ML

LECTURE-19

BY
Dr. Ramesh Kumar Thakur
Assistant Professor (II)
School Of Computer Engineering

Support Vector Machine Algorithm

- Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is **used for Classification as well as Regression problems.**

- However, primarily, it is used for Classification problems in Machine Learning.

- The **goal of the SVM** algorithm is to **create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category** in the future.

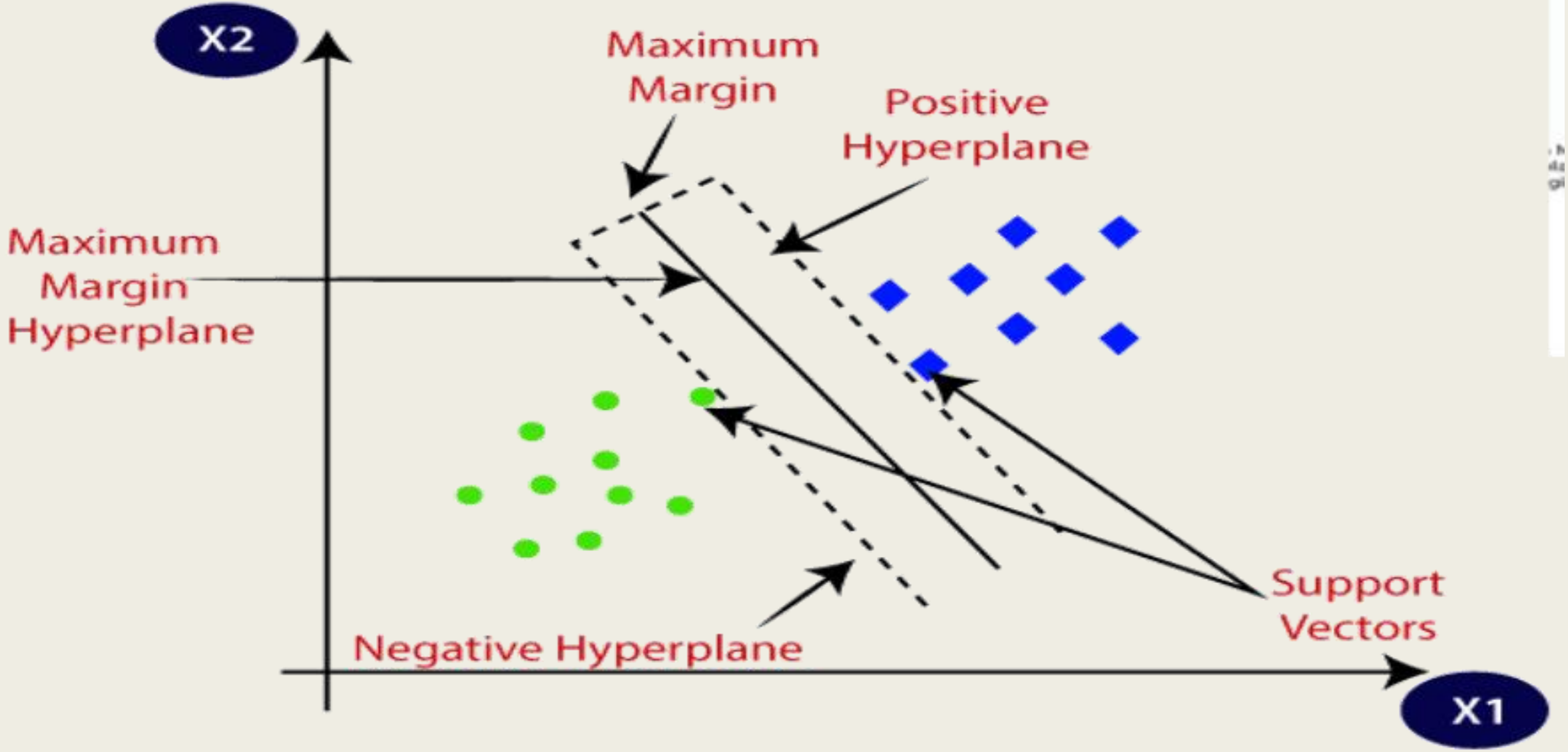
- This **best decision boundary is called a hyperplane.**

- SVM **chooses the extreme points/vectors that help in creating the hyperplane.**

- These **extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.**

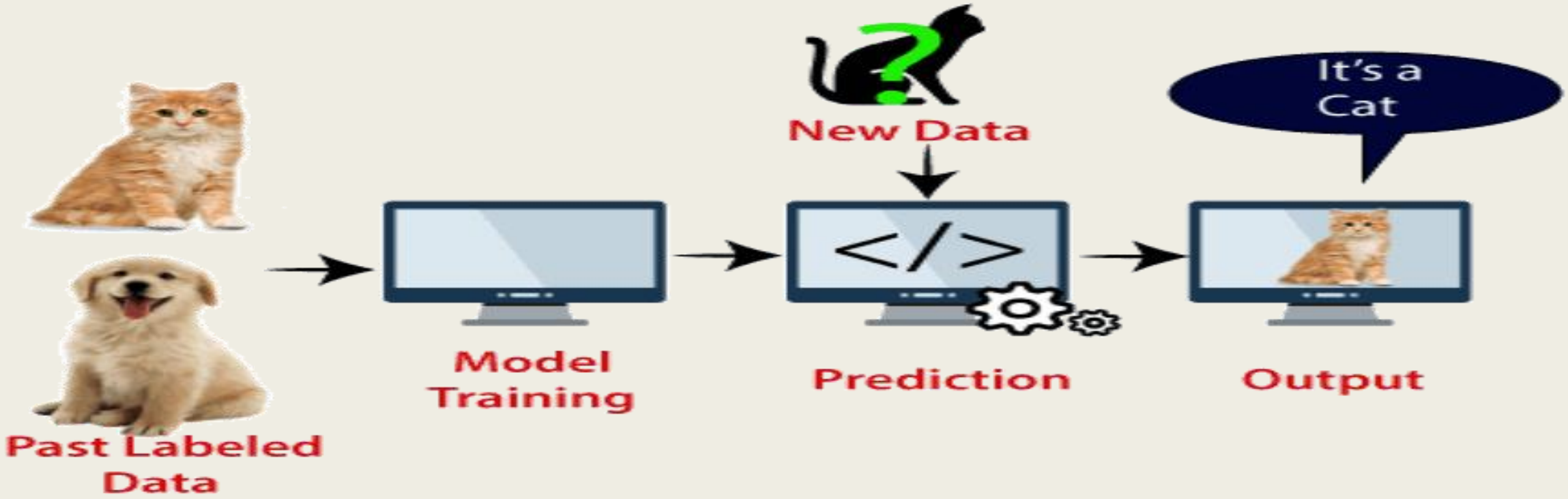
Support Vector Machine Algorithm

- Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



Support Vector Machine Algorithm

- **Example:** Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, a model can be created by using the SVM algorithm.
- We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature.
- So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog.
- On the basis of the support vectors, it will classify it as a cat. Consider the below diagram:



- SVM algorithm can be used for **Face detection, image classification, text categorization,** etc.

