Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 10 (10): 831-836, 2010
ISSN 1812-5654

Contribution of Kernels on the SVM Performance

H. Fizazi Izabatene, W. Benhabib and S. Ghardaoui
Laboratoire Signal Image Parole SIMPA, Faculté des Sciences,
Université des Sciences et de la Technologie d'Oran USTO-MB, BP 1505 Oran El M'naouer, Algérie

Abstract: The Support Vector Machines (SVM) are learning supervised techniques developed by Vapnik. Their learning, has its roots in the statistical theories with discrimination based on a linear separation in an adequate dimension space. The change of dimension is done through kernel function, which must be chosen from several. In order to evaluate the contribution of the choice of kernel on the SVMs performance, we conducted a classification of a satellite image representing the western region of ORAN, in ALGERIA, with varying kernels and their parameters.

Key words: Satellite imaging, classification, support vector machines, pixel, supervised learning, kernel

INTRODUCTION

The Support Vector Machines (SVM), are statistical technique uses on pattern recognition. Their decision or discrimination function represented by a linear separator (straight line, plane or hyper plane). Among the multitude of linear separators may be used, the SVM seek the most optimal one, presenting the bigger distance between the two classes (Vapnik, 1998; Harris *et al*., 1999; Massih, 2001).

However, data can be not linearly separable; in this case, the force of SVM resides on the use of kernels functions that make a projection, implicitly, to a larger space. In other way, these functions must be selected and configured by the user (Cornujols, 2002; Callut, 2003; Lauer and Bloch, 2006; Joachims, 1998, 2002).

Thus and to assess the influence of kernel in the SVM performance, we conducted a classification of a satellite image (Fizazi *et al*., 2001, 2008) by varying kernels and their parameters from Laplacian to polynomial passing through the Gaussian.

MATHEMATICAL FOUNDATIONS OF SVM

The SVM are, by their nature, a binary classification technique. However, the use of different strategies or approaches has extended their use to the multi-class classification (Devy, 2003; Moutarde, 2007).

Binary SVM: In the case of a binary classification, we consider a set of m data vectors which are associated to labels ti ε (-1, 1) representing their classes. On this dataset

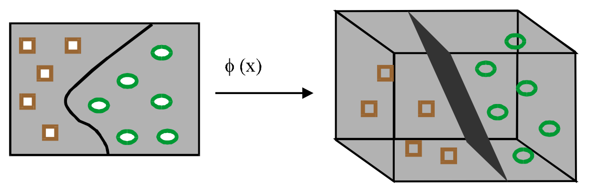

Fig. 1: Linear separation after projection into a bigger space

and to ensure a linear separation, the SVMs perform a projection to a larger space F as shown Fig. 1. This projection is done through a non-linear function φ (x) as:

$$\varphi: R^n \rightarrow F$$

The separator line h(x) of all projected data will be characterized by a normal vector w, a threshold b and a function h (x) = wt. φ (x) + b which will be used to define the class ti of a new example e as: ti = sign (wt. φ (e) + b).

The construction of the separator line h(x) must meet to a primary requirement: the maximization of the margin. The margin represents the distance between the nearest vectors called the support vectors. In another way, margin value is inversely proportional to the norm of w (1) (Vapnik, 1995; Fizazi *et al*., 2008). Consequently, the search of the optimal linear separator can be resumed by an optimization problem: Min 1/2 w^2 under the constraint that all data are correctly positioned relative to the linear separator, which gives: ∀ i ti.h (x) = 1.

In order to simplifying the constraints that are considered heavy, we solve the optimization problem by its dual using the method of Lagrange with introducing

Corresponding Author: H. Fizazi Izabatene, Laboratoire Signal Image Parole SIMPA, Faculté des Sciences, Université des Sciences et de la Technologie d'Oran USTO-MB, BP 1505 Oran El M'naouer, Algérie Tel: 213 41 42 52 07

variables αi called the Lagrangian variable. The αi representing the contribution of an xi element in the construction of the separator line h (x), So, only the αi corresponding to the support vector are not nil .Thus, We obtain L (w, b, α):

$$\begin{cases} \text{Min}\,\frac{1}{2}\|w\|^2 - \sum_{i=1}^{m}\alpha_i.\left[\left(w^t.\,\phi(x_i)+b\right).t_i - 1\right] \\ \alpha_i \geq 0 \;\; i=1...m \end{cases} \quad (1)$$

