AIE
ACADEMY OF INNOVATIVE EDUCATION

DAY 06

INTRODUCTION TO ML CLASSIFIERS - III

SUPPORT VECTOR MACHINE

Thakshila Dasun
BSc. Eng in Mechatronics Eng
CIMA (UK)

Support Vector Machines (1)

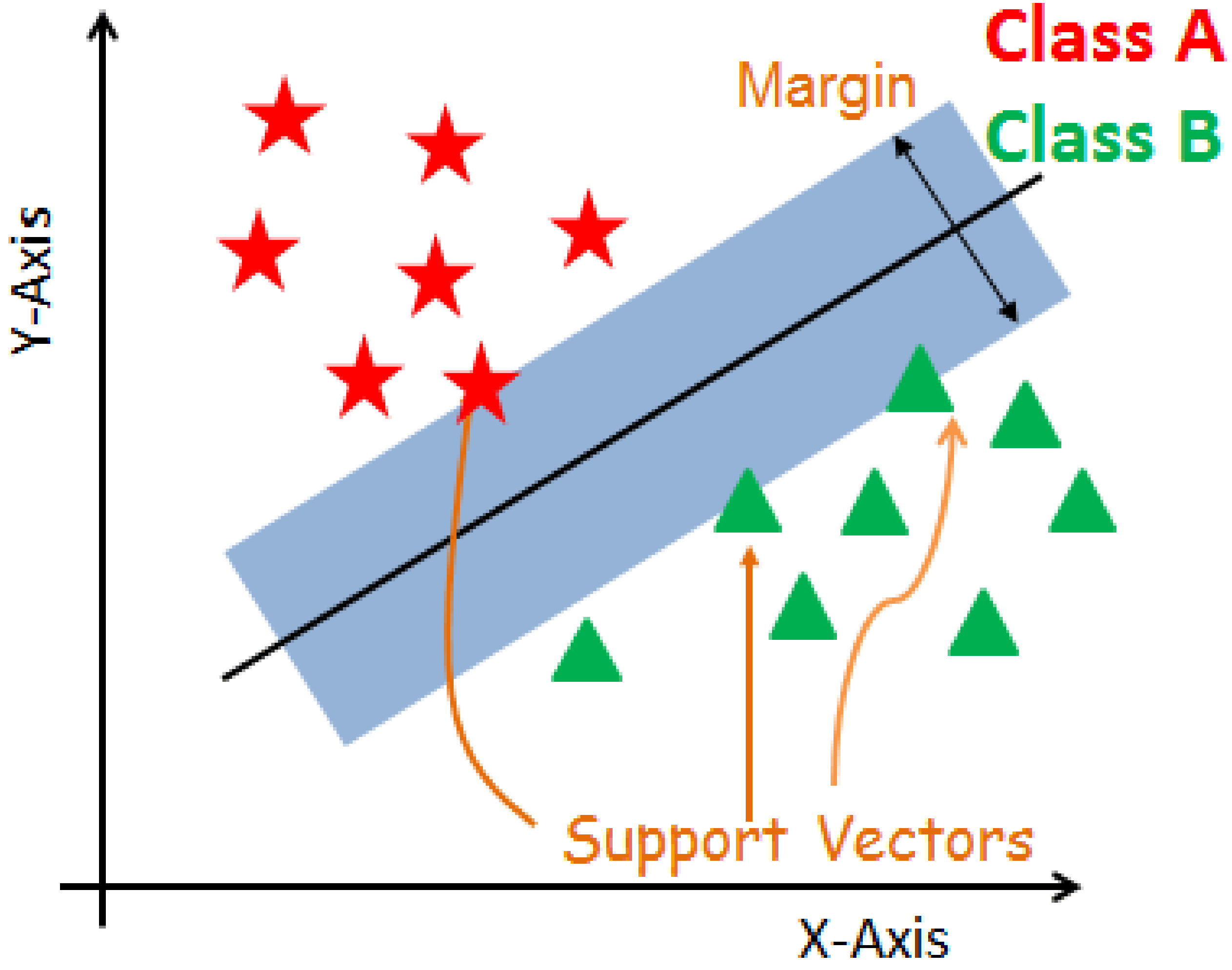
- SVM offers very high accuracy compared to other classifiers such as logistic regression, and decision trees.
- Very good in handling nonlinear input spaces.
- Used in a variety of applications such as face detection, intrusion detection, classification of emails, news articles and web pages, classification of genes, and handwriting recognition.

