# Simultaneous Relevant Feature Identification and Classification in High-Dimensional Spaces

L.R. Grate[1], C. Bhattacharyya[23], M.I. Jordan[23], and I.S. Mian[1]

[1] Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley CA 94720
[2] Department of EECS, University of California Berkeley, Berkeley CA 94720
[3] Department of Statistics, University of California Berkeley, Berkeley CA 94720

## 1 Abstract

Molecular profiling technologies monitor thousands of transcripts, proteins, metabolites or other species concurrently in biological samples of interest. Given two-class, high-dimensional profiling data, nominal LIKNON [4] is a specific implementation of a methodology for performing simultaneous relevant feature identification and classification. It exploits the well-known property that minimising an $l_1$ norm (via linear programming) yields a sparse hyperplane [15, 26, 2, 8, 17]. This work (i) examines computational, software and practical issues required to realise nominal LIKNON, (ii) summarises results from its application to five real world data sets, (iii) outlines heuristic solutions to problems posed by domain experts when interpreting the results and (iv) defines some future directions of the research.

## 2 Introduction

Biologists and clinicians are adopting high-throughput genomics, proteomics and related technologies to assist in interrogating normal and perturbed systems such as unaffected and tumour tissue specimens. Such investigations can generate data having the form $\mathcal{D} = \{(\mathbf{x}_n, y_n), n \in (1, \ldots, N)\}$ where $\mathbf{x}_n \in \mathbb{R}^P$ and, for two-class data, $y_n \in \{+1, -1\}$. Each element of a data point $\mathbf{x}_n$ is the absolute or relative abundance of a molecular species monitored. In transcript profiling, a data point represents transcript (gene) levels measured in a sample using cDNA, oligonucleotide or similar microarray technology. A data point from protein profiling can represent Mass/Charge (M/Z) values for low molecular weight molecules (proteins) measured in a sample using mass spectrocopy.

In cancer biology, profiling studies of different types of (tissue) specimens are motivated largely by a desire to create clinical decision support systems for accurate tumour classification and to identify robust and reliable targets, "biomarkers", for imaging, diagnosis, prognosis and therapeutic intervention [14, 3, 13, 27, 18, 23, 9, 25, 28, 19, 21, 24]. Meeting these biological challenges includes addressing the general statistical problems of classification and prediction, and relevant feature identification.

Support Vector Machines (SVMs) [30, 8] have been employed successfully for cancer classification based on transcript profiles [5, 22, 25, 28]. Although mechanisms for reducing the number of features to more manageable numbers include discarding those below a user-defined threshold, relevant feature identification is usually addressed via a filter-wrapper strategy [12, 22, 32]. The filter generates candidate feature subsets whilst the wrapper runs an induction algorithm to determine the discriminative ability of a subset. Although SVMs and the newly formulated Minimax Probability Machine (MPM) [20] are good wrappers [4], the choice of filtering statistic remains an open question.

Nominal LIKNON is a specific implementation of a strategy for performing simultaneous relevant feature identification and classification [4]. It exploits the well-known property that minimising an $l_1$ norm (via linear programming) yields a sparse hyperplane [15, 26, 2, 8, 17]. The hyperplane constitutes the classifier whilst its sparsity, a weight vector with few non-zero elements, defines a small number of relevant features. Nominal LIKNON is computationally less demanding than the prevailing filter–(SVM/MPM) wrapper strategy which treats the problems of feature selection and classification as two independent tasks [4, 16]. Biologically, nominal LIKNON performs well when applied to real world data generated not only by the ubiquitous transcript profiling technology, but also by the emergent protein profiling technology.

