A Hybrid AANN-KPCA Approach to Sensor Data Validation

REZA SHARIFI, REZA LANGARI
Department of Mechanical Engineering
College Station TX-77843
USA

http://www.mengr.tamu.edu/People/facultyinfo.asp?LastName=langari

Abstract: - In this paper two common methods for nonlinear principal component analysis are compared. These two methods are Auto-associative Neural Network (AANN) and Kernel PCA (KPCA). The performance of these methods in sensor data validation are discussed, finally a methodology which takes advantage of both of these methods is presented. The result is a unique approach to nonlinear component mapping of a given set of data obtained from a nonlinear quasi-static system. This method is finally compared with AANN and KPCA for sensor data validation and shows a better performance in terms of predicting/reconstructing the missing or corrupted channels of data.

Key-Words: -Nonlinear PCA, Kernel PCA, Sensor Fault Detection, Auto associative Neural Network

1. Introduction

Principal component analysis (PCA) is a mathematical technique that is used for compression of linearly correlated data. Introduced by Pearson [1] in 1901 and individually developed by Hotteling [2] in 1933, this technique finds the directions with maximum variance in n-dimensional set of data, and transforms the data into new coordinate which is a set of "n" orthonormal vector that have the maximum variance. When transformed, the directions with least variance which are called as noise directions are eliminated and the remaining dimensions contain the maximum information of original data with minimum possible dimension.

Because of the capability of compression of data, this PCA is very common in data processing methods. It is also widely been used for noise filtering and data validation in linear systems. PCA based data-driven fault diagnosis method, which only depends on the input and output data of the monitored process, have found broad applications since 1990's, especially in process industry [3-6]. The core basis of all these methods is that the data obtained from sensors is initially compressed and subsequently uncompressed or de-mapped into the original set of data. Since PCA is a linear technique its performance in sensor fault diagnosis is

acceptable as long as we have adequate co-linearity between different sensor readings. But, most real systems do not satisfy this property. Therefore, we need to find an alternative nonlinear method.

Nonlinear principal component analysis was introduced by Kramer in 1991 [7]. He used a specific architecture of neural network to train a unity network. The architecture of an Autoassociative neural network (AANN) is shown in Figure1. AANN is a five layer feed forward neural network where the second and forth layers incorporate nonlinear transfer functions while the third layer as well as the output layer have linear transfer function. In this architecture the third layer has the least number of neurons. In principle this network provides and identity map [7]; i.e. output of the network is the same as its input. A key element in the operation of the network is the fact that it finds the nonlinear correlation between different channels of input.

This is accomplished mainly via the so called bottleneck layer which has fewer nodes than the input and output layers. This layer in effect acts like a feature detector. In other word the original data are nonlinearly mapped into fewer dimensions than their original form and subsequently de-mapped to their original form. Therefore we can

expect that the data in the reduced dimension section act as a nonlinear principal component of the original data. The nonlinear relation for the first principal component is as follows.

$$Pc_1 = F_1(x) = \sum_{i=1}^{p} w_{1i}^{<2>} \sigma(\sum_{j=1}^{n} w_{ij}^{<1>} x_j) \qquad (1)$$

where $w_{ij}^{<k-1>}$ are the elements of the weight matrix of the kth layer in the neural network, x is the vector of inputs with dimension of n, p is the number of neurons in the mapping layer and σ is the activation function of the relevant neuron.
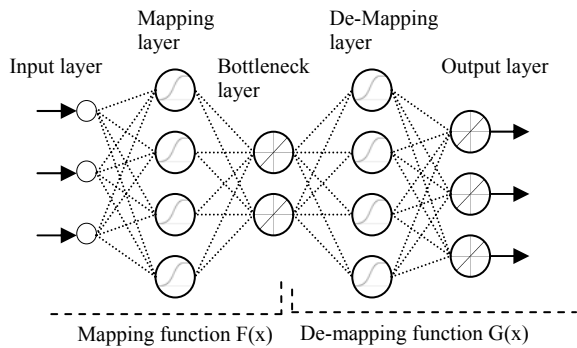

