Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT

Signal Processing 83 (2003) 729–743

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Simultaneous classification and relevant feature identification in high-dimensional spaces: application to molecular profiling data

C. Bhattacharyya[a,e], L.R. Grate[b], A. Rizki[b], D. Radisky[b], F.J. Molina[b,c], M.I. Jordan[a,d], M.J. Bissell[b], I.S. Mian[b,*]

[a] *Division of Computer Science, University of California Berkeley, Berkeley, CA 94720, USA*
[b] *Lawrence Berkeley National Laboratory, Life Sciences Division, Berkeley, CA 94720, USA*
[c] *Department of Mathematics, University of California Santa Cruz, Santa Cruz, CA 95064, USA*
[d] *Department of Statistics, University of California Berkeley, Berkeley, CA 94720, USA*
[e] *Current address: Department of CSA, Indian Institute of Science, Bangalore 560012, India*

Received 26 May 2002; received in revised form 6 September 2002

Abstract

Molecular profiling technologies monitor many thousands of transcripts, proteins, metabolites or other species concurrently in a biological sample of interest. Given such high-dimensional data for different types of samples, classification methods aim to assign specimens to known categories. Relevant feature identification methods seek to define a subset of molecules that differentiate the samples. This work describes LIKNON, a specific implementation of a statistical approach for creating a classifier and identifying a small number of relevant features simultaneously. Given two-class data, LIKNON estimates a sparse linear classifier by exploiting the simple and well-known property that minimising an L_1 norm (via linear programming) yields a sparse hyperplane. It performs well when used for retrospective analysis of three cancer biology profiling data sets, (i) small, round, blue cell tumour transcript profiles from tumour biopsies and cell lines, (ii) sporadic breast carcinoma transcript profiles from patients with distant metastases <5 years and those with no distant metastases $\geqslant 5$ years and (iii) serum sample protein profiles from unaffected and ovarian cancer patients. Computationally, LIKNON is less demanding than the prevailing filter-wrapper strategy; this approach generates many feature subsets and equates relevant features with the subset yielding a classifier with the lowest generalisation error. Biologically, the results suggest a role for the cellular microenvironment in influencing disease outcome and its importance in developing clinical decision support systems.

Keywords: L_1 norm minimisation; Molecular profiling data; Feature selection; Classification; Cancer biology; LIKNON; Minimax probability machine

1. Introduction

Molecular profiling studies of different types of biological specimens are both increasingly widespread and important. In cancer biology for example, commonplace investigations include monitoring the abundances of transcripts and/or proteins in normal and aberrant (tumour) tissues, sera or cell lines [19,6,18,39,25,34,15,36,40,26,29,35]. The adoption of profiling technologies is motivated largely by a desire to create clinical decision support systems for accurate

* Corresponding author.
E-mail address: smian@lbl.gov (I.S. Mian).

0165-1684/03/$ - see front matter
doi:10.1016/S0165-1684(02)00474-7

cancer classification and a need to identify robust and reliable molecular targets ("biomarkers") for intervention, diagnosis and imaging. The first attendant analytical task is classification and prediction, estimating a classifier from profiling data which assigns accurately samples to known classes. The second task, relevant feature identification, involves defining a small subset of the molecules monitored which best differentiate classes.

The subject of this work is the tasks of classification and relevant feature identification in the context of two-class molecular profiling data, i.e. samples are assigned to one of two categories such as normal or tumour specimens. Statistical challenges associated with solving these problems include the large number of features in an example vector ($\sim 10^3$–10^4 molecular abundances) and the small number of high-dimensional example vectors ($\sim 10^1$–10^2 samples). The classifier underlying a clinical decision support system would be expected to make precise diagnoses for many more and diverse patient samples than had been used for its estimation. This requirement for systems with good predictive capability necessitates classifiers which minimise misclassifications on future data, namely those with low generalisation error.

For two-class data, the classification and prediction problem is to learn a discriminating surface which separates the classes using a criterion such as generalisation error. Support vector machines (SVMs) [14] are good classifiers which achieve low generalisation error by minimising an associated quantity termed the margin. SVMs have been employed successfully for cancer classification using transcript profiles [10,31,36,40]. In contrast to SVMs, the newly formulated minimax probability machine (MPM) minimises directly an upper bound on the generalisation error [28]. As shown here, MPMs provide a viable alternative to SVMs for addressing classification and prediction problems related to profiling data.

MPMs and SVMs cannot define biomarkers in their own right because each feature in an example vector contributes to delineating the discriminating surface. In transcript profiling studies, relevant feature identification has oft been addressed via a filter-wrapper strategy [17,31,42]. The filter generates candidate gene subsets whilst the wrapper runs an induction algorithm to determine the discriminative ability of a subset. This procedure computes a statistic from the empirical distribution of genes in the two classes and orders genes according to this metric. Forward or backward selection creates subsets by adding or deleting genes successively. Each subset is used to estimate a classifier and to determine its generalisation error. A priori, the number of genes and which subset will produce a classifier with the lowest generalisation error is unknown. Thus, many runs are required to converge upon a subset that constitutes biomarkers. Although MPMs and SVMs are good wrappers, the choice of filtering statistic remains an open question.

This study shows the potential of sparse (linear) classifiers as a framework for addressing simultaneously the aforementioned problems of classification and relevant feature identification. In so doing, considerable prior statistical research is exploited in a new application domain. Here, the focus is sparse hyperplanes estimated by minimising an L_1 norm via linear programming [4,14,16,20,24,38,3]. LIKNON, [1] a specific implementation of this strategy, is used for retrospective analysis of data from three exemplars of transcript [26,41] and protein [35] profiling studies. LIKNON has non-trivial computational advantages over the prevailing filter-wrapper strategy because it creates a classifier and identifies relevant features in one pass through two-class data. Reexamination of the transcript profiles generates biological predictions for subsequent experimental and clinical investigation of two types of cancer and cellular microenvironments. Finally, the results reveal the ability of published data to answer unanticipated questions.

2. Materials and methods

2.1. Transcript profiling data: small, round, blue cell tumours

Previously [26], cDNA microarrays were used to monitor tumour biopsy and cell line samples from

[1] LIKNON is a word for a winnowing basket used in ancient Greece.

four distinct classes of small, round, blue cell tumours (SRBCTs) of childhood: neuroblastoma (NB), rhabdomyosarcoma (RMS), the Ewing family of tumors (EWS) and non-Hodgkin lymphoma (NHL) [26]. The transcript profiles consisted of 2308 nucleic acid sequences or "genes" monitored in 84 samples. These data were used to categorise samples on the basis of their cancer class (EWS, RMS, NHL or NB) and to define 96 genes which distinguished the four classes. Each class consisted of a mixture of tumour biopsy and cell lines samples, i.e. the origin of a specimen was ignored during categorisation.

Here, transcript profiles for the 84 SRBCT samples were downloaded from http://www.nhgri.nih.gov/DIR/Microarray/Supplement/. For each sample, the features in the 2308-dimensional example vectors were the log ratios of transcripts in the sample of interest compared to a common reference [26]. To determine whether profiling data have the potential to answer unanticipated questions, the SRBCT samples were partitioned so as to probe the interplay between tissue and cell culture cellular microenvironments, and cancer class. The seven, new, two-class data sets formulated by repartitioning the samples were as follows: Partition A, 46 EWS/RMS tumour biopsies and 38 EWS/RMS/NHL/NB cell lines; Partition B, 21 EWS/RMS cell lines and 30 EWS/RMS tumour biopsies; Partition C, 28 EWS tumour biopsies/cell lines and 23 RMS tumour biopsies/cell lines; Partition D, 17 EWS tumour biopsies and 13 RMS tumour biopsies; Partition E, 11 EWS cell lines and 10 RMS cell lines; Partition F, 17 EWS tumour biopsies and 11 EWS cell lines; and Partition G, 13 RMS tumour biopsies and 10 RMS cell lines (for NHL and NB, only cell lines were available). The seven two-class data sets were analysed using LIKNON and a Fisher score filter-MPM/SVM wrapper strategy.

