SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade
ASPproved by AICTE, New Delhi & Affiliated to Anna University, Chennai

Redesigning Common Mind & Business Towards Excellence

Build an Entrepreneurial Mindset Through Our Design Thinking FrameWork

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

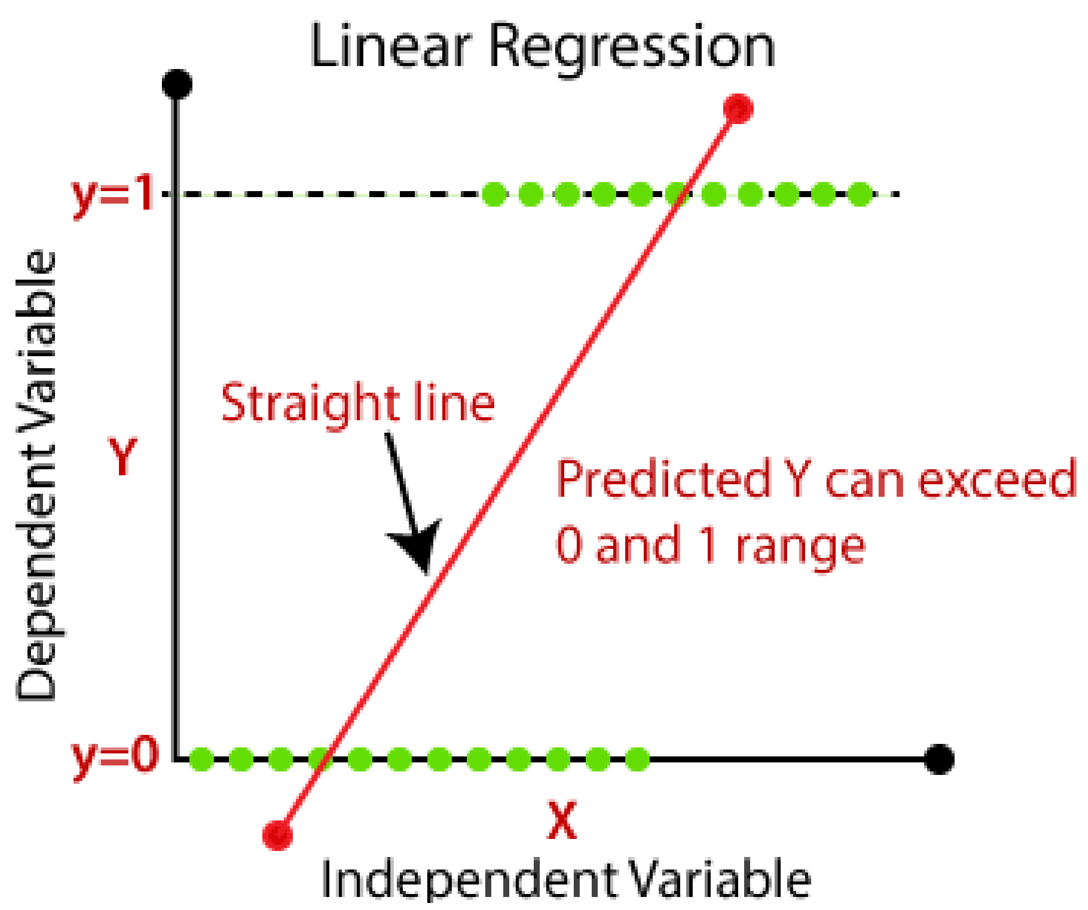
23AMB201 - MACHINE LEARNING

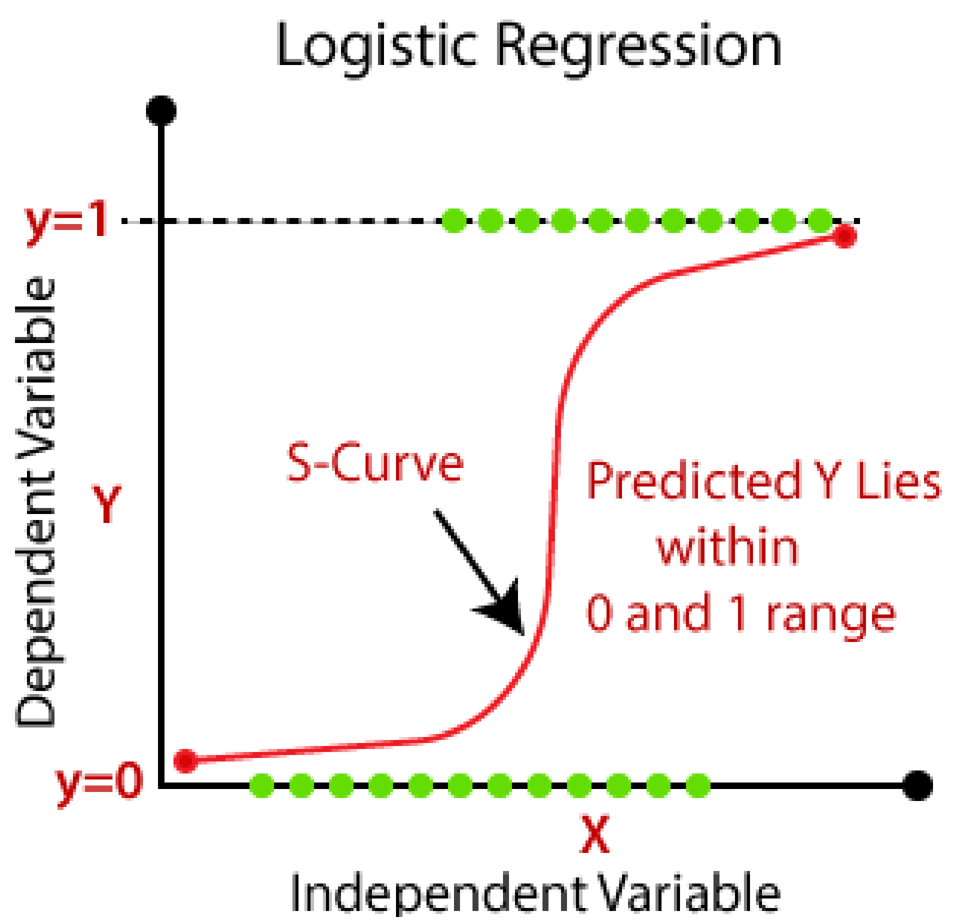
II YEAR IV SEM

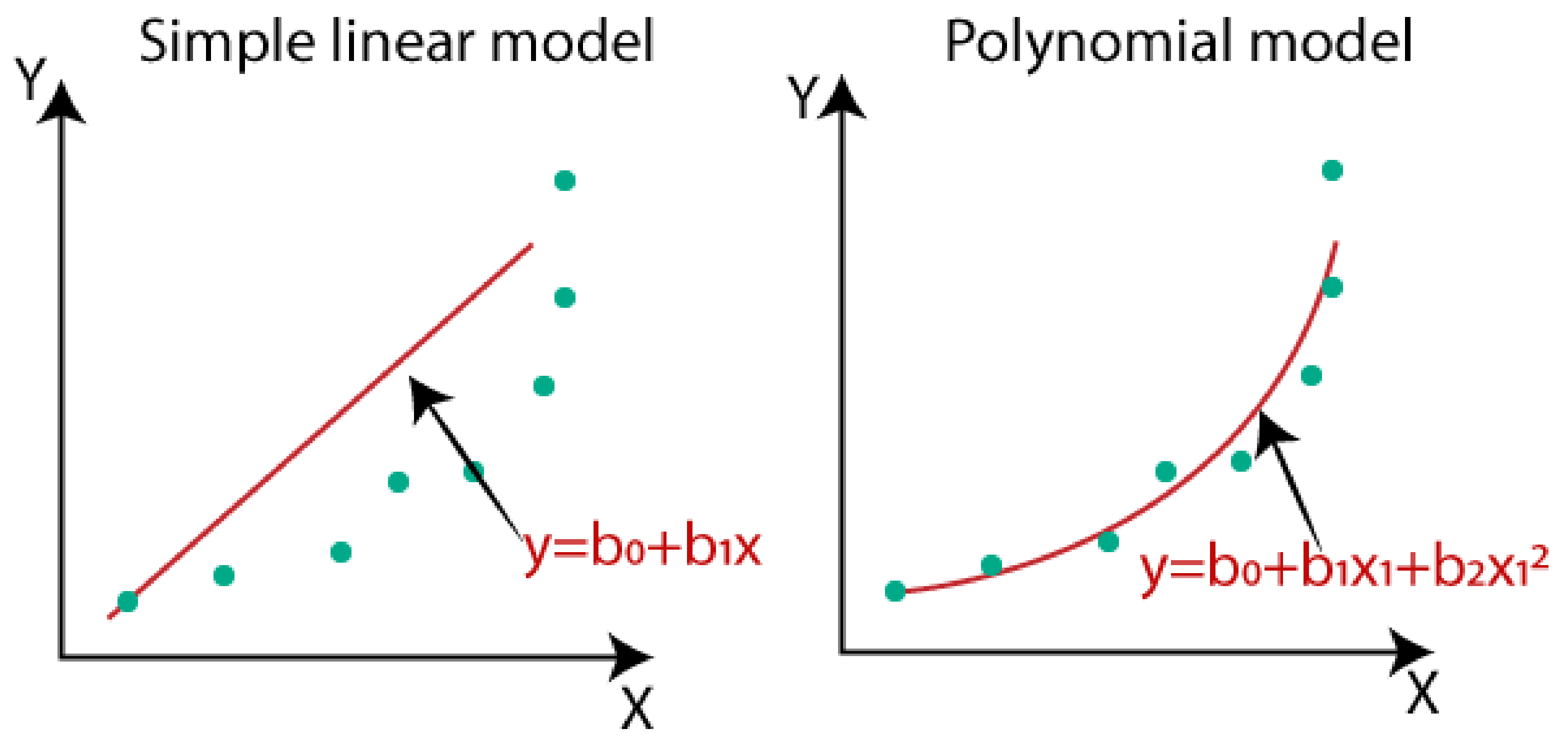