Figure 1 - Architecture of Autoassociative Neural Network

Several successful applications of AANN in sensor fault diagnosis have been presented since its birth in 1991. Hines, et al used AANN for sensor data validation in nuclear power plant [8]. Mattern, et al Applied this method for turbofan engine [9]. Antory, et al. have used this method for industrial process monitoring [10]. In spite of these, there are several difficulties in using AANN for sensor data validation. First of all, like any other neural networks, we need to find the minimum number of adjustable parameters in the network that can model our function within a certain error. This job is much more challenging in AANN because we have a large network of five layers. Another problem is to find the best architecture of the neural network. Kramer [7] in his introductory paper on AANN has presented some upper limits for the number of neurons in each layer but in reviewing the application of AANNs in different areas, it is evident that the best architecture usually has far less parameters (neurons)

than the proposed "upper limit" by Kramer. [11-13]

This problem is even more considerable when it comes to the number of neurons in bottleneck. In fact, the number of neurons in bottleneck is the number of principal components and can be expressed as the number of independent variables in the observed parameters. Therefore selecting more neurons for the bottleneck than the number of independent parameters questions the rationality of using principal components. In next part a method for finding the size of bottleneck in AANN is presented.

The second problem is that there is no specific training methodology for AANN. All of the studied papers in the literature have used the regular backpropagation algorithm for training. Having a neural network with 5 layers at least two of which have nonlinear transfer functions makes it very difficult to train network. In fact training the network is an optimization problem with a very complicated and nonlinear objective function. With the real observed variables which usually are noisy, objective functions have multiple local minimums around the minimum. This problem limits the application of AANNs for sensor fault diagnosis to a very narrow category of systems which have simple forms of nonlinear correlation.

The most important problem with AANNs is that there is no unique solution for the trained network. In other words each time we train the network with different initial conditions, we get different final weights for the network. Therefore we have different sets of nonlinear principal components. Apparently some of these mappings have better performance of fault detection than others. Therefore, the immediate question is that how should we train the network to perform well in sensor fault diagnosis. This problem is also addressed in this paper in the next section.

2. Kernel Principal Component Analysis

Method of Kernel Principal Component Analysis (KPCA) is introduced by Scholkopf et al [14] in 1998. The simplicity of method and its rational mathematical base has made this method very favorable. Numerous applications of this method have been presented Sine its publication,

especially in pattern recognition [15-20], but few of them are in sensor data validation [21],[22].

In this method, Instead of solving principal components in original space, x, we find the principal component of a nonlinear transformation of this space $\Phi(x)$ where $\Phi(x)$ usually have higher dimension than x and even can have infinite dimension. But as we will see, we never need to do any calculation in this high dimensional space.

Assuming the mapping data are centered i.e. $\sum_M \Phi(\mathbf{x}_k) = 0$, we can write the covariance matrix:

$$C = \frac{1}{M} \sum_{j=1}^{M} \Phi(\mathbf{x}_j)\Phi(\mathbf{x}_j)^{\mathrm{T}} \qquad (2)$$

where M is number of observed samples. In order to find the principal values we need to find the eigenvalues and eigenvectors of covariance matrix:

$$\lambda V = CV \qquad (3)$$

Knowing the fact that all solutions of V lie in the span of $\Phi(\mathbf{x}_1),...,\Phi(\mathbf{x}_M)$ [10]. We can write the vector V as follow:

$$V = \sum_{i=1}^{M} \alpha_i (\Phi(\mathbf{x}_i)) \qquad (4)$$

Solving this equation lead to the following Eigenvalue problem

$$M\lambda\boldsymbol{\alpha} = K\boldsymbol{\alpha} \qquad (5)$$

where K is called kernel matrix and is defined as following

$$K_{ij} := (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) \qquad (6)$$

and $\boldsymbol{\alpha}$ is the vector of α_i which are coefficients of eigenvectors used in equation 4.
After normalizing the eigenvectors, in order to find the principal components we have to project the data into the normalized eigenvectors. Therefore, the kth principal value is

$$Pc_k = (V^k \cdot \Phi(\mathbf{x})) = \sum_{i=1}^{M} \alpha_i^k (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})) \qquad (7)$$

One interesting characteristic of Kernel PCA is that we never need to calculate the nonlinear mapping Φ . In fact, we are always dealing with the dot product of two mapping function. Therefore, Instead of selecting function Φ , we can select a kernel function in the following form

$$k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \qquad (8)$$
$$k : R^n \to R$$

In other word, the dimension of nonlinear transfer function can be infinitely large because we never actually work in that space.