Types of SVM

- SVM can be of two types:
- **Linear SVM:** Linear SVM is **used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line,** then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

- **Non-linear SVM:** Non-Linear SVM is **used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line,** then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

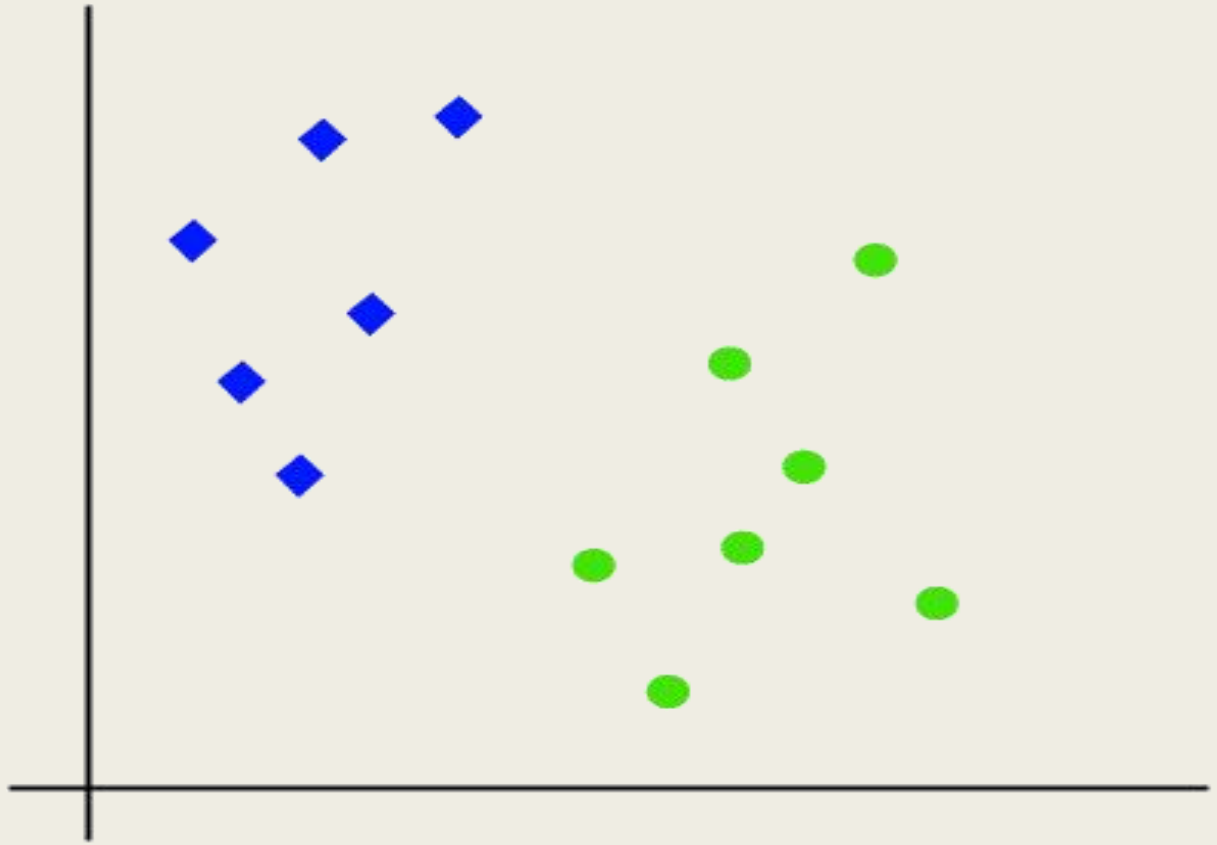
Hyperplane and Support Vectors in the SVM algorithm

- **Hyperplane:** There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points.
- This best boundary is known as the hyperplane of SVM.
- **The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features, then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.**
- We always **create a hyperplane that has a maximum margin, which means the maximum distance between the data points.**

- **Support Vectors:**
- The data points or vectors that are the **closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector.**
- Since **these vectors support the hyperplane, hence called a Support vector.**

