Support Vector Machines

(SVM)

by : Mercy Ogero

Definition

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks by finding the optimal hyperplane that maximizes the margin between classes.

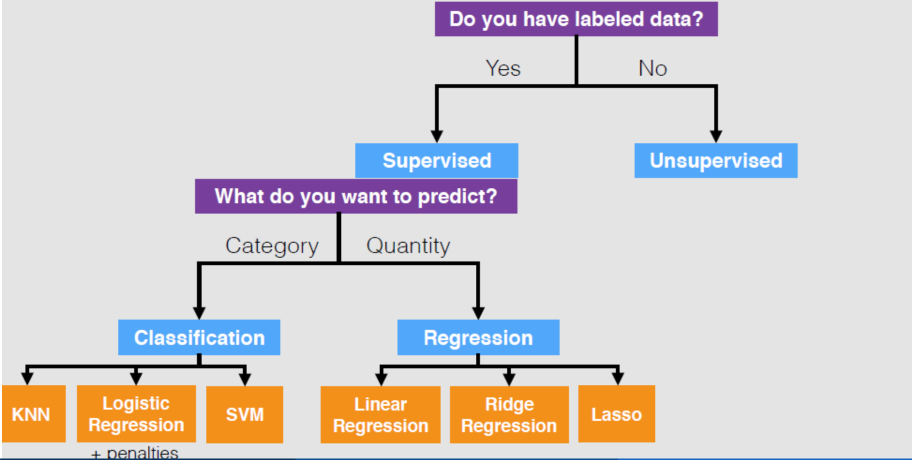
Support Vector Machine tries to find the best boundary known as a hyperplane that separates different classes in the data.

SVM is preferred when doing binary classification.

Key Characteristics

- Effective in high-dimensional spaces
- Memory efficient (uses support vectors only)
- Versatile through kernel functions
- Robust to overfitting

Machine Learning Methods



Goal

The primary goal of SVM is to find the best decision boundary called hyperplane that separates data points of different classes with the maximum possible margin.

The larger enough the margin, the better and the model performs better on some new and unseen data.

When we use SVM

SVM is commonly used during classification where the data has many features and it is not possible to separate the data with a straight line.

However, SVM is not preferred when the sample size is very large or there are overlapping classes present in the data. Additionally, it is not suitable when there are high number of outliers or when native probabilistic outputs are needed.

Therefore, SVM is used when the data set is unstructured and high-dimensional.

History of SVM

- **1960s:** Invetion of SVM by Vapnik and Chervonenkis at 1964
- **1990s:** Modern SVM developed by Vapnik and colleagues at AT&T Bells Labs to make the maximum-margin non-linear classifiers using the kernel trick.
- **1995:** Publication of key paper "Support-Vector Networks" proposed by Carina Cortes.
- **2000s:** Widespread adoption in various fields
- **Present:** Continues to be a popular and effective algorithm. The algorithim has been used in multiclass classification, regression and also intergrated with other machine learning models. The model has been popular with datasets due to its high interpretability and high accuracy.

Definition of Terms

1. Hyperplane

A decision boundary that separates different classes in the feature space. It classifies different classes by creating a separation boundary.

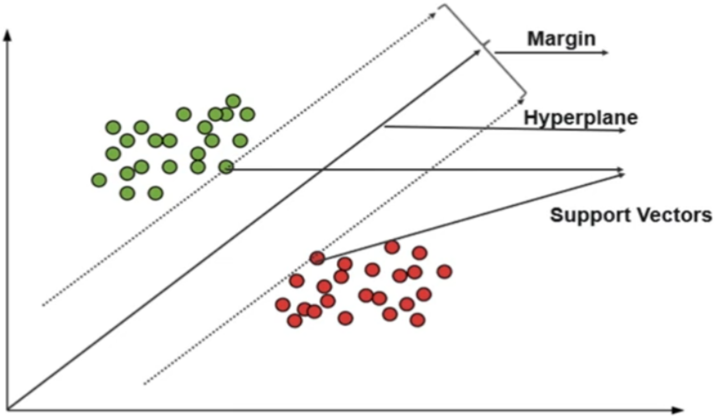
2. Margin

The distance between the hyperplane and the nearest data points from each class. A larger margin means better generalization to new data.