Several kernel functions are discussed by Boser et al. [23] and Vapnic et al. [24], but the most kernel function used in KPCA applications is Radial Basis Function defied as

$$k(\mathbf{x},\mathbf{y}) = \exp(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}) \qquad (9)$$

The proposed method of sensor fault diagnosis with KPCA is to find the principal components of observed data using a specific kernel function, then, find inverse transform of these principal components back to the original space.

This process, however, has some shortcomings. First of all we need to know what is the best kernel function for this and what are the parameters of that function for example if we select Radial Basis Function, defined in equation 12, The value of σ is important as well.

Also, size of kernel matrix is square of the number of observed variables. Therefore, with high number of training variables, we have a big kernel matrix and very high volume of calculations.

Another problem is that using KPCA in this form is in fact a "lazy" method in which we are using all of the training data during the online classification of data. Evidently, these types of algorithms take a lot of processing time and memory space and with a high volume of training data this method fails in performance.

3. Sensor Fault Diagnosis with Kernel PCA

If we find the nonlinear principal components of the data, with KPCA algorithm, we can train a simple network to map the data to the found principal components. It is also very simple to train another network to reconstruct the original data out of the principal values. Using these two networks serially works like an Autoassociative Neural Network but the difference is that we have a predefined mapping function which is found based on a rational logic. Figure 2 shows this procedure.

Using KPCA can also help in finding the number of independent variables or the number of neurons in the bottleneck layer of AANN, because the number of principal components is actually the number of independent valuables. Therefore, this method can be used as a supporting method for AANN. This procedure is explained in the next section as an example. Another indirect appliance of KPCA is to use the number of principal components found in this algorithm as the number of neurons in bottleneck layer in AANN. In other word before we apply AANN, we have an insight about the number of independent variables in the training dataset.
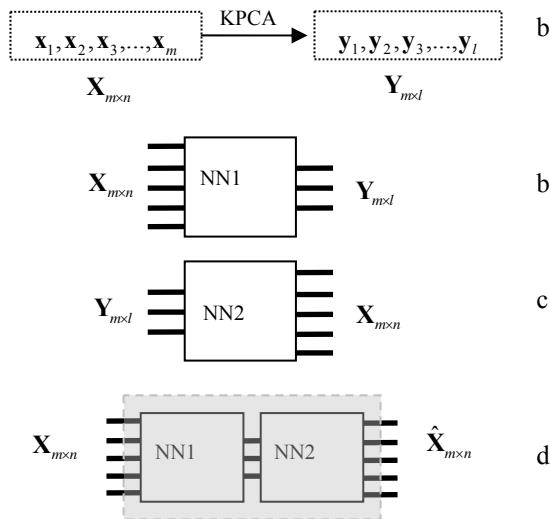

Figure 2- Schematic diagram of sensor fault diagnosis using KPCA

Table 1 compares briefly the two methods of AANN and KPCA. Phrases in the shaded table cells are disadvantages of the method

Table 1 Comparison of KPCA and AANN

Kernel Principal component Analysis	Autoassociative Neural Network
The algorithm gives the number of principal components	We have to assume the number of principal components
The training is a simple linear procedure	The training needs nonlinear optimization
principal components are unique	Non unique solution
Mapping function is limited to the selected kernel function	Capable of wide category of mapping functions
Slow algorithm with large number of observations	No problem with large number of observations
Lazy algorithm	The training data is not needed when trained

4. Application of Method in a Sample Problem

In order to compare these methods quantitatively, a sample problem is solved using AANN and KPCA. In this problem we assume that we have five measurements which are functions of two independent variables. Defined as follows:

$$\begin{cases} x_1 = t_1 + t_2 + e_1 \\ x_2 = \sin(t_1 - t_2) + e_2 \\ x_3 = \cos(t_1 + t_2) + e_3 \\ x_4 = \cos(t_1 - t_2) + e_4 \\ x_5 = (\sin t_1)^2 + e_5 \end{cases}$$