2.2. Transcript profiling data: sporadic breast carcinomas

Previously [41], cDNA microarrays were used to monitor 5192 genes in 97 sporadic breast carcinoma samples. These data were used to define 70 genes which discriminated between patients with distant metastases <5 yr and those with no distant metastases $\geqslant 5$ yr.

Here, transcript profiles for the 97 sporadic breast carcinoma samples were downloaded from http://www.rii.com/publications/vantveer.htm. For each sample, the features in the 5192-dimensional example vectors were the log ratios of the transcripts in the sample of interest compared to a common reference [41]. The two-class data set, 46 patients with distant metastases <5 yr and 51 patients with no distant metastases $\geqslant 5$ yr, was analysed using LIKNON.

2.3. Protein profiling data: ovarian cancer

Previously [35], SELDI-TOF mass spectroscopy was used to generate spectra for serum samples from unaffected and ovarian cancer patients. The protein profiles consisted of 15,154 Mass/Charge (M/Z) values measured in 200 samples. These data were used to define 5 "proteins" which differentiated non-maligant from ovarian cancer samples.

Here, protein profiles for the 200 serum samples were downloaded from http://clinicalproteomics.steem.com/. For each sample, the features in the 15,154-dimensional example vectors were SELDI-TOF mass spectrum amplitudes representing $15,154$ M/Z values in the sample of interest [35]. Each M/Z value represents a low molecular weight molecule. The two-class data set, 100 unaffected and 100 ovarian cancer serum samples, was analysed using LIKNON.

2.4. LIKNON: simultaneous classification and relevant feature identification

Consider two-class data, $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, consisting of N example vectors, $\mathbf{x}_i \in \mathbb{R}^P$. The label, $y_i \in \{+1, -1\}$, indicates whether the example vector $\mathbf{x}_i$ is equated with class 1 or with class 2. For the two-class profiling data described above, the number of example vectors, N, and their dimensionality, P, are (i) small round blue cell tumours, $N = 84, 51, 51, 30, 21, 28$ and 23, and $P = 2308$, (ii) sporadic breast carcinomas, $N = 97$ and $P = 5192$ and (iii) ovarian cancer, $N = 200$ and $P = 15{,}154$. Each feature x_p in a P-dimensional example vector corresponds to an observed transcript level or M/Z value.

If two-class data can be separated by a linear decision boundary, the discriminating surface has the form

of a hyperplane, $\mathbf{w}^{\mathrm{T}}\mathbf{x} = b$, parameterised in terms of a weight vector, $\mathbf{w} \in \mathbb{R}^P$, and offset term, $b \in \mathbb{R}$. A classifier is the hyperplane which satisfies the N inequalities $y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - b) \geqslant 0\ \forall i = \{1,\ldots,N\}$. The learning problem is to estimate the optimal weight vector $\mathbf{w}_*$ and offset b_*. Given this hyperplane, a vector $\mathbf{x}$ is assigned to a class based on the sign of the corresponding decision function. If $\mathrm{sign}(\mathbf{w}_*^{\mathrm{T}}\mathbf{x} - b_*) = +1$, $\mathbf{x}$ is identified with class 1, otherwise it is assigned to class 2.

The problems of classification and relevant feature identification can be solved concurrently by considering a sparse hyperplane, one for which the weight vector $\mathbf{w}$ has few non-zero elements. Recall that the class of a vector $\mathbf{x}$ is assigned according to $\mathrm{sign}(z)$.

$$z = \mathbf{w}^{\mathrm{T}}\mathbf{x} - b = \sum_{p=1}^{P} w_p x_p - b = \sum_{w_p \neq 0} w_p x_p - b.$$

If a weight vector element is zero, $w_p = 0$, then feature p in the example vector does not decide the class of $\mathbf{x}$ and is thus "irrelevant". Only a feature for which the element is non-zero, $w_p \neq 0$, contributes to $\mathrm{sign}(z)$ and is thus useful for discrimination. Thus, the problem of defining a small number of relevant features (biomarkers) can be thought of as synonymous with identifying a sparse hyperplane.

Learning a sparse hyperplane can be formulated as an optimisation problem. Minimising the L_0 norm of the weight vector, $\|\mathbf{w}\|_0$, minimises the number of non-zero elements. The L_0 norm is $\|\mathbf{w}\|_0 =$ number of $\{p : w_p \neq 0\}$. Unfortunately, minimising an L_0 norm is NP-hard. However, a tractable, convex approximation is to replace the L_0 norm with the L_1 norm [16]. Minimising the L_1 norm of the weight vector, $\|\mathbf{w}\|_1$, minimises the sum of the absolute magnitudes of the elements and sets most of the elements to zero. The L_1 norm is $\|\mathbf{w}\|_1 = \sum_{p=1}^{P} |w_p|$. The optimisation problem becomes

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \|\mathbf{w}\|_1 \\ \text{s. t.} \quad & y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - b) \geqslant 1 \\ & \forall i \in \{1,\ldots,N\}, \end{aligned} \tag{1}$$

$$\|\mathbf{w}\|_1 = \sum_{p=1}^{P} |w_p|, \quad |w_p| = \mathrm{sign}(w_p) w_p.$$

Problem 1 can viewed as a special case of minimising a weighted L_1 norm, $\min_{\mathbf{w}} \sum_{p=1}^{P} a_p |w_p|$, in which the vector of weighting coefficients $\mathbf{a}$ is a unit vector, $a_p = 1; \forall p \in \{1,\ldots,P\}$. In other words, all genes are presumed to be equally good relevant feature candidates. Prior knowledge about the (un)importance of feature p can be encoded by specifying the value of a_p.

If the data are not linearly separable, misclassification can be accounted for by adding a non-negative slack variable ξ_i to each constraint and introducing a weighted penalty term to the objective function

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \|\mathbf{w}\|_1 + C \sum_{i=1}^{N} \xi_i \\ \text{s. t.} \quad & y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - b) \geqslant 1 - \xi_i, \\ & \xi_i \geqslant 0 \quad \forall i \in \{1,\ldots,N\}. \end{aligned} \tag{2}$$

The term $\sum_{i=1}^{N} \xi_i$ is an upper bound on the number of misclassifications. The parameter C represents a tradeoff between misclassification and sparseness. The higher the value of C, the less sparse the solution. Here, setting $C = 1$ classified correctly all the points in the data sets encountered. However, the value of C can be chosen more systematically via cross validation.

Problem (2) can be recast as a linear programming problem by introducing extra variables u_p and v_p where $w_p = u_p - v_p$ and $|w_p| = u_p + v_p$. These variables are the pth elements of $\mathbf{u}, \mathbf{v} \in \mathbb{R}^P$. The L_1 norm becomes $\|\mathbf{w}\|_1 = \sum_{p=1}^{P}(u_p + v_p) = \mathbf{u} + \mathbf{v}$ and the problem can rewritten in a standard form as follows:

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v},b,\xi} \quad & (\mathbf{u} + \mathbf{v}) + C \sum_{i=1}^{N} \xi_i \\ \text{s. t.} \quad & y_i((\mathbf{u} - \mathbf{v})^{\mathrm{T}}\mathbf{x}_i - b) \geqslant 1 - \xi_i, \\ & \xi_i \geqslant 0 \quad \forall i \in \{1,\ldots,N\} \\ & u_p \geqslant 0, \ \ v_p \geqslant 0 \quad \forall p \in \{1,\ldots,P\}. \end{aligned} \tag{3}$$

Problem (3) minimises a linear function subject to linear constraints. This type of linear programming problem has been well studied in optimisation theory. There are efficient algorithms for solving problems involving $N \sim 10^4$ constraints and $(2 * P + 1) \sim 10^4$ variables. The code for LIKNON

is available at http://www.cs.berkeley.edu/~jordan/liknon/.

