Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 26 (2005) 1029–1038
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Support vector machine-based image classification for genetic syndrome diagnosis

Amit David, Boaz Lerner *

Pattern Analysis and Machine Learning Lab, Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, P.O. Box 653, 84105 Beer-Sheva, Israel

Received 31 August 2004
Available online 11 November 2004

Abstract

We implement structural risk minimization and cross-validation in order to optimize kernel and parameters of a support vector machine (SVM) and multiclass SVM-based image classifiers, thereby enabling the diagnosis of genetic abnormalities. By thresholding the distance of patterns from the hypothesis separating the classes we reject a percentage of the miss-classified patterns reducing the expected risk. Accurate performance of the SVM in comparison to other state-of-the-art classifiers demonstrates the benefit of SVM-based genetic syndrome diagnosis.

Keywords: Support vector machine (SVM); Multiclass classification by error correcting output code (ECOC); Rejection; Fluorescence in situ hybridization (FISH); Genetics.

1. Introduction

Since introduced to the machine learning community by Vapnik (1995) and especially in the recent years (Burges, 1998; Cristianini and Shawe-Taylor, 2000), the support vector machine (SVM) has been studied extensively for classification, regression and density estimation. Based on the principle of structural risk minimization (SRM), the SVM upper bounds the expected risk (mean error rate measured on the test set) using the sum of the empirical risk (mean error rate measured on the training set) and a bound, called the VC confidence. In order to construct an optimal decision hypothesis, the SVM holds the empirical risk fixed and minimizes the VC confidence by limiting the flexibility of the set of candidate functions searched by the machine (classifier). The minimum of this upper bound is reached by maximizing the margin between the decision hypothesis and the classes as defined using support vectors, thereby enabling the improvement of the

* Corresponding author.
E-mail addresses: boaz@ee.bgu.ac.il (B. Lerner), davidami@ee.bgu. ac.il (A. David).

0167-8655/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2004.09.048

classifier generalization ability. A quick survey of publications in Pattern Recognition Letters of the last four years alone, reveals a variety of applications the SVM is applied to, spanning from texture (Barzilay and Brailovsky, 1999) and face recognition (Déniz et al., 2003; Pang et al., 2003) to ultrasonic image segmentation (Kotropoulos and Pitas, 2003) and text detection (Lee et al., 2003). Additional applications studied are object detection in real-time streaming video (Genov and Cauwenberghs, 2003), spam message filtering (Drucker et al., 1999) and classification of cancer tissues (Furey et al., 2000). The SVM, having superior generalization capability and reputation of a highly accurate paradigm, is applied in this study to the classification of cytogenetic images.

In recent years, fluorescence in situ hybridization (FISH) has emerged as one of the most significant new developments in the analysis of human chromosomes. FISH offers numerous advantages compared with conventional cytogenetic techniques since it allows detection of numerical chromosome abnormalities during normal cell interphase. An important application of FISH is dot counting, i.e., the enumeration of signals (dots) within the nuclei, as the dots in the image represent the inspected chromosomes. Dot counting is used for diagnosing numerical chromosomal aberrations in, e.g., haematopoietic neoplasia, solid tumors and prenatal diagnosis (Tanke et al., 1995). However, a major limitation of the FISH technique for dot counting is the need to examine large numbers of cells. This is required for an accurate estimation of the distribution of chromosomes over cell population, especially in applications involving a relatively low frequency of abnormal cells. As visual evaluation by a trained cytogeneticist of large numbers of cells and enumeration of hybridization signals is expensive and time consuming, FISH analysis for dot counting can be expedited by automating the procedure.

A neural network (NN) two-class classifier has recently been proposed (Lerner et al., 2001) in order to discriminate between real signals and artifacts resulting from noise and out-of-focus images, thus enabling FISH dot counting. Aiming at clinical diagnosis we are interested here in improving the NN classification accuracy by studying an SVM classifier. To achieve the highest SVM accuracy for the cytogenetic problem, we empirically investigate using the SRM principle and the cross-validation methodology, not so commonly applied in SVM experimentation, the configuration, parameters and normalization method most appropriate for the SVM. Since clinical diagnosis requires studying a several genetic syndromes simultaneously, we also extend the suggested framework to investigate multiclass SVM-based FISH signal classification. Furthermore, we threshold the distance of tested data points from the optimal separating hypothesis in order to reject a user-defined percentage of the miss-classified patterns, allowing reduction of the SVM expected risk. Finally, we compare classification of cytogenetic images using the SVM to that based on other state-of-the-art machine learning paradigms. This study is an extension of a previous one (David and Lerner, 2004).