Where $x_1, ..., x_5$ are measurements and t_1 and t_2 are independent variables. In order to provide data for this problem, 100 random numbers in the range $[0, \pi / 2]$ were generated. Values of $e_1, ..., e_5$ are normalized Gaussian random numbers with the mean of zero and standard deviation defined as follows

$$\sigma_i = 0.05 \cdot \text{Range}(x_i)$$
where
$$\text{Range}(x_i) = \text{Max}(x_i) - \text{Min}(x_i)$$

In order to find the best architecture of AANN, 16 different architectures has been tested, and for each one process of training has been done 25 times. The algorithm of training is Backpropagation with Levenberg-Marquardt optimization

algorithm. Two category of architecture are in format of 5-J-2-J-5 and 5-J-3-J-5 for the values of J from 4 to 12

Minimum training error is shown in figure 3. As you see for values of J>7 the amount of training error does not change considerably with further increasing of the number of neurons. Therefore, the number of neurons in the second and forth layer is selected as 7 for both cases.

In order to compare the performance of networks with different size of bottleneck, it is tested with 2 and 3 neurons in the bottleneck. The input values are generated with the following values of t_1 and t_2

$$t_1 = t$$
$$t_2 = \cos(t)$$, $t \in [0,1]$

After generation of the correct values of all 5 sensors, one of them is corrupted by in different regions in order to check the ability of the AANN to reconstruct the faulty data.
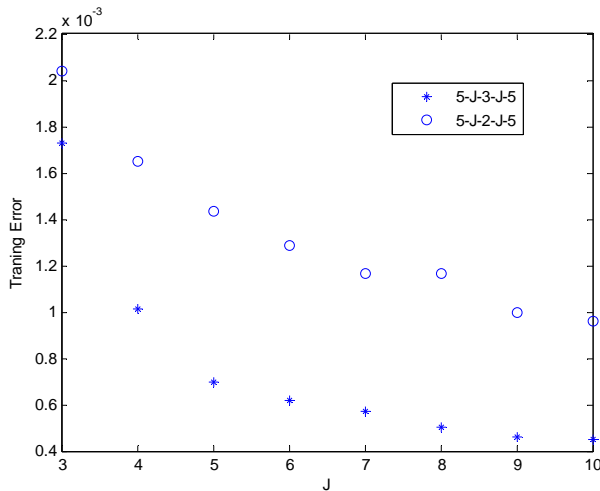

Figure 3 -Value of training error for different structures

The results are shown in Figure 5 and 6 which are the responses of a 5-7-3-7-5 and 5-7-2-7-5 network respectively. Comparing these two figures, we see that although the network with 3 neurons in the bottleneck has less training error -see figure 4-, it has less ability to reconstruct error in the channels. It is also noticeable that in the process of reconstruction for the network with 3 neurons in the bottleneck, the values of other channels are also more deviated which is unfavorable and

might mislead us to detection of different sensor as the faulty sensor.
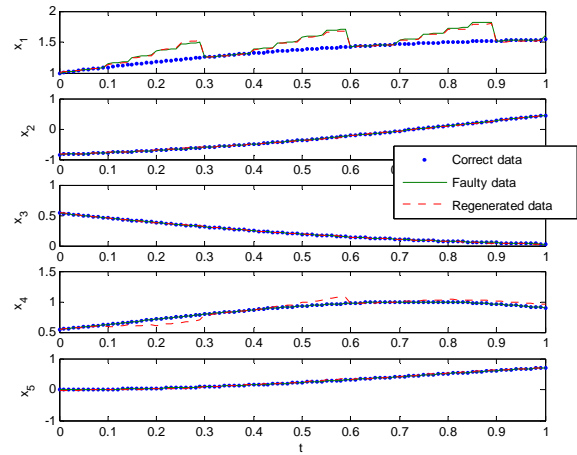

Figure 4 –Sensor value before and after the neural network in for a 5-7-3-7-5 network
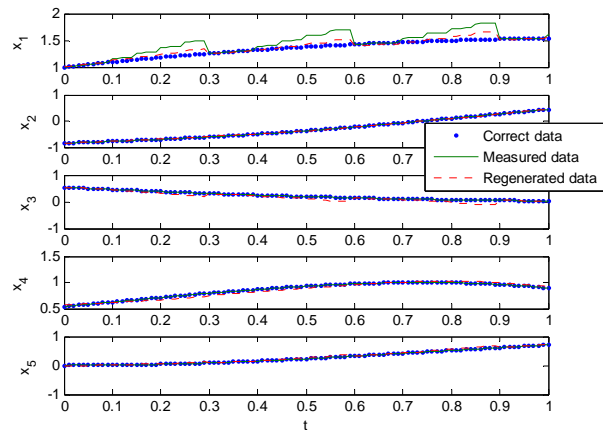

