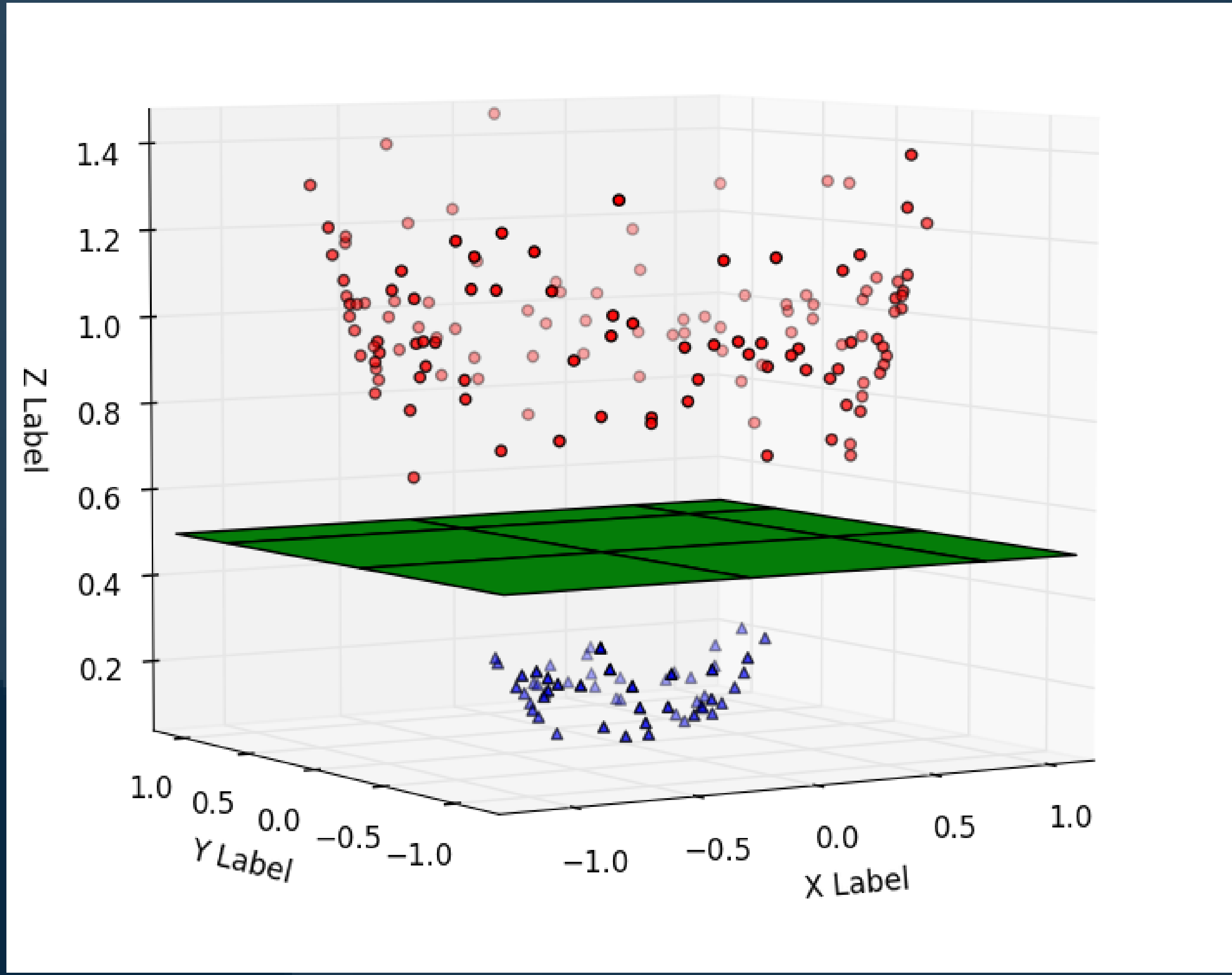
- The goal of SVMs is to find a hyperplane in a high-dimensional space that separates the different classes of data with the largest margin possible.



What is Hyperplane?

- In math, a hyperplane is a flat surface that's one dimension less than the space it exists in. For example, in 3D space, a hyperplane is a flat 2D surface, like a piece of paper.