Support Vector Machines (2)

- The SVM classifier separates data points using a hyperplane with the largest amount of margin.
- Also known as a discriminative classifier.
- SVM finds an optimal hyperplane which helps in classifying new data points.
- Can be employed in both types of classification and regression problems
- It can easily handle multiple continuous and categorical variables. SVM constructs a hyperplane in multidimensional space to separate different classes

SVM Classifier

- SVM generates optimal hyperplane in an iterative manner, which is used to minimize an error.
- The core idea of SVM is to find a maximum marginal hyperplane(MMH) that best divides the dataset into classes.



Support Vectors

- Support vectors are the data points, which are closest to the hyperplane.
- These points will define the separating line better by calculating margins. These points are more relevant to the construction of the classifier.

Hyperplane

- A hyperplane is a decision plane which separates between a set of objects having different class memberships

Margin

- A margin is a gap between the two lines on the closest class points. This is calculated as the perpendicular distance from the line to support vectors or closest points.
- If the margin is larger in between the classes, then it is considered a good margin, a smaller margin is a bad margin.

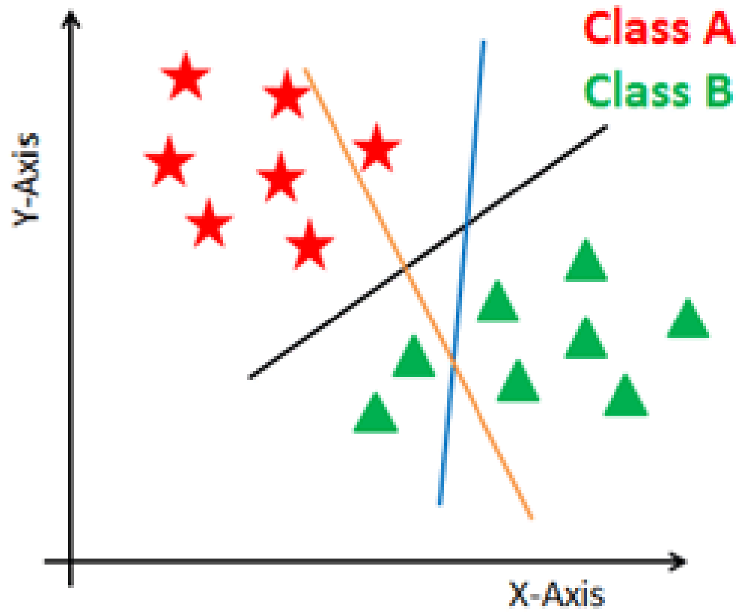
How does SVM work? (1)

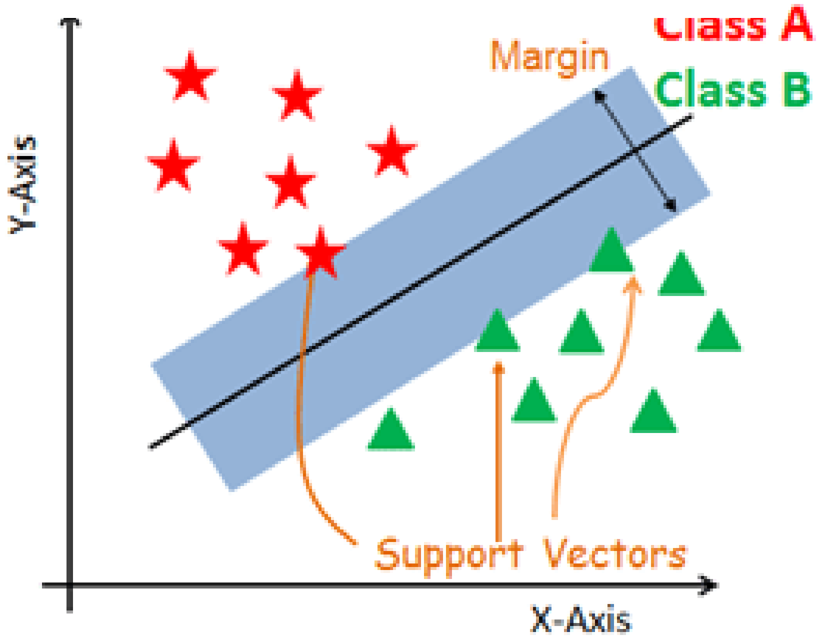
- The main objective is to segregate the given dataset in the best possible way.
- The distance between the either nearest points is known as the margin.
- The objective is to select a hyperplane with the maximum possible margin between support vectors in the given dataset