The SVM classifier and principles are described in Section 2. The cytogenetic domain represented by FISH images is illustrated in Section 3. The experimental methodology for evaluating the SVM and its results in classifying FISH signals are, respectively, given in Sections 4 and 5, before a discussion concluding the paper is provided in Section 6.

2. Support vector classifiers

2.1. Structural risk minimization

In statistical learning theory (Vapnik, 1995), we bound the difference between the expected risk (or risk), $R(\boldsymbol{\alpha})$, and the empirical risk, $R_{\text{emp}}(\boldsymbol{\alpha})$, when both training and test sets are assumed to be generated from the same underlying probability distribution $P(\boldsymbol{x}, y)$. The empirical risk is calculated by

$$R_{\text{emp}}(\boldsymbol{\alpha}) = \frac{1}{2l}\sum_{i=1}^{l}|y_i - f(\boldsymbol{x}_i, \boldsymbol{\alpha})|, \tag{1}$$

where l is the size of the training set, $\boldsymbol{\alpha}$ are the model parameters and $f(\boldsymbol{x}_i, \boldsymbol{\alpha})$ is the classifier output for a training vector (pattern) $\boldsymbol{x}_i$ having a cor-

responding label $y_i \in \{-1, 1\}$. The risk for an unseen test vector $\boldsymbol{x}$ is

$$R(\boldsymbol{\alpha}) = \int \frac{1}{2}|y - f(\boldsymbol{x}, \boldsymbol{\alpha})| \, \mathrm{d}P(\boldsymbol{x}, y). \tag{2}$$

Vapnik (1995) showed that with probability (confidence level) $1 - \eta$ for some $0 \leqslant \eta \leqslant 1$ a bound on the risk can be calculated as

$$R(\boldsymbol{\alpha}) \leqslant R_{\mathrm{emp}}(\boldsymbol{\alpha}) + \sqrt{[h(\log(2l/h) + 1) - \log(\eta/4)]l^{-1}} \tag{3}$$

trading between tight and reliable bounds for η close to 1 and 0, respectively. The parameter h is a non-negative integer called the Vapnik–Chervonenkis (VC) dimension defined as the largest number of patterns that can be separated (shattered) in all possible ways using the family of functions $f(\boldsymbol{\alpha})$ implemented by the classifier. Thus, the VC dimension is a measure of the capacity (complexity) of the class of functions in relation to the available amount of data. Bounding the risk manifests a well known pattern recognition trade off where a highly complex hypothesis is able to discriminate between much more complex data structures but at the cost of increasing the difference between the empirical and the expected risk, i.e., the VC confidence, in a phenomenon usually referred to as "over fitting". To find an optimal hypothesis reducing the VC confidence, the concept of SRM is introduced. It establishes nested subsets of hypotheses in which each subset has a smaller VC dimension than that of its outer subset. The concept builds upon finding the hypothesis subset that minimizes the bound on the risk while training the machine for each subset to minimize the empirical risk. The machine having the minimal sum of empirical risk and VC confidence is then being selected (Fig. 1).

Furthermore, the SVM selects the optimal hypothesis as the one yielding the maximum margin of separation between two classes (Vapnik, 1995). A large margin leads the SVM to an error bound depending on an "effective" VC dimension smaller than the VC dimension, thus providing a classifier having higher generalization capability (Cristianini and Shawe-Taylor, 2000).