2.5. Fisher score filter-MPM/SVM wrapper: independent classification and relevant feature identification

Given linearly separable two-class data, the task is to determine a hyperplane $\mathbf{w}^{\mathrm{T}}\mathbf{z} = b$ which separates example vectors belonging to class 1 ($\mathbf{x}$) and class 2 ($\mathbf{y}$). Both MPMs and SVMs attempt to minimise the generalisation error, i.e. misclassification on future data. The MPM framework seeks the hyperplane for which the misclassification probabilities for class 1, $P(\mathbf{w}^{\mathrm{T}}\mathbf{x} \leqslant b)$, and class 2, $P(\mathbf{w}^{\mathrm{T}}\mathbf{y} \geqslant b)$, are low. The SVM framework seeks the unique discriminating hyperplane which maximises the margin separating the classes. MPMs and SVMs are comparable in complexity (detailed descriptions of these techniques are available in Appendix A). Preliminary results (data not shown) indicated that the two-class profiling data examined here were indeed linearly separable. Hence, the use of LIKNON and SVMs with linear kernels was justified.

MPMs and SVMs only address the problem of classification and prediction. In the filter-wrapper strategy, relevant feature identification is an independent, data preprocessing step. For simplicity and illustrative purposes, SVM/MPM wrappers were employed in conjunction with a Fisher score filter. Given example vectors assigned to class $\mathbf{x}$ or class $\mathbf{y}$, the Fisher score for feature p is given by $F_p = (\bar{\mathbf{x}}_p - \bar{\mathbf{y}}_p)^2/(\boldsymbol{\Sigma}_{\mathbf{x}_p} + \boldsymbol{\Sigma}_{\mathbf{y}_p})$; $\bar{\mathbf{x}}_p$ and $\bar{\mathbf{y}}_p$ are the means of feature p in the respective classes, whereas $\boldsymbol{\Sigma}_{\mathbf{x}_p}$ and $\boldsymbol{\Sigma}_{\mathbf{y}_p}$ are standard deviations. Higher values signify more discriminative features. Given P features ranked in descending order according to their score $F_1, \ldots, F_P$, the Fisher score top-r ranked features are those ranked $1, \ldots, r$. A particular value of r signifies a specific feature subset for use in estimating a classifier. Forward selection creates feature subsets by progressively increasing the value of r in a user-defined manner.

Although the recursive feature selection approach utilises a separating hyperplane $\mathbf{w}$ [23], it is closer to a filter-wrapper strategy than to LIKNON. Features are ordered based on $|w_p|$, the absolute magnitude of the elements of the weight vector (the range of values for each feature are assumed to be the same). Backward elimination creates feature subsets by recursively removing the bottom 10% of features. The feature subsets are used as input to a wrapper of choice.

2.6. Computational experiments: LIKNON and Fisher score filter-SVM/MPM wrapper

LIKNON creates a classifier and identifies relevant features in a single pass through two-class data. The Fisher score filter-MPM/SVM wrapper strategy has distinct feature subset generation and classification steps. A relevant subset is equated with the feature subset of smallest cardinality that yields a classifier with the lowest generalisation error. These simultaneous and independent classification and relevant feature identification strategies were compared by means of the leave-one-out error, a surrogate for generalisation error.

Given the choice of leave-one-out error as the performance metric, LIKNON needs to be run twice for a given two-class data set: first to identify relevant features (a small subset l of the P input features) and second as a classifier which uses the resultant l-dimensional vectors as input. Use of the conventional error, number of misclassifications on a test set, would require one pass through data. Results (data not shown), indicated that for SRBCT Problem A, all leave-one-out partitionings gave the same set of LIKNON relevant genes as when all N example vectors were used.

For each of the seven partitionings of the SRBCT samples, Fisher scores for the P features in the example vectors were computed. The Fisher score top-r ranked features were used to generate 13 gene subsets where $r = 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048$ and 2308. Thus, LIKNON and the Fisher score filter defined 14 feature subsets that differed only in their dimensionality and precise nature of the genes. The leave-one-out error of SVMs/MPMs trained using example vectors derived from every subset was ascertained.

For a two-class data set, the leave-one-out error was determined as follows. The N example vectors were divided into an estimation set consisting of $N-1$ example vectors and a test set composed of the remaining example. The MPM/SVM or LIKNON classifier was

used to predict the (known) class of the example in the test set. This estimation and evaluation procedure was repeated N times so that the class of each example was assigned by a classifier estimated using all other examples. The leave-one-out error is the number of misclassifications out of N.

3. Results

3.1. Transcript profiles: small round blue cell tumours

Two approaches for performing classification and relevant feature identification given two-class high-dimensional molecular profiling data were evaluated. LIKNON creates a classifier and ascertains a small number of relevant features simultaneously. The widely used filter-wrapper strategy estimates a classifier for every feature subset generated by an independent filtering step. The performances of LIKNON and a Fisher score filter-MPM/SVM wrapper were assessed by means of the leave-one-out error, a common proxy for generalisation error when there are few example vectors.

MPMs are a viable alternative to SVMs for solving the classification and prediction problem in a filter-wrapper strategy. Table 1 presents the leave-one-out error when the seven problems were analysed using these classifiers as the wrapper. Irrespective of the feature subset, MPMs and SVMs had similar performance and generalised equally well. Whereas SVMs and MPMs could operate directly in high-dimensional spaces, the original study used the 10 dominant Principal Component Analysis components of the 2308-dimensional example vectors as input to artificial neural networks (ANNs) [26]. Since MPMs and SVMS solve convex optimisation problems, they avoid the local minima problems which plague ANNs.

LIKNON is computationally less demanding than the filter-MPM/SVM wrapper strategy in identifying relevant features. Table 2 tabulates the relevant features giving zero leave-one-out error for the seven two-class SRBCT data sets. For a given data set, similar numbers of Fisher score and LIKNON relevant genes are required and these are generally one to two orders of magnitude smaller than the 2308 input features. However, whilst LIKNON required one pass through data, the filter-wrapper approach required many runs to pinpoint its subset. For Partition A, the 23 LIKNON relevant features gave zero out of 84 leave-one-out error, whereas the top-16 or top-32 Fisher score ranked genes gave low, but not zero out of 84 leave-one-out error (Table 1).

LIKNON relevant features should be regarded as a small, though not necessarily unique set of biomarkers. Fig. 1 shows a histogram of Fisher scores for all 2308 genes overlaid with the Fisher scores of relevant genes. LIKNON relevant genes are not necessarily associated with high Fisher scores yet they yield classifiers with zero leave-one-out error. Higher Fisher scores correspond to larger differences in the empirical distributions of transcript levels (more discriminative features) so classifiers trained with top-ranked genes might be expected to generalise well. The results reinforce the notion that many distinct relevant feature subsets can fit the data equally well (see for example [12]).

From a numerical perspective, the 84 SRBCT transcript profiles are sufficiently informative that biological questions not considered in the original study can be posed and answered (see also [32]).

3.2. Cellular microenvironment and SRBCT classification

A biological assessment of the LIKNON relevant features reveals that the tissue or cell culture origin of a sample affects the nature and number of relevant genes. Table 3 lists these genes for the seven two-class data sets. Four of these compared tumour biopsies with tumour-derived cell lines in the context of different numbers of SRBCT classes, four (Partition A: EWS, RMS, NHL, NB), two (Partition B: EWS, RMS) and one (Partition F: EWS; G: RMS). There were 23 relevant genes for Partition A, 21 for B, 12 for F and 13 for G. Tissue and cell culture microenvironments are manifestations of variations in cell shape and cell-extracellular matrix interactions. This difference is reflected in relevant genes such as actin α 2, SMA3 (smooth muscle actin 3), and collagen type III α. The relevant genes represent good targets for studying how tumour cells escape quiescence and evade cell cycle arrest in vivo and in vitro.