Figure 6 –Sensor value before and after the neural network in for a 5-7-2-7-5 network

Since we have generated these data with a known function, we know that these data are originated from two independent sources but this is not the case for actual application of sensor fault diagnosis. In these cases, Nonlinear PCA with KPCA algorithms can be used. In fact, the number of major principal components represents the number of independent variables. In order to show this fact in our sample data, the algorithm of KPCA is done with radial basis function as defined in Equation 12. The parameter of radial basis function, σ, is changed discretely from 1 to 12 and with each value the algorithm of KPCA is applied and the first 5 principal values are calculated. Figure 6

shows these values. It is obvious from this graph that the first two components have much more contribution to the variance of total data.
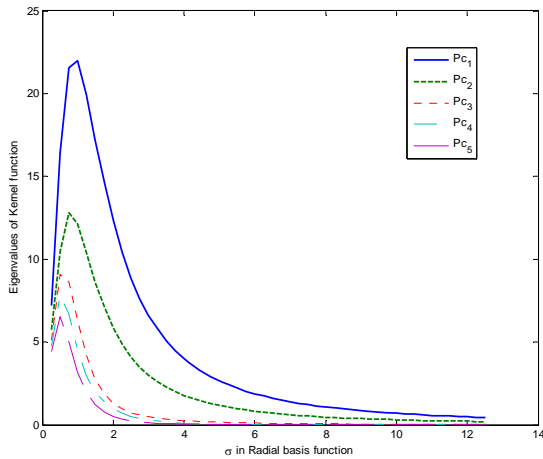

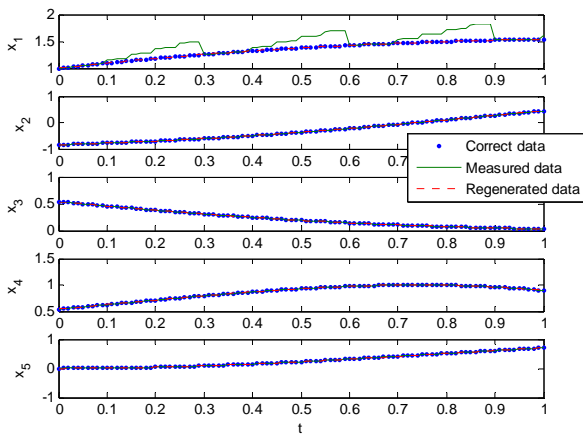Figure 6 –Nonlinear Principal values vs. σ in radial basis function

Figure 7 –Sensor values before and after the fault diagnosis algorithm with KPCA

5. Conclusion

In summary the problems of AANN as a method of Nonlinear PCA and as a technique for Sensor Fault Diagnosis are discussed and some useful strategies for finding the best architecture of AANN are suggested. Specially for the number of neurons in bottleneck layer which represents the number of independent imbedded variables we showed that if this number is more than the number of independent variables the network does not work effectively for the purpose of sensor fault

diagnosis. Method of Kernel PCA is also discussed and its pros and cons compared to AANN are studied. Since in the method of KPCA is a lazy algorithm and the reconstruction of the data form embedded variables needs very complicated computations and requires a nonlinear optimization, we suggested training of two neural network, for mapping and reconstructing respectively. Using these two networks in serial emulates an AANN and the numerical examples shows the better performance of this method.