UNIT II – SUPERVISED LEARNING ALGORITHMS

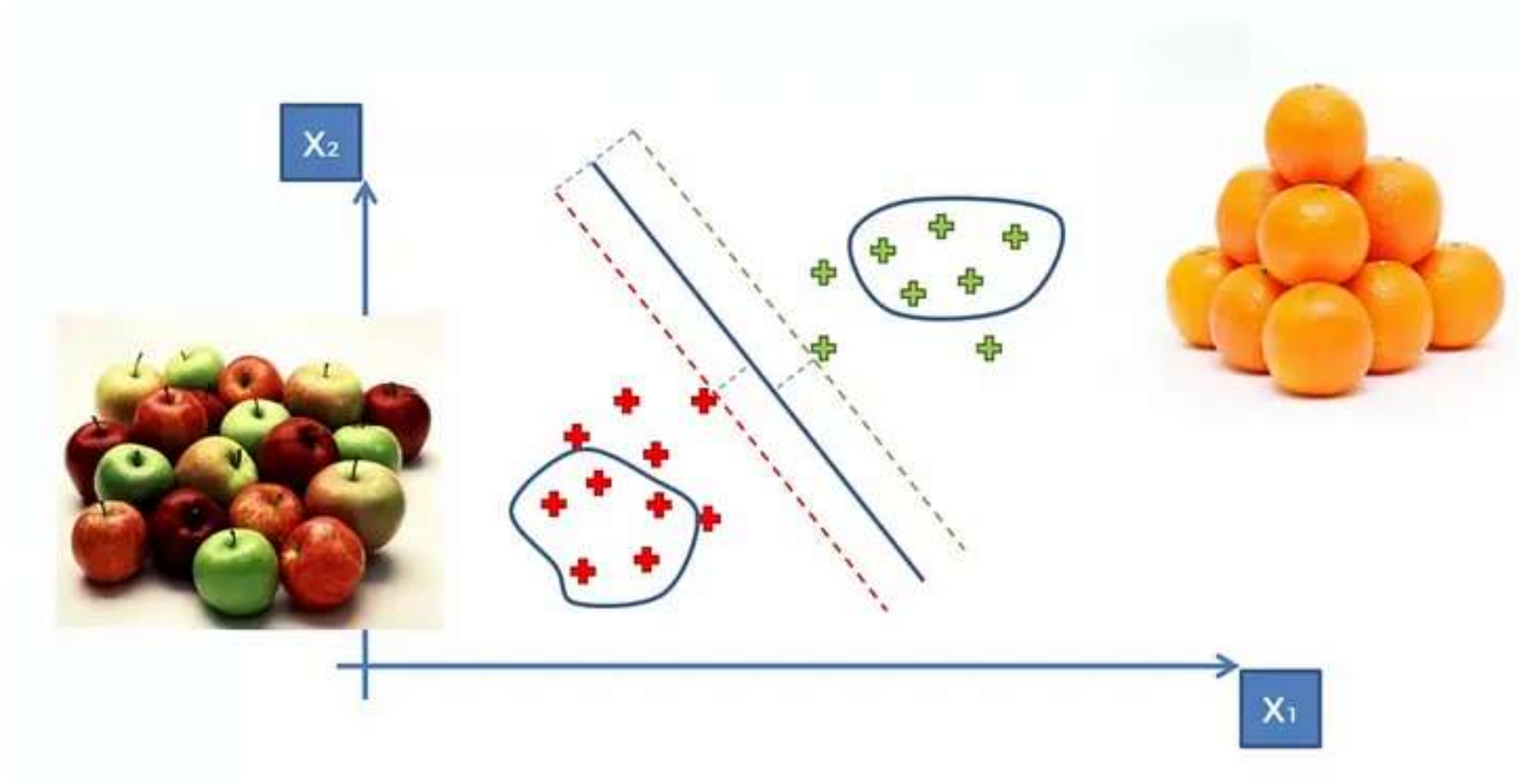
TOPIC 7 – SVM and Hyperparameter tuning

Recall


Linear Regression
y=1
y=0
Dependent Variable
Y
Straight line
Predicted Y can exceed
0 and 1 range
X
Independent Variable