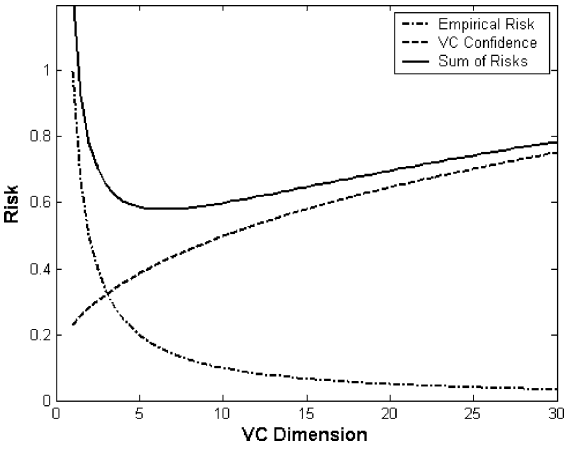

Fig. 1. An example of a decreasing function empirical risk, VC confidence and the sum of these two risks for increasing VC dimensions, 100 training patterns and 0.95 confidence level. The valley in the sum suggests the VC dimension value providing the best generalization for the SVM.

2.2. Linear and non-linear SVM classifiers

In the separable case, the linear SVM achieves the optimal separating hypothesis (hyperplane), i.e., the one maximizing the margin, by minimizing $\|\boldsymbol{w}\|^2/2$ while keeping

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1 \geqslant 0, \; i = 1, \ldots, l, \tag{4}$$

where $\boldsymbol{w}$ is the vector of hyperplane coefficients (orientation), b is a bias term and $\frac{b}{\|\boldsymbol{w}\|}$ indicates the distance of the hyperplane from the origin. Training an SVM is a quadratic optimization problem (Burges, 1998) in which

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) \tag{5}$$

is maximized w.r.t. $\alpha_i \geqslant 0$ subject to $\sum_{i=1}^{l} \alpha_i y_i = 0$, where the constraint on the Lagrange multipliers $\alpha_i \geqslant 0$ are derived from the Lagrangian formalization (Burges, 1998). When the optimization completes, the normal to the optimal hyperplane, $\boldsymbol{w}$, is computed by

$$\boldsymbol{w} = \sum_{i=1}^{l} \alpha_i y_i \boldsymbol{x}_i \tag{6}$$

and the bias b is computed from the Karush–Kuhn–Tucker conditions $\alpha_i \cdot [y_i((\boldsymbol{x}_i \cdot \boldsymbol{w}) + b) - 1] = 0$

for any i for which $\alpha_i \neq 0$, or averaged over all training points. Because the problem is quadratic, thereby convex, the optimization process always results in a global maximum. The maximal margin classifier is then derived using

$$f(\boldsymbol{x}) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i (\boldsymbol{x} \cdot \boldsymbol{x}_i) + b\right). \tag{7}$$

The support vectors, i.e., the vectors defining the maximal margin during training, are the only vectors necessary for shaping the decision boundary and they cannot be overlooked during classifier training. Correspondingly, their Lagrange multipliers are the only coefficients which do not vanish during training and thus participating in shaping the solution (6). The fewer the number of support vectors the better generalization can be expected (Cristianini and Shawe-Taylor, 2000).

For the non-separable case, no particular separating hyperplane can be found as some of the α_i's will get increasingly large values. The solution for this problem is to allow errors of the SVM in certain amount $\xi_i \geqslant 0$, $i = 1, \ldots, l$ in (4), i.e.,

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - (1 - \xi_i) \geqslant 0, \ i = 1, \ldots, l \tag{8}$$

and penalize each error by a user-defined parameter C determining the degree of penalty assigned to an error. The optimal hyperplane is then determined by minimizing

$$\|\boldsymbol{w}\|^2/2 + C\left(\sum_{i=1}^{l} \xi_i\right)^k, \quad k > 0. \tag{9}$$

Repeating the optimization procedure as for the linear separable case leads to similar equations and constraints except that $0 \leqslant \alpha_i \leqslant C$, and the support vectors can now lie either on the margin ($y_i f(\boldsymbol{x}_i) = 1$) having $0 < \alpha_i < C$ or inside the margin area ($y_i f(\boldsymbol{x}_i) < 1$) having $\alpha_i = C$.