So, we have to minimize L (w, b, α) according to w, b and maximizing it according to 'α'. Thus, we search the extreme of L (w, b, α):

$$\frac{\delta L(w,b,\alpha)}{\delta w} = 0 \quad \Rightarrow w^* = \sum_{i=1}^{m}\alpha_i.t_i.\phi(x_i) \quad (2)$$

$$\frac{\delta L(w,b,\alpha)}{\delta b} = 0 \quad \Rightarrow w^* = \sum_{i=1}^{m}\alpha_i.t_i = 0 \quad (3)$$

From Eq. 2 and 3 we can rewrite the decision function h (x) as h (x) $\Sigma^n_{i=0}$∝i.ti.φ $(xi)^t$.φ (x)+b and transcribe, through L (w, b, α) the optimization problem to their dual:

$$\begin{cases} \text{Max}\,\sum_{i=1}^{m}\alpha_i - \frac{1}{2}\cdot\sum_{i,j=1}^{m} t_i\cdot t_i\cdot\alpha i\cdot\alpha_j\cdot\phi(x_i)^t\cdot\phi(x_i) \\ \sum_{i=1}^{m}\alpha_i\cdot t_i = 0 \\ \alpha_i \geq 0 \;\; i=1..m \end{cases} \quad (4)$$

Thus, the search of the optimal linear separator became a quadratic programming problem where the αi are computable and b, w deducted.

However, we can note a difficulty in calculating the scalar product φ $(x)^t$·φ (x). This difficulty will increase with increasing the dimension of projection space. For this raison, the use kernel functions k (x, y) as: K (x, y) = Φ $(x)^t$. Φ (y) which do the projection implicitly. The expression of optimization problem expressed in Eq. 4 became as follows:

$$\begin{cases} \text{Max}\,\sum_{i=1}^{m}\alpha_i - \frac{1}{2}\cdot\sum_{i,j=1}^{m} t_i\cdot t_i\cdot\alpha i\cdot\alpha_j\cdot K(x_i,x_j) \\ \sum_{i=1}^{m}\alpha_i\cdot t_i = 0 \\ \alpha_i \geq 0 \;\; i=1..m \end{cases} \quad (5)$$

The main kernel functions used are given in the Table 1.

In practice, it's impossible to classify correctly all data. For this reason Vapnik proposed an upper limit for

Table 1: The main kernel form

Kernel	Generic form	Parameters
Laplacian	K (x, y) = exp (-‖x-y‖/δ)	δ: Standard deviation
Polynomial	K (x, y) = exp $(x{\cdot}y+1)^p$	p: Polynomial order
Gaussian	K (x, y) = exp (-‖x-y‖^2/2.δ^2)	δ: Standard deviation

Lagrangian multipliers αi (Vapnik, 1998) .This limit, called the regularization constant C, represents a compromise between the maximization of the margin and the minimization of the error. So, the optimization problem becomes:

$$\begin{cases} \text{Max}\,\sum_{i=1}^{m}\alpha_i - \frac{1}{2}\cdot\sum_{i,j=1}^{m} t_i\cdot t_i\cdot\alpha i\cdot\alpha_j\cdot K(x_i,x_j) \\ \sum_{i=1}^{m}\alpha_i\cdot t_i = 0 \\ \alpha_i \geq 0 \;\; i=1..m \end{cases} \quad (6)$$

THE MULTI-CLASS SVM

The application of SVMs to classification containing k classes, k>2 requires an increase in the number of binary classifiers (Fizazi *et al*., 2001; Ouamri *et al*., 2009). The management of different classifiers obtained can be done thought two strategies one against one or one against all (Guermeur, 2007):

- The one against one strategy, consists to design all the possible binary classifiers, so, for k classes we have k (k-1)/2 classifiers. To assign an element e to a class, e must be tested with all classifiers designed, whenever e is assigned to a class; we increment a counter i associated to it (i is initially set to zero). e will be assigned to the class that presents a maximum value of a counter
- The one against all approach consists to oppose each class i to the k-1 other classes and, consequently, develop k binary classifiers. To assign an element e to a class, e must be tested with all the classifiers. e will be assigned to the class which presents the maximum value of the decision function at the point e

IMPLEMENTATION AND RESULTS

In present study, we applied SVM to the classification of a satellite image captured by LANDSAT5 Thematic Mapper TM. This satellite image represents the western region of ORAN in ALGERIA on March 15, 1993 at 9 am 45 mn. Data of this study area have been provided by the Spatial Researches Center Arzew-Algeria. The area concerned by our studies is marked in red in Fig. 2.

