MAXIMISING MARGIN IS A OPTIMISATION PROBLEM

1. Classify each data point correctly
2. Maximise the margin with subject to constraints as follows

$$M = \arg\max_{w} \frac{2}{\sqrt{(w^T \cdot w)}} \qquad (3)$$

subject to constraint in Equations below such that:

$$y^i = +1 \longrightarrow w \cdot x^i + b \geq 1 \qquad (4)$$

$$y^i = -1 \longrightarrow w \cdot x^i + b \leq 1 \qquad (5)$$

where Equation 4 and Equation 5 can written as:

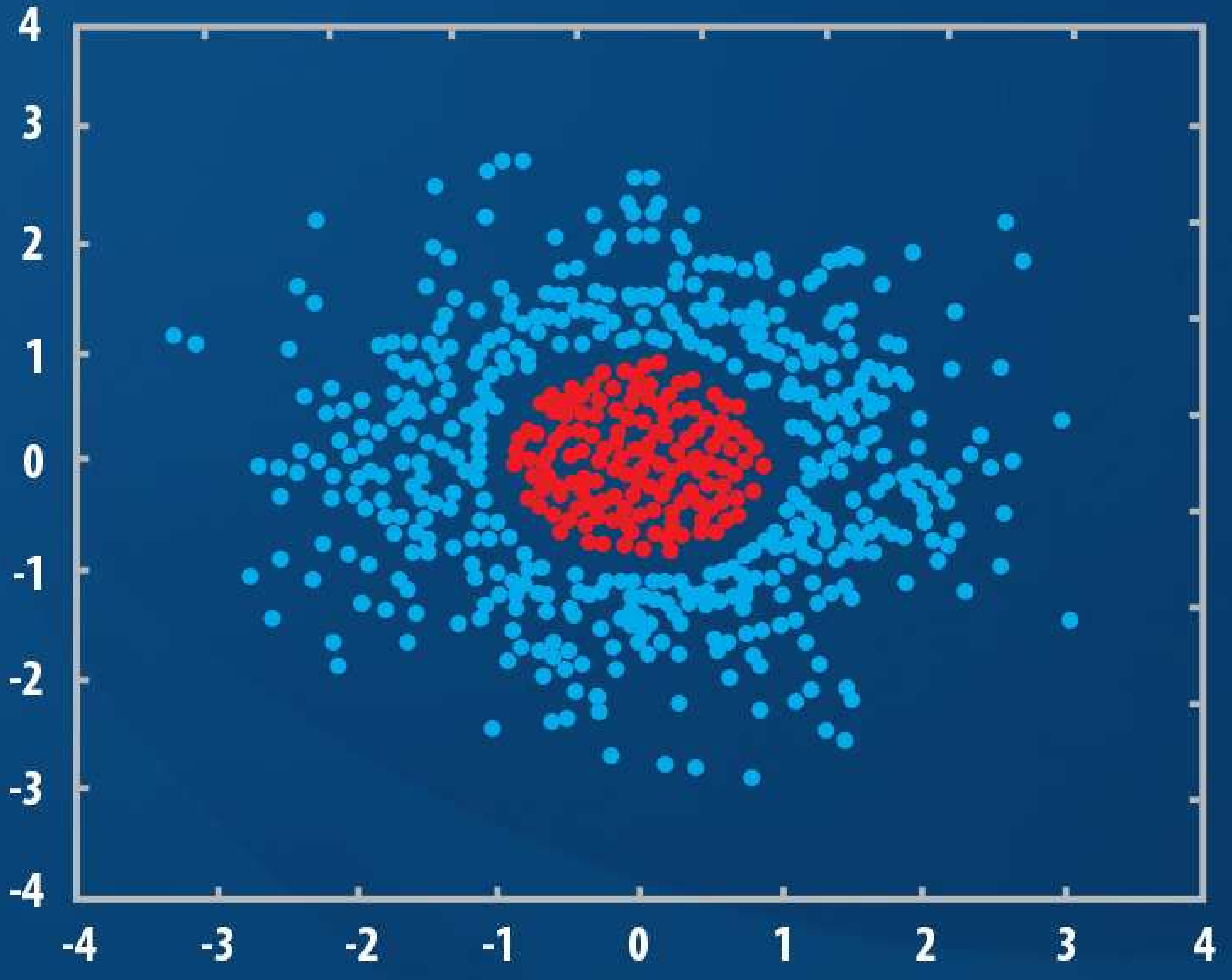
$$y^i (w \cdot x^i + b) \geq +1 \qquad (6)$$