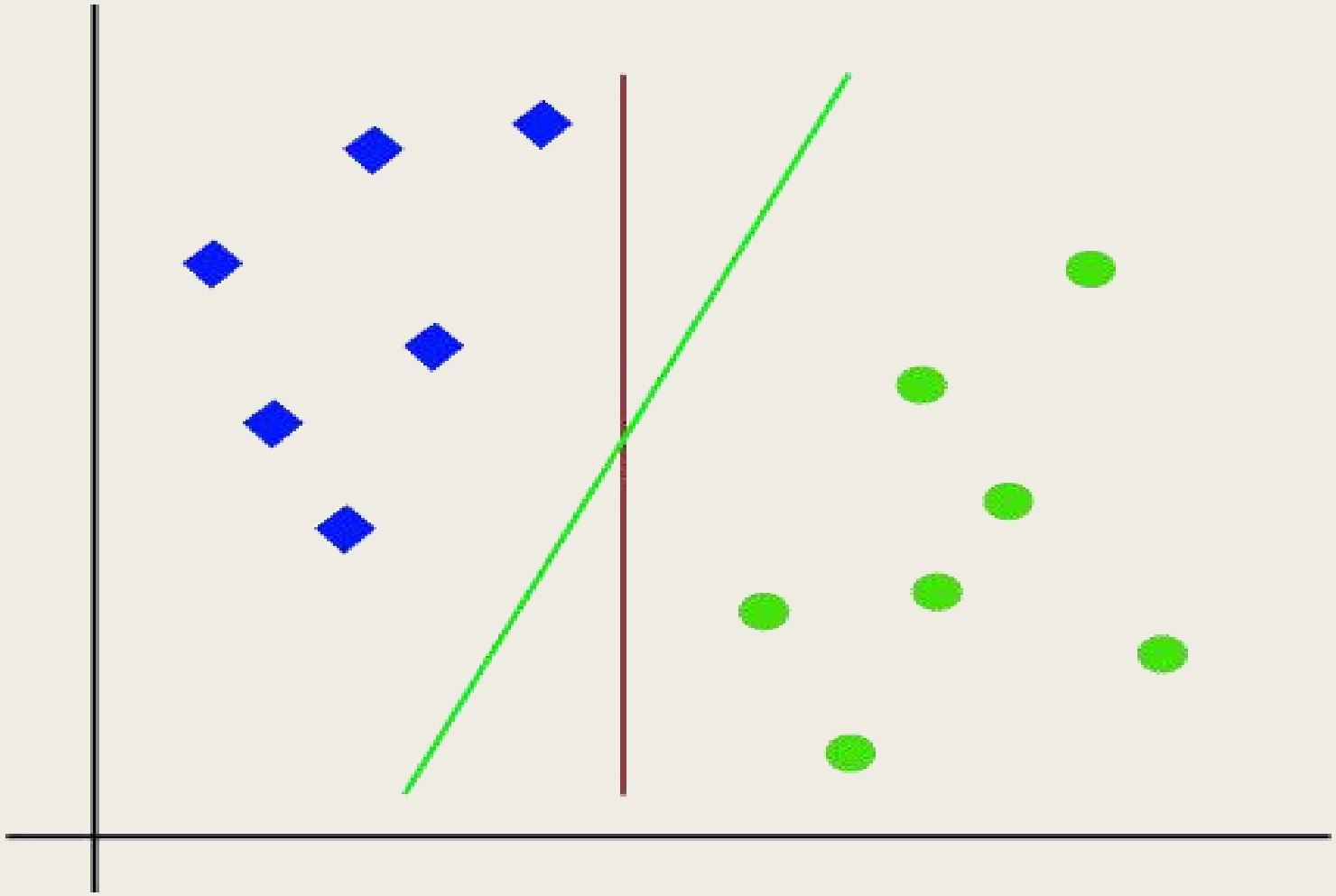
How does SVM works?

- **Linear SVM:**
- The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x1 and x2.
- We want a classifier that can classify the pair(x1, x2) of coordinates in either green or blue.
- Consider the below image:



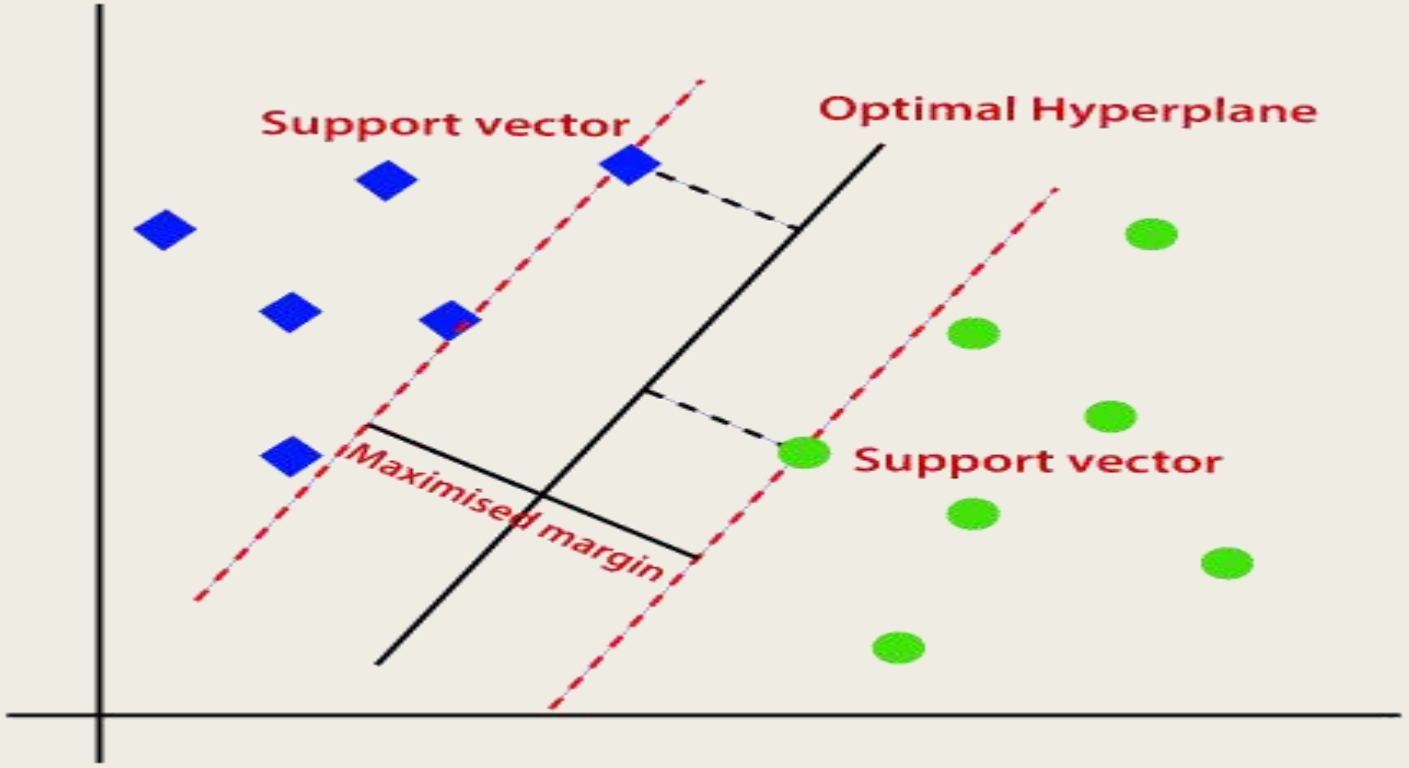
How does SVM works?