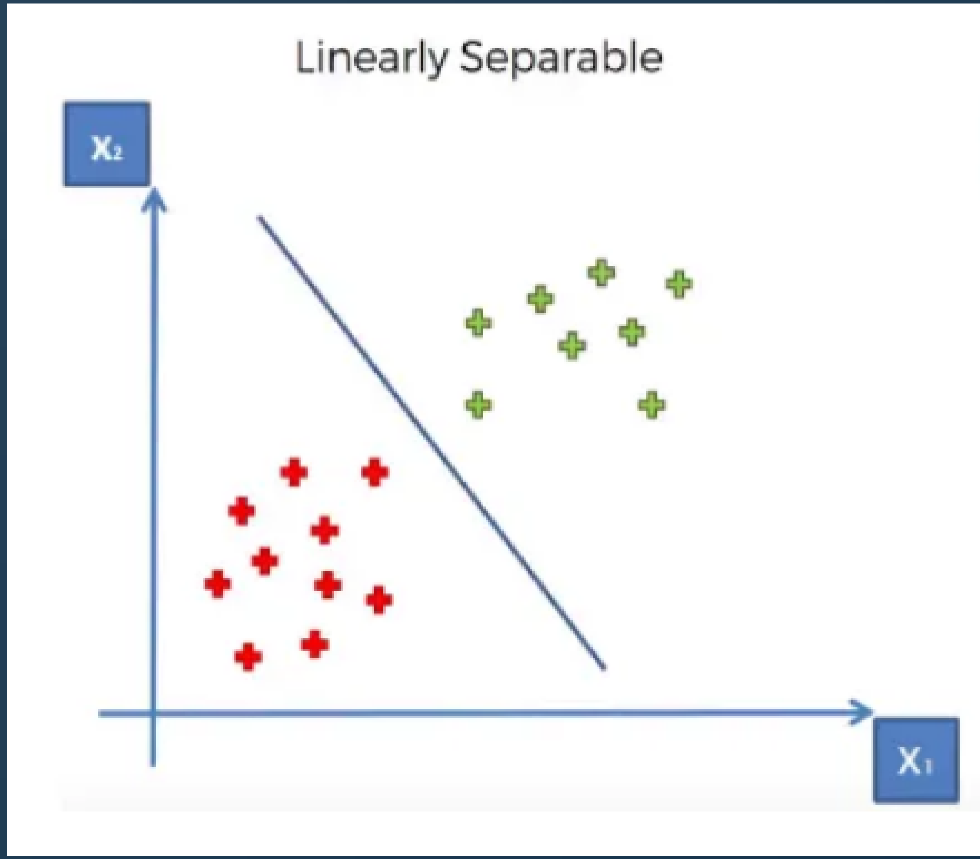
- A hyperplane can be defined by an equation that uses constants and variables. The variables represent coordinates on the hyperplane, and the constants determine its position and orientation in space.



What is Hyperplane?

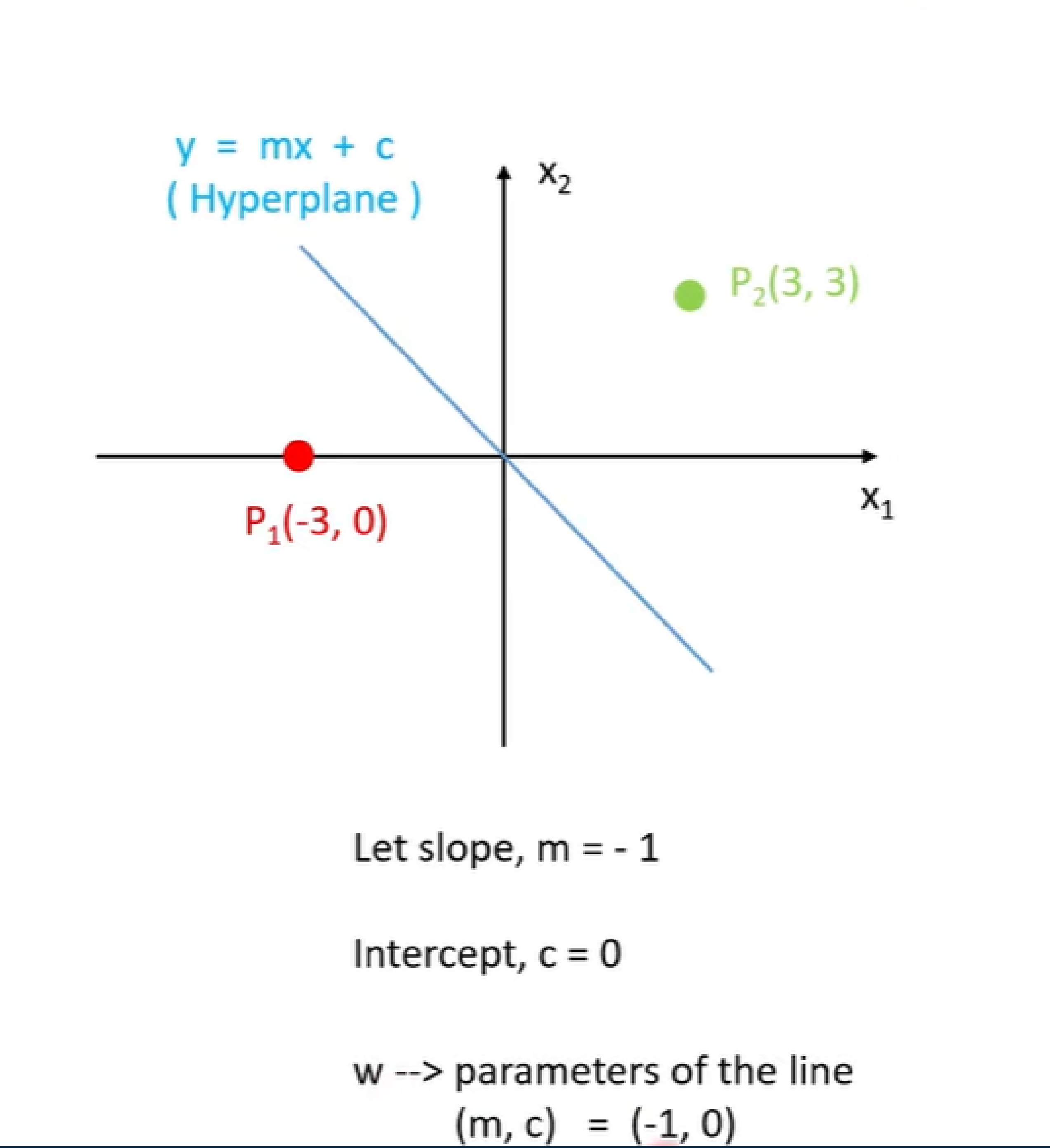
- The hyperplane divides the space it exists in into two regions, with the points on one side of the hyperplane being different from the points on the other side.