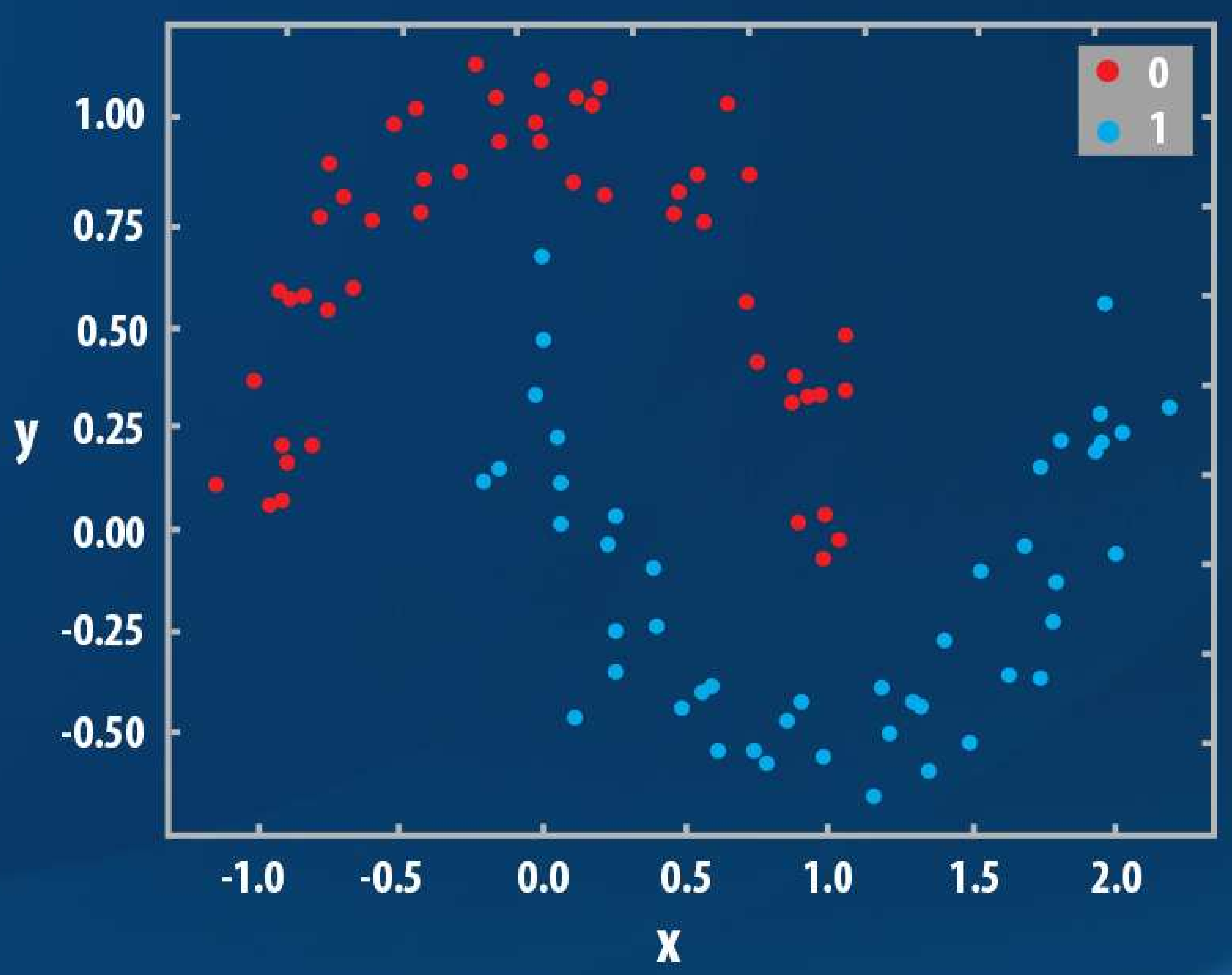
It is optimisation problem solved using Lagrange multipliers *method.*

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

SUPPORT VECTOR MACHINES FOR NON-LINEARLY SEPARABLE PROBLEMS

Examples of Non-Linearly separable problems





Kernel Trick

- Used to solve non-Linearly separable problems
- In SVM, the kernel trick means transforming data into higher dimensional space that has a dividing margin between classes of data.