Table 1
Prediction of SRBCT transcript profiles using MPMs and SVMs.

Rank	A ($N=84$)		B ($N=51$)		C ($N=51$)		D ($N=30$)		E ($N=21$)		F ($N=28$)		G ($N=23$)	
	SVM	MPM	SVM	MPM	SVM	MPM	SVM	MPM	SVM	MPM	SVM	MPM	SVM	MPM
1	10	10	7	8	3	5	2	2	1	1	17	2	2	2
2	9	9	1	3	3	3	2	1	0	0	2	1	1	1
4	6	5	2	3	4	2	2	1	0	0	0	0	0	1
8	9	4	2	4	4	1	2	0	0	2	1	0	0	0
16	8	4	3	0	0	0	2	2	0	6	2	2	0	5
32	5	2	2	11	1	4	2	12	0	1	1	4	0	1
64	3	15	3	2	0	6	0	0	0	0	1	0	0	3
128	1	2	2	3	0	3	3	1	0	0	1	2	0	1
256	0	1	0	0	0	1	3	1	0	1	1	0	0	0
512	0	0	0	0	0	1	3	2	0	2	1	1	0	0
1024	0	0	0	0	0	0	3	1	0	1	2	2	0	0
2048	2	1	0	0	0	1	2	1	1	0	2	2	0	0
2308	3	1	2	1	4	1	1	1	0	1	2	4	0	0

The seven two-class data sets and numbers of example vectors in each class (total N) are Partition A, 46 EWS/RMS tumour biopsies and 38 EWS/RMS/NHL/NB cell lines; Partition B, 21 EWS/RMS cell lines and 30 EWS/RMS tumour biopsies; Partition C, 28 EWS tumour biopsies/cell lines and 23 RMS tumour biopsies/cell lines; Partition D, 17 EWS tumour biopsies and 13 RMS tumour biopsies; Partition E, 11 EWS cell lines and 10 RMS cell lines; Partition F, 17 EWS tumour biopsies and 11 EWS cell lines; and Partition G, 13 RMS tumour biopsies and 10 RMS cell lines. For each partition, the table gives the leave-one-out error out of N for an SVM or MPM estimated using the feature subset indicated. The first 12 feature subsets are the Fisher score top-r ranked genes where r takes on the value given. The final "subset" corresponds to all features in the original 2308-dimensional example vectors [26].

Table 2
Identification of relevant genes in SRBCT transcript profiling data using LIKNON and a Fisher score filter-MPM/SVM wrapper strategy

Name	Class 1 samples	Class 2 samples	N	Classifier		
				SVM	MPM	LIKNON
A	EWS/RMS tumour	EWS/RMS/NHL/NB cell line	84	256	512	23
B	EWS/RMS cell line	EWS/RMS tumour	51	256	16	21
C	EWS tumour/cell line	RMS tumour/cell line	51	16	8	8
D	EWS tumour	RMS tumour	30	64	64	8
E	EWS cell line	RMS cell line	21	2	2	2
F	EWS tumour	EWS cell line	28	4	4	12
G	RMS tumour	RMS cell line	23	4	8	13

For each of the seven two-class data sets, the total number of example vectors N is listed. "SVM" and "MPM" give the Fisher score feature subset of smallest cardinality that yielded a classifier with zero out of N leave-one-out error (taken from Table 1). "LIKNON" gives the cardinality of the relevant features identified; each feature subset yielded a LIKNON classifier with zero out of N leave-one-out error.

Transcriptional differences between tumour biopsies and cell lines confound attempts to define biomarkers for classifying SRBCTs. Three Partitions compared EWS and RMS in the context of tumour biopsies and cell lines (Partition C), tumour biopsies (Partition D) and cell lines (Partition E). The relevant genes for EWS tumour biopsies and RMS tumour biopsies (Partition D) may constitute clinically useful biomarkers for fine-grained cancer class diagnosis and/or imaging. Of the 96 EWS, RMS, NHL and NB cancer class markers identified originally [26], 11 are markers for the cellular microenvironment of the sample.

The results reiterate the view that information provided by interactions with neighbouring cells, the composition and organisation of the surrounding

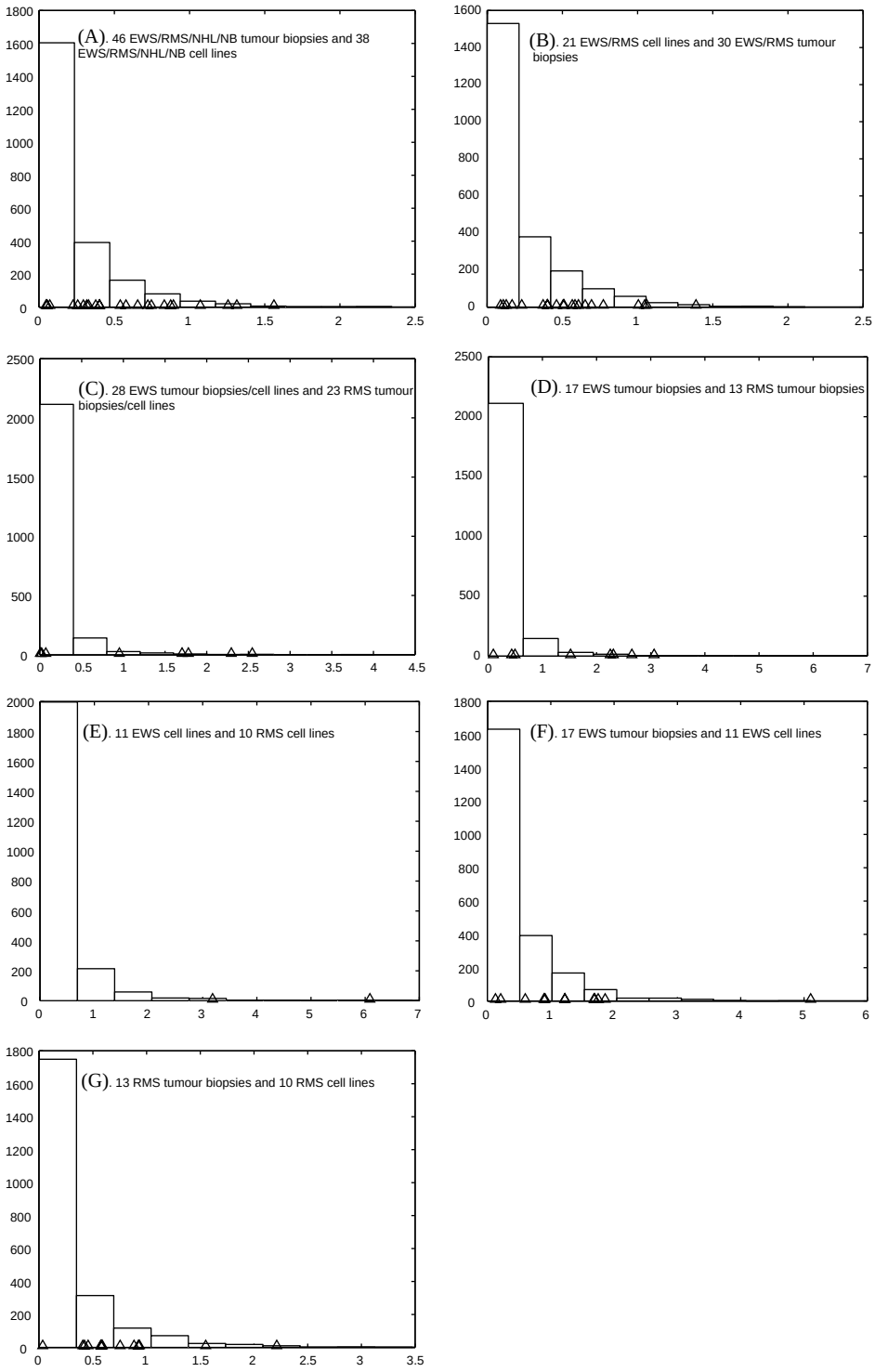

Fig. 1. Histograms of Fisher scores for all 2308 genes in the seven SRBCT binary problems. The abscissa represents the Fisher score and the ordinate the number of genes with that score. As might be expected, most genes have low scores and far fewer genes have high scores. Open triangles mark the Fisher scores of the LIKNON relevant genes. For each problem, the number of triangles is the same as the entry in the "LIKNON" column of Table 2.