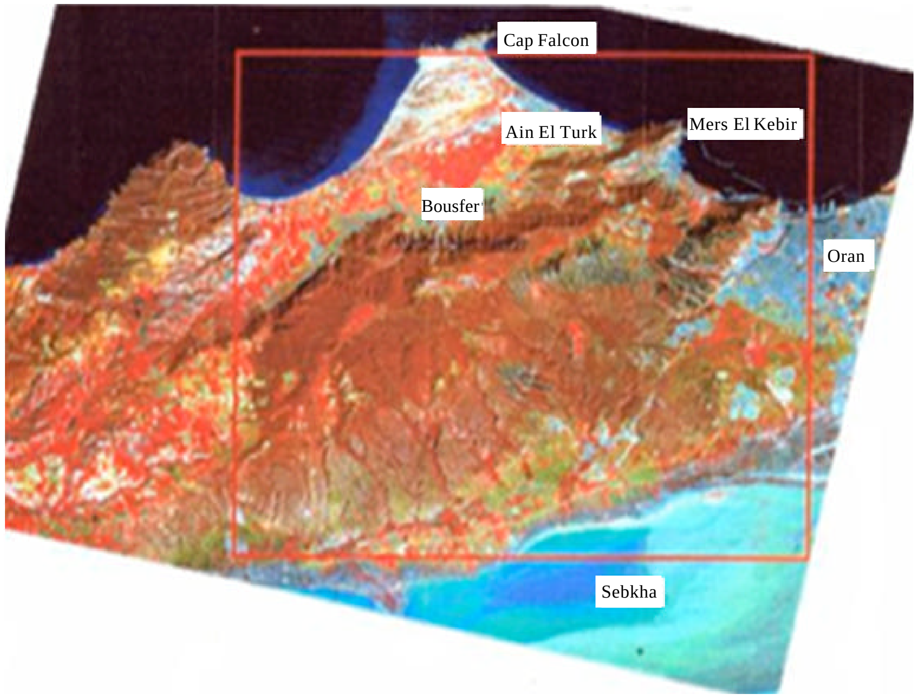

Fig. 2: Satellite image of Landsat5 TM representing the region of Oran in the west of Algeria

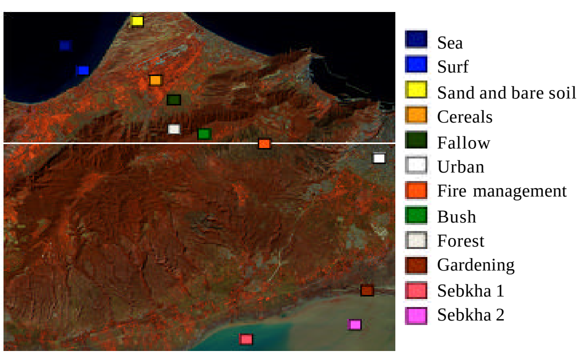

Fig. 3: The different the regions (classes) in the study area

This study area was chosen for its varied classes (12 classes) which can present's an interest for our studies (Fizazi *et al*., 2008).

We started our application by loading tree images corresponding to the channels TM1 TM3 and TM4. To facilitate the use of these images, we started the treatment by madding a contrast enhancement and a color composite with combining the blue filter to channel TM1, the green filter to channel TM3 and the red filter to channel TM4. On, the resulting image we can identified twelve different classes identified using the thematic knowledge. The different classes present on the image are shown in Fig. 3.

Table 2: Results obtained with the Polynomial kernel

Tests	Parameters	Recognition rate (%)
1	p = 2	62.82
2	p = 3	63.50
3	p = 5	63.93

Using our thematic knowledge, we built too, our training set which contained the same number of representatives from each class presents in our area study. Subsequently, our training set contained 840 tri- dimensional vectors (TM1 TM3 TM4).