- So as it is 2-d space so by just using a straight line, we can easily separate these two classes.
- But there can be multiple lines that can separate these classes. Consider the below image:



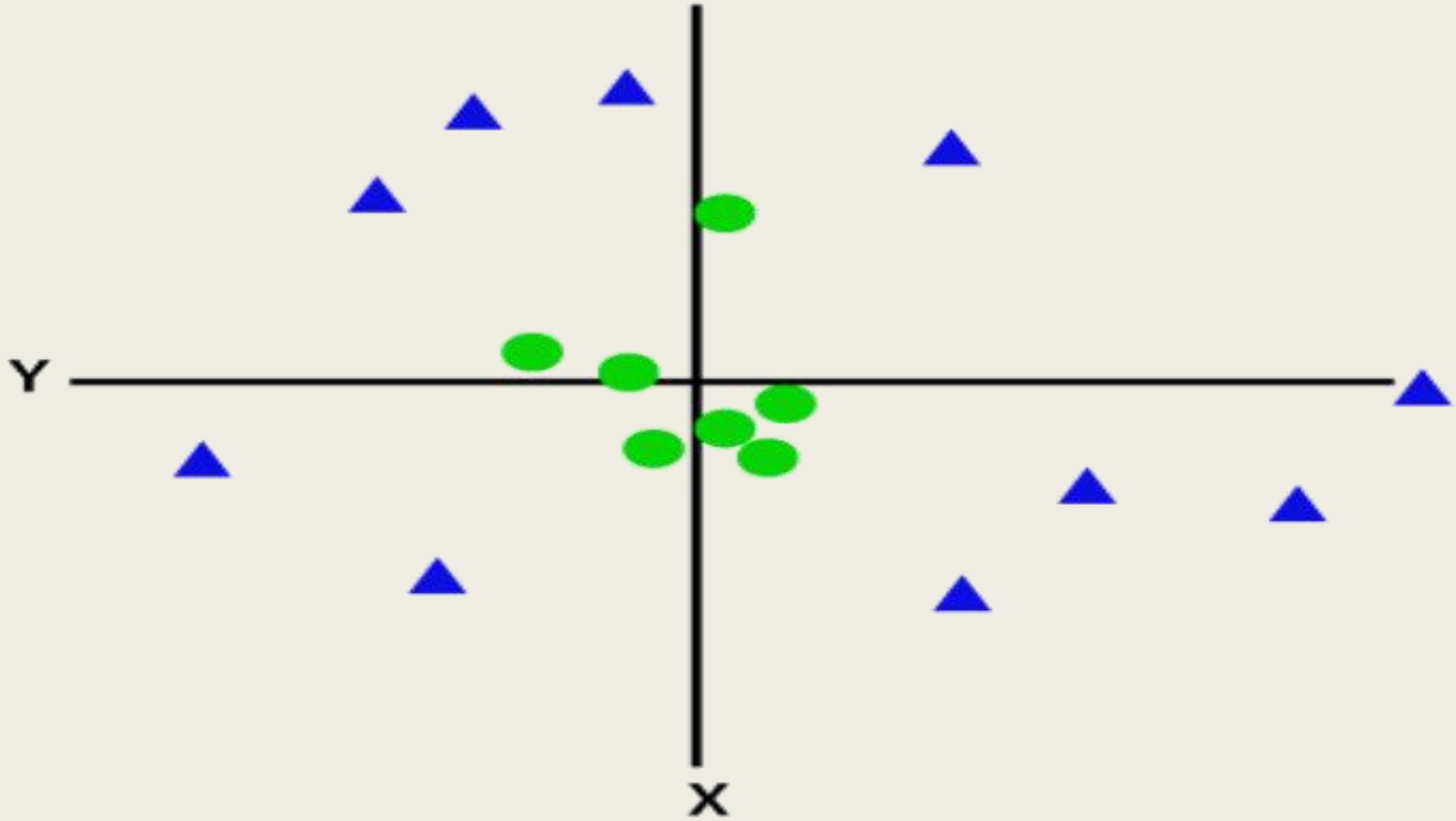
How does SVM works?

- The SVM algorithm helps to **find the best line or decision boundary; this best boundary or region is called as a hyperplane.**
- SVM algorithm **finds the closest point of the lines from both the classes. These points are called support vectors.**
- The **distance between the vectors and the hyperplane is called as margin.**
- And the **goal of SVM is to maximize this margin.**
- The **hyperplane with maximum margin is called the optimal hyperplane.**



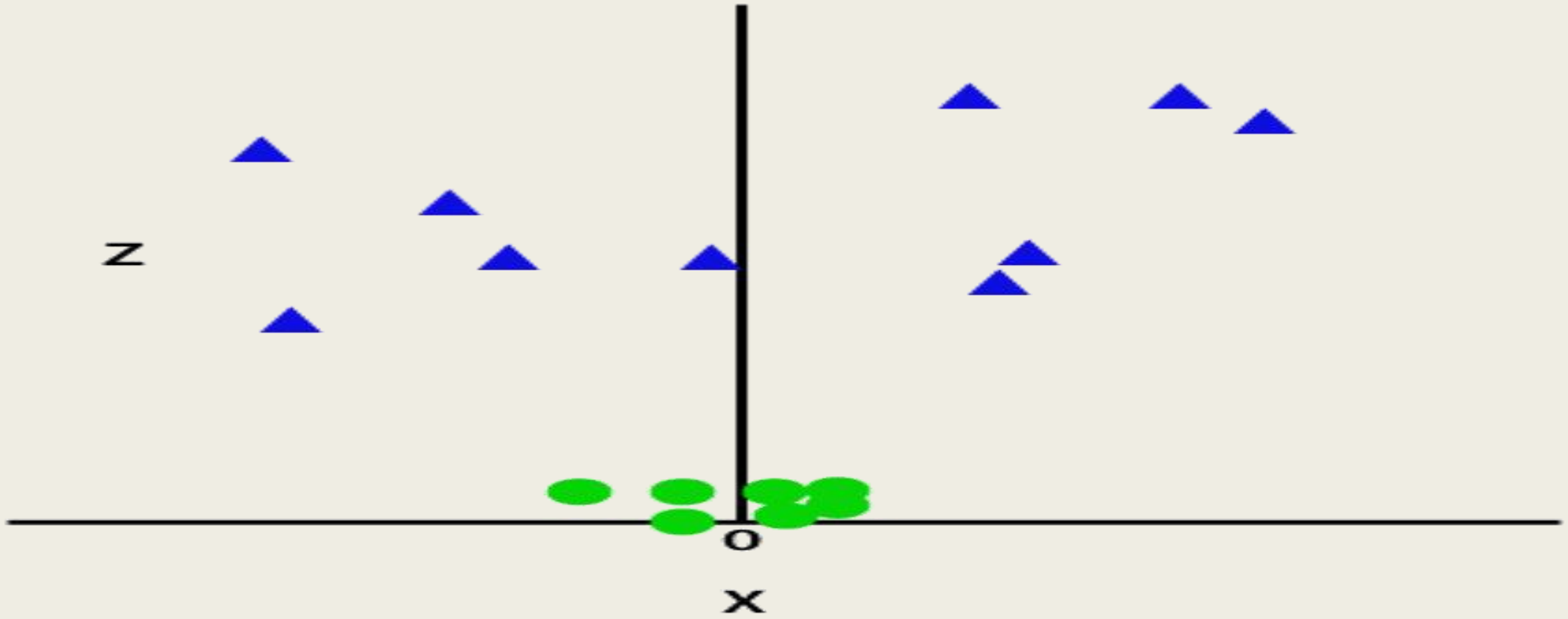
How does SVM works?

