ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

Study on Machine Learning Techniques using SVM

M.Selvanayaki[1], T.S.AnushyaDevi[2]

[1,2]Assistant Professor, Department of BCA, PSGR Krishnammal College for Women, Coimbatore, India

selvim@psgrkc.ac.in[1], sanushya@psgrkc.ac.in[2]

Abstract: Machine Learning is the study of computer algorithms that improve automatically through experience. Applications range from data mining programs that discover general rules in large data sets, to information filtering systems that automatically learn user's interests. An important task of machine learning is classification, also referred as pattern recognition; where one attempts to build algorithms capable of automatically constructing methods for distinguish between different exemplars. This paper deals about different machine learning techniques for the prediction process.

Keywords: Machine Learning, Supervised Learning, Unsupervised Learning, Classification, Prediction, Support Vector Machine.

I. INTRODUCTION

Data mining is often used to apply to the two separate processes of Knowledge discovery and prediction. Knowledge discovery provides explicit information that has a readable form and can be understood by the user (ex: Association Rules Mining). Forecasting, or predictive modeling provides predictions of future events and may be transparent and readable in some approaches (ex: Rule-based systems) and opaque in others such as neural networks. Data Mining relies on the use of real world data [2]. The primary aim of the data mining is to discover patterns in the data that lead to better understanding of the data generating process and to useful predictions.

Data mining algorithms include algorithms for Association rules, Sequential Patterns, Classification (Decision Trees, SVM, Bayesian Classification, Neural Networks), Prediction (regression analysis, Regression trees, Model Trees), Clustering, Collaborative filtering (Web mining) and so on.

II. MACHINE LEARNING TECHNIQUES

Machine Learning, computational learning theory used in the context of Data Mining, to denote the application of generic model-fitting or classification algorithms for predictive data mining. Machine learning is the study of computer algorithms that improve automatically through "experience". It extracts information from data automatically by computational and statistical methods [3]. Machine Learning is a technique that can discover previously unknown regularities and trends from diverse datasets. A major focus of machine learning research is to automatically learn to recognize complex patters and make intelligent decisions based on the data. Hence Machine learning is closely related to fields such as statistics, probability theory, data mining, pattern recognition, artificial intelligence, adaptive control and theoretical computer science. The applications for machine learning include machine perception, computer vision, natural language processing, syntactic pattern recognition, search engines, medical diagnosis, bioinformatics, brain-machine interfaces and cheminformatics, detecting credit card fraud, stock market analysis, classifying DNA sequences, speech and handwriting recognition, object recognition in computer vision, game playing, software engineering and robot locomotion.

A. Supervised Learning

Input data is called training data and has a known label or result such as spam/not-spam or a stock price at a time. A model is prepared through a training process where it is required to make predictions and is corrected when those predictions are wrong. The training process continues until the model achieves a desired level of accuracy on the training data [4]. Example problems are classification and regression. Example algorithms include Logistic Regression and the Back Propagation Neural Network.

B. Unsupervised Learning

Input data is not labelled and does not have a known result. A model is prepared by deducing structures present in the input data. This may be to extract general rules. It may through a mathematical process to systematically reduce redundancy, or it may be to organize data by similarity. Example problems are clustering, dimensionality reduction and association rule learning. Example algorithms include: the Apriori algorithm and k-Means.

C. Semi-Supervised Learning

Input data is a mixture of labelled and unlabeled examples. There is a desired prediction problem but the model must learn the structures to organize the data as well as make predictions. Example problems are classification and regression. Example algorithms are extensions to other flexible methods that make assumptions about how to model the unlabeled data.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

III. Supervised Learning Algorithms

In this section we see about the four supervised learning algorithms such as Decision trees, Classification rule, Artificial Neural Networks and Support Vector Machine.

A. Decision trees

Decision tree methods construct a model of decisions made based on actual values of attributes in the data. Decisions fork in tree structures until a prediction decision is made for a given record. Decision trees are trained on data for classification and regression problems. Decision trees are often fast and accurate and a big favorite in machine learning.In Artificial Intelligence (AI) field, Quinlan has contributed through his ID3 and C4.5 algorithms. C4.5 is one of the most popular and the efficient method in decision tree-based approach. Here C4.5 [9] algorithm creates a tree model by using values of only one attribute at a time. According to authors, the decision tree induction, which was initially designed to solve classification problems, has been extended to deal with single or multi-dimensional regression. The major benefits of decision trees are i) produce intensive results, ii) easy to understand, iii) and holds well-organized knowledge structure.

Decision Trees (DT) are trees that classify instances by sorting them based on feature values, where each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. The feature that best divides the training data would be the root node of the tree.

B. Classification Rules

The goal of classification is to build a set of models that can correctly predict the class of the different objects. The inputs to these models is a set of objects (i.e., training data), the classes that these objects belong to (i.e., dependent variables), and a set of variables describing different characteristics of the objects (i.e., independent variables). Once such a predictive model is built, it can be used to predict class of the objects for which the class information is not known priori. The key advantage of supervised learning methods over unsupervised methods (clustering) is that by having an explicit knowledge of the classes the different objects belong to, these algorithms can perform an effective feature selection if that leads to better prediction accuracy.