Table 3
LIKNON relevant genes for the seven SRBCT two-class data sets

A	B	C	D	E	F	G	Image Id	Gene description
•							23019	Guanine nucleotide binding protein (G protein), alpha stimulating activity polypeptide 1
•							27549	Heterogeneous nuclear ribonucleoprotein A1
•							745343 (57)	Regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
•							809910 (44)	Interferon-inducible
•							1461138	H4 histone family, member G
•							159455	Similar to vaccinia virus HindIII K4L ORF
•							845363	Non-metastatic cells 1, protein (NM23A) expressed in
•							511091	Human protein immuno-reactive with anti-PTH polyclonal antibodies mRNA, partial cds
•							782811	High-mobility group (nonhistone chromosomal) protein isoforms I and Y
•							244652	SET translocation (myeloid leukemia-associated)
•		•	•				814260 (75)	Follicular lymphoma variant translocation 1
•					•		1493527	Asparagine synthetase
•	•						839736 (79)	Crystallin, alpha B
•	•						882522	Argininosuccinate synthetase
•	•					•	755474	Isoleucine-tRNA synthetase
•	•					•	882484	Chaperonin containing TCP1, subunit 7 (eta)
•	•					•	122159 (40)	Collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)
•	•				•	•	45542 (62)	Human insulin-like growth factor binding protein 5 (IGFBP5) mRNA
•	•				•		379708	
•	•				•		530814	Selenoprotein P, plasma, 1
•	•				•		470261	SMA3
•	•				•		868304 (83)	Actin, alpha 2, smooth muscle, aorta
•	•	•			•	•	51293	Aminoacylase 1
	•						1492147	Ribosomal protein S4, X-linked
	•						1492412	Ubiquitin A-52 residue ribosomal protein fusion product 1
	•						731308	Citrate synthase
	•						234376	Homo sapiens mRNA; cDNA DKFZp564F112 (from clone DKFZp564F112)
	•	•	•				878798	Beta-2-microglobulin
	•				•		757489	Tubulin, alpha 2
	•				•		43733 (9)	Glycogenin 2
	•					•	1492104	Tubulin, beta, 2
	•					•	22040	Matrix metalloproteinase 9 (gelatinase B, 92kD gelatinase, 92kD type IV collagenase)
	•	•	•	•		•	296448 (1)	Insulin-like growth factor 2 (somatomedin A)
		•					307660	Fatty acid binding protein 4, adipocyte
		•					377461 (18)	Caveolin 1, caveolae protein, 22kD
		•					1476065	Leukemia-associated phosphoprotein p18 (stathmin)
		•	•	•			207274 (2)	Human DNA for insulin-like growth factor II (IGF-2); exon 7 and additional ORF
			•				214990	Gelsolin (amyloidosis, Finnish type)
			•				52076 (19)	Olfactomedinrelated ER localized protein
			•				51448	Activating transcription factor 3
			•				842784	Phosphate carrier, mitochondrial
					•		878833	Ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)
					•		743230	Human silencing mediator of retinoid and thyroid hormone action (SMRT) mRNA
					•		309864	Jun B proto-oncogene
						•	128302	Parathymosin
						•	298062 (25)	Troponin T2, cardiac
						•	823851	AE-binding protein 1
						•	785847	Ubiquitin-conjugating enzyme E2M (homologous to yeast UBC12)
						•	244618 (7)	ESTs

Partition A, EWS/RMS tumour biopsies and EWS/RMS/NHL/NB cell lines; Partition B, EWS/RMS tumour biopsies and EWS/RMS cell lines; Partition C, EWS tumour biopsies/cell lines and RMS tumour biopsies/cell lines; Partition D, EWS tumour biopsies and RMS tumour biopsies; Partition E, EWS cell lines and RMS cell lines; Partition F, EWS tumour biopsies and EWS cell lines; and Partition G, RMS tumour biopsies and RMS cell lines. For 11 clones, the number in parenthesis denotes its rank in the 96 genes defined by the original study as biomarkers for the four SRBCT cancer classes irrespective the origin of the sample (ranks are taken from Table 3 of Supplementary Methods [26]).

extracellular matrix, and signals from soluble factors lead to cells in tissue and cell culture microenvironments operating in distinctly different contexts [7]. Whereas most tissues are normally in a state of low proliferation, cell culture systems have been designed to study and to propagate cells in a more defined and simplified environment. Thus, identifying clinically relevant biomarkers for cancer classification requires that the material assayed capture critical determinants of in situ cellular phenotypes.

3.3. Transcript profiles: sporadic breast carcinomas

LIKNON performed well when applied to sporadic breast carcinoma transcript profiles and a two-class data set for patients with distant metastases <5 yr and those with no distant metastases $\geqslant 5$ yr [41]. The leave-one-out error for the 97 5192-dimensional example vectors was 1 out of 97.

The cellular microenvironment appears to be a factor in determining disease outcome. Table 4 lists the 72 LIKNON relevant genes. These prognostic markers include genes involved in cell structure (troponin T1, keratin 13, keratin 15, keratin 19, actin $\gamma 2$), cell–cell communication (cadherin 7, cadherin 18), and cell signalling (cysteine-rich angiogenic inducer 61, tissue factor pathway inhibitor 2, small inducible cytokine subfamily B).

The relevant genes include known biologically and clinically useful biomarkers. For example, the neuropeptide Y receptor Y1 has been found with high incidence in situ invasive and metastatic breast cancer [37]. Screening of axillary lymph nodes for mammaglobin expression increased the detection of breast cancer metastases compared with routine histology [9]. Breast carcinoma amplified sequence 1 is highly expressed in three amplified breast cancer cell lines and in one breast tumor [13]. Keratin 19 is a characteristic of a human breast epithelial cell line with stem cell properties [22]. The open reading frame on human chromosome 12 (HsC12orf3) is a LIKNON relevant feature [21] for a 1987-dimensional transcript profiling data set consisting of 13 gastrointestinal stromal tumours and 6 spindle cell tumours from locations outside the gastrointestinal tract [1].

The original study [41] defined 70 genes as prognostic markers. Three of these, two unannotated genes (AL080059, Contig 48328RC) and CEGP1, are LIKNON relevant genes. A PSI-BLAST [2] search using the CEGP1 protein sequence revealed significant similarity to matrilin-2, a member of a filament forming family of proteins distributed in extracellular matrices [30]. Thus, the CEGP1 gene may encode a new matrilin with a role in breast cancer.

Seven genes designated as prognostic markers by two independent studies may be noteworthy candidates for subsequent experimental and clinical study. The 70 original prognostic markers were amongst 231 genes identified as significantly correlated with disease outcome [41]. Four of the 72 LIKNON relevant genes in addition to the three mentioned above were found in this set of 231 genes. These genes were phosphatidylinositol $(4,5)$ bisphosphate 5-phosphatase A, paired basic amino acid cleaving system 4, preferentially expressed antigen in melanoma, and ESTs (Contig 48328, Contig 55725).