- Think of it like a sheet of paper dividing a room into two parts - the points on one side are on one side of the paper, and the points on the other side are on the other side of the paper.



The Math

- Let us take 2 points namely P1 and P2 as given in the figure.
- Let the equation of the hyperplane be $y=m*x+c$.
- Let us assume that the line passes through origin and has slope of -1.
- Since 'm' and 'c' affect the characterstics of the line let them be the parameters
- Thus let W be a 1d Vector containing the value of these parameters.



The Math

$P_1(-3, 0)$

$$w^Tx = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -3 & 0 \end{bmatrix}$$

$$w^Tx = 3$$

(Positive)

Inference: For all the points which lie in the left side of the hyperplane, w^Tx value will be Positive

$P_2(3, 3)$

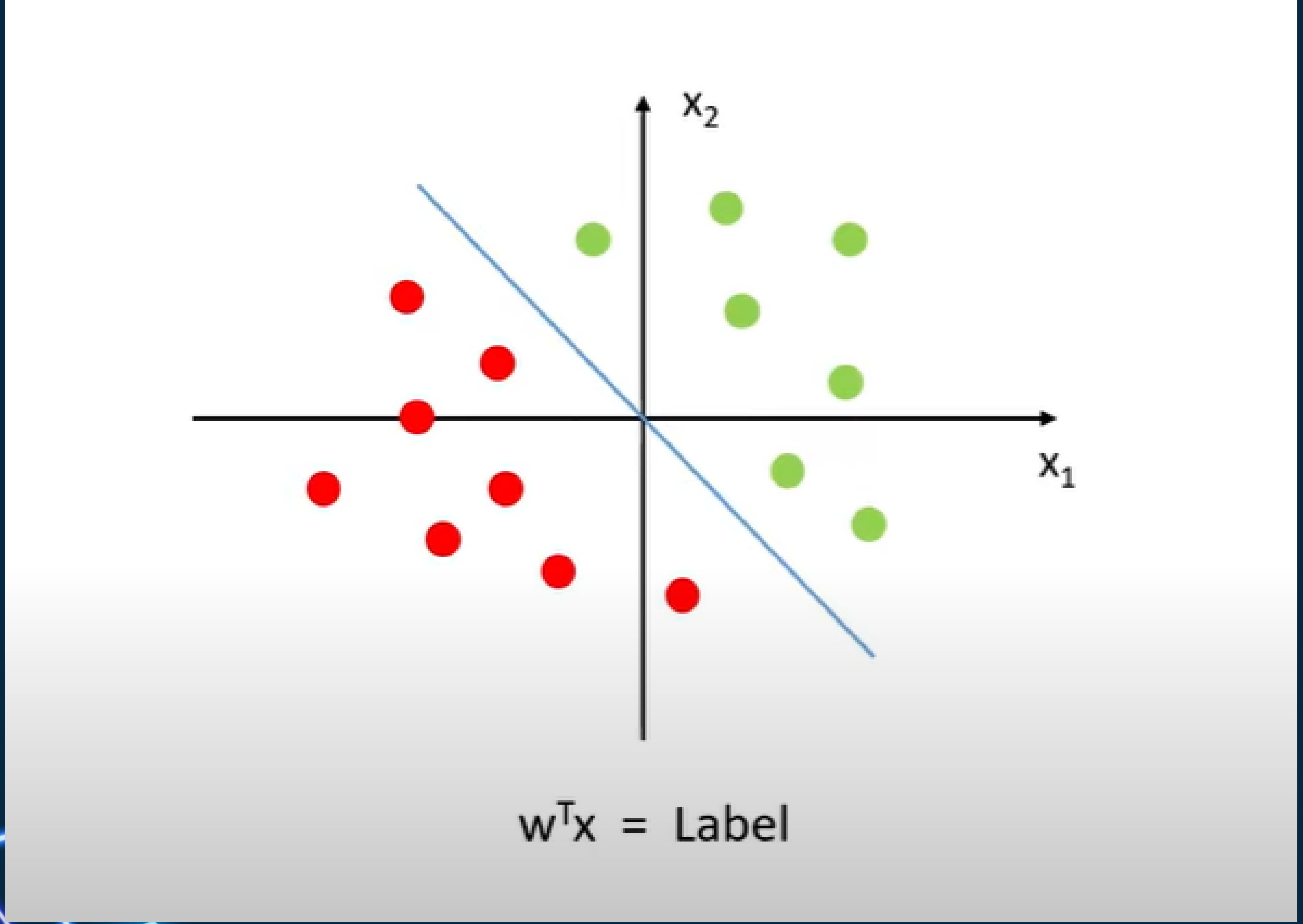
$$w^Tx = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} 3 & 3 \end{bmatrix}$$

