Supervised Learning - I

By:
Dr. Shikha Deep

Contents:

(1) Logistic Regression
(2) Support Vector Machines
(3) Random Forest
(4) Linear Discriminant Analysis

Logistic Regression:

- A logistic regression is a supervised **classification** model.
- It "squash" the values into a 0–1 range using a Sigmoid Function.

The sigmoid (or logistic) function is:

$$y = \frac{1}{1 + e^{-(a_0 + a_1 x)}}$$

where a_0 is intercept value, a_1 coefficient of x increment

x is independent variable and y is dependent variable

- This converts any real number into a value between 0 and 1.
- To predict the probability which ranges from 0 to 1.
- This converts any real number into a value between 0 and 1.

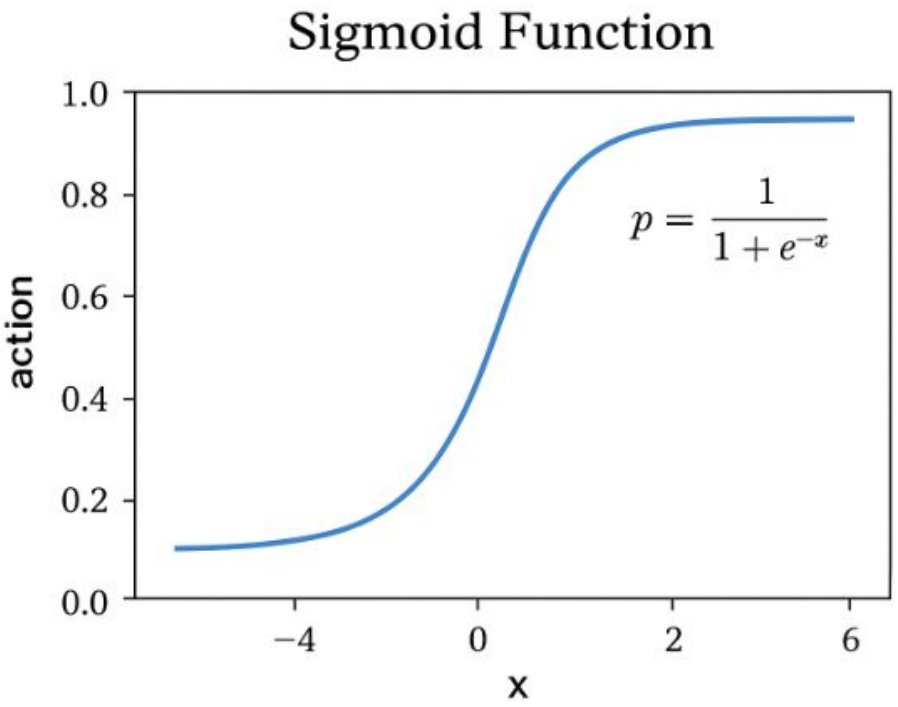
x value	Output of sigmoid
$-\infty$	0
0	0.5
$+\infty$	1

Sigmoid Curve

- It looks like an **S-shaped curve**.
- When x increases (e.g., study hours), probability of passing increases gradually, not suddenly.

Interpretation Example:

- If $p > 0.5 \rightarrow$ predict 1 (Pass)
- If $p < 0.5 \rightarrow$ predict 0 (Fail)
- So logistic regression doesn't just say *"yes/no"* — it says *"how confident we are"*.


Sigmoid Function
1.0
0.8
0.6
0.4
0.2
0.0
action
$p = \frac{1}{1+e^{-x}}$
−4
0
2
6
x

Example:

1. Disease diagnose
2. Fraud detection
3. Emergency Detection
4. Email classification - Spam and Not Spam email
5. Bank Loan – Default & Not default

Support Vector Machines (SVMs)

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks.

- It's goal is to find the best boundary (line or hyperplane) that separates the data into classes.
- The main goal of SVM is to maximize the margin between the two classes. The larger the margin the better the model performs on new and unseen data.
- SVM is often called the maximum margin classifier.

Key Concepts of SVMs

(1) Hyperplane: A hyperplane is a decision boundary that separates data points of different classes in the feature space. Mathematically, it is represented as

$$w.x + b = 0$$

If have multiple features then, it becomes $w_1x_1 + w_2x_2 + b = 0$

(2) Support Vectors: These are the data points that are closet to the hyperplane. These points influence the orientation and position of the hyperplane.