The training set constructed, we fix the compromise C between the maximization of the margin and the minimization of the error at 500 in order to compare equitably the different kernels functions. In another way and since to the number of classes (12 classes) we have chosen the one against all approach to minimize the number of classifier

Tests and results obtained with polynomial kernel: In this test, we varied the order of the polynomial kernel from 2 to 5. The obtained results are resumed in the Table 2. In another way the Fig. 4 shows the better and the worse resulting images.

Thus, the various tests did with the polynomial kernel, show that increasing of the polynomial order increases the recognition rate. However, it is important to note that this increase is minimal (from p = 2 to p = 5 we

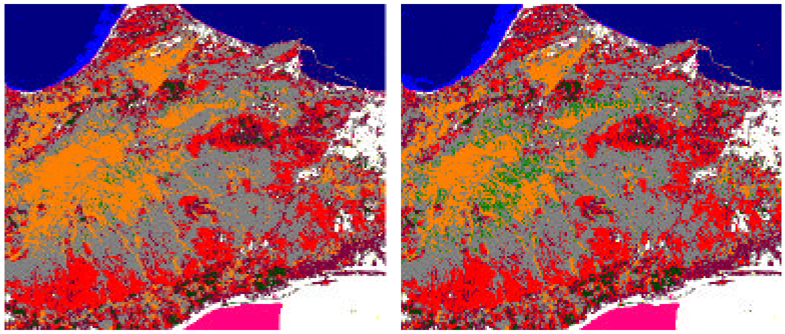

Fig. 4: Resulting images of test 1 and 3 using the polynomial kernel (from left to right). (a) Result of test 1 (much confusion) and (b) result of test 3 (very little confusion)

obtained an increase of 1.11%). These results can be explained by the fact that an order p of the polynomial kernel induce that the vectors will be projected in a space with (n+p)!/n!p! dimension such as n represents the initial space dimension (Ben Ishak, 2007). Thus, the test 3 gave the better results since over the projection space is increasing the separability of data increases too. In another way, the test 1 gave poorer results because its projection space is less than the test 3 one.

From the Fig. 4a resulting from the test 1 of the Polynomial kernel, we can note, a lot of confusions and that specially in the classes: urban, fire management, fallow sebkha2, sand and bare soil with an over estimation of the class forest.

From the Fig. 4b resulting from the test 2 of the Polynomial kernel, we can see confusions in the classes sebkha2, sand and bare soil with a decrease of the over estimation of the class forest and the augmentation of the recognition on the class bush.

To summarize obtained the results, we can say that the augmentation of polynomial order, increases the recognition rate but did not remove the confusions. This relation between the polynomial order and the recognition rate confirmed the results obtained by other applications such as: the recognition of handwritten numbers by SVM (Nemmour and Dibounce, 2007) or the results obtained by hybridization of SVM with other methods such as K-Means (Houcini, 2009).

Tests and results obtained with gaussian kernel: For this test, we varied the Gaussian standard deviation from 0.1 to 1. The obtained results are resumed in the Table 3. In another way the Fig. 5 shows the better and the worse resulting images.

From the Table 3, we can note that the increase of Gaussian standard deviation decrease the recognition rate. Thus, the best recognition rate was done with a Gaussian standard deviation fixed at 0.1. These results can be explained by the fact that the number of the support vector is inversely proportional to the value of standard deviation (Loosli, 2004). Thus, a standard deviation at 0.1 allows us to draw a separator with greater precision. In another way, a standard deviation at 1 gives a less accurate separator

Table 3: Results obtained with the Gaussian kernel

Tests	Parameters	Recognition rate (%)
1	$\delta = 0.1$	74.51
2	$\delta = 0.5$	64.98
3	$\delta = 0\ 1$	62.83

On the Fig. 5a resulting from the test 1 of the Gaussian kernel, we can note, that the fixation of the Gaussian standard deviation at 0.1 gave confusions principally situated at Sebkha2, market-gardening, sand and bare soil.

On the Fig. 5b resulting from the test 2 of the Gaussian kernel, we can see the disappearance of the classes Sebkha2, sand and bare soil with an overestimation of the cereal at the expense of the classes fallow and bush.

To resume the obtained the results with this kernel, we can say that increasing the Gaussian standard deviation has decreased the recognition rate with the disappearance of some classes and an over estimation of another. This relation between the Gaussian standard deviation and the recognition rate confirmed another results obtained on different applications such as: the recognition of handwritten numbers by SVM (Nemmour and Diboune, 2007), the interpretation of medical image (Achour and Sahli, 2006) or the results obtained by hybridization of SVM with K-Means (Houcini and Iarkani, 2009).