A powerful approach to gain linear separation in the non-separable case is to map the original data vectors and all resulting vector inner products needed for the above training algorithm into a higher-dimensional (even infinite) space, where linear separation may be achieved by a maximal margin classifier. A motivating example is shown in Fig. 2 in which non-separable one-dimensional data are mapped onto a two-dimensional space

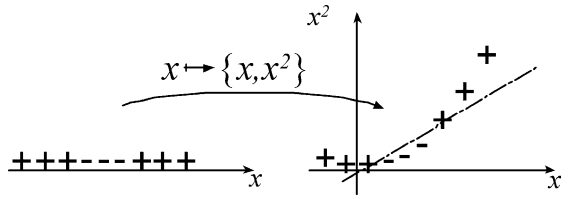

Fig. 2. Achieving linear separation by projecting non-separable one-dimensional data onto a two-dimensional space.

using the mapping $x \mapsto \{x^2, x\}$ before being separated linearly.

Vapnik (1995) showed that if there exists a kernel function K in the original space that describes the inner product between vectors x_i and x_j in the space projected by Φ

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) \tag{10}$$

then the SVM formalization leads to the same training procedure and solution as for the linear case independently of feature space dimension. The kernel replaces the inner product modifying L_D (5), as well as the decision boundary (7)

$$f(\boldsymbol{x}) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b\right). \tag{11}$$

Again, the (support) vectors having non-zero α_i are the only vectors to shape the decision boundary, i.e., a non-linear optimal separating hypothesis. The choice of a kernel function for the SVM has been researched extensively in the literature although usually the implementation is being performed in an ad-hoc fashion. An empirical evaluation of three kernel types for the cytogenetic domain is described in Section 4.2.

2.3. Multiclass SVM

The SVM is a binary classifier which can be extended by fusing several of its kind into a multiclass classifier. In this study, we fuse SVM decisions using the error correcting output codes (ECOC) approach, adopted from the digital communication theory (Dietterich and Bakiri, 1995). In the ECOC approach, up to $2^{n-1} - 1$ (where n is the number of classes) SVMs are trained, each of them aimed at separating a different combina-

tion of classes. For example, if we have 3 classes A, B and C we need three classifiers; One SVM classifies A from B and C, a second SVM classifies B from A and C and a third SVM classifies C from B and A. The multiclass classifier output (target) code for a pattern is a combination of targets of all the separate SVMs. That is in our example, vectors from classes A, B and C have codes $\{1,-1,-1\}$, $\{-1,1,-1\}$ and $\{-1,-1,1\}$, respectively. If each of the separate SVMs classifies a pattern correctly, the multiclass classifier target code is met and the ECOC approach reports no error for that pattern. However, if at least one of the SVMs misclassifies the pattern, the class selected for this pattern is the one its target code closest in the Hamming distance sense to the actual output code and this may be an erroneous decision. Table 1 gives an example of an ECOC for the four-class classifier employed in Section 4 in FISH signal discrimination using seven SVMs creating a 7-bit-code.

2.4. Rejection

Since most of the classification errors occur near the separating hypothesis where classes overlap, rejecting patterns located at this region can reduce the overall risk. However, this method rejects both wrongly and correctly classified vectors, thus reducing both the error and correct classification rates. This general pattern recognition principle allows regulating the number of rejected patterns in order to meet an error requirement. In this study, we apply this principle to the SVM by thresholding its output (the distance of patterns from the separating hypothesis), thus rejecting wrongly classified patterns close to the separating hypothesis. Moreover, changing the threshold provides a means to adjust the SVM to different requirements.

Table 1
The ECOC used in FISH signal classification

Class	Bit 1	Bit 2	Bit 3	Bit 4	Bit 5	Bit 6	Bit 7
A	−1	−1	−1	−1	−1	−1	−1
B	−1	−1	−1	1	1	1	1
C	−1	1	1	−1	−1	1	1
D	1	−1	1	−1	1	−1	1

"Bit *i*" represents the *i*th SVM.