3.4. Protein profiles: ovarian cancer

The performance of LIKNON on protein profiles is comparable to its performance on transcript profiling data. The two-class data set encompassed 100 unaffected and 100 ovarian cancer serum samples and features corresponded to M/Z values rather than transcript abundances [35]. The leave-one-out error for the 200 15,154-dimensional example vectors was 3 out of 200. The number of features in the example vectors is considerably greater than the 2308 and 5192 transcript levels in the SRBCT and breast carcinoma transcript profiles.

The origin and precise nature of the proteins or peptides corresponding to the 51 LIKNON relevant features awaits future experimental determination. Thus, it is not possible to comment on their biological significance in ovarian cancer.

4. Discussion

The success of LIKNON in solving classification and relevant feature identification problems associated with transcript and protein profiles augurs well for its

Table 4
LIKNON relevant genes for a two-class data set involving sporadic breast carcinomas and 46 patients with distant metastases <5 yr and 51 patients with no distant metastases $\geqslant 5$ yr

Id	Name	Description
U45975	PIB5PA	Phosphatidylinositol (4,5) bisphosphate 5-phosphatase, A
NM_001611	ACP5	Acid phosphatase 5, tartrate resistant
NM_001635	AMPH	Amphiphysin (Stiff-Mann syndrome with breast cancer 128kD autoantigen)
NM_000909	NPY1R	Neuropeptide Y receptor Y1
NM_001647	APOD	Apolipoprotein D
NM_001444	FABP5	Fatty acid binding protein 5 (psoriasis-associated)
NM_001657	AREG	Amphiregulin (schwannoma-derived growth factor)
NM_002411	MGB1	Mammaglobin 1
NM_002509	NKX2B	NK-2 (Drosophila) homolog B
NM_002570	PACE4	Paired basic amino acid cleaving system 4
NM_002652	PIP	Prolactin-induced protein
NM_002809	PSMD3	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 3
NM_002820	PTHLH	Parathyroid hormone-like hormone
NM_004291	CART	Cocaine- and amphetamine-regulated transcript
NM_012342	NMA	Putative transmembrane protein
NM_006186	NR4A2	Nuclear receptor subfamily 4, group A, member 2
NM_003657	BCAS1	Breast carcinoma amplified sequence 1
U56725	HSPA2	Heat shock 70kD protein 2
NM_005181	CA3	Carbonic anhydrase III, muscle specific
U17077	BENE	BENE protein
NM_006103	HE4	Epididymis-specific, whey-acidic protein type, four-disulfide core; putative ovarian carcinoma marker
NM_006115	PRAME	Preferentially expressed antigen in melanoma
NM_002362	MAGEA4	Melanoma antigen, family A, 4
NM_020974	CEGP1 (*)	Homo sapiens CEGP1 protein (CEGP1), mRNA
NM_006398	UBD	Diubiquitin
NM_004988	MAGEA1	Melanoma antigen, family A, 1 (directs expression of antigen MZ2-E)
NC_001807	ND1	Human mitochondrion, complete genome
NM_007359	MLN51	MLN51 protein
NM_004950	DSPG3	Dermatan sulphate proteoglycan 3
NM_000169	GLA	Galactosidase, alpha
NM_000239	LYZ	Lysozyme (renal amyloidosis)
NM_001267	CHAD	Chondroadherin
NM_001062	TCN1	Transcobalamin I (vitamin B12 binding protein, R binder family)
NM_000518	HBB	Hemoglobin, beta
NM_001321	CSRP2	Cysteine and glycine-rich protein 2
NM_000668	ADH1B	Alcohol dehydrogenase 2 (class I), beta polypeptide
NM_005794	HEP27	Short-chain alcohol dehydrogenase family member
NM_001554	CYR61	Cysteine-rich, angiogenic inducer, 61
NM_006528	TFPI2	Tissue factor pathway inhibitor 2
NM_001756	SERPINA6	Serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6
Contig53506	TMPRSS2	Transmembrane protease, serine 2
NM_006419	SCYB13	Small inducible cytokine B subfamily (Cys-X-Cys motif), member 13 (B-cell chemoattractant)
NM_004887	SCYB14	Small inducible cytokine subfamily B (Cys-X-Cys), member 14 (BRAK)
NM_001555	IGSF1	Immunoglobulin superfamily, member 1
M63438	IGKC	Immunoglobulin kappa constant
V00522	HLA-DRB3	Major histocompatibility complex, class II, DR beta 3
NM_002122	HLA-DQA1	Major histocompatibility complex, class II, DQ alpha 1
NM_004361	CDH7	Homo sapiens cadherin 7, type 2 (CDH7), mRNA
NM_004934	CDH18	Cadherin 18, type 2
NM_003283	TNNT1	Troponin T1, skeletal, slow
NM_002274	KRT13	Keratin 13
NM_002275	KRT15	Keratin 15

Table 4 (*continued*)

Id	Name	Description
NM_002276	KRT19	Keratin 19
NM_001615	ACTG2	Actin, gamma 2, smooth muscle, enteric
AL080059	(*)	Homo sapiens mRNA; cDNA DKFZp564H142 (from clone DKFZp564H142)
NM_014665	KIAA0014	KIAA0014 gene product
AK000451		Homo sapiens cDNA FLJ20444 fis, clone KAT05128
NM_020373	C12orf3	Chromosome 12 open reading frame 3
AB040886	KIAA1453	KIAA1453 protein
NM_017852	FLJ20510	Hypothetical protein FLJ20510
Contig7755_RC	MGC5395	ESTs
AI497657_RC	GNG4	ESTs
Contig48328_RC	(*)	ESTs
Contig55725_RC, ESTs; Contig50122_RC, ESTs; Contig29015_RC, ESTs; Contig44909_RC, ESTs; Contig38438_RC, ESTs Contig37946_RC, ESTs; Contig36499_RC, ESTs; Contig45511_RC, ESTs; Contig39285_RC, ESTs		

The leave-one-out error was 1 out of 97. The three genes present in the 70 prognostic markers defined by the original study [41] are denoted (*).

utility in analysing other types of high-dimensional molecular profiling data. Biologically, the results reveal a role for the cellular microenvironment in breast cancer prognosis and its importance in developing clinical decision support systems for cancer classification. The small number of relevant features defined here present tractable targets (putative biomarkers) for investigations of basic mechanisms, validation via high-density tissue microarrays [27], and eventual deployment in the clinic.

LIKNON is based on the L_1 norm optimal hyperplane so it is applicable only for linear decision boundaries. The L_2 norm optimal hyperplane, or SVM, is more general in that it can handle non-linear functions specified via a positive definite but otherwise arbitrary kernel function. Current evidence suggests that this restriction may not be limiting because linear separability is a facet of the two-class data sets examined here and elsewhere [21] (unpublished).

The statistical task addressed by LIKNON can be viewed as forward classification: given samples assigned to classes, estimate good classifiers. All seven two-class partitionings of the SRBCT samples yielded classifiers with zero leave-one-out error. The inverse classification problem can be thought of as identifying other, equally predictive partitionings of the data. These new classifications and attendant relevant features would require the formulation of novel biological hypotheses aimed at explaining the common aspects of samples identified with each class.

Acknowledgements

This work was supported by the National Science Foundation, National Institute on Aging, National Institute of Environmental Health Sciences, Department of Energy and California Breast Cancer Research Program.

Appendix A. Supplementary methods

The minimax probability machine (MPM) [28] is a newly formulated technique for handling classification and prediction problems. As discussed below, MPMs minimise directly an upper bound on the generalisation error whereas Support Vector Machines (SVMs) focus on the associated margin. MPMs and SVMs are comparable in complexity and possess the same advantage over ANNs. Both are less prone to overfitting and by solving convex optimisation problems, they avoid the local minima which plague ANNs. They can operate in high-dimensional spaces in contrast to ANNs where, for example, the dimensionality of the SRBCT 2308 feature input was reduced by

considering only the 10 dominant Principal Component Analysis components [26].