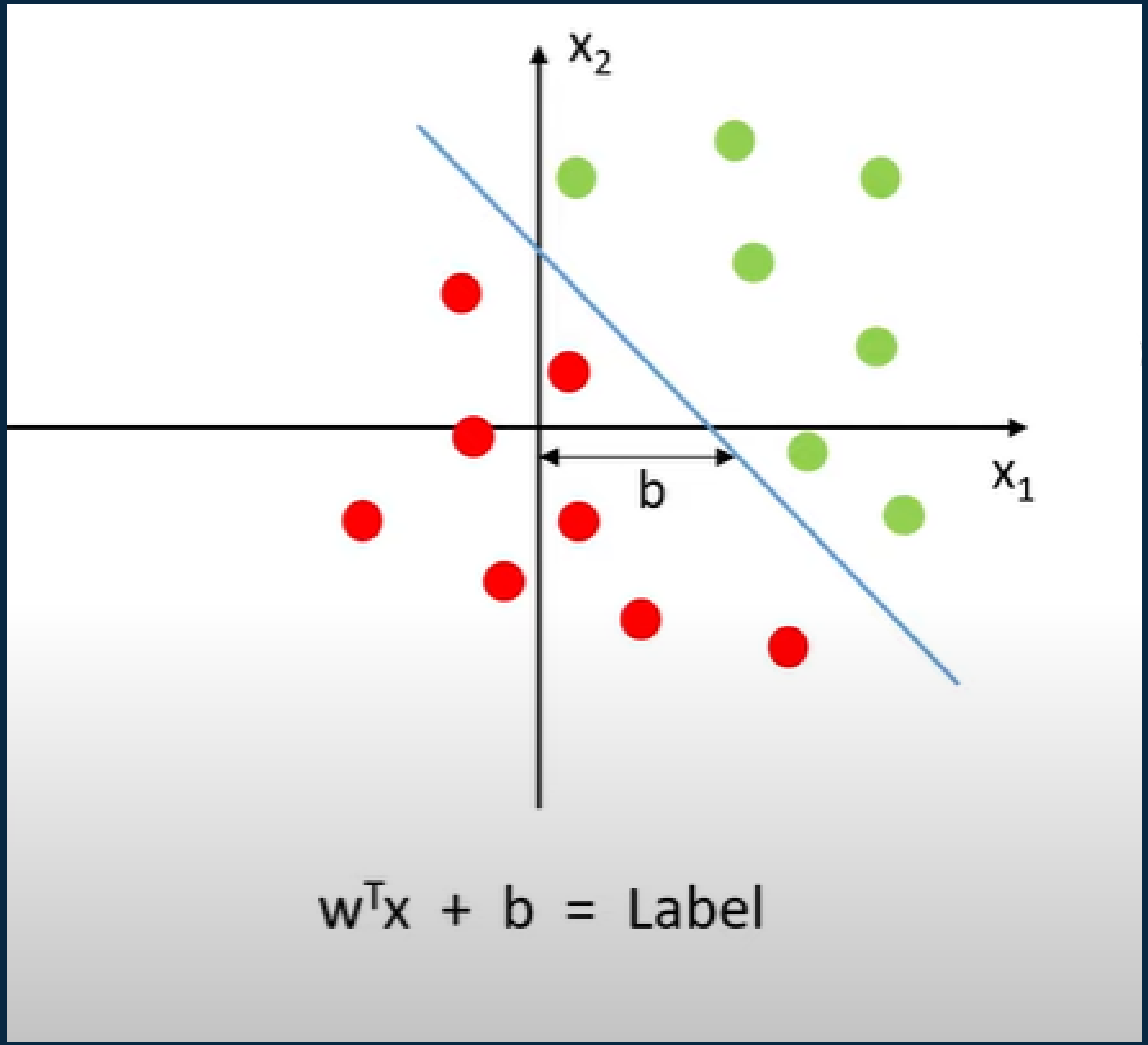
$$w^Tx = -3$$

(Negative)