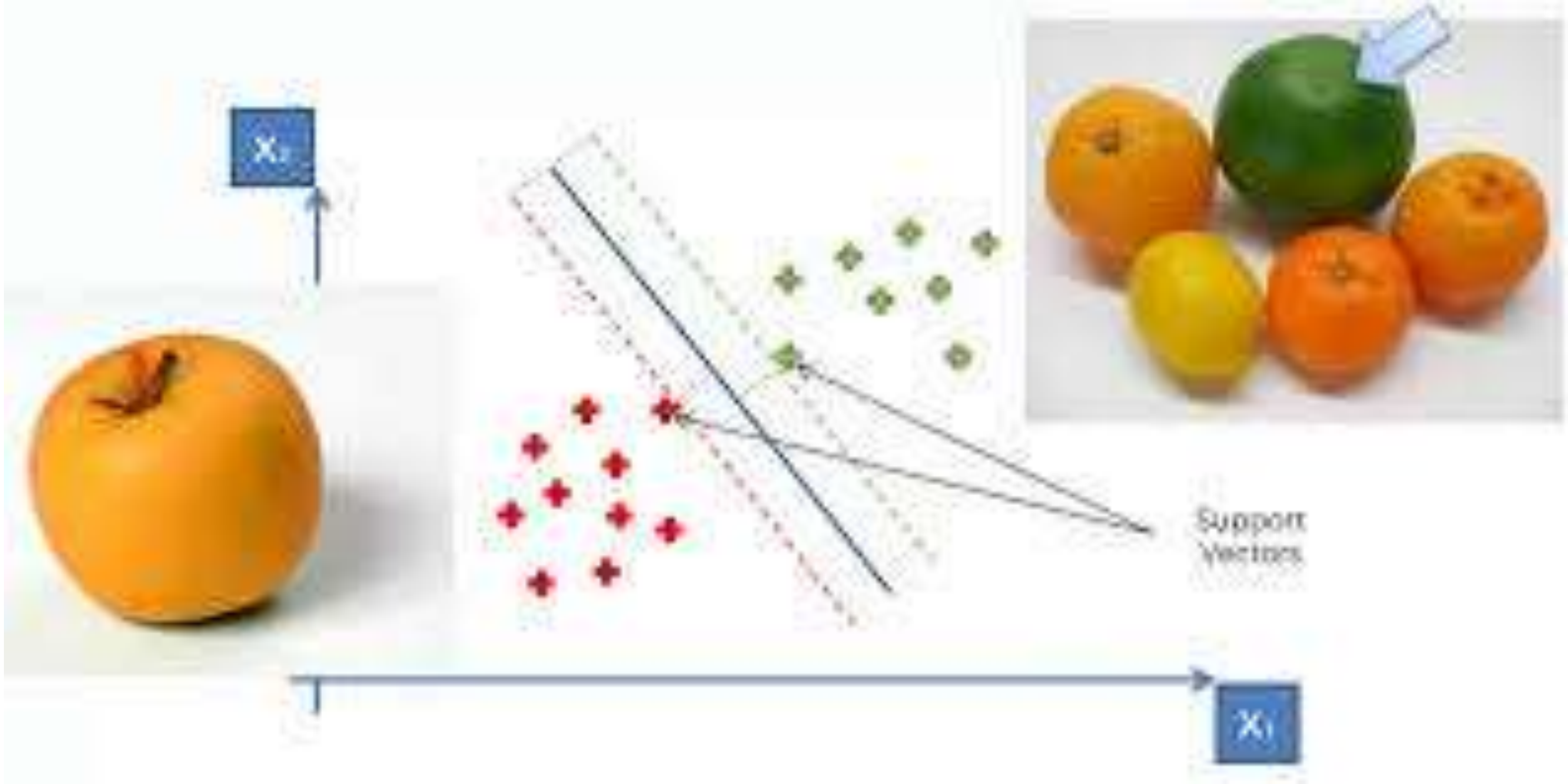

Logistic Regression
y=1
y=0
Dependent Variable
Y
S-Curve
Predicted Y Lies
within
0 and 1 range
X
Independent Variable


Simple linear model
Y
X
y=b0+b1x
Polynomial model
Y
X
y=b0+b1x1+b2x1²



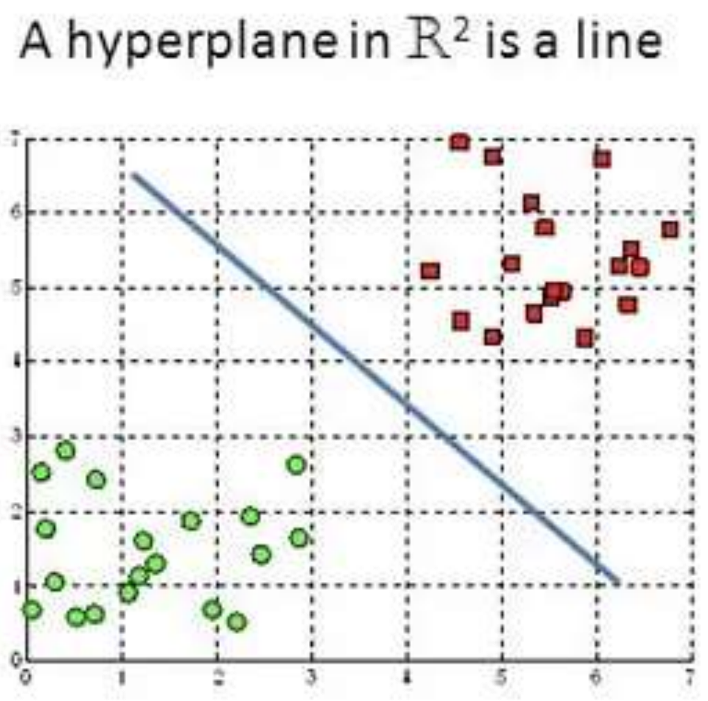
A machine would try to learn from the apples that are very much like apple so it would know what an apple and the same for orange.