HOW SUPPORT VECTOR MACHINES DEALS WITH NON-LINEARLY SEPARABLE PROBLEMS?

SVM uses "Kernel trick". This trick maps data to some higher-dimensional feature space.

Refer Example below:

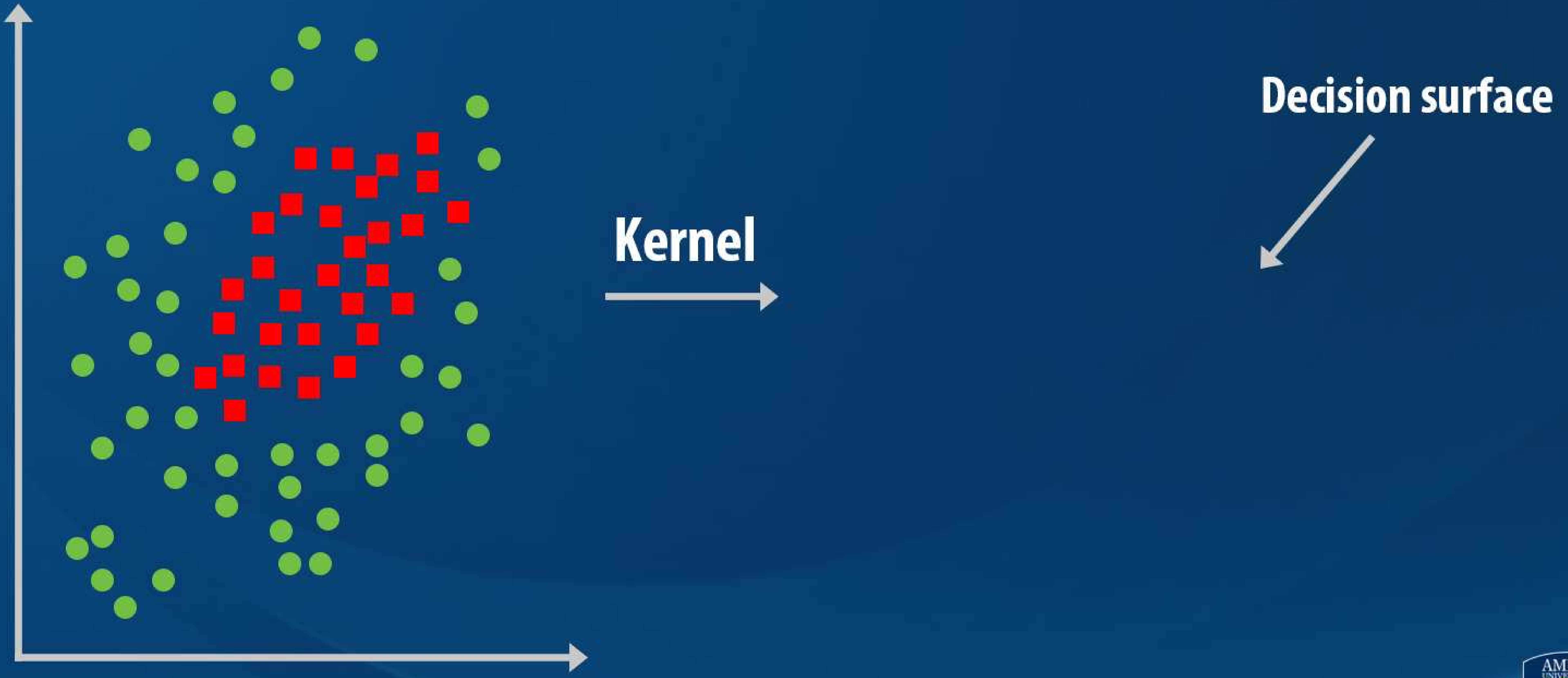


Figure 6: Kernel trick of SVM

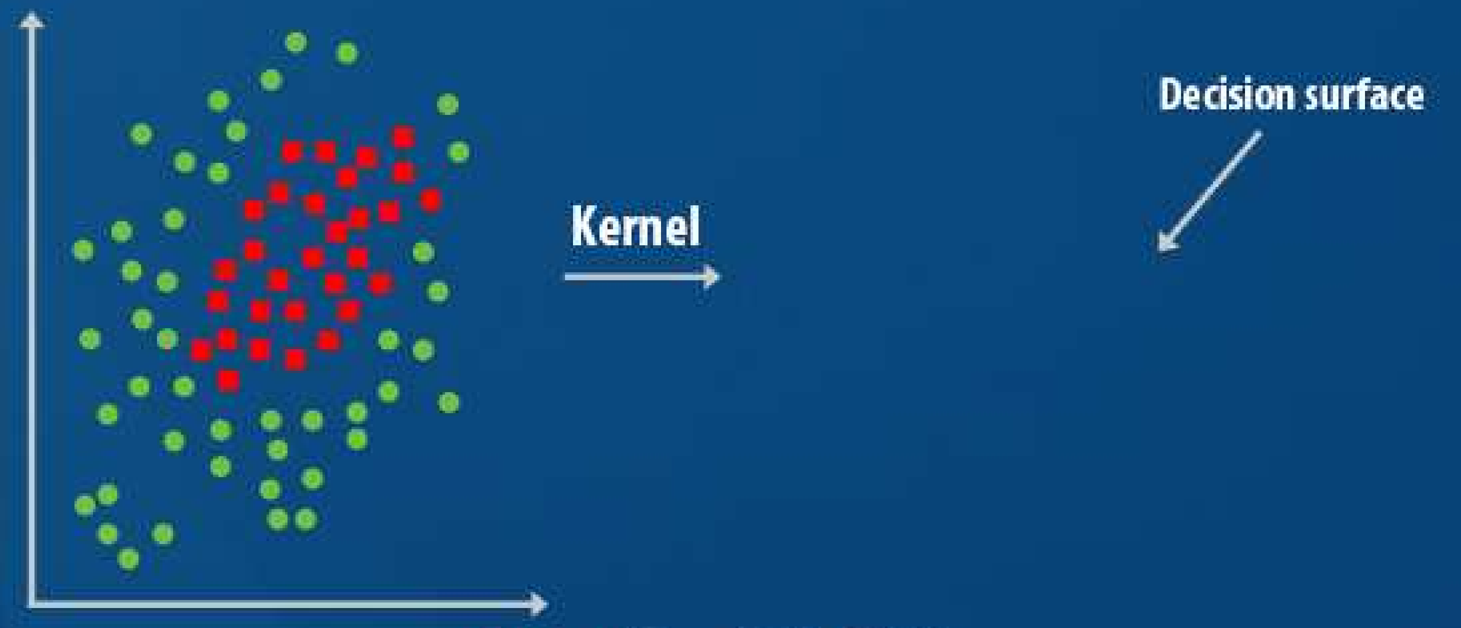
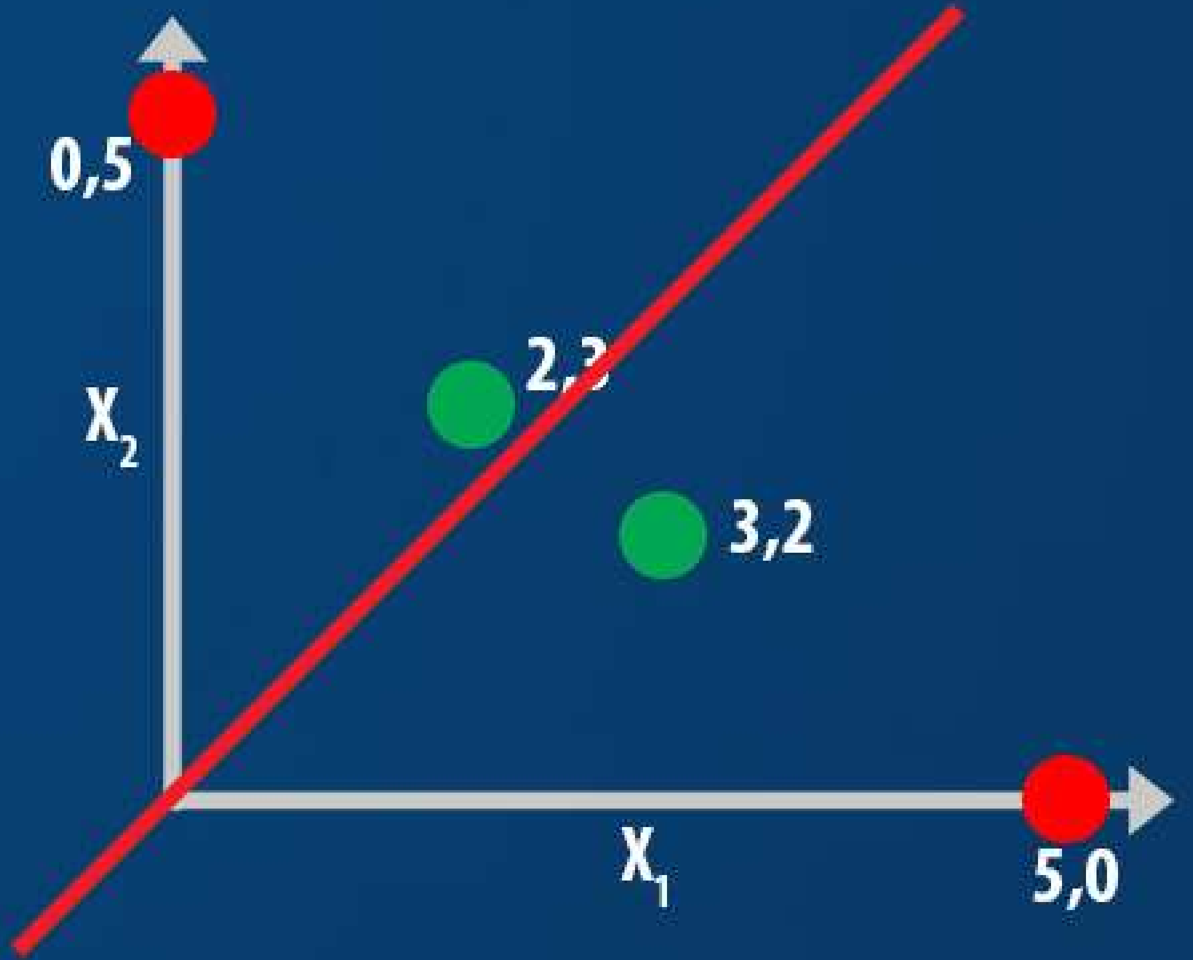