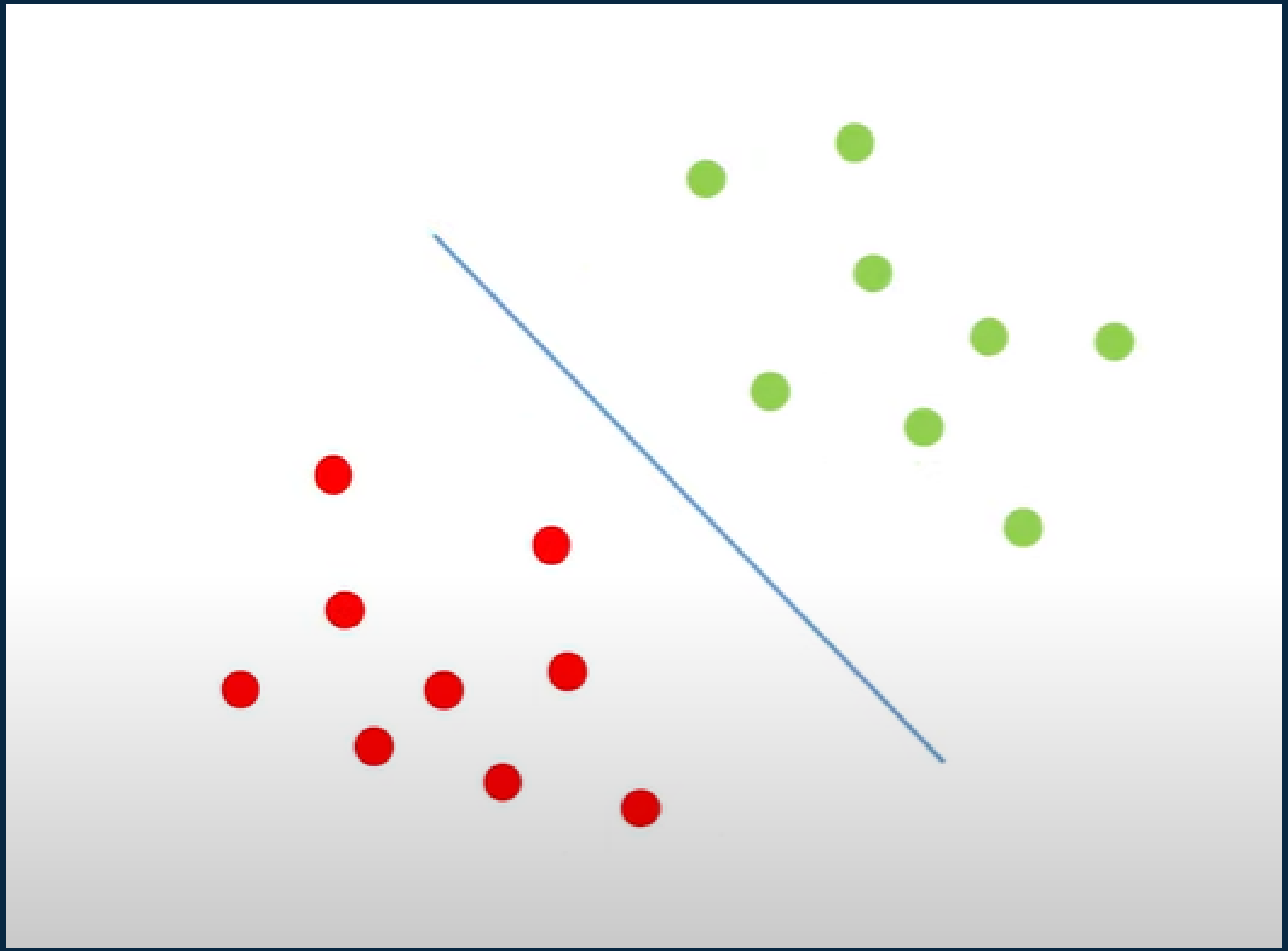
Inference: For all the points which lie in the right side of the hyperplane, w^Tx value will be Negative

- Since all the hyperplanes doesn't pass through origin neither all the points can't be classified with origin as reference point we can add some bias(a constant) to our line for a better fit of our model.





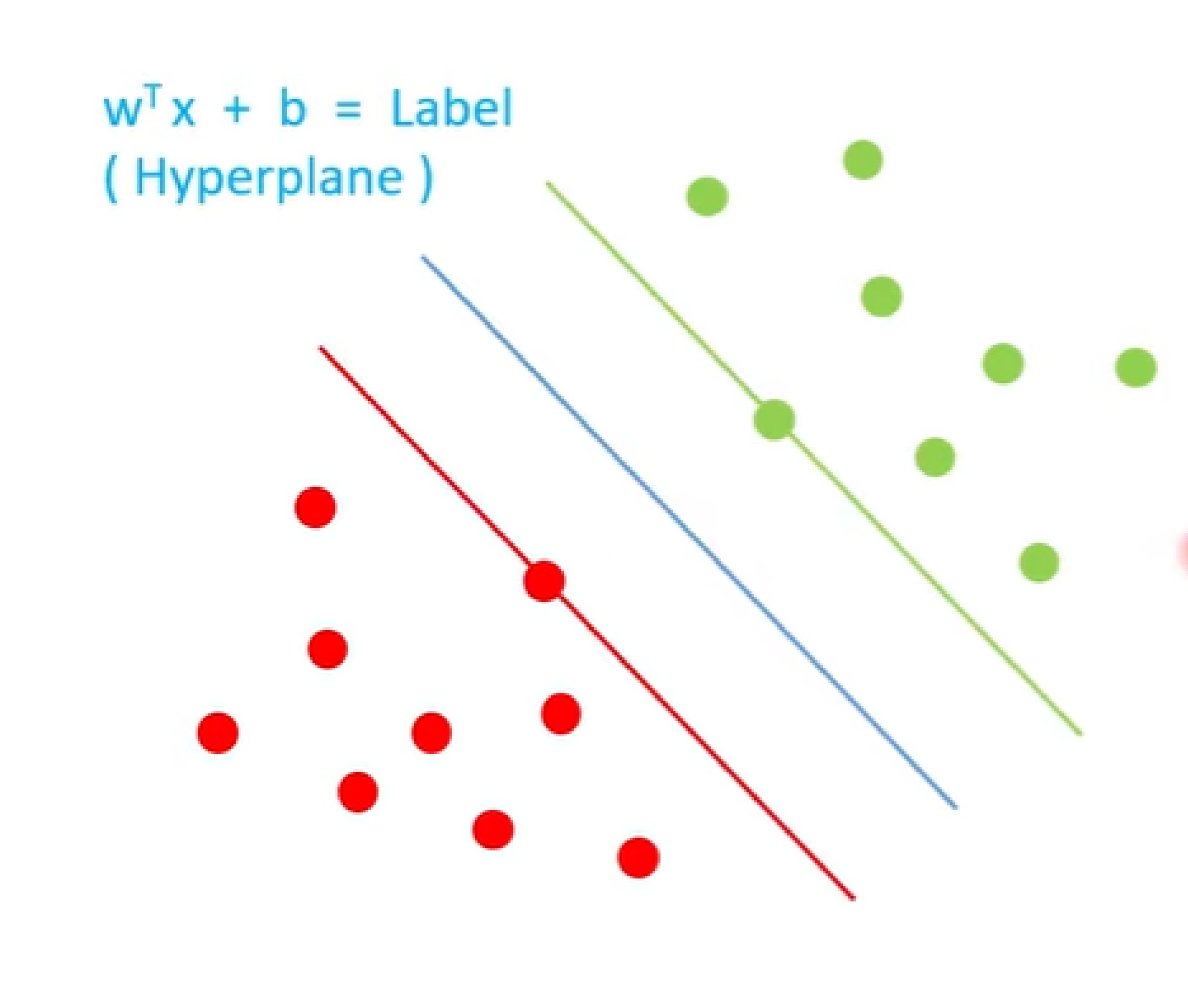
How to find the best hyperplane?