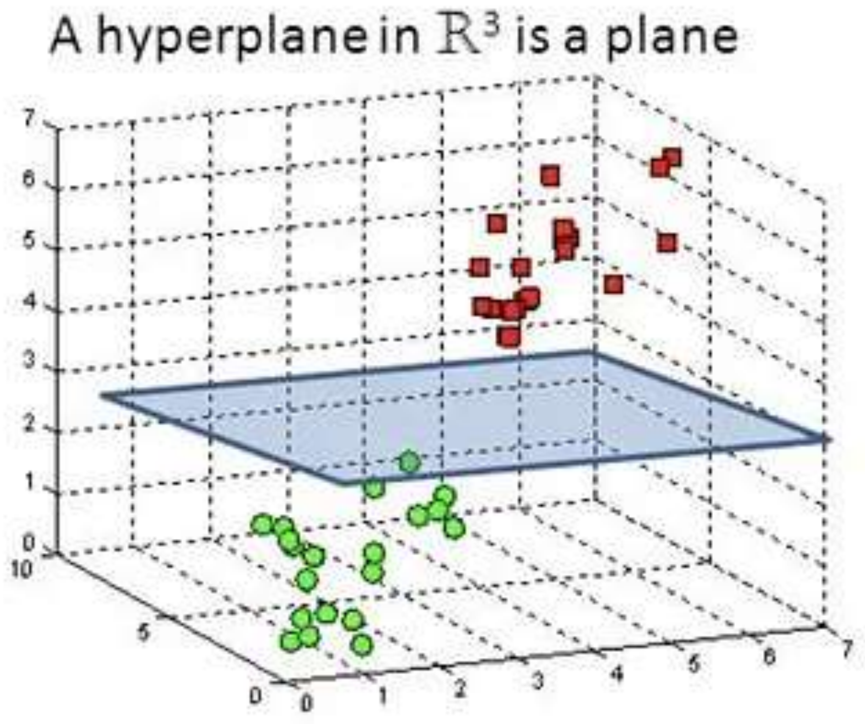
Look at apples that are very much like orange and oranges that are very much like an apple

