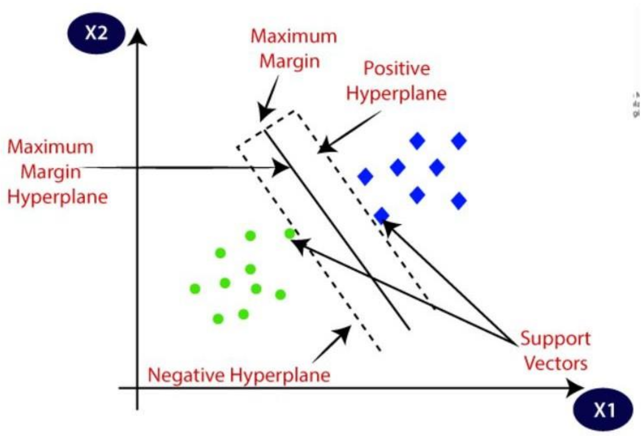
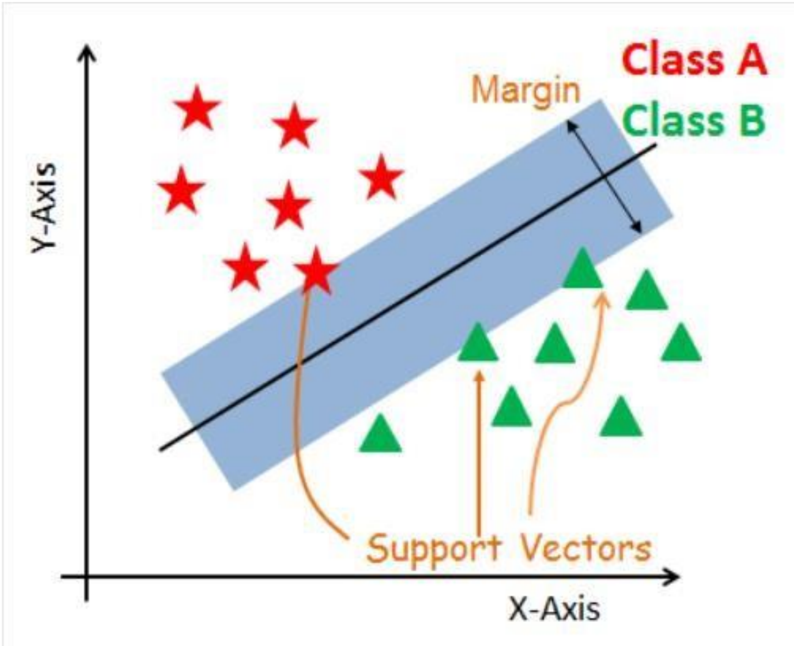
Support Vector Machines

- Support Vector Machines (SVM) are typically used for classifying things.
- SVM can also handle other tasks, like predicting values (regression).
- It's good at dealing with both numbers and categories.
- SVM uses a line (hyperplane) in space to separate different groups.
- The line is adjusted to minimize mistakes.
- The main idea is to find the best line (maximum margin) to split thegroups.
- SVMs are a supervised learning algorithm used for classification and regression problems.
- The goal of SVMs is to create a hyperplane that separates data points into different classes.
- The hyperplane is chosen to maximize the margin between the two classes.
- Support vectors are the data points that are closest to the hyperplane.



Goal of SVM:

The SVM algorithm's aim is to draw a special line (like a superhero line) thatcan split a space into different groups. This way, when new data comes along, we can quickly figure out which group it belongs to. This special line is called a "superplane."



Support Vectors

Support vectors are the data points closest to the separation line (hyperplane). They play a crucial role in defining the line and calculating the margins.

Hyperplane

The hyperplane is like a decision line that separates objects into different categories or classes.

Margin

Margin is the gap between the closest class points and the separation line. It's measured as the perpendicular distance from the line to the support vectors or nearest points.

