Machine Learning

Lecture 14: Support Vector Machines (SVM)

COURSE CODE: CSE451

2023

Course Teacher

Dr. Mrinal Kanti Baowaly

Associate Professor

Department of Computer Science and Engineering, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh.

Email: mkbaowaly@gmail.com

Support Vector Machines (SVM)

- SVM is one of the most popular and widely used supervised machine learning algorithms.
- It offers very high accuracy compared to other classifiers such as logistic regression, decision trees, and Naïve Bayes.
- It can be employed in both types of classification and regression problems.

Applications of SVM

- Face detection
- Intrusion detection
- Classification of emails, news articles and web pages
- Classification of genes
- Handwriting recognition

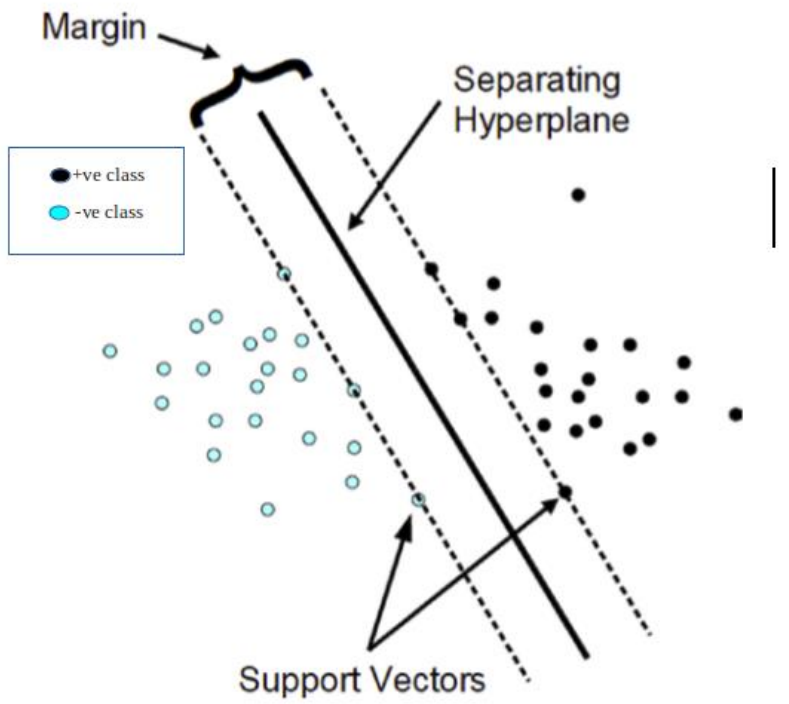
Support Vector Machines

- Support Vectors, hyperplane and margin
- The core idea of SVM
- How does it work
- Kernels
- Classifier building in Scikit-learn
- Tuning Hyperparameters
- Advantages and Disadvantages

Source: Study from DataCamp

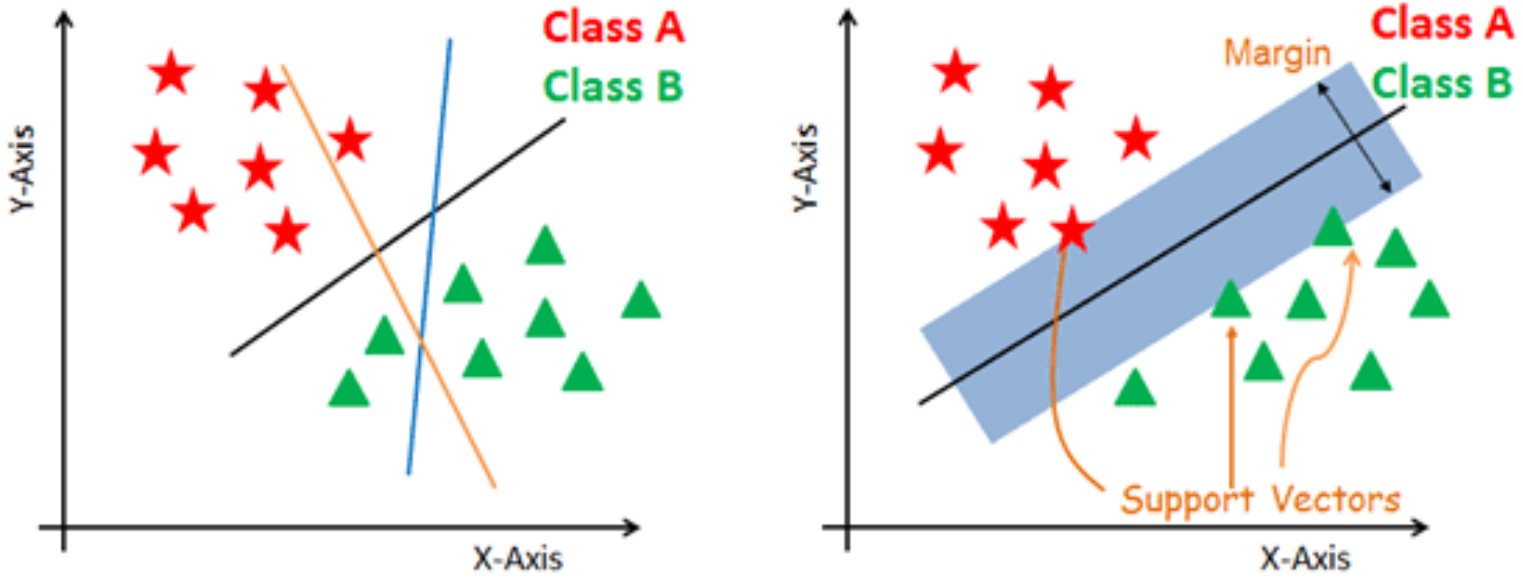
Support Vector Machines (SVM)