Tests and results obtained with laplacian kernel: For the Laplacian kernel tests, we varied the Laplacian standard deviation from 0.1 to 1. The obtained results are given in the Table 4. In another way the Fig. 6 shows the better and the worse resulting images.

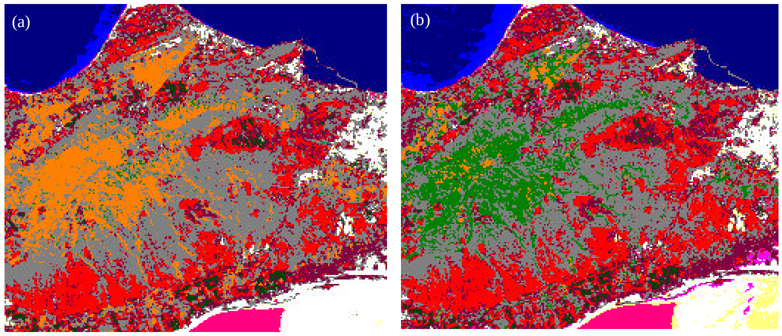

Fig. 5: Resulting images of test 1 and 3 using the Gaussian kernel (from left to right). (a) Result of test 1(presence of confusion) and (b) result of test 2 (no confusion)

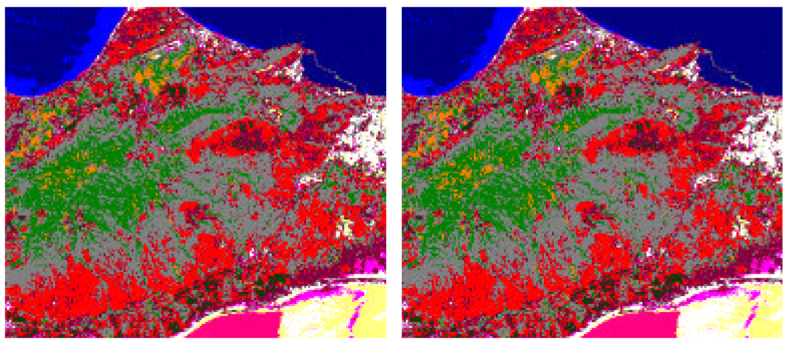

Fig. 6: Resulting images of test 1 and 3 using the Laplacian kernel (from left to right). (a) Result of test 1 (little confusion) and (b) result of test 2 (reduced quality of classification)

Table 4: Results obtained with the Laplacian kernel

Tests	Parameters	Recognition rate (%)
1	$\delta = 0.1$	75.29
2	$\delta = 0.5$	74.87
3	$\delta = 0\,1$	74.45

The first test of the Laplacian kernel was done with a standard deviation fixed at 0.1. This test gave the best recognition rate (75.28% against 74.51 for the Gaussian kernel). In another way; the increase of the Laplacian standard deviation has decreased the recognition rate. Such as the Gaussian kernel, the number of the support vector is inversely proportional to the value of standard deviation (Loosli, 2004). Thus, a standard deviation at 0.1 allows us to draw a separator with greater precision and the standard deviation fixed at 1 gives a less accurate separator.

From the Fig. 6a resulting from the test 1 of the Laplacian kernel, we can note that the test done with a standard deviation fixed at 0.1 presents some confusions at the classes Sebkha 2, market-gardening, urban, sand and bare soil.

From the Fig. 6b resulting from the test 2 of the Laplacian kernel, we can see that increasing the Laplacian standard has reduce the quality of the classification on the image and its recognition rate but the confusions are located in the same classes.

To summarize the obtained the results with this kernel, we can say that increasing the Laplacian standard deviation has decreased the recognition rate confusions located in the same classes. This relation between the Laplacian standard deviation and the recognition rate were confirmed by the hybridization of SVM with K-Means (Houcini and Iarkani, 2009).

CONCLUSIONS

In order to evaluate the importance of the kernels on the SVM performance. We applied the Support Vector Machines to the classification of satellite images (a combination of tree images corresponding to the channels TM1, TM3 and TM4) representing the western region of Oran, by varying the kernels and their parameters.