References:

[1] Pearson K., On lines and planes of closest fit to systems of points in space *Phil. Mag.* 2, 1901 , 559–72

[2] Hotelling H., Analysis of a complex of statistical variables into principal components *J. Educ. Psychol.*, 1933, 24 417–41

[3] Bakshi B. R., Multiscale PCA with application to multivariate statistical process monitoring, *AIChE J.*, Vol 44, 1998, 1596-1610

[4] Gertler G. W., Li Y., Huang, and McAvoy T. Isolation enhanced principal component analysis. *AIChE J.*, 1999, Vol.45, 323-334

[5] Dunia R., Qin S. J., Subspace approach to multi- dimensional fault identification and reconstruction. *AIChE J.*, 1998, Vol.44, 1813-1831

[6] Haiqing W., Zhihnan S., Ping L., Fault Detection Behavior and Performance Analysis of PCA- based Process Monitoring Methods. *Ind. Eng. Chem.* Res., 2002, Vol. 41, 2455-2464

[7] Kramer M. A., Nonlinear Principal Component Analysis using Autoassociative Neural Networks. *AIChE J.*, 1991, Vol.37, 233-243

[8] Hines J. W., Wrest D.J. Uhrig R.E. Plant Wide Sensor Calibration Monitoring. *Proceedings of the 1996 IEEE International Symposium on Intelligent Control*, Dearborn, MI September 15-18, 1996

[9] Mattern D.L., Jaw L.C., Guo T., Graham R. and McCoy W., Using Neural Networks for Sensor Validation, *34th IAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, July 13-15,1998 / Cleveland, OH

[10] Antory D., Kruger U, Irwin G.W., McCullough G., Industrial process monitoring using nonlinear principal component models, Intelligent Systems, 2004. *Proceedings. 2004 2nd International IEEE Conference*, Volume 1, 22-24 June 2004 Page(s):293 - 298 Vol.1

[11] Mattern D.L., Jaw L.C., Ten H.G., Graham R. and McCoy W., Simulation of an engine sensor validation scheme using an Autoassociative Neural Network, *AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, 33rd, Seattle, WA, July 6-9, 1997

[12] Huang J.; Shimizu H.; Shioya S. Data pre-processing and output evaluation of an Autoassociative neural network model for online fault detection in virginiamycin production Source: *Journal of Bioscience and Bioengineering*, Volume 94, Number 1, July 2002, pp. 70-77(8)

[13] Tan S., Mayrovouniotis M.L., Reducing data dimensionality through optimizing neural network inputs, *AIChE Journal* Volume 41, Issue 6, Date: June 1995, Pages: 1471-1480

[14] Scholkopf B. Smola A. Muller K.L., Nonlinear Component Analysis as Kernel Eigenvalue Problem. *Neural Computation*, Volume 10, Number 5, 1 July 1998, pp. 1299-1319(21)

[15] Li W. Gong W., Liang Y. Chen W., Feature Selection Based on KPCA, SVM and GSFS for Face Recognition *Lecture Notes in Computer Science*, 2005, no. 3687, pp. 344-350

[16] He Y. Zhao L. Zou C., Face Recognition Based on PCA/KPCA Plus CCA, *Lecture Notes in Computer Science*, 2005, no. 3611, pp. 71-74

[17] Zhao, Q; Lu, H GA-Driven LDA in KPCA Space for Facial Expression Recognition *Lecture Notes in Computer Science*, 2005, no. 3611, pp. 28-36

[18] Lu B., Bi D., Tan J., A Method of Image Denoising Based KPCA, *Infrared Technology*, 2004, vol. 26, no. 6, pp. 58-61

[19] Kaieda K., Abe S., KPCA-based training of a kernel fuzzy classifier with ellipsoidal regions, *International Journal of Approximate Reasoning*, 2004, vol. 37, no. 3, pp. 189-217

[20] Twining C.J., Taylor C.J., The use of Kernel Principal Component Analysis to model data distributions. *Pattern Recognition* 36, 2003, 217-227

[21] Cho J., Lee J. Choi S. W., Lee D., Lee I.B., Sensor fault identification based on Kernel Principal Component Analysis., *Proceedings of the 2004 IEEE International Conference on Control Applications*, Volume 2 , 2-4 Sept., 2004,1223 - 1228 Vol.2

[22] Choi S. W., C. Lee and I.-B. Lee, "Fault detection and isolation of nonlinear processes based on kemel PCA", *Chemometrics and intelligence Laboratory Systems*

[23] Boser B. E., Guyon I. M., Vapnik V. N., A Training Algorithm for Optimal Margin Classifiers., *Proceeding of the 5th Annual ACM Workshop on Computational Learning Theory.*, 1992, p144-152

[24] Vapnik V., Chervonenkis A., The Nature of *Statistical Learning Theory.*, 1995, New York Springer-Verlag.