(3) Margins: The margin is the distance between the hyperplane and the nearest data point of either class. SVM tries to maximizes this margin to improve classification confidence.

(4) Hard Margins: A maximum-margin hyperplane that perfectly separates the data without misclassifications.

(5) Soft Margins: Allow some misclassification or margin violations for non linearly separable data.

How does Support Vector Machine Algorithm Work?

- The key idea behind the SVM algorithm is to find the hyperplane that best separates two classes by maximizing the margin between them. This margin is the distance from the hyperplane to the nearest data points (support vectors) on each side.
- The best hyperplane also known as the "hard margin" is the one that maximizes the distance between the hyperplane and the nearest data points from both classes.

Kernel Function :

When data is not linearly separable i.e it can't be divided by a straight line, SVM uses a technique called kernels to map the data into a higher-dimensional space where it becomes linearly separable data.

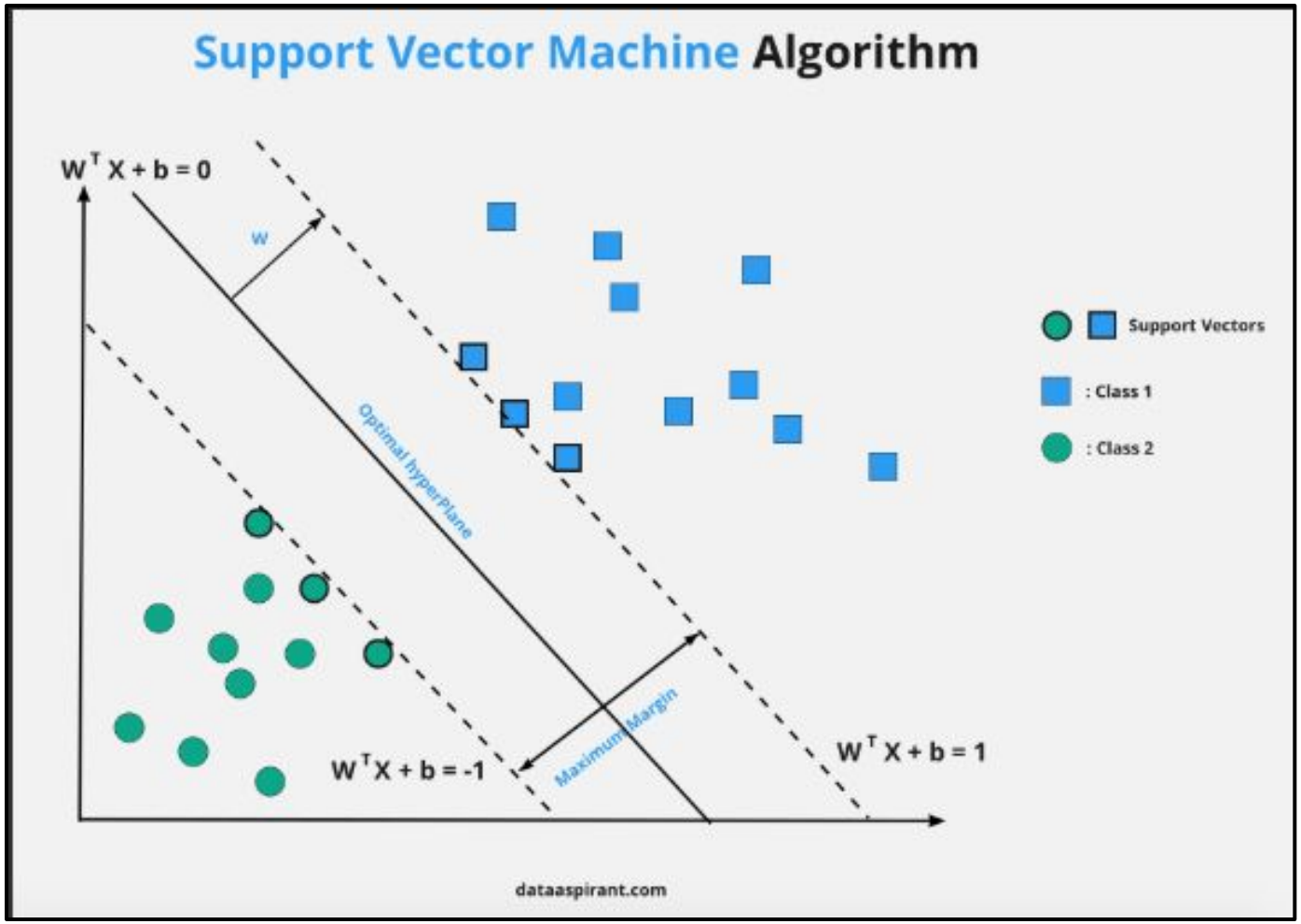
- This transformation helps SVM find a decision boundary even for non-linear data.