Types of SVM

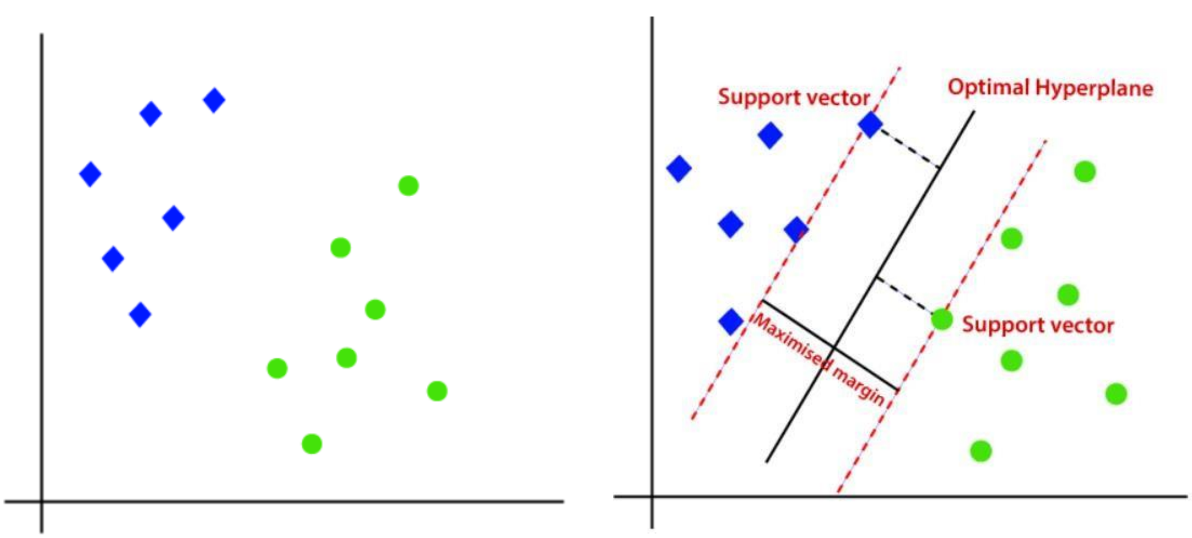
SVM comes in two types:

Linear SVM:

Use this when you can draw a straight line to separate your data into twogroups. It's for simple cases. Here's a deeper look:

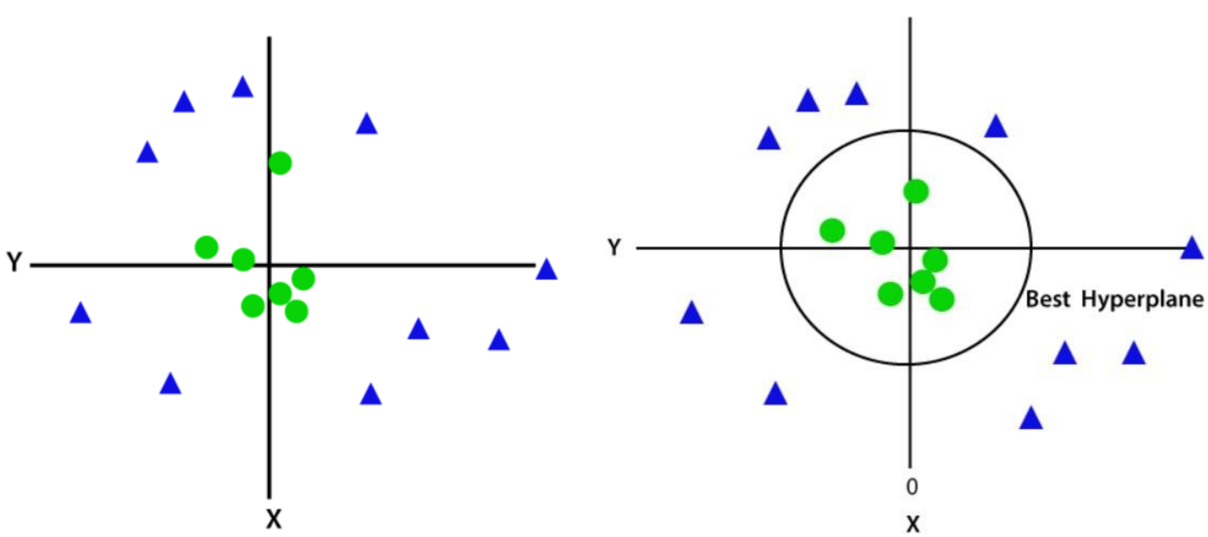
- ✓ *Decision Boundary*: Linear SVM aims to find a hyperplane that separates the classes in the input feature space. This hyperplane is a straight line in 2D, a plane in 3D, and a hyperplane in higher dimensions.
- ✓ *Applications:* Linear SVM is suitable for scenarios where the classes are linearly separable, meaning a single straight line can effectively separate the data points. It's commonly used in text classification, image classification (with linearly separable features), and binary classification tasks.