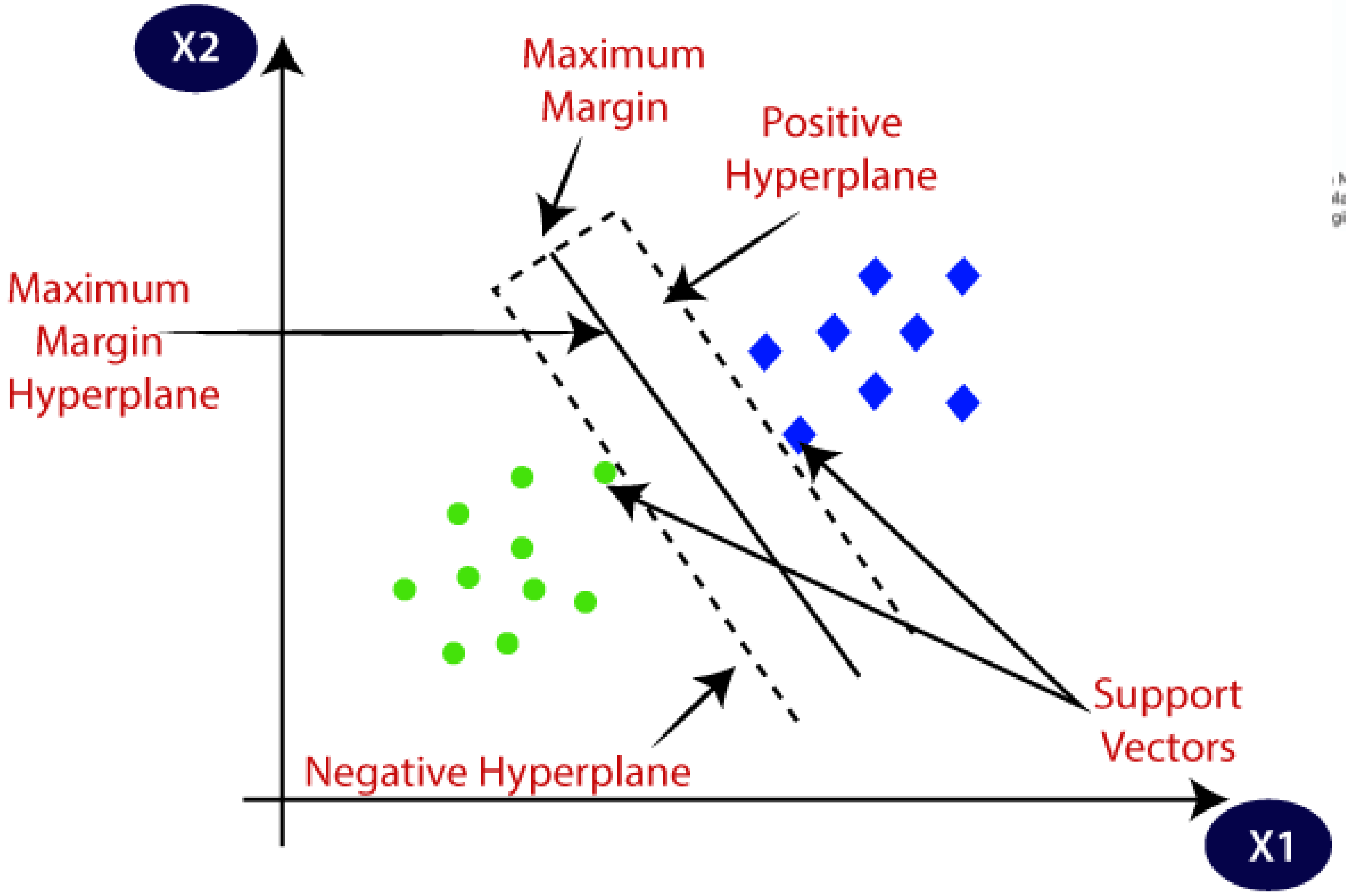


Why Support Vector Machine?

SVM searches the classifier line through maximum margin which means this line is drawn equidistant from both support vectors and the sum of margin from support vectors to the line is maximum.



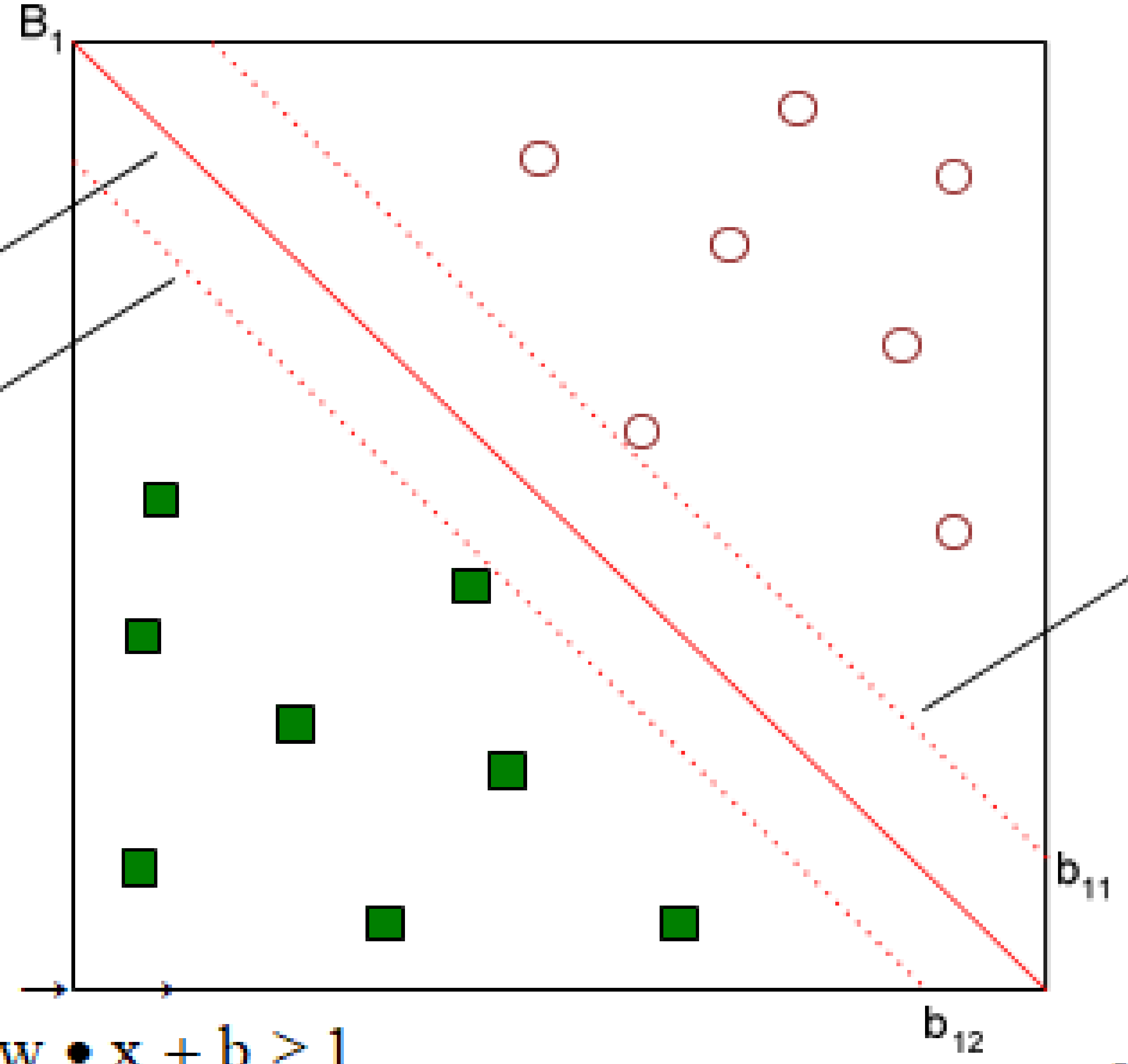




SVM: Calculation

$\vec{w} \bullet \vec{x} + b = 0$

$w \bullet x + b = -1$



$\vec{w} \bullet \vec{x} + b = +1$

$$f(\vec{x}) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x} + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x} + b \leq -1 \end{cases}$$