3. Support Vectors

The data points closest to the hyperplane that actually define the margin. These critical points determine the position and orientation of the decision boundary.

SVM Concept



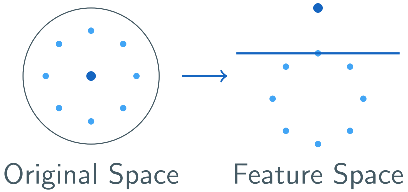
4. Kernel

Function that is essential in separating non-linear data by mapping it to a higher dimensional space. ie, it transforms non-linear data to higher dimensions.

5. Feature Space

Higher-dimensional space where data becomes linearly separable

Kernel Transformation



6. Soft Margin

Allows some errors. Commonly used when the data is not perfectly separable. The errors allow better generalization and is common in the real world data.

7. Hard Margin

SVM finds a hyperplane that separates all the data points without any errors. Hard margin works best when the data is clean.

Soft vs Hard Margin

Types of SVM

1. Linear SVM

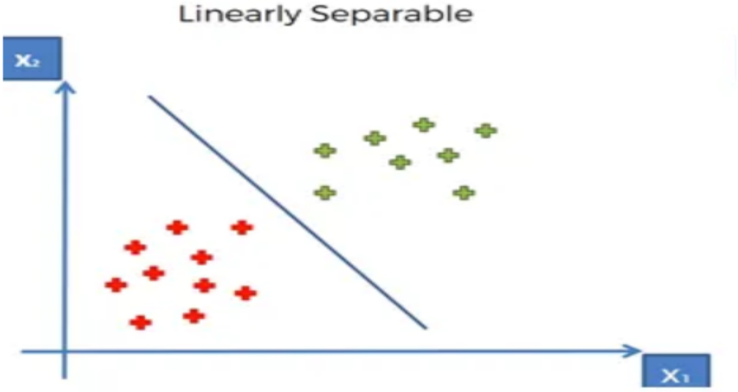
Linear SVM uses a linear decision boundary to separate data points of different classes. The following are the characteristics of linear SVM:

- Separates data with straight line
- Useful when the data is linearly separable
- Points near the boundary affect the model, other points do not have an impact on the hyperplane.
- Easier to interpret and explain, however, its sensitive to outliers.
- Uses low computation power

When to Use Linear SVM

- If the data can be linearly separated.
- if the number of features are many.

Linear SVM



2. Non-Linear SVM

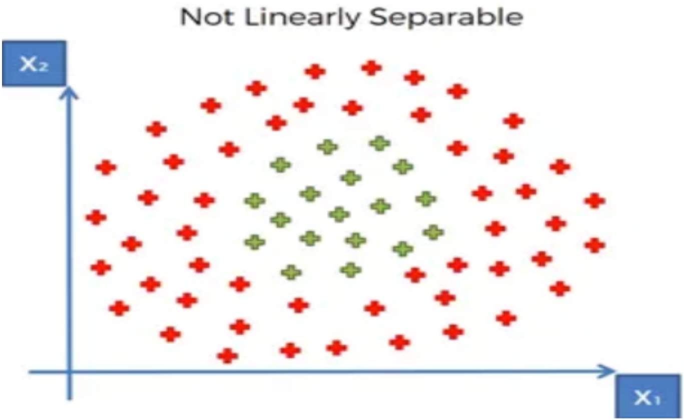
Non-linear SVM is used when the data cannot be separated to two classes by a straight line. By using the non-linear function, the data can be transformed to a higher-dimensional feature space using the kernel function where the data points can be linearly separated.