A wide range of classifiers are available, each with its strengths and weaknesses. Classifier performance depends greatly on the characteristics of the data to be classified. The most widely used classifiers are the Neural Network (Multi-layer Perceptron), Support Vector Machines, k-Nearest Neighbors [8], Gaussian Mixture Model, Gaussian, Naive Bayes, Decision Tree and RBF classifiers.

C. Artificial Neural Networks

Artificial neural network (ANN) is suitable for modeling nonlinear relationship. It is a powerful data modeling tool that is able to capture and represent any kind of input – output relationships. The theory of ANN and application in textile engineering exactly in spinning process and modeling yarn properties [5]. The ANN employed in this study was a three layer Back- Propagation Multilayer Perceptron network, the input layer, one hidden layer, and the output layer. The Multilayer Perception artificial neural network (MLP) is the most popular and widely used nonlinear network for solving many practical problems in applied sciences, including textile engineering. The reason for the popularity of the (MLP) network is that it is very flexible and can be trained to assume the shape of the patterns in the data, regardless of the complexity of these patterns [6].

D. Support Vector Machines

Support vector machines (SVM) was originally developed for the classification problem by Vapnik and co-workers. This technique was built on the structural risk minimization principle. Now, with the introduction of □-insensitive loss function, SVM has been extended to solve nonlinear regression estimation. By using the kernel function [1], SVM plays role to map the data to high dimensional feature space and then find a linear separating hyper plane with the maximal margin in that high dimensional space. The technique has demonstrated much success in prediction studies in textile engineering exactly in fiber and yarn relationship area, and gives a powerful accuracy with good performance in many studies. Support vector machine for regression (SVMR) with the commonly kernel function, i.e., polynomial and (radial basis function (RBF)) in addition to Pearson VII Universal Kernel which was applied as a kernel function of SVM in [20], and referred to as PUK [7]. The parameters of support vector machine for regression (SVM) such as the complexity parameter C, and the value of □-insensitive loss function , and the kernel parameters such as the degree d of polynomial kernel, the width of RBF kernel function γ, and PUK kernel parameters (ω and σ) were optimized by using Grid search approach using 10 fold cross validation. The errors that were used as an indicator of the predictive performance of the models were Root mean-squared error(RMSE) & Correlation coefficient(R).

IV. Support Vector Machines – Pattern Classification Algorithm

Support Vector Machine represents a new approach to supervised pattern classification which has been successfully applied to a wide range of pattern recognition problems [1]. SVM as supervised machine learning technology is attractive because it has an extremely well developed learning theory, statistical learning theory. SVM is based on strong mathematical foundations and results in simple yet very powerful algorithms. SVM has a

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

number of interesting properties, including the solution of Quadratic Programming problem is globally optimized, effective avoidance of over fitting, the ability to handle large feature spaces, can identify a small subset of informative points called SV and so on. The SVM approach is superior in all practical applications and showing high performances. For the last couple of years, support vector machines have been successfully applied to a wide range of pattern recognition problems such as text categorization, image classification, face recognition, hand written character recognition, speech recognition, biosequence analysis, biological data mining, Detecting Steganography in digital images, Stock Forecast, Intrusion Detection and so on. In these cases the performance of SVM is significantly better than that of traditional machine learning approaches, including neural networks.

Support vector machine finds a nonlinear decision function in the input space by mapping the data into a higher dimensional feature space and separating it there by means of a maximum margin hyperplane. The computational complexity of the classification operation does not depend on the dimensionality of the feature space, which can even be infinite. Over fitting is avoided by controlling the margin. The separating hyperplane is represented sparsely as a linear combination of points. The system automatically identifies a subset of informative points and uses them to represent the solution. Finally, the training algorithm solves a simple convex optimization problem.

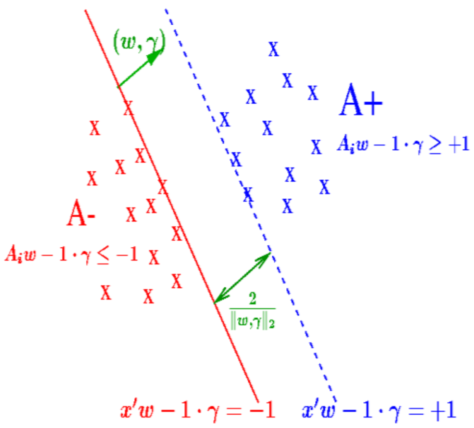
A. Mathematical formulation of Support Vector Machine

Support vector machines (SVM) are learning systems that use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. The support vector machine is a training algorithm for learning classification and regression rules from data. The geometrical interpretation of support vector classification is that the algorithm searches for the optimal separating surface, i.e. the hyperplane that is, equidistant from the two classes (This can be extended to multi-class problems). Kernel functions are then introduced in order to construct non-linear decision surfaces.

