INTRODUCING LINEAR SUPPORT VECTOR ALGORITHM AND ITS IMPLEMENTATION

Hamid Shafie*

Amin Samimi*

Abstract: *The support vector machine is an algorithm that is innovated base on the approximately new theory of statistical teaching. Its aim is to be resistant against noise and high generation on new samples. In teaching SVM, Kernels, Kernel parameters and selecting the characteristic have an important role. So, they should be selected precisely to improve the accuracy of SVM classification. One of the background related to SVM, is important in the structure of the SVM algorithm which reduce the time of algorithm performance in teaching. Combination of SVM with other methods is desirably increased the accuracy of this algorithm and made it more powerful in data classification. Its fame is due to its success in recognizing handwritten letters with is equal to exactly arranged nervous networks.*

Keywords: *Support vector machine, kernel, classification, training*

*Department of Computer Engineering, Behbahan Branch, Islamic Azad University, Behbahan, Iran

1- INTRODUCTION

The original SVM algorithm was invented by Vladimir Vapnik in 1963 and was extended#by Vapnik and Corinna Cortes #to the nonlinear case in 1995. The support vector machine (Support vector machines) is one of the supervised learning methods (Supervised learning) that is used for classification and regression. SVM uses a technique called kernel trick for #data #transformation and then finds the optimal boundary between the possible outcomes based on this transformation [1]##

#n simple term, it does quite complex transformations then determines how to disconnect your data based on tags or outputs you have defined.

One of the techniques currently used widely for the classification is the method of support vector machine (SVM). In some way, it may be possible to compare today's popularity of the SVM method to the neural networks popularity in the last decade. The reason of this is the applicability of this method to solving various problems, while methods such as decision-making tree may not be easily used in various issues. SVM algorithm is classified as the pattern identification (recognition) algorithms. #SVM algorithm may be used wherever there is a need for pattern recognition or object classification in particular classes.

2- SVM CLASSIFIER

This algorithm is one of the supervised learning methods used for classification and regression. This is one of the relatively new methods that have indicated good performance in recent years compared to the older methods for classification such as the classification of Perceptron neural networks. The SVM classifier works base on#the linear classification of data and during this classification, it is trying to select the line with more confidence margin. The equation for finding the optimal line for the data is solved by the QP methods known in restricted problem solving (solving problems with significant limitations). Before dividing the line,#the data must be transferred to a space with higher dimensions by the phi functions so that the machine can#classify the highly complex data[2].

The algorithm has strong and perfect theoretical bases and is not sensitive to the number of problem dimensions. The efficient methods for learning SVM are growing rapidly. In a learning process consisting of two classes#the SVM purpose is #to find the best function for classification in the way that the members of the two classes can be identified in the data set. The reason SVM insists on the largest boundary hyper-plane is because it provides better the capability to make generalizations about the algorithm. Not only does it help the performance of the classification and its accuracy on the test data, but also it provides the space for better classification of the future data [3].

One of the problems of SVM is its computational complexity. However, this problem has been acceptably solved. A solution is that a great optimization problem is divided into a series of smaller problems each of which contains a carefully chosen pair of variables which the problem can effectively take advantage of. This process continues until all of the analyzed (divided) parts will be solved.

3- THE SUPPORT VECTOR MACHINE THEORY

The support vector machine is one of the supervised learning methods that can be used for both classification and regression. Originally, it's a two-class classifier that separates the classes by a linear boundary. In this method, the samples closest to the decision-making boundary are called support vectors. These vectors define the decision-making boundary equation. Unlike the case of the empirical risk minimization that is trying to minimize the training error, this method shows better performance on data which have not built the model, for it uses the Structural Risk Minimization Principle applied by the maximization of the distance between two page-passing clouds (hyperplanes) from support vectors of both classes [4].

In order to easily understand the support vector machine theory, the simplest possible expression is used for the two-class classification in a linearly separable case.

This method assumes that the samples are labeled as $y_i = \{-1, +1\}$

Each sample is represented as a vector. The maximum margin method is used to find the optimal decision-making boundary.

So, in addition to correctly dividing all samples of both classes into two classes (groups), the decision-making boundary should find the decision-making boundary (hyper-plane) with the maximum distance from support vectors. Decision-making boundary in a vector space can be mathematically expressed as:

$$f(\bar{x}) = \mathrm{sgn}(\bar{w}.\bar{x} + b)$$

In which w is the normal vector of the hyper-plane and b is its intercept. As mentioned, the decision-making boundary must properly classify the samples. It's mathematically expressed as:

$$y_i(\overline{w}.\overline{x}+b) \geq 1$$

On the other hand, the decision-making boundary should have the maximum distance from the samples of each class and according to figure 1, it means to maximize $\frac{2}{\overline{w}}$

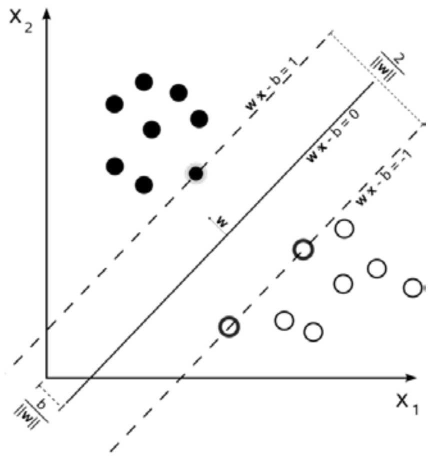

Figure 1: separator hyper-plane and support vectors

Therefore an optimization problem can be defined as:

$$\min \frac{1}{2}\|\overline{w}\|^2$$
$$\text{s.t.} \quad y_i(\overline{w}.\overline{x_i}+b) \geq 1$$

Lagrange multiplier method is used to solve the optimization problem. So the problem becomes:

$$\min_{w,b} \max_{\alpha \geq 0} \{\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\overline{w}.\overline{x_i}+b)-1]\}$$
$$\text{s.t.} \quad \alpha_i \geq 0$$
$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

Where α_i is the Lagrange coefficient.