A.1. Minimax probability machine

Let $\mathbf{x}$ and $\mathbf{y}$ denote random vectors in a two-class classification and prediction problem with means and covariance matrices given by $\mathbf{x}\sim(\bar{\mathbf{x}},\mathbf{\Sigma_x})$ and $\mathbf{y}\sim(\bar{\mathbf{y}},\mathbf{\Sigma_y})$. "$\sim$" signifies that the random variable has the specified mean and covariance matrix but that the distribution is otherwise unconstrained $(\mathbf{x},\bar{\mathbf{x}},\mathbf{y},\bar{\mathbf{y}}\in\mathbb{R}^P;\mathbf{\Sigma_x},\mathbf{\Sigma_y}\in\mathbb{R}^{P\times P})$. The hyperplane $\mathbf{w}^{\mathrm{T}}\mathbf{z}=b$ separates the two classes with maximal probability with respect to all distributions having the specified means and covariance matrices. The minimax framework minimises the generalisation error by seeking the hyperplane for which the misclassification probabilities, $\Pr(\mathbf{w}^{\mathrm{T}}\mathbf{x}\leqslant b)$ and $\Pr(\mathbf{w}^{\mathrm{T}}\mathbf{y}\geqslant b)$, are low. The optimisation problem becomes

$$\begin{aligned}\min_{\alpha,\mathbf{w},b}\quad & \alpha\\ \text{s. t.}\quad & \alpha\geqslant \sup_{\mathbf{x}\sim(\bar{\mathbf{x}},\mathbf{\Sigma_x})}\Pr\{\mathbf{w}^{\mathrm{T}}\mathbf{x}\leqslant b\},\\ & \alpha\geqslant \sup_{\mathbf{y}\sim(\bar{\mathbf{y}},\mathbf{\Sigma_y})}\Pr\{\mathbf{w}^{\mathrm{T}}\mathbf{y}\geqslant b\}.\end{aligned}\tag{A.1}$$

The quantity α can be interpreted as an upper bound on the generalisation error. The supremum in both constraints are computed via a theorem stated in Bertsimas and Sethuraman [5]:

$$\sup_{\mathbf{y}\sim(\bar{\mathbf{y}},\mathbf{\Sigma_y})}\Pr\{\mathbf{w}^{\mathrm{T}}\mathbf{y}\geqslant b\}=\frac{1}{1+d^2}\quad\text{with}$$

$$d^2=\inf_{\mathbf{w}^{\mathrm{T}}\mathbf{y}\geqslant b}(\mathbf{y}-\bar{\mathbf{y}})^{\mathrm{T}}\mathbf{\Sigma_y}^{-1}(\mathbf{y}-\bar{\mathbf{y}}).\tag{A.2}$$

Problem (A.1) can be recast as

$$\begin{aligned}&\min_{\mathbf{w}}\sqrt{\mathbf{w}^{\mathrm{T}}\mathbf{\Sigma_x}\mathbf{w}}+\sqrt{\mathbf{w}^{\mathrm{T}}\mathbf{\Sigma_y}\mathbf{w}}\\ &\text{s.t. }\mathbf{w}^{\mathrm{T}}(\bar{\mathbf{x}}-\bar{\mathbf{y}})=1.\end{aligned}\tag{A.3}$$

The optimal b_* is obtained from

$$b_*=\mathbf{w}_*^{\mathrm{T}}\bar{\mathbf{x}}-\kappa_*\sqrt{\mathbf{w}_*^{\mathrm{T}}\mathbf{\Sigma_x}\mathbf{w}_*}=\mathbf{w}_*^{\mathrm{T}}\bar{\mathbf{y}}+\kappa_*\sqrt{\mathbf{w}_*^{\mathrm{T}}\mathbf{\Sigma_y}\mathbf{w}_*},$$

$$\kappa_*=1/(\sqrt{\mathbf{w}_*^{\mathrm{T}}\mathbf{\Sigma_x}\mathbf{w}_*}+\sqrt{\mathbf{w}_*^{\mathrm{T}}\mathbf{\Sigma_y}\mathbf{w}_*}),\tag{A.4}$$

where $\mathbf{w}_*$ is the optimal $\mathbf{w}$ in (A.3). The optimal $\alpha_*=1/(1+\kappa_*^2)$. Problem (A.3) is a second-order cone programme (SOCP). Efficient algorithms for solving this type of convex optimisation problem are available [8,33]. MPMs were implemented using an iterative scheme [2].

A.2. Support vector machines

The SVM framework also seeks a hyperplane but its geometric underpinning results in the introduction of a quantity termed a margin [14]. Amongst all separating hyperplanes, there is a unique hyperplane which yields the maximum margin of separation between the two classes. SVMs minimise the generalisation error by finding this optimal hyperplane, one which maximises this attendant margin. The final optimisation problem is

$$\max_{\beta}L(\beta)=-\frac{1}{2}\sum_{ij}\beta_i\beta_jK(\mathbf{z}_i,\mathbf{z}_j)+\sum_i\beta_i$$

$$\text{s. t.}0\leqslant\beta_i\leqslant C,\tag{A.5}$$

$$\sum_{i\in\text{ Class1}}\beta_i=\sum_{j\in\text{Class2}}\beta_j,\tag{A.6}$$

where β_i and β_j are dual variables. C is a user-defined penalty determining the number of permissible misclassifications; higher values signify that fewer outliers are ignored ($C\to\infty$ corresponds to the hard margin case). Here, C was fixed at 100. Preliminary experiments indicated that the two-class data sets were linearly separable so only linear kernels, $K(\mathbf{z}_i,\mathbf{z}_j)=\mathbf{z}_i^{\mathrm{T}}\mathbf{z}_j$, were considered. Thus, the optimal values for $\mathbf{w}_*$ and b_* are $\mathbf{w}_*=\sum_{i\in\text{Class1}}\beta_i\mathbf{z}_i-\sum_{j\in\text{Class2}}\beta_j\mathbf{z}_j$ and b_* is computed from the KKT conditions (see [11]).

References

[1] S. Allander, N. Nupponen, M. Ringner, G. Hostetter, G. Maher, N. Goldberger, Y. Chen, C.J., A. Elkahloun, P. Meltzer, Gastrointestinal Stromal Tumors with KIT mutations exhibit a remarkably homogeneous gene expression profile, Cancer Res. 61 (2001) 8624–8628.

[2] S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.