$$\text{Margin} = \frac{2}{\|\vec{w}\|}$$

Important terms in Support Vector Machine

Support Vector Machines are a set of supervised learning methods used for classification, regression, and outliers detection problems.

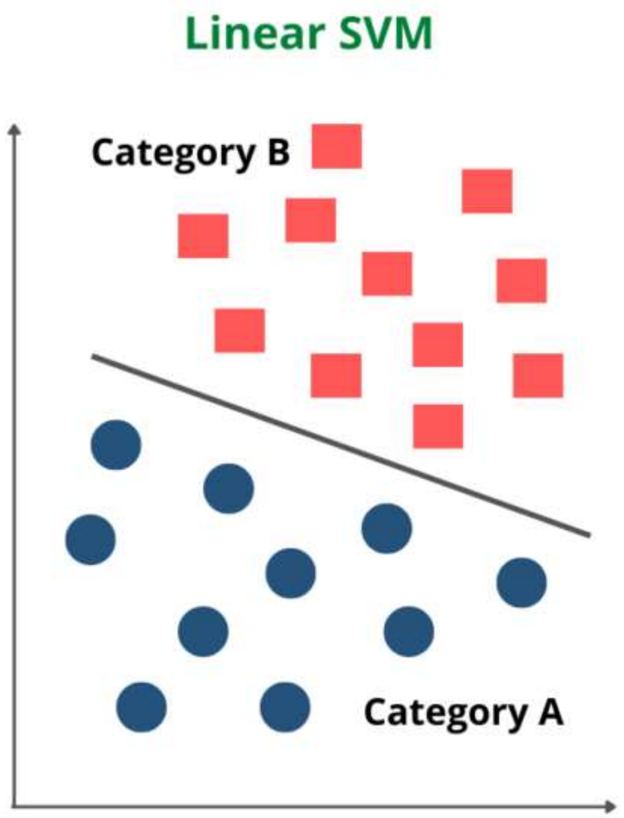
1. **Hyperplane** – It is a decision plane or space which is divided between a set of objects having different classes.
2. **Support Vectors** – Datapoints that are closest to the hyperplane are called support vectors. The separating line will be defined with the help of these data points.
3. **Kernel** – A kernel is a function used in SVM for helping to solve problems. They provide shortcuts to avoid complex calculations.
4. **Margin** – It may be defined as the gap between two lines on the closet data points of different classes. A large margin is considered a good margin and a small margin is considered as a bad margin.

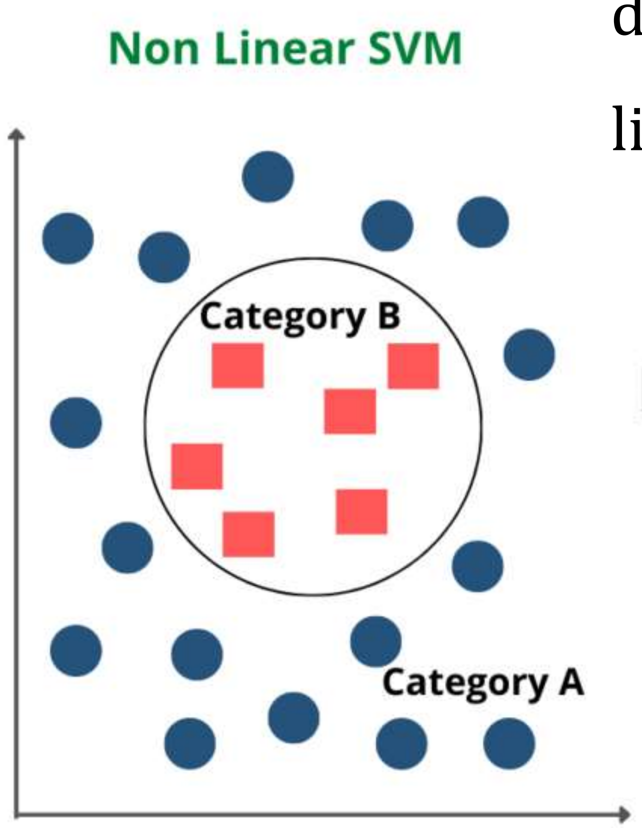
Types of Support vector Machine

Linear SVM is used in the case of linearly separable data. It means if a dataset can be classified into two classes by using a straight line, then such data is termed linearly separable data, and we can use the Linear SVM classifier.

$$K(x, x_i) = \text{sum}(x * x_i)$$

The product between two vectors is the sum of the multiplication of each pair of input values.