How does SVM work? (2)

- SVM searches for the maximum marginal hyperplane in the following steps,

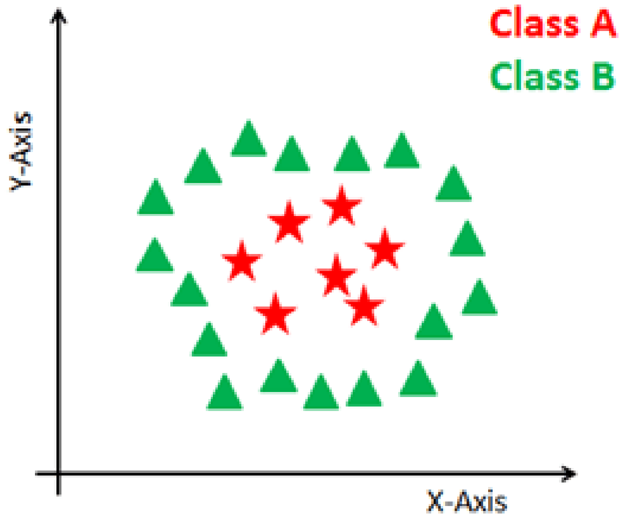
  1. Generate hyperplanes which segregates the classes in the best way
  2. Select the right hyperplane with the maximum segregation from the either nearest data points as shown in the right-hand side figure.

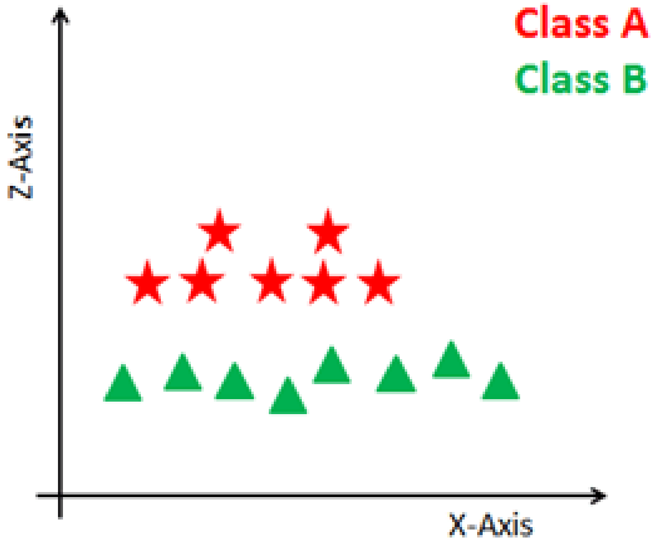




Dealing with non-linear and inseparable planes

- SVM uses a kernel trick to transform the input space to a higher dimensional space as shown on the right.
- The data points are plotted on the x-axis and z-axis (Z is the squared sum of both x and y: z=x^2+y^2). Now you can easily segregate these points using linear separation.





Types of Kernels

- SVM algorithms use a set of mathematical functions that are defined as the kernel.
- The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions.
- These functions can be different types. For example ***linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.***
  Introduce Kernel functions for sequence data, graphs, text, images, as well as vectors.
- The most used type of kernel function is **RBF.** Because it has localized and finite
- response along the entire x-axis.
  The kernel functions return the inner product between two points in a suitable feature space. Thus by defining a notion of similarity, with little computational cost even in very high-dimensional spaces.

Examples of SVM Kernels

1. Polynomial kernel $$k(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i} \cdot \mathbf{x_j} + 1)^d$$

2. Gaussian kernel $$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

3. Gaussian radial basis function (RBF) $$k(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\gamma\|\mathbf{x_i} - \mathbf{x_j}\|^2)$$

4. Laplace RBF kernel $$k(x, y) = \exp\left(-\frac{\|x - y\|}{\sigma}\right)$$

5. Sigmoid kernel $$k(x, y) = \tanh(\alpha x^T y + c)$$