Optimal Hyperplane



Non-linear SVM: If a straight line can't split your data, go for this one. Ithandles more complicated, curvy data divisions. Instead, it maps the data into a higher-dimensional space where separation becomes possible. Here's more detail:

- ✓ *Kernel Trick:* Non-linear SVM utilizes the kernel trick to implicitly map the input features into a higher-dimensional space. This transformation enables the algorithm to find a linear decision boundary in the transformed space, which corresponds to a non-linear decision boundary in the original feature space.
- ✓ *Applications*: Non-linear SVM is useful when the relationship between features and classes is non-linear. It's applied in various domains such as image recognition, sentiment analysis, bioinformatics, and financial forecasting, where complex decision boundaries are common.



Best Hyperplane

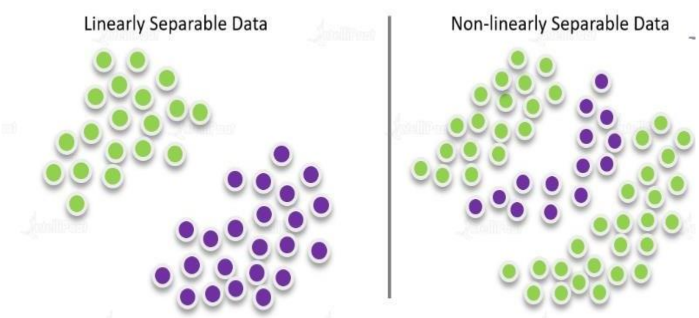
Comparison:

Linear SVM:

- ✓ Advantages: Faster training time, simpler model interpretation, suitable for linearly separable data.
- ✓ Disadvantages: Limited to linear decision boundaries, may not perform well with non-linearly separable data.

Non-linear SVM:

- ✓ Advantages: Flexibility to capture complex relationships between features and classes, can handle non-linearly separable data.
- ✓ Disadvantages: Slower training time due to the need for feature space transformation, potential overfitting with high-dimensional feature spaces.



SVM Figure 1: Linearly Separable and Non-linearly Separable Datasets

Choosing Between Linear and Non-linear SVM:

- ✓ Use Linear SVM when the data is linearly separable or when simplicity and speed are important.
- ✓ Use Non-linear SVM when dealing with complex data distributions or when linear separation is not feasible.

Working flow of Support Vector Machines (SVM)

The working flow of Support Vector Machines (SVM) involves several steps, from data preprocessing to model evaluation. Here's a comprehensive procedure:

1. Data Collection:

- Gather the dataset containing features and corresponding labels or target variables.

2. Data Preprocessing:

- Handle missing values: Impute or remove missing values.
- Feature scaling: Scale numerical features to ensure they have similar ranges.
- Feature encoding: Encode categorical variables into numerical representations if needed.
- Feature selection: Optionally, select relevant features to reduce dimensionality.

3. Training and Testing Split:

- Split the dataset into training and testing sets to evaluate the model's performance on unseen data.

4. Model Training:

- Select the appropriate SVM algorithm (linear or non-linear) based on the problem's characteristics.
- Tune hyper Parameters such as the choice of kernel, regularization parameter (C), and kernel parameters (gamma for non-linear kernels).
- Train the SVM model on the training data using the selected hyper Parameters.

