Chapter 12
Support Vector Machines and Applications
Vandana M. Ladwani
PESIT-BSC, India

ABSTRACT
Support Vector Machines is one of the powerful Machine learning algorithms used for numerous applications. Support Vector Machines generate decision boundary between two classes which is characterized by special subset of the training data called as Support Vectors. The advantage of support vector machine over perceptron is that it generates a unique decision boundary with maximum margin. Kernelized version makes it very faster to learn as the data transformation is implicit. Object recognition using multiclass SVM is discussed in the chapter. The experiment uses histogram of visual words and multiclass SVM for image classification.

INTRODUCTION TO PATTERN RECOGNITION
What is Pattern Recognition
Pattern recognition deals with discovering the classes in a dataset. The dataset consists of objects of different classes. Objects can be images, signals or any other details of the data depending on the application. Pattern recognition can be used in various areas such as character recognition, anomaly detection, face recognition, signal classification, medical decision making etc. Pattern recognition is the fundamental building block of intelligent system.

Features
Features represent the distinguishing information retrieved from the data. This information helps to differentiate between the data from various classes. For character data the features can be pixel intensities. For signal classification the features can be energy levels at different frequency. Feature detection itself is a vast area. Features impact the classification performance. Classification model is trained learns boundaries between various classes using these features.
Classification

Classification is the task of distinguishing the data into different classes to which it belongs. For classification data is preprocessed then appropriate features are extracted and depending upon the application and the data available supervised, unsupervised or semi-supervised classification is used.

Supervised, Unsupervised, and Semi-Supervised Classification

In case of supervised classification training data has the correct label for the class to which belongs to. Features are extracted from the training data, these features along with the labels are presented to the classifier, using a learning algorithm the classifier learns the decision boundaries between various classes. Features are extracted from the test data and then presented to the classifier and the classifier predicts the label for it. Unsupervised learning is used for clustering, in case of unsupervised learning dataset no information regarding which class the data belongs to and even the number of classes dataset contains. Unsupervised learning assigns the patterns to the various classes based on the similarity of the features. Semi-supervised learning on the other hand has data points with labels as well as data points for which the class is not known. Semi-supervised helps to cluster the unlabeled data by using the constraints derived from the labeled data.

WHY SUPPORT VECTOR MACHINES

Support Vector Machine Introduction

Support vector machine is one the most powerful machine learning algorithms used for classification or regression tasks. Support Vector Machines are used in various applications such as face recognition, medical decision making, regression tasks etc. with high accuracy.

Perceptron Algorithm

If we consider a 2 dimensional data the linear discriminant function is given as

\[ w_1x_1 + w_2x_2 + b = 0 \]

Where \( b \) represents the bias term. For a multidimensional case the equation of linear discriminant function is \( f(x) = w'x + b = 0 \) This represents a hyperplane.

Perceptron algorithm tries to find this hyperplane and uses the hyperplane for classifying the linearly separable data. All the data points satisfying the relation \( w'x + b > 0 \) lie to one side of the hyperplane, whereas data points satisfying the relation \( w'x + b < 0 \) belong to other side of the hyperplane.