Figure 6: Kernel trick of SVM

EXAMPLE ILLUSTRATION OF KERNEL TRICK IN SVM

Consider two dimensional data set with 4 data points as shown in figure below.



Which equation best represents the data set above? It is $X_1 \times X_2$

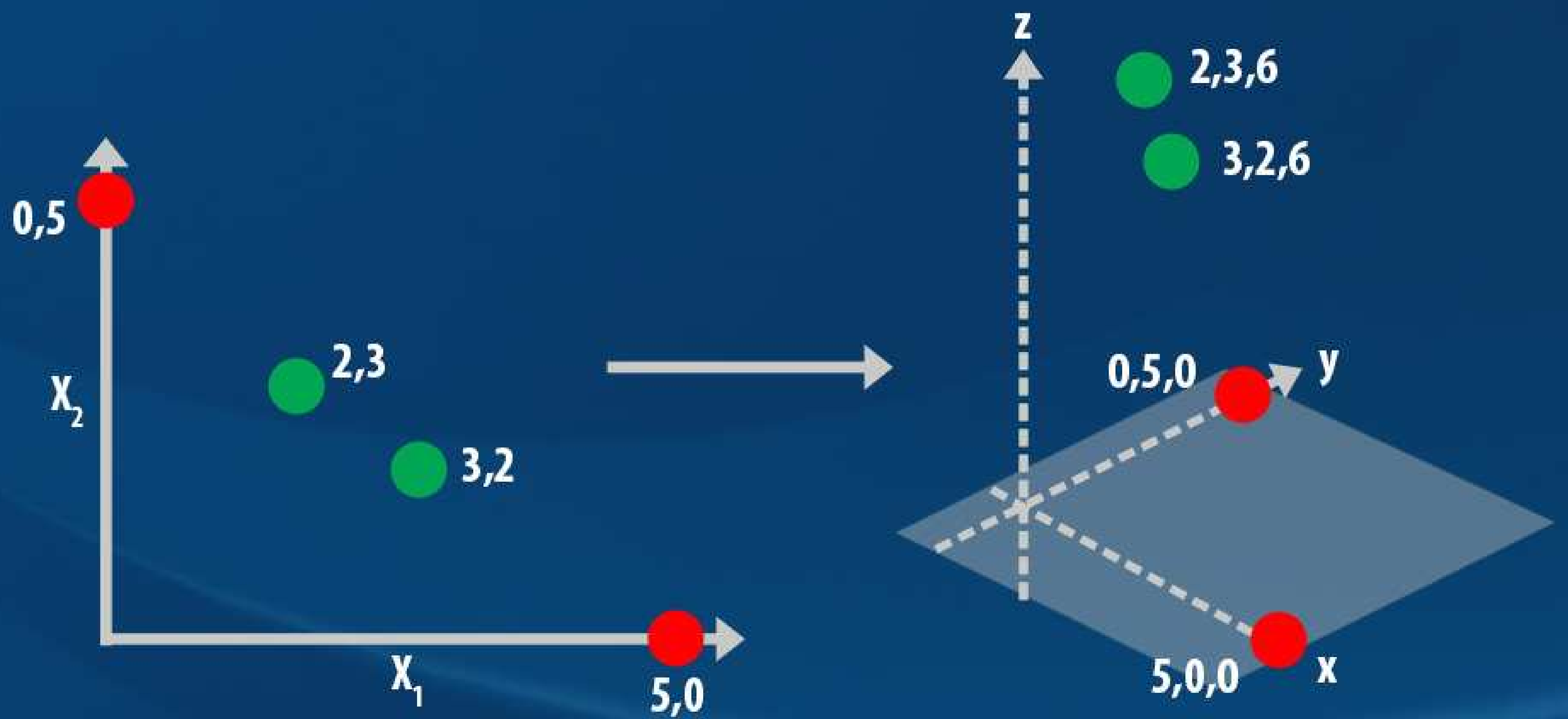
Data Points	(0,5)	(3,2)	(2,3)	(5,0)
$X_1 \times X_2$	0	6	6	0
$X_1 + X_2$	5	5	5	5