3. The FISH domain

One approach to automatic FISH dot counting relies on an auto-focusing procedure (Netten et al., 1996) to select the 'clearest' two-dimensional projected image of the three-dimensional nuclei. However, basing dot counting on auto-focusing mechanisms has some shortcomings and it has been recently proposed (Lerner et al., 2001) to base FISH dot counting on a classifier discriminating between in and out-of-focus images taken at different focal planes of the same field-of-view (FOV) composing a stack of images. Each stack image is analyzed and its signals are discriminated by the classifier as valid (real) data or artifacts, which are the result of out-of-focusing. Following this discrimination in each stack image, an image containing no artifacts is selected to represent the stack (FOV). The procedure is then repeats itself for other FOVs until the entire slide is covered or the required number of (in-focus) images (or nuclei) is obtained. Proportion estimation of the number of cells having a specific number of signals (chromosomes) can be then performed using these images in order to reflect possible numerical abnormalities identifying a genetic disease.

Indeed, previous research (Lerner et al., 2001; Lerner and Lawrence, 2001) has already indicated the accuracy of FISH signal classification for dot counting using machine learning classifiers. The preparation of the FISH data as well as image acquisition and analysis are described in (Lerner et al., 2001) and thus avoided here. Over 3000 signals are segmented from 400 FISH images and represented by twelve features of size, shape, intensity and color before being introduced to the SVM classifying the signals as real or artifact of two genetic syndromes—trisomies 21 (Down syndrome) and 13 (Patau syndrome).

4. The empirical study

4.1. Methodology

We have studied and applied an SVM in classifying real and artifact FISH signals of two genetic syndromes. SRM for the SVM is explored

empirically by employing different families of functions, characterized by their kernels and different settings of classifier parameters. The SVM parameters are usually determined as those minimizing the expected risk (Kotropoulos and Pitas, 2003) or an upper bound on the expected risk measured on the training set using leave-one-out cross-validation (Cristianini and Shawe-Taylor, 2000; Kotropoulos and Pitas, 2003). However, parameters determined using the test set are biased and may lead to overfitting that set at the expense of satisfactory generalization ability. Furthermore, the use of a bound established using the training set cannot guarantee reliable estimation of the risk (measured on the test set). Thus, in this study we evaluate the SVM parameters using a third set, a validation set, as part of a cross-validation (CV) procedure. Although both validation and cross-validation are uncommon in the experimentation of SVM, they achieve reliable error estimation (Dietterich, 1998). That is, training and validating SVMs having different combinations of kernels and settings of parameters are evaluated on a partition of the data including 70% of the signal patterns using a 5-fold cross-validation (CV5) experiment. The setting corresponding to the SVM achieving on average the minimum risk on the validation set is employed to train an SVM on that partition and then to test it on the other partition of the data containing the remaining 30% of the patterns. Measured on the test set, the SVM risk is compared to that of other machine learning state-of-the-art classifiers.

4.2. SVM model selection

In exploring the projection best suitable for the FISH data, we have evaluated three different models (kernels). The linear kernel composes of an inner product between data vectors

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j \tag{12}$$

leading to a separating hyperplane. The only parameter that should be set for this kernel is the cost parameter C. A more complicated kernel is a polynomial of degree D in the data

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j + 1)^D \tag{13}$$

leading to a polynomial of degree D separating hypothesis. D is determined independently of the dimensionality of the data. The third kernel used is a symmetric hyper Gaussian, called a radial basis function,

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left\{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}\right\}. \tag{14}$$

The separating hypothesis is a linear combination of symmetric hyper Gaussians which their widths are controlled by σ.

The SVM parameters are usually selected in an ad-hoc manner (Cristianini and Shawe-Taylor, 2000; Kotropoulos and Pitas, 2003; Déniz et al., 2003). Here we optimize these parameters for the cytogenetic data simultaneously over a grid of values. For example, values for C, D and σ are checked in the $[1, 10{,}000]$, $[0.5, 10]$ and $[0.5, 1000]$ ranges, respectively. Fig. 3 shows an example of this simultaneous optimization for the Gaussian SVM through the presentation of contours of equal risk for different (C, σ) parameter settings.