## 3 Simultaneous relevant feature identification and classification

Consider a data set $\mathcal{D} = \{(\mathbf{x}_n, y_n), n \in (1, \ldots, N)\}$. Each of the $N$ data points (profiling experiments) is a $P$-dimensional vector of features (gene or protein abundances) $\mathbf{x}_n \in \mathbb{R}^P$ (usually $N \sim 10^1 - 10^2; P \sim 10^3 - 10^4$). A data point $n$ is assigned to one of two classes $y_n \in \{+1, -1\}$ such a normal or tumour tissue sample. Given such two-class high-dimensional data, the analytical goal is to estimate a sparse classifier, a model which distinguishes the two classes of data points (classification) *and* specifies a small subset of discriminatory features (relevant feature identification). Assume that the data $\mathcal{D}$ can be separated by a linear hyperplane in the $P$-dimensional input feature space. The learning task can be formulated as an attempt to estimate a hyperplane, parameterized in terms of a weight vector $\mathbf{w}$ and bias $b$, via a solution to the following $N$ inequalities [30]:

$$\begin{aligned} y_n z_n = y_n(\mathbf{w}^T \mathbf{x}_n - b) \geq 0 \\ \forall n = \{1, \ldots, N\} \ . \end{aligned} \tag{1}$$

The hyperplane satisfying $\mathbf{w}^T \mathbf{x} - b = 0$ is termed a classifier. A new data point $\mathbf{x}$ (abundances of $P$ features in a new sample) is classified by computing $z = \mathbf{w}^T \mathbf{x} - b$. If $z > 0$, the data point is assigned to one class otherwise it belongs to the other class.

Enumerating relevant features at the same time as discovering a classifier can be addressed by finding a sparse hyperplane, a weight vector $\mathbf{w}$ in which most components are equal to zero. The rationale is that zero elements do not contribute to determining the value of $z$:

$$z = \sum_{p=1}^{P} w_p x_p - b \ .$$

If $w_p = 0$, feature $p$ is "irrelevant" with regards to deciding the class. Since only non-zero elements $w_p \neq 0$ influence the value of $z$, they can be regarded as "relevant" features.

The task of defining a small number of relevant features can be equated with that of finding a small set of non-zero elements. This can be formulated as an optimisation problem; namely that of minimising the $l_0$ norm $\|\mathbf{w}\|_0$, where $\|\mathbf{w}\|_0 =$ number of$\{p : w_p \neq 0\}$, the number of non-zero elements of $\mathbf{w}$. Thus we obtain:

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \|\mathbf{w}\|_0 \\ \text{subject to } & y_n(\mathbf{w}^T\mathbf{x}_n - b) \geq 0 \\ & \forall n = \{1, \ldots, N\} \ . \end{aligned} \tag{2}$$

Unfortunately, problem (2) is NP-hard [10]. A tractable, convex approximation to this problem can be obtained by replacing the $l_0$ norm with the $l_1$ norm $\|\mathbf{w}\|_1$, where $\|\mathbf{w}\|_1 = \sum_{p=1}^{P} |w_p|$, the sum of the absolute magnitudes of the elements of a vector [10]:

$$\begin{aligned} \min_{\mathbf{w},b} \ & \|\mathbf{w}\|_1 = \textstyle\sum_{p=1}^{P} |w_p| \\ \text{subject to } & y_n(\mathbf{w}^T\mathbf{x}_n - b) \geq 0 \\ & \forall n = \{1, \ldots, N\} \ . \end{aligned} \tag{3}$$

A solution to (3) yields the desired sparse weight vector $\mathbf{w}$.

Optimisation problem (3) can be solved via linear programming [11]. The ensuing formulation requires the imposition of constraints on the allowed ranges of variables. The introduction of new variables $u_p, v_p \in \mathbb{R}^P$ such that $|w_p| = u_p + v_p$ and $w_p = u_p - v_p$ ensures non-negativity. The range of $w_p = u_p - v_p$ is unconstrained (positive or negative) whilst $u_p$ and $v_p$ remain non-negative. $u_p$ and $v_p$ are designated the "positive" and "negative" parts respectively. Similarly, the bias $b$ is split into positive and negative components $b = b_+ - b_-$. Given a solution to problem (3), either $u_p$ or $v_p$ will be non-zero for feature $p$ [11]:

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v},b_+,b_-} \quad & \textstyle\sum_{p=1}^{P} (u_p + v_p) \\ \text{subject to } \quad & y_n((\mathbf{u} - \mathbf{v})^T\mathbf{x}_n - (b_+ - b_-)) \geq 1 \\ & u_p \geq 0; v_p \geq 0; b_+ \geq 0; b_- \geq 0 \\ & \forall n = \{1, \ldots, N\}; \forall p = \{1, \ldots, P\} \ . \end{aligned} \tag{4}$$

A detailed description of the origins of the $\geq 1$ constraint can be found elsewhere [30].