This form (which is called the primary form) changes into a twin form (Dual Form) by the Placement of the Lagranzhyn derivatives with respect to the primary variables (*w,b*).# Lagrange multiplier values are achieved by solving the optimization problem by Dual Form. Terms Karvsh-Kuhn-Tucker (Karush-Kuhn-Tucker) implies that the optimal value of w is as follows:

$$w = \sum_{i=1}^{n} \alpha_i \, x_i y_i$$

It's also easily proved that the value of b is calculated in the following way:

$$b = \frac{1}{N_{sv}} \sum_{i=1}^{n} y_i - \overline{w}.\overline{x_i}$$

That N_{sv} is the number of support vectors. Finally, the decision-making function can be expressed as follows:

$$f(\overline{x}) = \mathrm{sgn}(\overline{w}.\overline{x} + b)$$

4- SEPARABLE STATE WITH THE EXCEPTION

In this case, taking exceptions to some parts, the samples can be separated linearly [5]. A variable called slack is defined whose value equals to the distance of exception point from the decision-making boundary. In this case, the objective function becomes:

$$\min \frac{1}{2} \|\overline{w}\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad y_i(\overline{w}.\overline{x_i} + b) \geq 1 - \xi_i$$

In which c is the parameter that must be specified by the user and controls the amount of fines (penalty) applied on the objective function for each exception.

5-Nonlinear Separable State

In this state which is the closest state to the real cases, the samples are non-linearly separated (illustration of linear and nonlinear filters can be seen in Figure 2). Input vectors are mapped to a space (feature space) with the dimensions higher than those of the input sample space. Generally, as the sample size increases, the possibility of the linearly sample separation becomes higher.

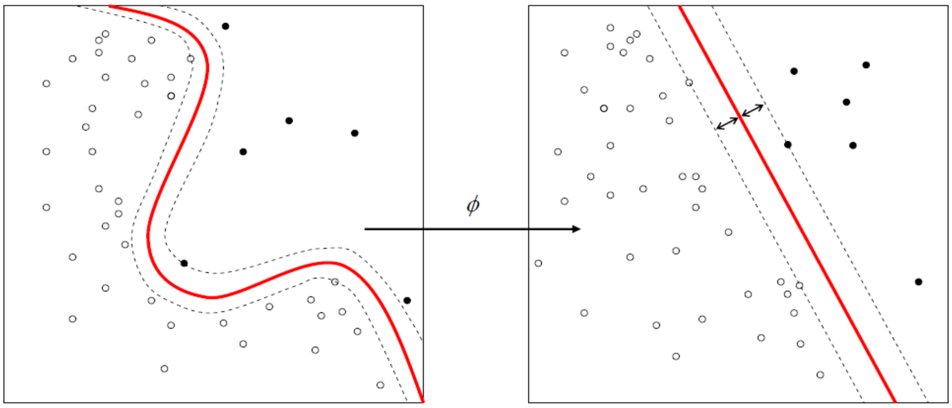

Figure 2: separating linear and nonlinear

Support vector machine finds the optimal decision-making boundary in the feature space and the decision-making boundary is determined by #the mapping of hyper-plane to the input space of the equation. Due to the large dimensions,#he kernel trick is used in order to reduce computation volume (computation complexity) [6]. The equation of inner product of two vectors mapped in the feature space is calculated by the kernel without the need to mapping every single one of those two [7]##

In this case, the dual form is:

$$\max L_d(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j K(\bar{x}_i, \bar{x}_j)$$

$$\text{s.t.} \quad c \geq \alpha i \geq 0$$

$$\sum_{i=1}^{n} \alpha_i \, y_i = 0$$

And the decision-making function is:

$$D(x) = \text{sgn}\left(\sum_{i=1}^{n} y_i \alpha_i K(\bar{x}, \bar{x}_i) + b\right)$$

5- CONCLUSION

Support Vector Machine is a relatively new method of machine learning techniques that is applied in classification problems. This algorithm will converge to a general solution, because of optimization nature that is proposed in that. The Models that have been built with this method because of the structural Minimize and risk-taking have a good performance in the face of data that models are not built based on them.

REFERENCES

[1] C.W.Hsu and C.J.Lin, "*A simple decomposition method for support vector machines*", To appear in Machine Learning, 2001.

[2] G.Fung, O.L.Mangasarian and A.J.Smola, "*Minimal Kernel Classifiers. DataMining Institute Technical*" Report 00-08, November 2000. IEEE Transactions on Pattern Analysis and Machine Intelligence, submitted.

[3] Norikazu Takahashi and Tetsuo Nishi, "*Rigorous proof of termination of SMO algorithm for support vector machines*", IEEE Transactionson Neural Networks, 16(3):774–776, 2005.

[4] Gaelle Loosli and Stephane Canu, Comments on the "*Core Vector Machines: Fast SVM training on very larged at a sets*". Technical report ,INSA Rouen, July 2006. Submitted to JMLR.

[5] Danny Roobaert, "*DirectSVM: A simple support vector machine perceptron*". Journal of VLSI Signal Processing Systems, 2001. To appear.

[6] Gert Cauwenberghs and Tomaso Poggio. "*Incremental and decremental support vector machine learning*". In Advances in Neural Information Processing Systems, (NIPS*2000), volume 13. NIPS, Cambridge MA: MIT Press, 2001

[7] J. Weston, "*LOO-support vector machines*", in Proc. IJCNN'99, 1999.