- Uses kernel functions for complex patterns
- Handles non-linearly separable data
- More flexible decision boundaries

Common Types of Kernels in non-linear SVM

- **RBF (Radial Basis Function)**
- **Polynomial**
- **Sigmoid**

Non-linear SVM Example

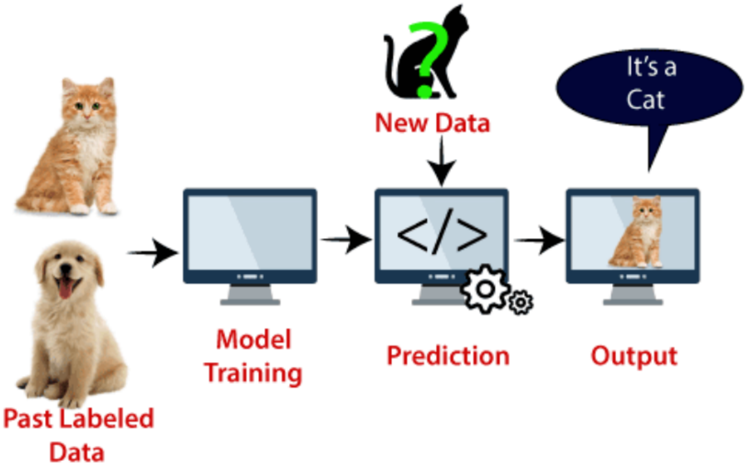


How SVM Works: Step by Step

Core Process:

1. **Data Preparation**
   - Collect and clean the dataset. The data should have labels (Y) and input features(X). Then, Normalize the data so that all variables are equal and the influence of unusual variables is minimized in the dataset
   - Split data into training and testing sets. Normally, the data is splitted according to the 70:30 ratio.
2. **Selecting the SVM Type**
   - Choose between Linear or Nonlinear SVM. If the data can be separable, use the linear SVM.
   - if the data cannot be separated, select an appropriate kernel function that will transform the data to a high dimensional space where it can be separated linearly.
3. **Model Training**
   - SVM finds the hyperplane that best separates the data
   - Maximizes the margin, that is the distance between the support vectors or the nearest data points.

SVM Visualization



Advanced Steps

4. **Support Vector Identification**
   - Identify the data points closest to the hyperplane and these points are called the support vectors
   - The support vectors define the decision boundary. Points far away from the hyperplane do not affect the boundary. The support vectors are the most important in the model.
5. **Model Prediction**
   - Use the trained SVM to classify new data points
   - Decision function determines the side of the hyperplane
   - Output is the predicted class label
6. **Model Evaluation and Optimization**
   - Evaluate performance and accuracy of the model using the testing data that was initially divided.
   - Metrics such as accuracy, F1-Score, confusion matrix and precision and recall are used to evaluate the model

Key Advantages

- **Effective in high dimensions:** Can handle datasets with many features efficiently
- **Memory efficient:** Only support vectors are used to define the model
- **Flexible through kernels:** Can model linear and non-linear relationships
- **Robust to overfitting:** Especially with the right choice of parameters
- **Mathematically sound:** Finds the global optimum for the margin

Practical Benefits

- Works well when classes are clearly separable
- Handles small and medium datasets effectively
- Strong theoretical foundation supports reliability
- Can be applied to a wide range of problems (text, images,

Key Challenges

- **Not ideal for large datasets:** Training time grows quickly with dataset size
- **Sensitive to noise:** Outliers and overlapping classes can affect performance
- **Parameter tuning required:** C, kernel, and gamma must be chosen carefully
- **Computationally intensive:** Non-linear SVMs especially can be slow to train

Practical Limitations

- Selecting the right kernel is not always straightforward
- Less intuitive when using complex kernels (black-box effect)
- Hard to get probability estimates directly
- Multi-class problems require additional strategies

Main Limitations

- **Interpretability:** Hard to explain decisions with non-linear kernels
- **Scalability:** Slow on very large datasets
- **Parameter Sensitivity:** Performance depends heavily on correct parameter selection
- **Data Requirements:** Needs representative and clean training data
- **Imbalanced Data:** Performance drops if one class dominates

Despite these limitations, SVM remains widely used for its robustness, accu