- **Non-Linear SVM:** If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line.
- Consider the below image:



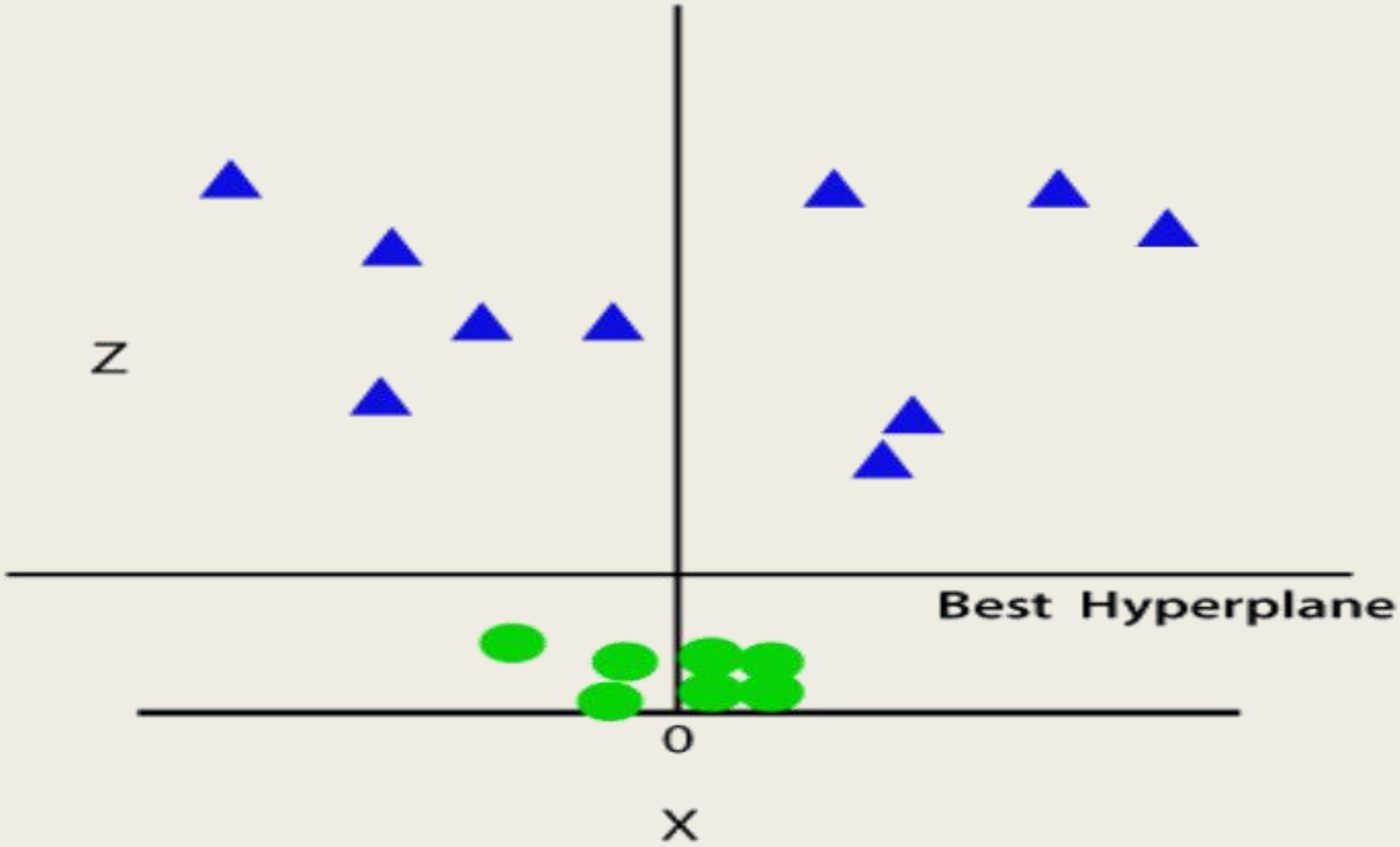
How does SVM works?