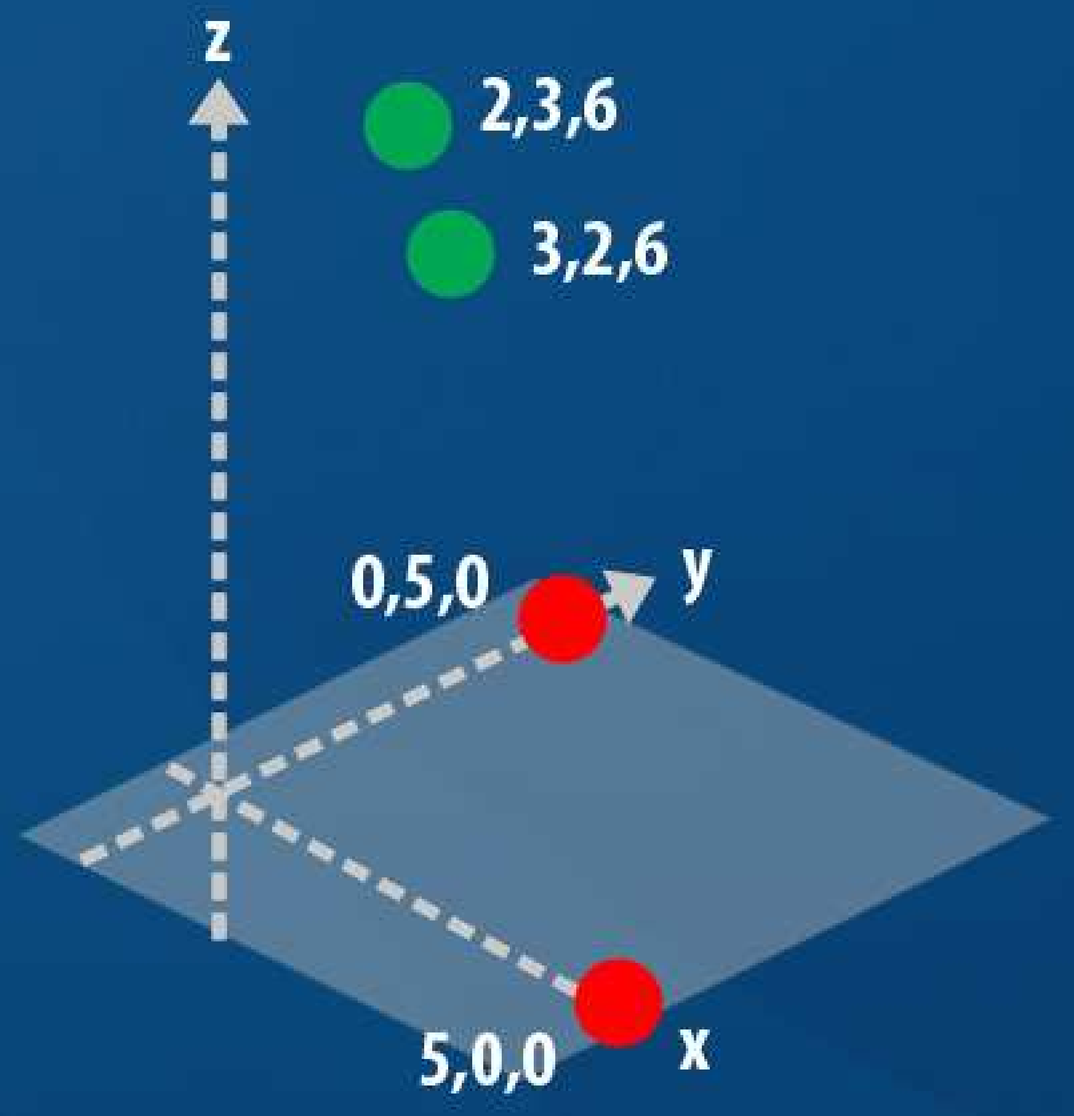
Data Points	(0,5)	(3,2)	(2,3)	(5,0)
$X_1 \times X_2$	0	6	6	0
$X_1 + X_2$	5	5	5	5