Thus, we did different tests using the Polynomial, Gaussian and Laplacian kernel. The main conclusion of our experimentation is the sensitivity of the SVM to the choice of the kernel function and its configuration. So, the kernel function must be considered as an influent parameter on the performance of the SVM.

Consequently, the most important difficulty on using SVM is also its greatest strength, the kernel functions. This difficulty is amplified by the absence of algorithms that can configure the type of the kernel function automatically.

REFERENCES

Achour, M. and F. Sahli, 2006. Interpretation of brain images by hybrid approach neuro-SVM. Project of End of Studies of Engineer (June 2006) Department of Data Processing. University of Sciences and the Technology of Oran Algeria.

Ben Ishak, A., 2007. Séléction de variables par les machines á vecteurs supports pour la discrimination binaire et multiclasse en grande dimension. Thése en Cotutelle Franco Tunisienne, l'Uéniversité de la Méditerranée le 06 Décembre 2007.

Callut, J., 2003. Efficient implementation of support vector machines for classification. Memory Presented to Obtain Master in Data Processing. Department of Computer Science, Free University of Brussels.

Cornujols, A., 2002. A new method of learning: SVM support vector machines. Newsletter of AFIA, No. 51, June 2002.

Devy, M., 2003. Introduction to the support vectors machines and other kernels algorithms. Seminars of Classification Cycle. December 2003, CNRS, Group ADTS.

Fizazi, H., N.T. Derraz and F. Marakache, 2001. Classification of satellites images by Dempster-shafer theory. Project of End of Studies of Engineer (June 2001) Director of Project Dr. H. Fizazi. Department of Data Processing. University of Sciences and the Technology of Oran Algeria.

Fizazi, H., S. Ghardaoui and W. Benhabib, 2008. The SVMs for segmenting the content of satellite images. Scientific Days on Computing and Applications, March 2009. Guelma, Algeria.

Guermeur, Y., 2007. vSVM multi-class theory and application. Department of Doctoral Education in Computer Science, Graduate School IAEM Loraine.

Harris, D., D.H. Wu and N.V. Vladimir, 1999. Support vector machines for spam categorization. IEEE Trans. Neural Network, 10: 1048-1054.

Houcini, A. and N. Iarkani, 2009. Segmentation of the satellite image by K-means hybridization and separators has wide margin. Project of End of Studies of Engineer (June 2001) Director of Project Dr. H. Fizazi. Department of Data Processing. University of Sciences and the Technology of Oran Algeria.

Joachims, T., 1998. Making Large-Scale Support Vector Machine Learning Practical. In: Advances in Kernel Methods: Support Vector Machines, Scholkopf, B., C. Burges and A. Smola (Eds.). MIT Press, Cambridge, MA., pp: 386.

Joachims, T., 2002. Learning to Classify Text using Support Vector Machines: Methods, Theory and Algorithms. Kluwer Academic Publishers, Norwell, MA, USA., pp: 224.

Lauer, F. and G. Bloch, 2006. SVM for identification. Days of Identification and Modeling Experiment JIME'2006, Poitiers, France (JIME'2006).

Loosli, G., 2004. Une boîte á outils rapide et simple pour les SVM. Conférence d'Apprentissage 2004 (CAp'04), Presses Universitaires de Grenoble.

Massih, M., 2001. Machine learning and information retrieval: Application to the extraction of surface information and the summary text. Ph.D. Thesis, University of Paris.

Moutarde, F., 2007. Brief introduction to SVMs. Mines School of Paris. http://condor.depaul.edu/~ntomuro/courses/578/notes/SVM-overview.pdf.

Nemmour, H. and R. Diboune, 2007. Use of machines support vector (SVM) for recognition of handwritten digits. Department of Instrumentation and Automation. University of Sciences and the Technology of Houari Boumediene, Algiers, Algeria.

Ouamri, W., A. Tekkouk and H. Fizazi, 2009. Fusion methods of region segmentation of satellite images (FCM-SA). Proceedings of the 1st International Conference on Image and Signal Processing and their Applications, ISPA 2009 University of Abdelahamid Ibn Badis, Mostaganem, Algeria.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. 1st Edn., Springer-Verlag, New York, ISBN: 0-387-94559-8.

Vapnik, V.N., 1998. Statistical Learning Theory. 1st Edn., John Wiley and Sons, New York.