- So to separate these data points, we need to add one more dimension.
- For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as:
- $z=x^2 +y^2$
- By adding the third dimension, the sample space will become as below image:



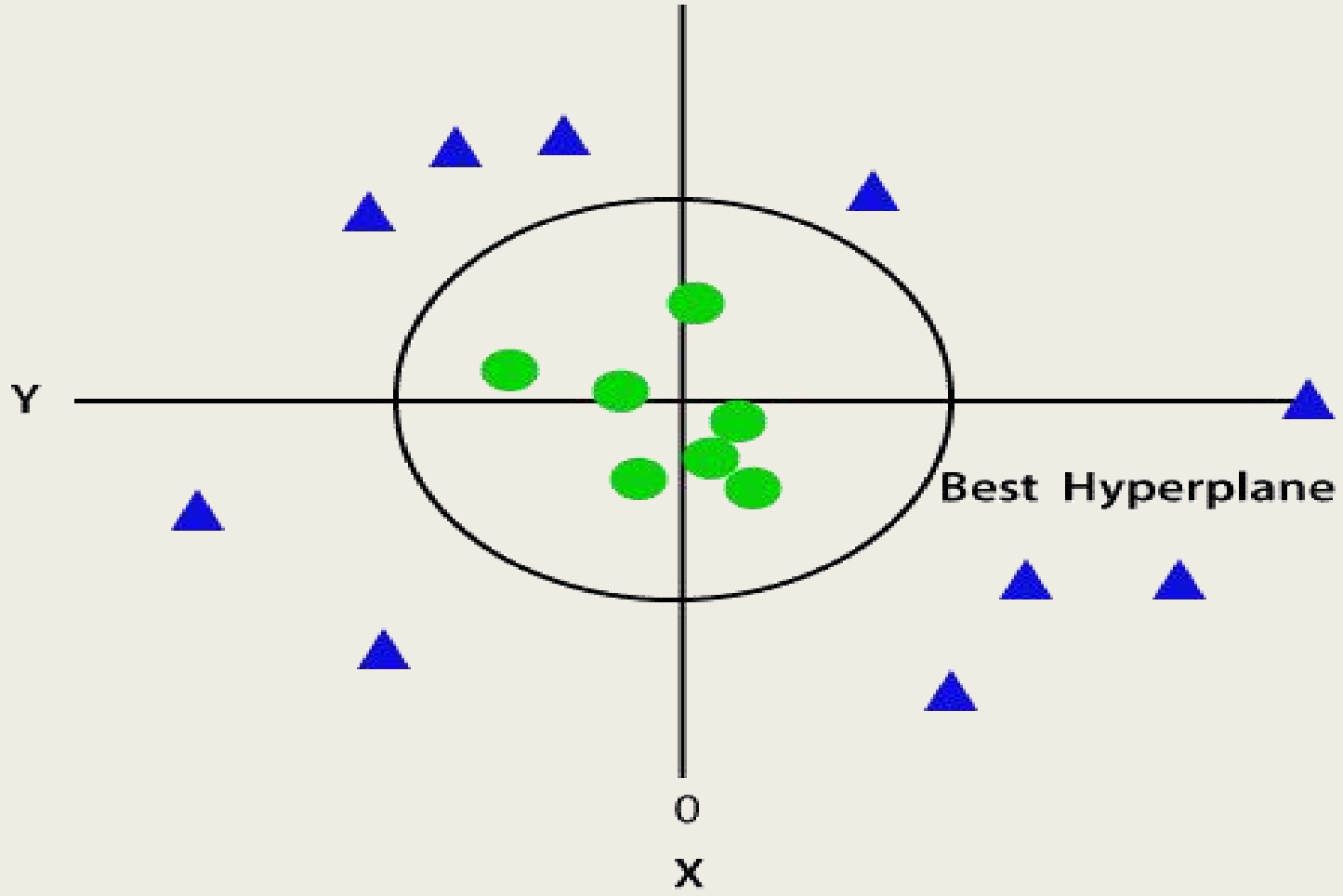
How does SVM works?

- So now, SVM will divide the datasets into classes in the following way.
- Consider the below image:



How does SVM works?

- Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis.
- If we convert it in 2d space with z=1, then it will become as:



- Hence we get a circumference of radius 1 in case of given non-linear data.

Numerical Example of Linear SVM

Suppose we are given the following positively labeled data points in $\Re^2$:

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

and the following negatively labeled data points in $\Re^2$ (see Figure 1):

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

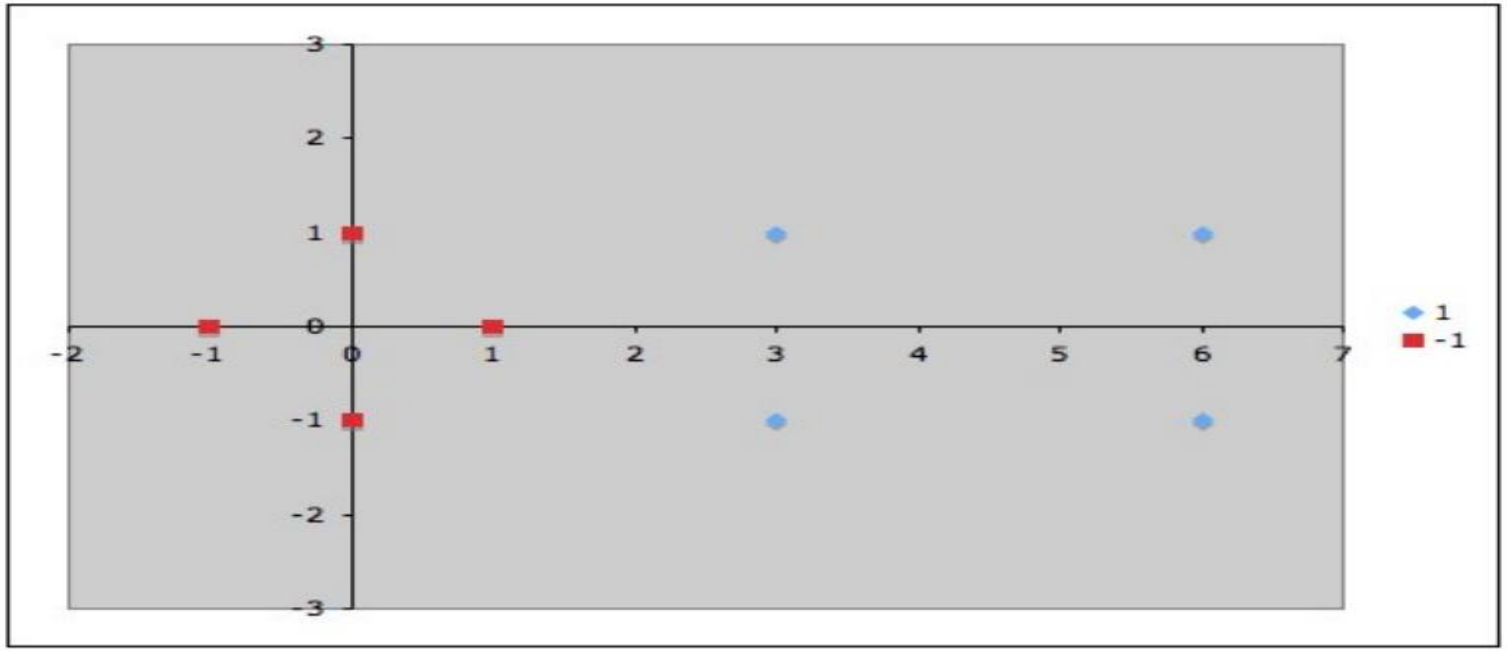


Figure 1: Sample data points in $\Re^2$. Blue diamonds are positive examples and red squares are negative examples.