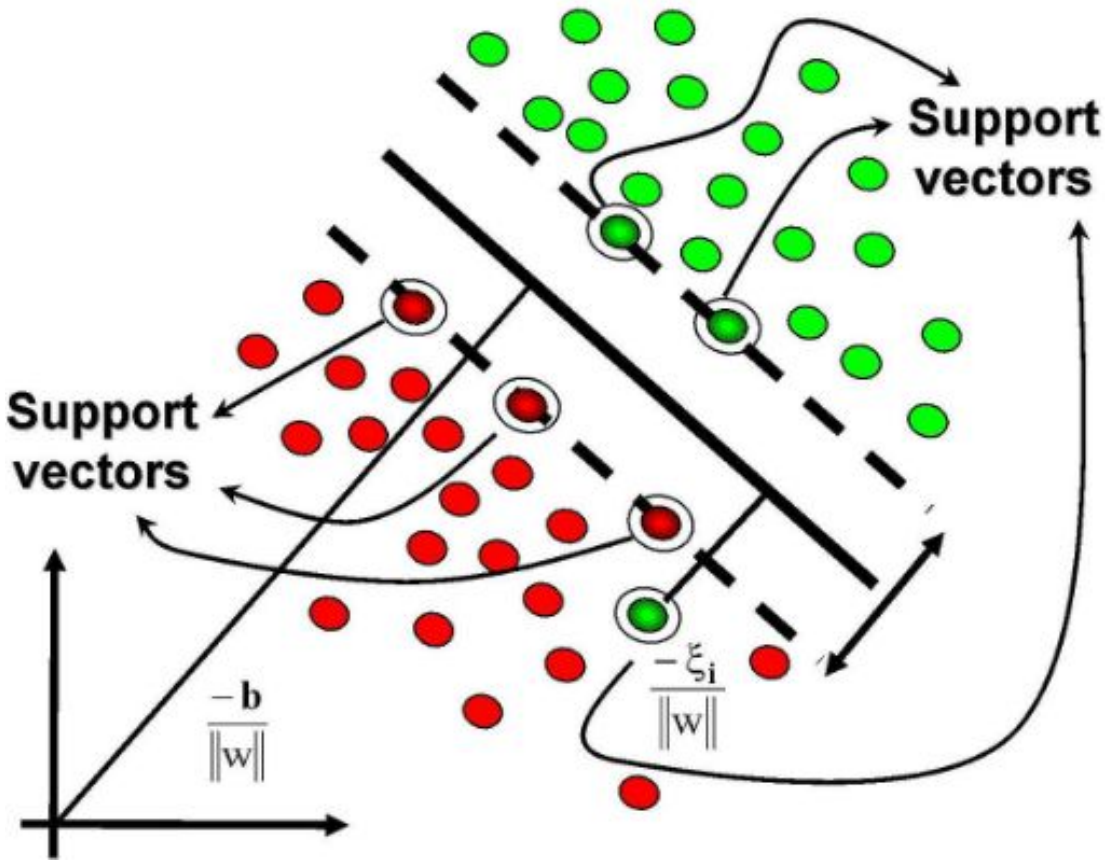
Types of Kernels :

(1) **Linear Kernel:** Suitable for linearly separable data.

(2) **Polynomial Kernel:** For Curved boundaries.

(3) **Radial Bias Function (RBF) Kernel:** Captures complex relationship.


Support Vector Machine Algorithm
W^T X + b = 0
w
Optimal hyperplane
Maximum Margin
W^T X + b = -1
W^T X + b = 1
Support Vectors
: Class 1
: Class 2
dataaspirant.com


Support vectors
Support vectors
$\frac{-b}{\|w\|}$
$\frac{-\xi_i}{\|w\|}$

Comparison with Logistic Regression

Feature	Logistic Regression	SVM
Boundary	Based on probability (sigmoid)	Based on maximum margin
Output	Probability (0–1)	Class label (+1 or -1)
Works well for	Linearly separable data	Complex and non-linear data
Flexibility	Simple	More powerful with kernels


Thanks!