- **Support Vectors:** the data points, which are closest to the hyperplane.
- **Hyperplane:** a decision boundary that divides the data points into two classes.
- **Margin:** a gap between the two lines on the closest class points.



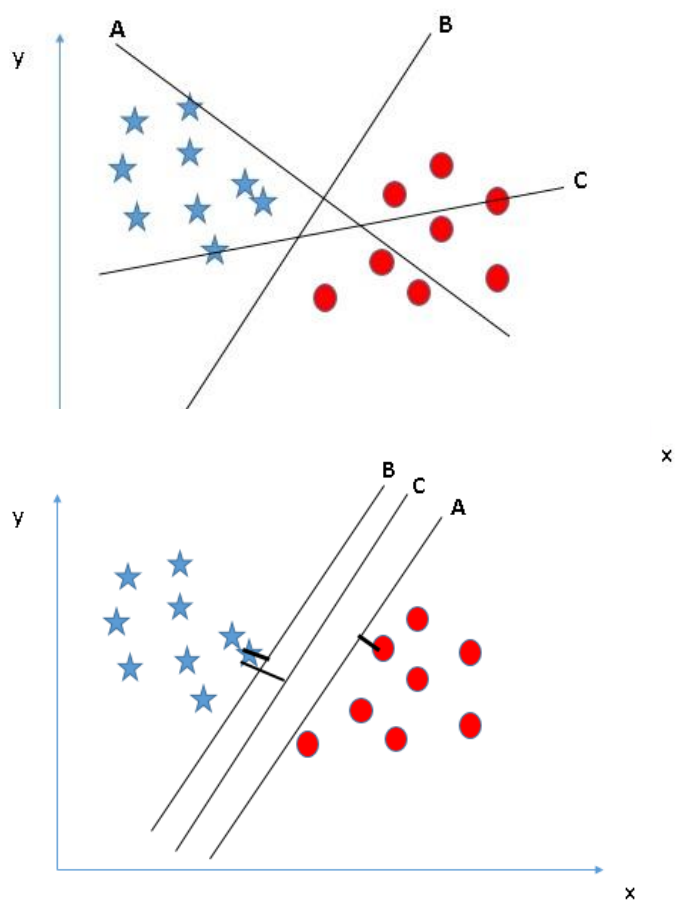
The core idea of SVM

The core idea of SVM is to find a maximum marginal hyperplane(MMH) that best divides the dataset into classes (hence also known as a discriminative classifier).



How does SVM work? Or How to Identify the right hyper-plane in SVM?

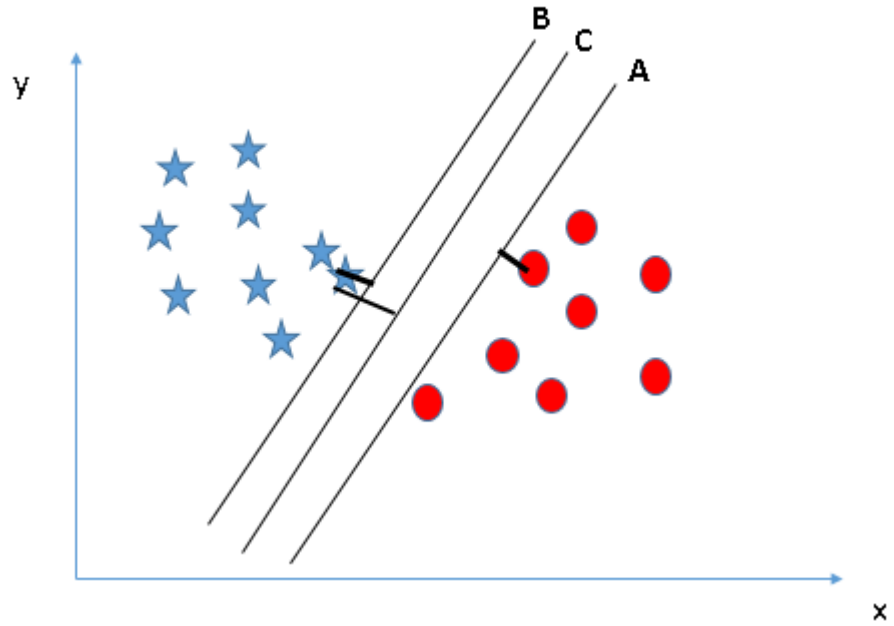
- Select the hyper-plane which segregates the classes better
- Select the hyperplane for which the margin is maximum
- SVM selects the hyper-plane which classifies the classes accurately prior to maximizing margin



Detail: Study from AnalyticsVidhya

Why does SVM maximize the margin?