4.3. Implementation

Training an SVM is mainly solving a quadratic optimization problem. However, for large databases it is computationally demanding causing slow convergence of the optimization process. Approaches to reduce the computational demands divide the problem into smaller problems, solving each one separately and combining the results into a complete solution. We examine here two of these approaches. The first approach is a MATLAB wrapper (Schwaighofer, 2002) to SVM*light* (Joachims, 1999) and the second is a routine (Schwaighofer, 2002) utilizing the MATLAB optimization toolbox. The first approach has been found superior both in the classification error it achieves and speed of convergence.

4.4. Data normalization

Three data normalization methods have been checked in comparison to no normalization. The first method ('scaling') normalizes the data to have zero mean and unit variance in each dimension. The second method ('linear') maps the data linearly onto the $[-1, 1]$ range. The third normalization

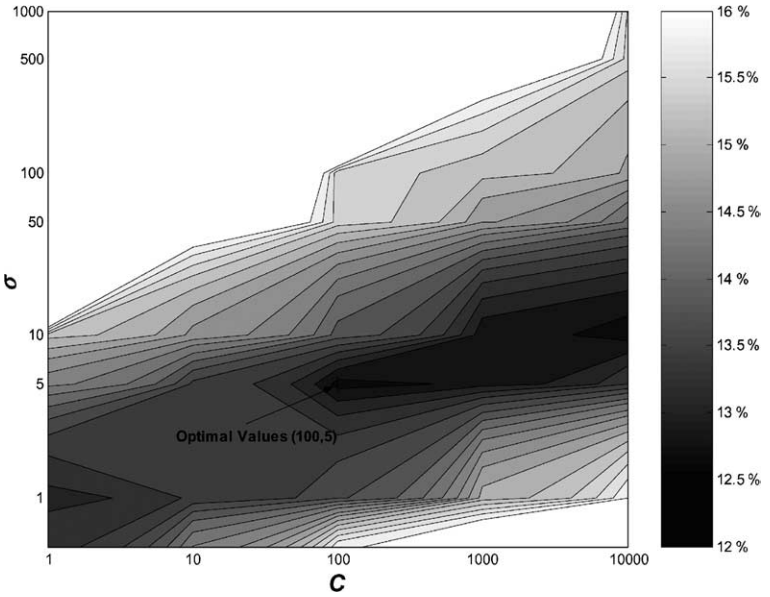

Fig. 3. Optimization using the risk with respect to (C, σ) of the Gaussian SVM for the cytogenetic data.

Table 2
Comparison of the Gaussian SVM risk for different normalization methods of the cytogenetic data, each utilizing an optimal setting of (C, σ)

Normalization method	Expected risk (%)	C	σ
None	13.77	10,000	50
Scaling	12.82	100	5
Linear	13.03	100	1
Whitening	13.24	100	5

method ('whitening') applies to the data a whitening transformation $A = \Lambda^{-1/2}U^{\mathrm{T}}$, where U and Λ are the matrices of eigenvectors and eigenvalues of the data covariance matrix, respectively, coinciding the transformed data coordinates with the principal axes (similar to the well-known principal component analysis (PCA) Fukunaga, 1990) and also turning the covariance matrix to the identity matrix. Our experiments revealed that the 'scaling' normalization method fitted the data to each of the kernels better than the other normalization methods, as reflected in Table 2 for the Gaussian kernel.

4.5. Nearest-neighbor classifier

To establish a reference classifier to the SVM, we implement the k-nearest-neighbor (KNN) model classifying a test data vector according to the most frequent class label of the k closest training vectors. Based on Table 3, showing the KNN classification error of the cytogenetic data for different k, we choose the 7-nearest-neighbor classifier as a reference for the SVM.

Table 3
Error rates achieved by the KNN classifier having different k values

k	Error rate (%)
1	18.75
3	13.98
5	14.19
7	13.24
9	14.41

5. Results

5.1. Binary classification

First we test the SVM in classifying FISH signals as real or artifacts. For each of the evaluated kernels, we optimize the relevant parameters as

Table 4
The SVM risk for FISH signal classification using different kernels and corresponding optimal settings

Kernel	Kernel parameter	C	Expected risk (%)
Linear	–	10	15.45
Polynomial	$D = 3$	7	13.14
Gaussian	$\sigma = 5$	100	12.82

described before. Table 4 depicts the SVM risk for the optimal parameters of each of the kernels. Since the Gaussian kernel demonstrates superior performance over the other kernels, it is the kernel of choice for the remaining experiments.