Numerical Example of Linear SVM

We would like to discover a simple SVM that accurately discriminates the two classes. Since the data is linearly separable, we can use a linear SVM (that is, one whose mapping function $\Phi()$ is the identity function). By inspection, it should be obvious that there are three support vectors (see Figure 2):

$$\left\{ s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, s_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, s_3 = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \right\}$$

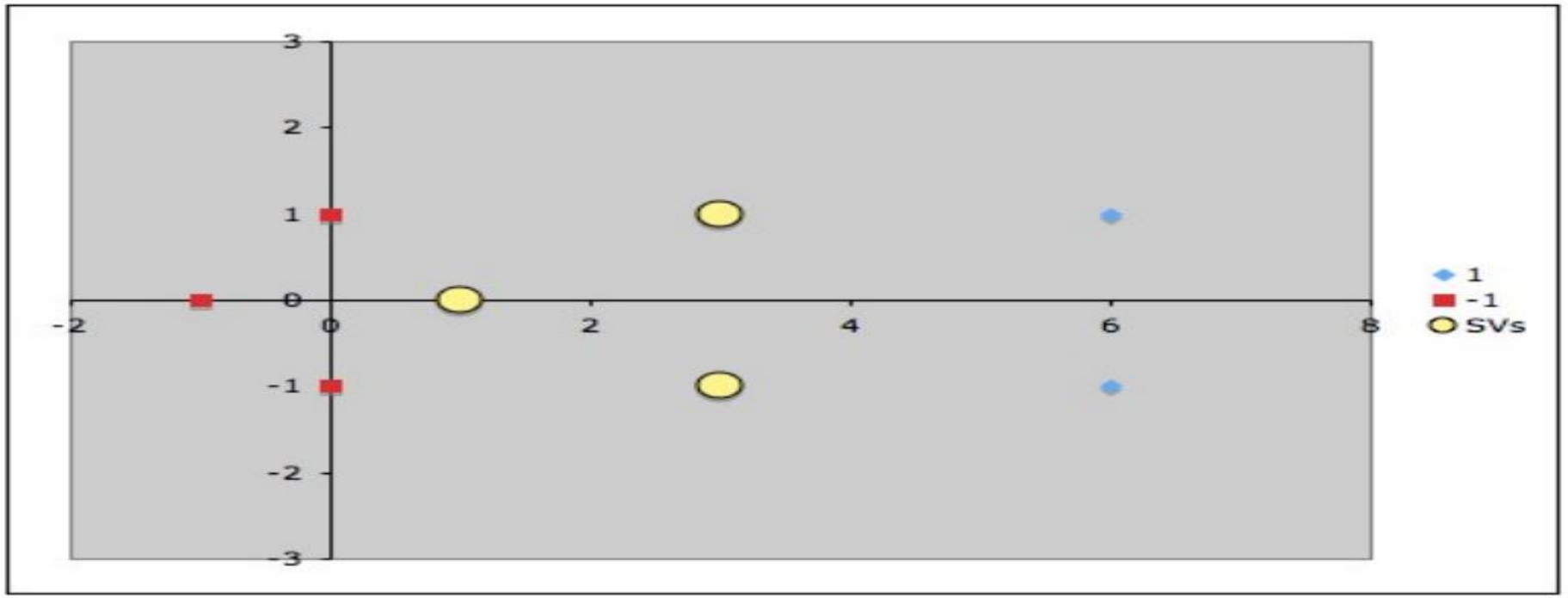


Figure 2: The three support vectors are marked as yellow circles.

Numerical Example of Linear SVM

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

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

- Similarly,

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

Numerical Example of Linear SVM

our task is to find values for the α_i such that

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

Since for now we have let $\Phi() = I$, this reduces to

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

Numerical Example of Linear SVM

$$\alpha_1 \tilde{s}_1 \cdot \tilde{s}_1 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_1 + \alpha_3 \tilde{s}_3 \cdot \tilde{s}_1 = -1 \qquad \alpha_1(1+0+1)+\alpha_2(3+0+1)+\alpha_3(3+0+1) = -1$$

$$\alpha_1 \tilde{s}_1 \cdot \tilde{s}_2 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_2 + \alpha_3 \tilde{s}_3 \cdot \tilde{s}_2 = +1 \qquad \alpha_1(3+0+1)+\alpha_2(9+1+1)+\alpha_3(9-1+1) = 1$$

$$\alpha_1 \tilde{s}_1 \cdot \tilde{s}_3 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_3 + \alpha_3 \tilde{s}_3 \cdot \tilde{s}_3 = +1 \qquad \alpha_1(3+0+1)+\alpha_2(9-1+1)+\alpha_3(9+1+1) = 1$$

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

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

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

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

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

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

$$\alpha_1 = -3.5$$

$$\alpha_2 = 0.75$$

$$\alpha_3 = 0.75$$

Numerical Example of Linear SVM

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

Finally, remembering that our vectors are augmented with a bias, we can equate the last entry in $\tilde{w}$ as the hyperplane offset b and write the separating hyperplane equation $y = wx + b$ with $w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $b = -2$. Plotting the line gives the expected decision surface (see Figure 4).