[3] K. Bennett, C. Campbell, Support Vector Machines: Hype or Hallelujah?, SIGKDD Explorations 2 (2000) 1–13.
[4] K. Bennett, A. Demiriz, Semi-supervised support vector machines, in: Neural and Information Processing Systems, Vol. 11, MIT Press, Cambridge, MA, 1999, pp. 368–374.
[5] D. Bertsimas, J. Sethuraman, Moment problems and semidefinite optimization, in: Handbook of Semidefinite Optimization, Kluwer Academic Publishers, Dodrecht, 2000, pp. 469–509.
[6] A. Bhattacharjee, W. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. Mark, E. Lander, W. Wong, B. Johnson, T. Golub, D. Sugarbaker, M. Meyerson, Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses, Proc. Nat. Acad. Sci. 98 (2001) 13790–13795.
[7] M. Bissell, D. Radisky, Putting tumours in context, Nat. Rev. Cancer 1 (2001) 46–54.
[8] S. Boyd, L. Vandenberghe, Convex optimization, course notes for EE364, Stanford University, The notes are available at http://www.stanford.edu/class/ee364/ (2001).
[9] G. Branagan, D. Hughes, M. Jeffrey, C. Crane-Robinson, P. Perry, Detection of micrometastases in lymph nodes from patients with breast cancer, Br. J. Surg. 89 (2002) 86–89.
[10] M. Brown, W. Grundy, D. Lin, N. Cristianini, C. Sugnet, T. Furey, M. Ares Jr., D. Haussler, Knowledge-based analysis of microarray gene expression data by using support vector machines, Proc. Nat. Acad. Sci. 97 (2000) 262–267.
[11] C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (1998) 121–167.
[12] M. Chow, E. Moler, I. Mian, Identifying marker genes in transcription profile data using a mixture of feature relevance experts, Physiol. Genomics 5 (2001) 99–111.
[13] C. Collins, J. Rommens, D. Kowbel, T. Godfrey, M. Tanner, S.-I. Hwang, D. Polikoff, G. Nonet, J. Cochran, K. Myambo, K. Jay, J. Froula, T. Cloutier, W.-L. Kuo, P. Yaswen, S. Dairkee, J. Giovanola, G. Hutchinson, J. Isola, O.-P. Kallioniemi, M. Palazzolo, C. Martin, C. Ericsson, D. Pinkel, D. Albertson, W.-B. Li, J. Gray, Positional cloning of ZNF217 and NABC1: genes amplified at 20q13.2 and overexpressed in breast carcinoma, Proc. Nat. Acad. Sci. 95 (1998) 8703–8708.
[14] N. Cristianini, J. Shawe-Taylor, Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, Cambridge, England, 2000.
[15] S. Dhanasekaran, T. Barrette, D. Ghosh, R. Shah, S. Varambally, K. Kurachi, K. Pienta, M. Rubin, A. Chinnaiyan, Delineation of prognostic biomarkers in prostate cancer, Nature 432 (2001) 822–826.
[16] D. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, Technical Report, Statistics Department, Stanford University, 1999, The report is available at http://www-stat.stanford.edu/~donoho/reports.html.
[17] T. Furey, N. Cristianini, N. Duffy, D. Bednarski, M. Schummer, D. Haussler, Support vector machine classification and validation of cancer tissue samples using microarray expression data, Bioinformatics 16 (2000) 906–914.
[18] M. Garber, O. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyana-Gengelbach, M. van de Rijn, G. Rosen, C. Perou, R. Whyte, R. Altman, P. Brown, D. Botstein, I. Petersen, Diversity of gene expression in adenocarcinoma of the lung, Proc. Nat. Acad. Sci. 98 (2001) 13784–13789.
[19] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, C. Bloomfeld, E. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[20] T. Graepel, B. Herbrich, R. Schölkopf, A. Smola, P. Bartlett, K. Müller, K. Obermayer, R. Williamson, Classification on proximity data with lp-machines, in: Ninth International Conference on Artificial Neural Networks, Vol. 470, IEE, London, 1999, pp. 304–309.
[21] L. Grate, C. Bhattacharyya, M. Jordan, I. Mian, Integrated analysis of transcript profiling and protein sequence data, Mechanisms Ageing Dev., in press.
[22] T. Gudjonsson, R. Villadsen, H. Lind Nielsen, L. Rønnov-Jessen, M. Bissell, O. Petersen, Isolation immortalization, and characterization of a human breast epithelial cell line with stem cell properties, Genes Dev. 16 (2002) 693–706.
[23] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning 46 (2002) 389–422.
[24] T. Hastie, R. Tibshirani, F.J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2000.
[25] I. Hedenfalk, D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, M. Raffeld, Z. Yakhini, A. Ben-Dor, E. Dougherty, J. Kononen, L. Bubendorf, W. Fehrle, S. Pittaluga, S. Gruvberger, N. Loman, O. Johannsson, H. Olsson, B. Wilfond, G. Sauter, O.-P. Kallioniemi, A. Borg, J. Trent, Gene-expression profiles in hereditary breast cancer, N. Engl. J. Med. 344 (2001) 539–548.
[26] J. Khan, J. Wei, M. Ringner, L. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. Antonescu, C. Peterson, P. Meltzer, Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nat. Med. 7 (2001) 673–679.
[27] J. Kononen, L. Bubendorf, A. Kallioniemi, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M. Mihatsch, G. Sauter, O.P. Kallioniemi, Tissue microarrays for high-throughput molecular profiling of tumor specimens, Nat. Med. 4 (1998) 844–847.
[28] G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, M. Jordan, Minimax probability machine, Adv. Neural Process. Systems 14.
[29] L. Liotta, E. Kohn, E. Perticoin, Clinical proteomics, personalized molecular medicine, J. Am. Med. Assoc. 14 (2001) 2211–2214.
[30] L. Mates, E. Korpos, F. Deak, Z. Liu, D. Beier, A. Aszodi, I. Kiss, Comparative analysis of the mouse and human genes (matn2 and matn2) for matrilin-2, a filament-forming protein widely distributed in extracellular matrices, Matrix Biol. 21 (2002) 163–174.

[31] E. Moler, M. Chow, I. Mian, Analysis of molecular profile data using generative and discriminative methods, Physiol. Genomics 4 (2000) 109–126.

[32] E. Moler, D. Radisky, I. Mian, Integrating naïve Bayes models and external knowledge to examine copper and iron homeostasis in Saccharomyces cerevisiae, Physiol. Genomics 4 (2000) 127–135.

[33] Y. Nesterov, A. Nemirovsky, Interior Point Polynomial Methods in Convex Programming: Theory and Applications, SIAM, Philadelphia, PA, 1994.

[34] D. Notterman, U. Alon, A. Sierk, A. Levine, Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays, Cancer Res. 61 (2001) 3124–3130.

[35] E. Petricoin III, A. Ardekani, B. Hitt, P. Levine, V. Fusaro, S. Steinberg, G. Mills, C. Simone, D. Fishman, E. Kohn, L. Liotta, Use of proteomic patterns in serum to identify ovarian cancer, The Lancet 359 (2002) 572–577.

[36] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. Mesirov, T. Poggio, W. Gerald, M. Loda, E. Lander, T. Golub, Multiclass cancer diagnosis using tumor gene expression signatures, Proc. Nat. Acad. Sci. 98 (2001) 15149–15154.

[37] J. Reubi, M. Gugger, B. Waser, J. Schaer, Y(1)-mediated effect of neuropeptide y in cancer: breast carcinomas as targets, Cancer Res. 61 (2001) 4636–4641.

[38] A. Smola, T. Frieß, B. Schölkopf, Semiparametric support vector and linear programming machines, in: Neural and Information Processing Systems, Vol. 11, MIT Press, Cambridge, MA, 1999.

[39] T. Sorlie, C. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. Eisen, M. van de Rijn, S. Jeffrey, T. Thorsen, H. Quist, J. Matese, P. Brown, D. Botstein, P. Lonning, A.-L. Borresen-Dale, Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications, Proc. Nat. Acad. Sci. 98 (2001) 10869–10874.

[40] A. Su, J. Welsh, L. Sapinoso, S. Kern, P. Dimitrov, H. Lapp, P. Schultz, S. Powell, C. Moskaluk, H. Frierson Jr., G. Hampton, Molecular classification of human carcinomas by use of gene expression signatures, Cancer Res. 61 (2001) 7388–7393.

[41] L. van't Veer, H. Dai, M. van de Vijver, Y. He, A. Hart, M. Mao, H. Peterse, K. van der Kooy, M. Marton, A. Witteveen, G. Schreiber, R. Kerkhoven, C. Roberts, P. Linsley, R. Bernards, S. Friend, Gene expression profiling predicts clinical outcome of breast cancer, Nature 415 (2002) 530–536.

[42] J. Weston, M.S.O., Chapelle, M. Pontil, T. Poggio, V. Vapnik, Feature Selection for SVMs, in: Advances in Neural Information Processing Systems, Vol. 13, MIT Press, Cambridge, MA, 2000, pp. 668–674.