If the data $\mathcal{D}$ are not linearly separable, misclassifications (errors in the class labels $y_n$) can be accounted for by the introduction of slack variables $\xi_n$. Problem (4) can be recast yielding the final optimisation problem,

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v},b_+,b_-} \quad & \textstyle\sum_{p=1}^{P}(u_p+v_p) + C\sum_{n=1}^{N}\xi_n \\ \text{subject to } & y_n((\mathbf{u}-\mathbf{v})^T\mathbf{x}_n - (b_+ - b_-)) \geq 1-\xi_n \\ & u_p \geq 0; v_p \geq 0; b_+ \geq 0; b_- \geq 0; \xi_n \geq 0 \\ & \forall n = \{1,\ldots,N\}; \forall p = \{1,\ldots,P\} \ . \end{aligned} \tag{5}$$

$C$ is an adjustable parameter weighing the contribution of misclassified data points. Larger values lead to fewer misclassifications being ignored: $C = 0$ corresponds to no outliers being ignored whereas $C \to \infty$ leads to the hard margin limit.

## 4 Computational, software and practical issues

Learning the sparse classifier defined by optimisation problem (5) involves minimising a linear function subject to linear constraints. Efficient algorithms for solving such linear programming problems involving ∼10,000 variables ($N$) and ∼10,000 constraints ($P$) are well-known. Standalone open source codes include `lp_solve`[4] and `PCx`[5].

Nominal LIKNON is an implementation of the sparse classifier (5). It incorporates routines written in Matlab[6] and a system utilising `perl`[7] and `lp_solve`. The code is available from the authors upon request. The input consists of a file containing an $N \times (P+1)$ data matrix in which each row represents a single profiling experiment. The first $P$ columns are the feature values, abundances of molecular species, whilst column $P+1$ is the class label $y_n \in \{+1, -1\}$. The output comprises the non-zero values of the weight vector $\mathbf{w}$ (relevant features), the bias $b$ and the number of non-zero slack variables $\xi_n$.

The adjustable parameter $C$ in problem (5) can be set using cross validation techniques. The results described here were obtained by choosing $C = 0.5$ or $C = 1$.

## 5 Application of nominal Liknon to real world data

Nominal LIKNON was applied to five data sets in the size range ($N = 19$, $P =$ 1,987) to ($N = 200$, $P =$ 15,154). A data set $\mathcal{D}$ yielded a sparse classifier, $\mathbf{w}$ and $b$, and a specification of the $l$ relevant features ($P \gg l$). Since the profiling studies produced only a small number of data points ($N \ll P$), the generalisation error

[4] `http://www.netlib.org/ampl/solvers/lpsolve/`
[5] `http://www-fp.mcs.anl.gov/otc/Tools/PCx/`
[6] `http://www.mathworks.com`
[7] `http://www.perl.org/`

of a nominal LIKNON classifier was determined by computing the leave-one-out error for $l$-dimensional data points. A classifier trained using $N-1$ data points was used to predict the class of the withheld data point; the procedure repeated $N$ times. The results are shown in Table 1.

Nominal LIKNON performs well in terms of simultaneous relevant feature identification and classification. In all five transcript and protein profiling data sets a hyperplane was found, the weight vector was sparse ($< 100$ or $< 2\%$ non-zero components) and the relevant features were of interest to domain experts (they generated novel biological hypotheses amenable to subsequent experimental or clinical validation). For the protein profiles, better results were obtained using normalised as opposed to raw values: when employed to predict the class of 16 independent non-cancer samples, the 51 relevant features had a test error of 0 out of 16.

On a powerful desktop computer, a $> 1$ GHz Intel-like machine, the time required to create a sparse classifier varied from 2 seconds to 20 minutes. For the larger problems, the main memory RAM requirement exceeded 500 MBytes.