5. Model Evaluation:

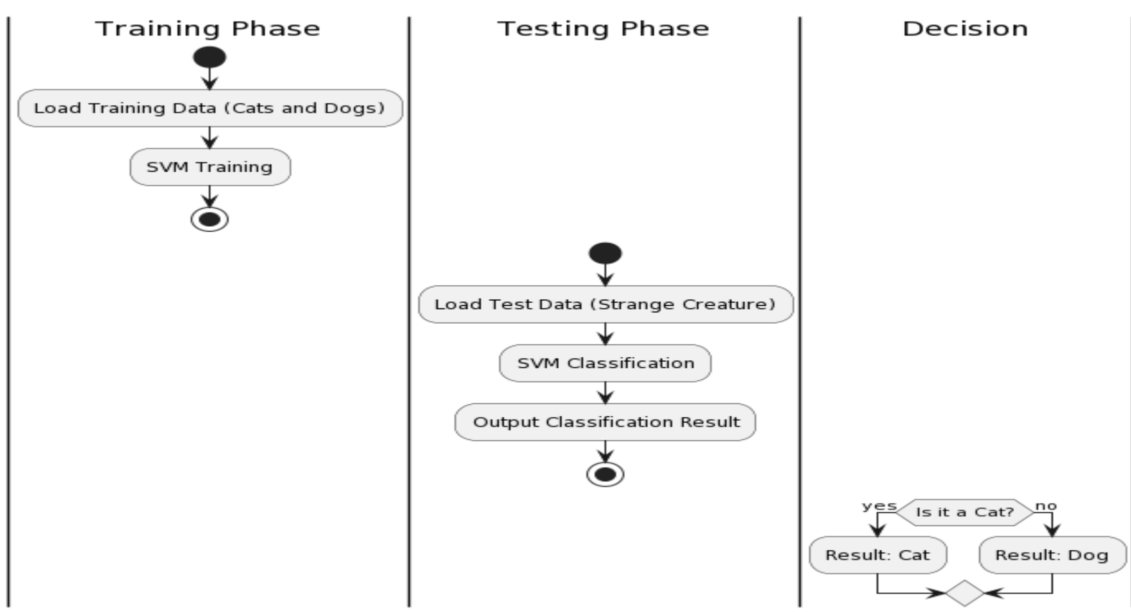
- Evaluate the trained model's performance on the testing set using appropriate metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve (AUC-ROC).
- Analyze the confusion matrix to understand the model's performance in classifying different classes.

6. Model Optimization:

- Fine-tune hyperparameters using techniques like grid search or randomized search to improve model performance.
- Perform feature engineering or selection to enhance model interpretability and generalization.

7. Deployment:

- Once satisfied with the model's performance, deploy it to production for making predictions on new, unseen data.
- Monitor the model's performance over time and retrain periodically if necessary.



Advantages:

- Good Accuracy: SVMs are good at making accurate predictions.
- Faster Predictions: They work quickly when predicting.
- Memory Efficiency: They use less computer memory since they only use a part of the training data.
- Clear Separation: SVMs work best when there's a clear gap between categories.
- High-Dimensional Data: They handle data with lots of features (dimensions) well.
- Kernel SVM contains a non-linear transformation function to convert the complicated non-linearly separable data into linearly separable data
- It is effective when the number of features are greater than the number of data points
- It employs a subset of training points in the decision function or support vectors, making SVM memory efficient
- Apart from common kernels, it is also possible to specify custom kernels for the decision function

Disadvantages:

- Not for Large Datasets: SVMs aren't great for big datasets because they take a long time to train.
- Slower Training: They take more time to train than Naïve Bayes.
- Poor with Overlapping Data: If categories overlap, SVMs struggle.
- Kernel Sensitivity: They can be sensitive to the type of mathematical function (kernel) used.
- If the number of features is significantly greater than the number of data points, it is crucial to avoid overfitting when choosing kernel functions and

regularization terms

- ❖ Probability estimates are not directly provided by SVMs; rather, they are calculated by using an expensive fivefold cross-validation
- ❖ It works best on small sample sets due to its high training time

Applications of SVM

- ❖ **Face Detection:** Classifies images of people's faces by creating a bounding box around them.
- ❖ **Bioinformatics:** Classifies genes to differentiate between proteins, identify biologicalproblems, and detect cancer cells.
- ❖ **Text Categorization:** Classifies documents into different categories based on their content.
- ❖ **Generalized Predictive Control (GPC):** Provides control over industrial processes.
- ❖ **Handwriting Recognition:** Recognizes handwritten characters by matching them against pre-existing data.
- ❖ **Image Classification:** Classifies images into different categories.