EXAMPLE ILLUSTRATION OF KERNEL TRICK IN SVM

Consider two dimensional data set with 4 data points as shown in figure below transformed in high dimension using Kernel.

Data points in 2D	(0,5)	(3,2)	(2,3)	(5,0)
$\downarrow (x,y) \rightarrow (x, y, x \times y) = (x, y, z)$				
Data points in higher dimensions	(0, 5, 0)	(3, 2, 6)	(2, 3, 6)	(5, 0, 0)




z
2,3,6
3,2,6
0,5,0
y
5,0,0
x

SVM KERNEL I

If the data set is not linearly separable, we can map the samples X into a feature space of higher dimensions:

$$f = \phi(x) \qquad (7)$$

The decision function in the new space becomes:

$$f(x) = \phi(x)^T \cdot w + b \qquad (8)$$

where, w

$$w = \sum_{i=1}^{n} \alpha_i y_i \varphi(x_i) \qquad (9)$$

Here αi is Lagrange multipliers and yi are class labels. By using Equation 9, Equation 8 becomes,

$$f(x) = \left(\sum_{i=1}^{n} \alpha_i y_i \cdot \varphi(x)^T \cdot \varphi(x_i) \right) + b \qquad (10)$$

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

SVM KERNEL II

- The function $\varphi(x)$ is a kernel-induced implicit mapping. All that is required is the inner product of the vectors in the new space.
- **Definition:** A kernel is a function that takes two vectors x_i and x_j as arguments and returns the value of the inner product of their images $\varphi(x_i)$ and $\varphi(x_j)$.

$$K(x_1,x_2) = \varphi(x_i)^T \cdot \varphi(x_2)$$

- The learning algorithm in the kernel space can be obtained by replacing all inner products in the learning algorithm in the original space with the kernels:

$$f(x) = \varphi(x)^T w + b = \sum_{j=1}^{m} \alpha_j y_j K(x, x_j) + b$$

TYPES OF SVM KERNELS

There are several Kernels to choose from depending upon the data set. Below are the names:

1. Linear :

$$K(x_i, x_j) = x^T . x_j \qquad (11)$$

2. Polynomial or order p

$$K(x_i, x_j) = (1 + x^T . x_j)^p \qquad (12)$$

3. Gaussian (radial-basis function network)

$$K(x_i, x_j) = \exp - \frac{(||x - x||)^2}{2^{\sigma 2}} \qquad (13)$$

$$f(x) = \left(\sum_{i=1}^{n} \alpha_i y_i \cdot \varphi(x)^T \cdot \varphi(x_i) \right) + b \qquad (10)$$

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

HOW TO CHOOSE BETWEEN SVM KERNELS?

1. Linear kernel:

- Most suitable in cases where data is linearly separable.
- However, if data is not linearly separable then this kernel is not a good choice.

2. Gaussian or RBF kernel:

- The most used kernel function is RBF. It is particularly used in situations where we do not have any prior knowledge of the data.
- It best suits to non-linearly separable problems.

3. Polynomial kernel:

- It is less useful and computationally expensive over RBF.