- For a Support Vector Machine Problem there can be N-number of Hyperplanes

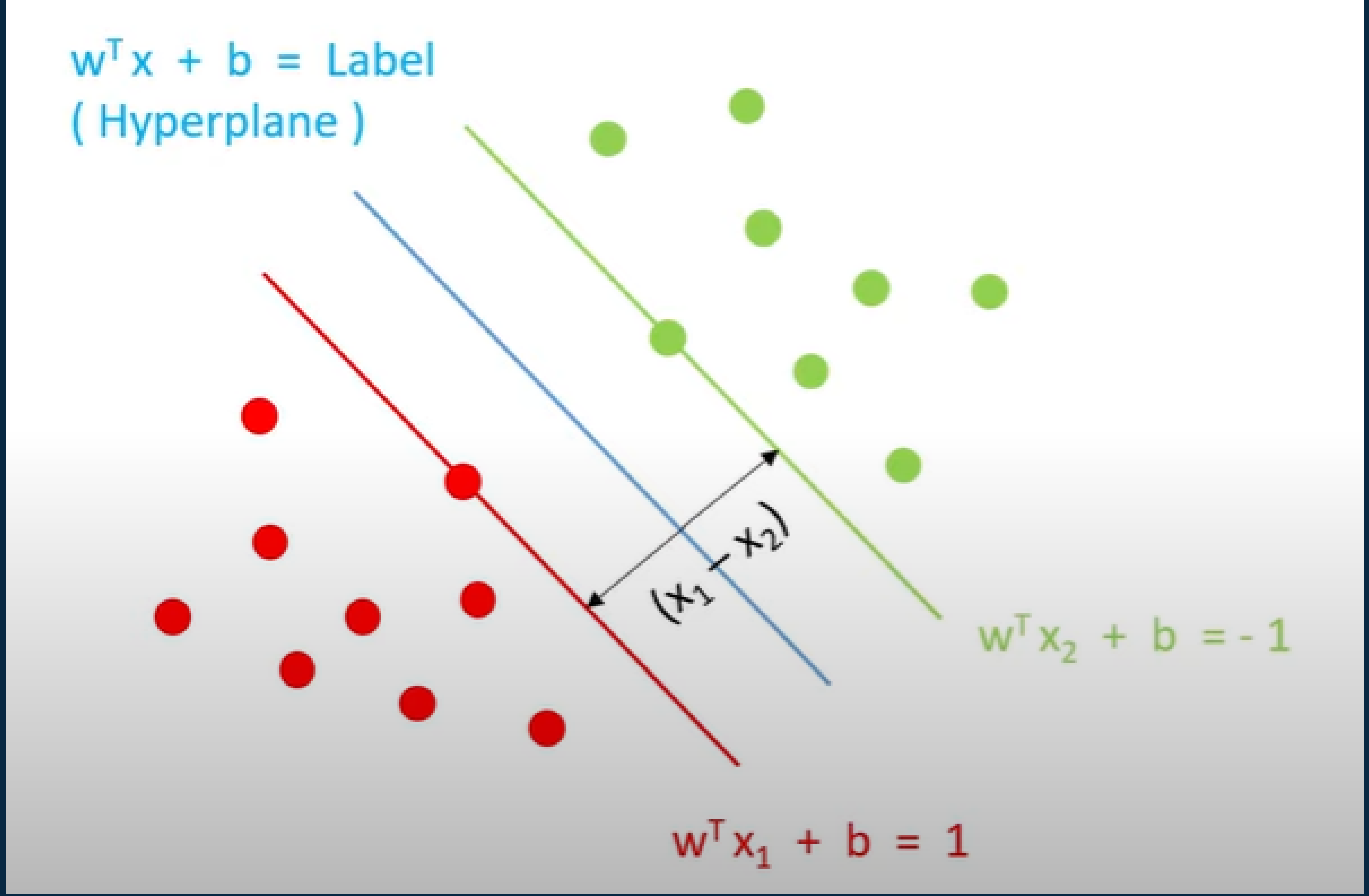
- But for our model to be robust and classify our data correctly the Hyperplane of the SVM should be a optimal one.

- The main goal of this model is to maximize the minimum distance between the Support Vector of 2 classes.



- Now we'll calculate the distance between those support vectors X1 and X2.

- The Equation W^T*X1+b=1 and W^T*X2+b=-1 be the hypothesis equations which classifies the data points into 2 different classes



$$w^T x_1 + b = 1$$

$$(-)\quad w^T x_2 \mp b = \pm 1$$

$$w^T (x_1 - x_2) = 2$$

- We'll subtract these 2 equations from each other.