5.2. Four-class classification

Extending previous experiment and according to Section 2.3, we employ seven SVMs and the ECOC algorithm to discriminate signals as real or artifact of two genetic syndromes indicated in the FISH images by red and green signals. Each of the seven SVMs of the four-class classifier uses a Gaussian kernel and an optimal setting for the parameters of that SVM. Table 5 gives the confusion matrix of the resulted SVM-based classifier for the four classes. The classifier overall risk is 15.77%.

5.3. Rejection

As explained in Section 2.4, we increase the number of rejected patterns leading to a reduced risk by raising the threshold on the SVM output. Fig. 4 demonstrates this procedure for the binary Gaussian SVM and threshold values in the range

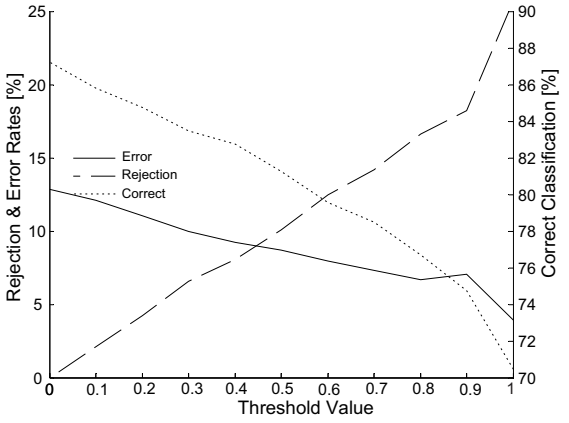

Fig. 4. Rejection implemented for the binary Gaussian SVM classifying the cytogenetic data.

[0, 1]. This range of threshold values is selected since "0" corresponds to no rejection and "1" to the maximum absolute value of the SVM output for the support vectors in the separable case, although other upper values having less intuitive meaning may do as well.

5.4. Comparison to other machine learning classifiers

A comparison of the Gaussian SVM optimized to the FISH problem with other state-of-the-art classifiers for the two-class classification problem is provided in Table 6. The 7-nearest-neighbor, which is the most accurate KNN according to Table 3, as well as a multilayer perceptron neural network, Bayesian neural network, naïve Bayesian classifier and linear classifier are the compared

Table 5
A confusion matrix for the accuracy of the SVM-based four-class classifier of FISH signals

True class	SVM-based classifier			
	Real red	Artifact red	Real green	Artifact green
Real red	90.79	9.2	0.00	0.00
Artifact red	13.45	84.26	0.00	2.28
Real green	0.62	0.00	81.13	18.24
Artifact green	0.42	5.02	12.55	82.00

Table 6
Comparison of the Gaussian SVM to other machine learning classifiers, all optimized to the FISH domain

Classifier	Error rate (%)
Gaussian SVM	12.8
7-Nearest-neighbor	13.2
Multilayer perceptron neural network	13.6
Bayesian neural network	11.8
Naïve Bayesian classifier	17.0
Linear classifier	15.9

classifiers. These four latter classifiers were optimized to the problem previously (Lerner and Lawrence, 2001). Besides the Bayesian neural network, the SVM is the most accurate classifier achieving higher than 87% accuracy in classifying over 3000 FISH signals into real and artifact classes.

6. Discussion

We have suggested an SVM for the classification of images required for genetic syndrome diagnosis. Using the principle of SRM and cross-validation procedure we selected a model for the SVM evaluating linear, polynomial and Gaussian kernels. Having the most accurate separating hypothesis, the SVM utilizing the Gaussian kernel was optimized to the FISH domain. The SVM, extended to multiclass problems using the ECOC algorithm, accurately classified FISH signals as real or artifacts of two genetic abnormalities.