Two key elements in the implementation of SVM are the techniques of mathematical programming and kernel functions. The parameters are found by solving a quadratic programming problem with linear equality and inequality constraints. The kernel functions used in case of nonlinear SVM, may be Polynomial, Gaussian, Neural networks.

B. Standard linear and non-linear Support vector Machine

Learning machine algorithms are implementations of statistical inference principles. Typically, the machine is presented with a set of training examples, (x_i, yi) where the x_i is the real world data instances and the y_i are the labels indicating which class the instance belongs to. For the two class pattern recognition problem, $y_i = +1$ or $y_i = -1$. A training example (x_i, yi) is called positive if $y_i = +1$ and negative k otherwise. From the geometric point of view, the support vector machine constructs an optimal hyperplane given by $w^T x - \gamma = 0$ between two classes of examples. The problem of selecting optimal hyperplane from the multiple possible hyper planes is an ill-posed one. The free parameters are a vector of weights **w** which is orthogonal to the hyperplane and a threshold value **γ**.



Each training example is related to a separating hyperplane by a quantity called margin. The functional margin is $\eta_i = y_i (w^T x - \gamma)$. If $\eta_I > 0$, then the data point is correctly classified by the hyperplane. When the weight vector w of the functional margin is normalized ($w=w/\|w\|$), the geometric margin is obtained, which measures the distance of data points from the hyperplane in Euclidean space. The expression margin of a training set is used to refer to the maximum geometric margin over all possible hyper planes. The hyperplane defined by the maximum geometric margin is unique and it is known as a maximal margin hyperplane. Separating the classes with a large margin minimizes a bound on the expected generalization error one of the simplest model of SVM based maximal margin is Maximal Margin classifier.

If w is weight vector realizing functional margin 1 on the positive point X^+ and on the negative point X^- , then the two planes parallel to the hyperplane which passes through one or more points called bounding hyper planes are given by

$$w^T x - \gamma = 1$$
$$w^T x - \gamma = -1$$

The margin between the optimal hyperplane and the bounding plane is $1/\|w\|$, and so the distance between the bounding hyperplanes is $2/\|w\|$. (Distance of the bounding plane $w^T x - \gamma = 1$ from the origin is $|- \gamma +$

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

1|/||w|| and the distance of the bounding plane $w^T x - \gamma = -1$ from the origin is |- γ - 1|/||w||) Fig 1.

The points falling on the bounding planes are called support vectors and these points play crucial role in the theory. The data points x belonging to two classes A+ and A- are classified based on the condition. These inequality constraints can be combined to give. The learning problem

$$\text{Minimize} \quad = \frac{1}{2}\|\mathbf{w}\|^2$$

$$\text{subject to} \quad D_{ii}(\mathbf{w}^T \mathbf{x}_i - \gamma) \geq 1, i = 1, \ldots, l.$$

is hence to find an optimal hyperplane <w, γ>, $w^T x - \gamma = 0$ which separates A^+ from A^- by maximizing the distance between the bounding hyperplanes. Then the learning problem is formulated as optimization problem.

V. Conclusion

By using the above machine learning techniques the pattern classification problems can be identified and solved through the Linear and non-linear support vector machine. The linear support vector machine is more useful in the pattern classification problems. The datasets of any pattern can be trained and tested through machine learning tools like WEKA, MONGODB, and HADOOP etc. The results can also be viewed through GUI methods. Machine Learning Techniques are more commonly used in Pattern classification problems and regression.

References

[1] Cristianini, N and Shawe – Taylor, J. (2000). "An Introduction to Support vector Machines", Cambridge University Press.

[2] Han J and M. Kamber, Morgaan Kaufmann, "Data Mining Concepts and Techniques".

[3] C. Tan, Q. Yu, J.H. Ang, A dual-objective evolutionary algorithm for rules extraction in data mining, Computational Optimization and Applications 34 (2006) 273–294.

[4] Mitchell T.M, "Machine Learning", McGraw-Hill, Boston, 1997.

[5] Yen, G. G. and Lu, H. (2000), Hierarchical genetic algorithm based neural network design, In: IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks, pp. 168–175.

[6] Neocleous, C. & Schizas, C., (2002), Artificial Neural Network Learning: A Comparative Review, LNAI 2308, pp. 300–313, Springer-Verlag Berlin Heidelberg.

[7] Roy, A. (2000), On connectionism, rule extraction, and brain-like learning. IEEE Transactions on Fuzzy Systems, 8(2): 222-227.

[8] Ruggieri, S. (2001). Efficient C4.5. IEEE Transactions on Knowledge and Data Engineering 14 (2): 438-444.

[9] Parekh, R., and Yang, J., and Honavar, V. (2000), Constructive Neural Network Learning Algorithms for Pattern Classification. IEEE Transactions on Neural Networks. 11(2), pp. 436-451.

[10] Cover, T., Hart, P. (1967), Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1): 21–7.