- Then we divide the Entire equation with the magnitude of W vector which is ||W||.

- Since Wt/||W|| is a unit vector it's magnitude becomes 1.

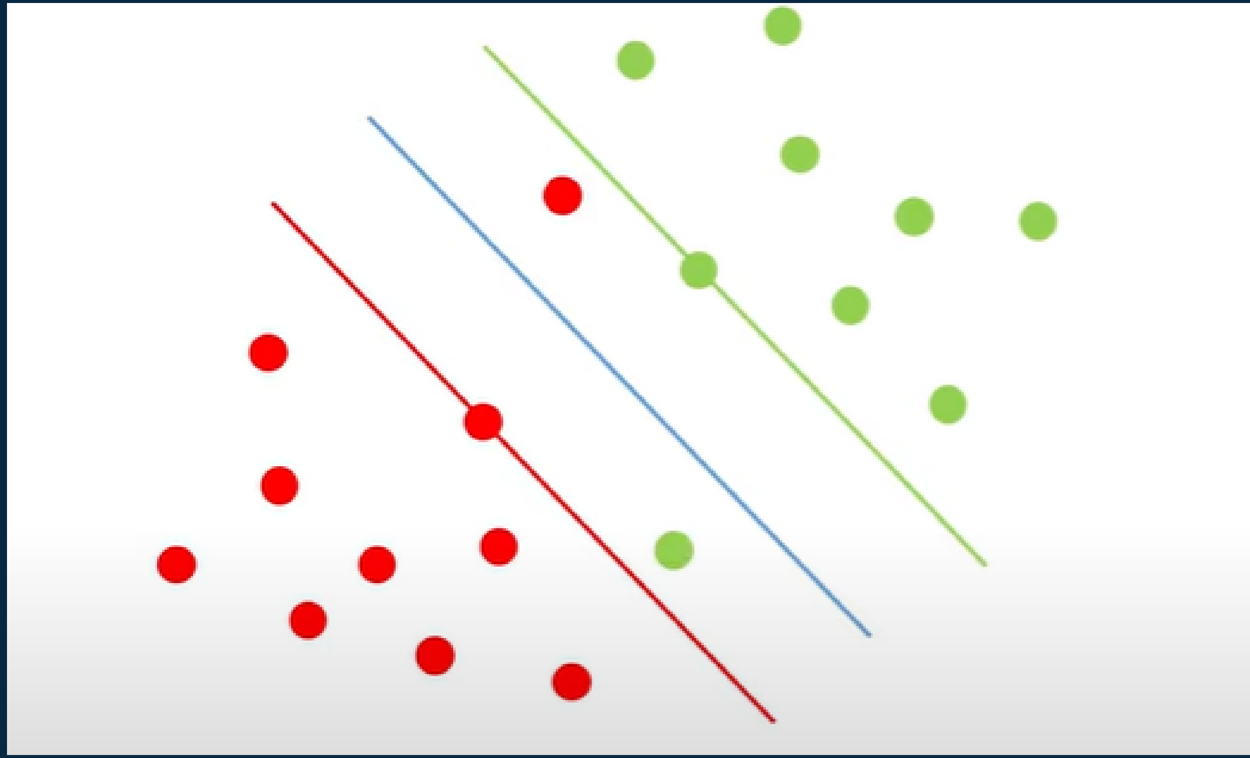
$$\frac{w^T}{||w||}(x_1 - x_2) = \frac{2}{||w||}$$

$$(x_1 - x_2) = \frac{2}{||w||} \quad (\text{margin})$$

- Now,our goal is to maximize this 2/||W|| this means in turn we have to minimize the ||w||.
- So Now we've defined our own hypothesis function inorder to label the datapoint for our classification.

$$y_i = \begin{cases} -1, & w^T x_1 + b \leq -1 \\ 1, & w^T x_1 + b \geq 1 \end{cases}$$

- Now that we've defined our hypothesis function,so for the following image to your right side , Do you think that could we able to classfiy the red and green points correctly?