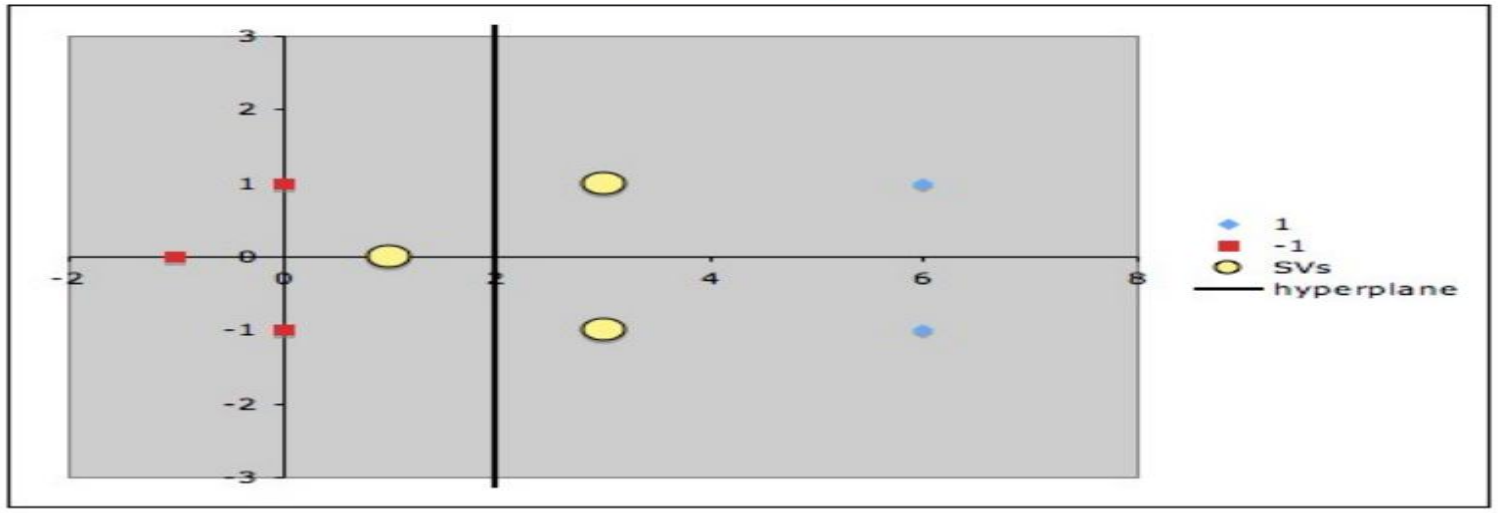
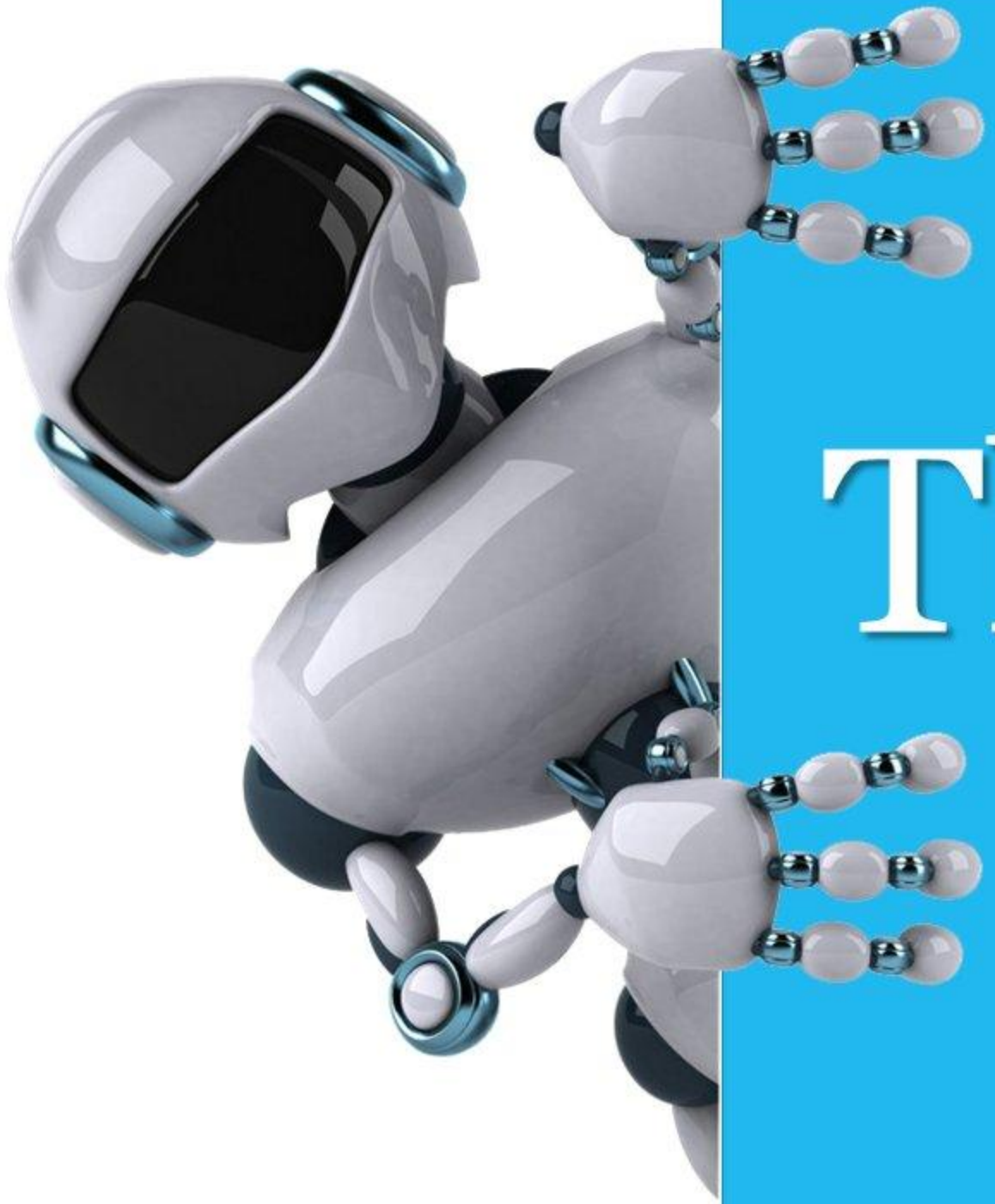


Figure 4: The discriminating hyperplane corresponding to the values $\alpha_1 = -3.5, \alpha_2 = 0.75$ and $\alpha_3 = 0.75$.



Thank you