Thresholding the SVM output provided a mechanism of trading off erroneous decisions with rejecting a fraction of the data. This rejection mechanism, rarely employed in the SVM literature, opens an opportunity to further hierarchical increase of model precision.

Accurately discriminating FISH signals, the SVM classifier is a major contribution to genetic syndrome diagnosis. The classifier outperforms most of other machine learning classifiers and is inferior only to the Bayesian neural network, the latter having pronounced computational requirements. As the superiority of the Bayesian neural network is attributed to the exploitation of a-priori knowledge additionally to that acquired from the data, we are interested in ways to combine this knowledge within the SVM framework in order to improve the classifier performance.

Acknowledgement

This work was supported in part by the Paul Ivanier Center for Robotics and Production Management, Ben Gurion University, Beer-Sheva, Israel.

References

Barzilay, O., Brailovsky, V.L., 1999. On domain knowledge and feature selection using a support vector machine. Pattern Recognition Lett. 20, 475–484.

Burges, C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. 2, 121–167.

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines. Cambridge University Press, Cambridge, UK.

David, A., Lerner, B., 2004. Signal discrimination using a support vector machine for genetic syndrome diagnosis. In: 17th Internat. Conf. on Pattern Recognition (ICPR2004), 23–26 August, Cambridge, UK.

Déniz, O., Castrillón, M., Hernández, M., 2003. Face recognition using independent component analysis and support vector machines. Pattern Recognition Lett. 24, 2153–2157.

Dietterich, T.G., 1998. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput. 10, 1895–1923.

Dietterich, T.G., Bakiri, G., 1995. Solving multiclass learning problems via error-correcting output codes. J. Artif. Intell. Res. 2, 263–286.

Drucker, H., Wu, D., Vapnik, V., 1999. Support vector machines for spam categorization. IEEE Trans. Neural Networks 10, 1048–1054.

Fukunaga, K., 1990. Statistical Pattern Recognition, second ed. Academic Press, San Diego.

Furey, T.S., Cristianini, N., Duffy, N., Bednarski, B.W., Schummer, M., Haussler, D., 2000. Support vector classification and validation of cancer tissue samples using microarray expression. Bioinformatics 16, 906–914.

Genov, R., Cauwenberghs, G., 2003. Kerneltron: Support vector "machine" in silicon. IEEE Trans. Neural Networks 14, 1426–1434.

Joachims, T., 1999. Making large-scale SVM learning practical. In: Schölkopf, B., Burges, C., Smola, A. (Eds.), Advances in Kernel Methods—Support Vector Learning. MIT Press http://svmlight.joachims.org/.

Kotropoulos, C., Pitas, I., 2003. Segmentation of ultrasonic images using support vector machines. Pattern Recognition Lett. 24, 715–727.

Lee, C.W., Jung, K., Kim, H.J., 2003. Automatic text detection and removal in video sequences. Pattern Recognition Lett. 24, 2607–2623.

Lerner, B., Lawrence, N.D., 2001. A comparison of state-of-the-art classification techniques with application to cytogenetics. Neural Comput. Appl. 10, 39–47.

Lerner, B., Clocksin, W.F., Dhanjal, S., Hult'en, M.A., Bishop, C.M., 2001. Feature representation and signal classification in fluorescence in situ hybridization image analysis. IEEE Trans. Systems Man Cybernet. A 31, 655–665.

Netten, H., van Vliet, L.J., Vrolijk, H., Sloos, W.C.R., Tanke, I.T., Young, I.T., 1996. Fluorescent dot counting in interphase cell nuclei. Bioimaging 4, 93–106.

Pang, S., Kim, D., Bang, S.Y., 2003. Membership authentication in the dynamic group by face classification using SVM ensemble. Pattern Recognition Lett. 24, 215–225.

Schwaighofer, A., 2002. SVM toolbox for Matlab. Available from: <http://www.cis.tugraz.at/igi/aschwaig/>.

Tanke, H.J., Florijn, R.J., Wiegant, J., Raap, A.K., Vrolijk, J., 1995. CCD microscopy and image analysis of cells and chromosomes stained by fluorescence in situ hybridization. Histochem. J. 27, 4–14.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.