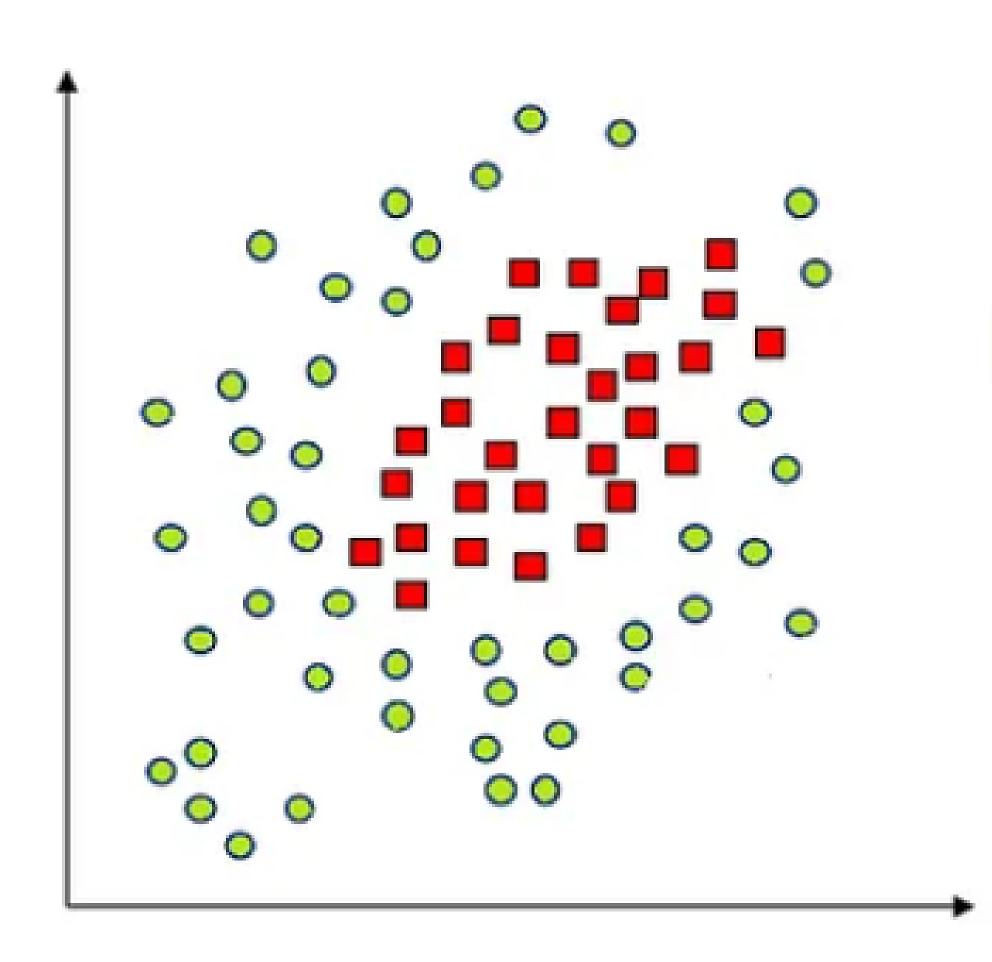
- These are the outliers in the data which affects our model's performance .

$$\min \left(\frac{||w||}{2} \right) + c * \sum \varepsilon_i$$

c --> Number of errors

ε_i --> Error magnitude

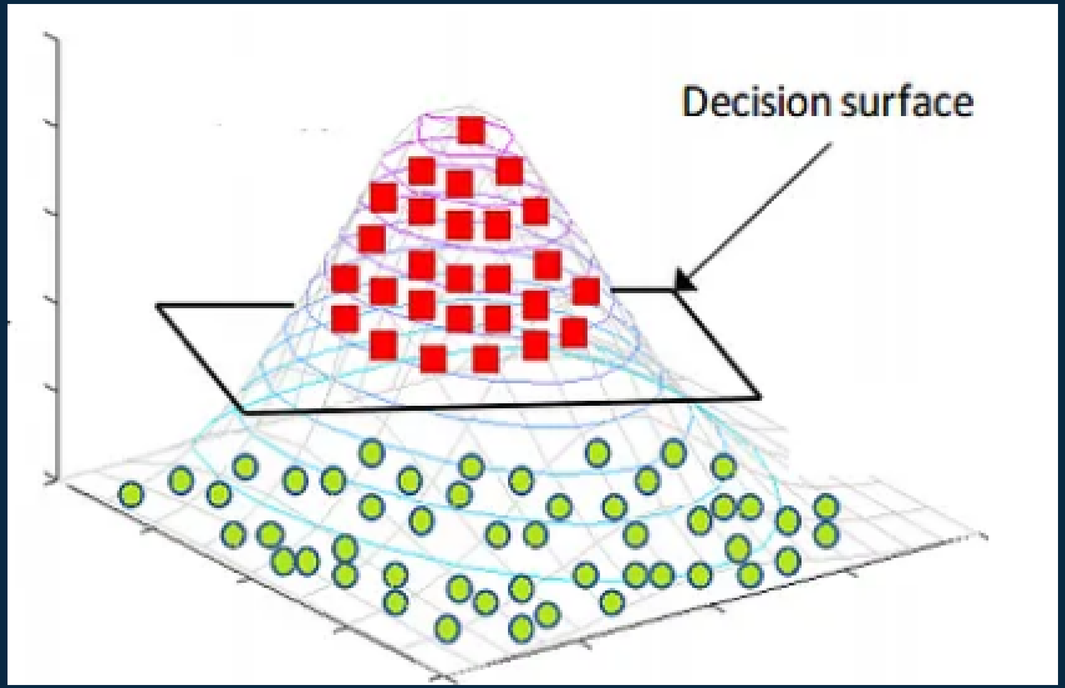
How do you seperate this data?



Hint:Always look for a diiferent angle!!

What is kernel trick?

- Kernel Trick is widely used in Support Vector Machines (SVM) model to bridge linearity and non-linearity.
- It converts non-linear lower dimension space to a higher dimension space thereby we can get a linear classification.
- So, we are projecting the data with some extra features so that it can convert to a higher dimension space.



Types of kernel in SVM?

- Linear Kernel
- Polynomial Kernel
- Gaussian Kernel
- Exponential Kernel
- Hyperbolic or the Sigmoid Kernel

Kernel Definition

- A function that takes as its inputs vectors in the original space and returns the dot product of the vectors in the feature space is called a *kernel function*
- More formally, if we have data $\mathbf{x}, \mathbf{z} \in X$ and a map $\phi : X \to \Re^N$ then

$$k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$$

is a kernel function