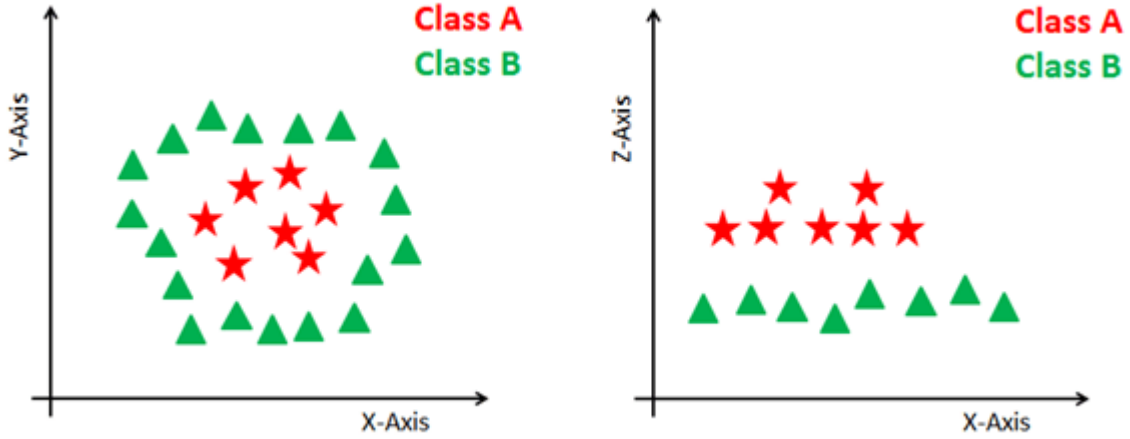
- Maximizing the margin help to decide the right hyper-plane
- Selecting the hyper-plane with a higher margin ensures robust classification



Detail: Study from AnalyticsVidhya

Kernel trick: Dealing with non-linear and inseparable planes

- Some problems can't be solved using linear hyperplane, as shown in the figure below (left-hand side).
- SVM uses a kernel trick to transform the input space to a higher dimensional space as shown on the right.



Here, we will add a new feature, z=x^2+y^2

Source: Study from DataCamp AnalyticsVidhya

SVM Kernels

- Linear Kernel
- Polynomial Kernel
- Radial Basis Function Kernel

Source: Study from DataCamp AnalyticsVidhya

Tuning Hyperparameters

- **Kernel:** There are various types of functions such as linear, polynomial (poly), and radial basis function (rbf). Here “rbf”(radial basis function) and “poly”(polynomial kernel) are useful for non-linear hyper-plane. It’s called nonlinear svm.
- **Regularization:** Here C is the penalty parameter, which represents misclassification or error term. A smaller value of C creates a small-margin hyperplane and a larger value of C creates a larger-margin hyperplane.
- **Gamma:** A lower value of Gamma will loosely fit the training dataset, whereas a higher value of gamma will exactly fit the training dataset, which causes over-fitting.
- Example: svm.SVC(kernel='rbf', C=1, gamma=0).fit(X, y)

Adv. & Disadv. of SVM

Advantage

- SVM offers very high accuracy compared to other classifiers such as logistic regression, decision trees, and Naïve Bayes.
- SVM works well with a clear margin of separation and with high dimensional space.
- It uses less memory because it uses a subset of training points in the decision phase.

Disadvantage

- Required training time is higher for large datasets.
- It works poorly with overlapping classes, e.g. when the data set has more noise
- It is sensitive to the type of kernel used.

LAB: Build Lung Cancer Prediction Model Using SVM

1. Let us investigate the Lung Cancer Dataset from here: https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link
2. It has 1000 patients and 24 predictor variables (age, gender, air pollution exposure, alcohol use, dust allergy, etc.) without index and ID. The variable (level) to predict the risk of lung cancer is encoded as 0 and 1 where 0 means low risk of lung cancer and 1 means medium or high risk of lung cancer.
3. Build a binary classification model using SVM to predict the risk of lung cancer (0, 1) of the patients. Estimate Accuracy, and F1 Score to evaluate the performance of the model.
4. Tune Hyperparameters (Kernel, C parameter, Gamma) to optimize the model performance

Study Materials of SVM

- Support Vector Machines with Scikit-learn
- Support Vector Machines (SVM)
- SVM using Scikit-Learn in Python