It means if a straight line cannot classify a dataset, then such data is termed non-linear data

$$K(x, x_i) = \exp(-\text{gamma} * \text{sum}(x - x_i^2))$$

Radial Basis Kernel(RBF) is a kernel function that is used to find a non-linear classifier or regression line. RBF kernel, mostly used in SVM classification, maps input space in indefinite dimensional space.

1️⃣ Hard Margin SVM

- **Strict separation**: Assumes data is **perfectly separable.**
- **No misclassification allowed**: Every data point must be on the correct side of the decision boundary.
- **Maximizes the margin** while ensuring 100% correct classification.

When to Use?

✅ If data is **clean** and **linearly separable** (no overlap between classes).

❌ **Not suitable** for noisy or overlapping data.

Mathematical Constraint:

For a given dataset (x_i, y_i), the decision boundary satisfies:

$$y_i(w \cdot x_i + b) \geq 1$$

where w is the weight vector and b is the bias.

2 Soft Margin SVM

- **Allows some misclassification:** Useful for data with **overlapping classes** or **outliers.**
- **Introduces a penalty (slack variable ξ)** for misclassified points.
- **Trade-off between margin width and classification accuracy.**

When to Use?

✅ If data is **not perfectly separable.**

✅ If dataset has **noise** or **outliers.**

✅ If slight misclassification is acceptable for better generalization.

Mathematical Constraint:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

Dataset

We will classify two sets of points into two classes:

✅ Class A (Positive class, +1):

(1,1), (2,2), (3,3)

❌ Class B (Negative class, -1):

(1,3), (2,4), (3,5)

We assume these two classes are **linearly separable.**

Step 1: Define the Decision Boundary

SVM tries to find the optimal **hyperplane** in the form of:

$$w_1 x + w_2 y + b = 0$$

where:

- w_1, w_2 are the weights (parameters)
- b is the bias
- (x, y) are the data points

The goal of SVM is to maximize the **margin** while ensuring that all points are correctly classified.

For a correctly classified point (x_i, y_i), the SVM condition is: $y_i(w_1 x_i + w_2 y_i + b) \geq 1$

Step 2: Identify the Support Vectors

Support Vectors are the closest points to the hyperplane.

In this case, the points **(2,2) from Class A** and **(2,4) from Class B** are the support vectors.

These points should satisfy:

$$w_1(2) + w_2(2) + b = +1$$

$$w_1(2) + w_2(4) + b = -1$$

Step 3: Solve for w_1, w_2, and b

We solve these two equations:

1. $2w_1 + 2w_2 + b = 1$
2. $2w_1 + 4w_2 + b = -1$

Step 3.1: Subtract the two equations

$$(2w_1 + 4w_2 + b) - (2w_1 + 2w_2 + b) = -1 - 1$$

$$2w_2 = -2$$

$$w_2 = -1$$

Step 3.2: Solve for w_1

Substituting $w_2 = -1$ into the first equation:

$$2w_1 + 2(-1) + b = 1$$

$$2w_1 - 2 + b = 1$$

$$2w_1 + b = 3$$

Step 3.3: Solve for b

Similarly, substituting $w_2 = -1$ into the second equation:

$$2w_1 + 4(-1) + b = -1$$

$$2w_1 - 4 + b = -1$$

$$2w_1 + b = 3$$

Since both equations are the same, we need another constraint to determine w_1.

From the margin condition:

$$\text{Margin} = \frac{1}{\|w\|}$$

$$\|w\| = \sqrt{w_1^2 + w_2^2}$$

For simplicity, we assume $w_1 = 1$, so:

$$\|w\| = \sqrt{1^2 + (-1)^2} = \sqrt{2}$$

Thus, $w_1 = 1$ and we substitute into $2(1) + b = 3$, giving:

$$b = 1$$

Step 4: Classifying New Points

Using our decision function: $f(x, y) = w_1 x + w_2 y + b$

- If $f(x, y) > 0 \rightarrow$ Class **A (+1)**
- If $f(x, y) < 0 \rightarrow$ Class **B (-1)**

Let's classify **(3,2)**:

$$f(3, 2) = (1)(3) + (-1)(2) + 1 = 3 - 2 + 1 = 2 > 0$$

✅ **Class A (+1)**

Let's classify **(3,4)**:

$$f(3, 4) = (1)(3) + (-1)(4) + 1 = 3 - 4 + 1 = 0$$

Since it lies on the decision boundary, it is a **support vector**.

References

1. Tom M. Mitchell, “Machine Learning”, McGraw-Hill Education (India) Private Limited, 2013.
2. Sebastian Raschka , Yuxi (Hayden) Liu Machine Learning with PyTorch and Scikit-Learn: Developmachine learning and deep learning models with Python Packt Publishing Limited (23 December 2022).