## 6 Heuristic solutions to problems posed by domain experts

Domain experts wish to postprocess nominal LIKNON results to assist in the design of subsequent experiments aimed at validating, verifying and extending any biological predictions. In lieu of a theoretically sound statistical framework, heuristics have been developed to prioritise, reduce or increase the number of relevant features.

In order to prioritise features, assume that all $P$ features are on the same scale. The $l$ relevant features can be ranked according to the magnitude and/or sign of the non-zero elements of the weight vector $\mathbf{w}$ ($w_p \neq 0$). To reduce the number of relevant features to a "smaller, most interesting" set, a histogram of $w_p \neq 0$ values can be used to determine a threshold for pruning the set. In order to increase the number of features to a "larger, more interesting" set, nominal LIKNON can be run in an iterative manner. The $l$ relevant features identified in one pass through the data are removed from the data points to be used as input for the next pass. Each successive round generates a new set of relevant features. The procedure is terminated either by the domain expert or by monitoring the leave-one-out error of the classifier associated with each set of relevant features.

Preliminary results from analysis of the gastrointestinal stromal tumour/spindle cell tumour transcript profiling data set indicate that these extensions are likely to be of utility to domain experts. The leave-one-out error of the relevant features identified by five iterations of nominal LIKNON was at most one. The details are: iteration 0 (number of relevant features = 6, leave-one-out error = 0), iteration 1 (5, 0), iteration 2 (5, 1), iteration 3 (9, 0), iteration 4 (13, 1), iteration 5 (11, 1).

Iterative LIKNON may prove useful during explorations of the (qualitative) association between relevant features and their behaviour in the $N$ data points. The

**Table 1.** Summary of published and unpublished investigations using nominal LIKNON [4, 16].

| | |
|---|---|
| Transcript profiles | Sporadic breast carcinoma tissue samples [29] |
| | inkjet microarrays; relative transcript levels |
| | `http://www.rii.com/publications/vantveer.htm` |
| Two-class data | 46 patients with distant metastases $< 5$ years |
| | 51 patients with no distant metastases $\geq 5$ years |
| Relevant features | 72 out of $P$=5,192 |
| Leave-one-out error | 1 out of $N$=97 |
| Transcript profiles | Tumour tissue samples [1] |
| | custom cDNA microarrays; relative transcript levels |
| | `http://www.nhgri.nih.gov/DIR/Microarray/selected_publications.html` |
| Two-class data | 13 *KIT*-mutation positive gastrointestinal stromal tumours |
| | 6 spindle cell tumours from locations outside the gastrointestinal tract |
| Relevant features | 6 out of $P$=1,987 |
| Leave-one-out error | 0 out of $N$=19 |
| Transcript profiles | Small round blue cell tumour samples (EWS, RMS, NHL, NB) [19] |
| | custom cDNA microarrays; relative transcript levels |
| | `http://www.nhgri.nih.gov/DIR/Microarray/Supplement` |
| Two-class data | 46 EWS/RMS tumour biopsies |
| | 38 EWS/RMS/NHL/NB cell lines |
| Relevant features | 23 out of $P$=2,308 |
| Leave-one-out error | 0 out of $N$=84 |
| Transcript profiles | Prostate tissue samples [31] |
| | Affymetrix arrays; absolute transcript levels |
| | `http://carrier.gnf.org/welsh/prostate` |
| Two-class data | 9 normal |
| | 25 malignant |
| Relevant features | 7 out of $P$=12,626 |
| Leave-one-out error | 0 out of $N$=34 |
| Protein profiles | Serum samples [24] |
| | SELDI-TOF mass spectrometry; M/Z values (spectral amplitudes) |
| | `http://clinicalproteomics.steem.com` |
| Two-class data | 100 unaffected |
| | 100 ovarian cancer |
| Relevant features | 51 out of $P$=15,154 |
| Leave-one-out error | 3 out of $N$=200 |

gastrointestinal stromal tumour/spindle cell tumour transcript profiling data set has been the subject of probabilistic clustering [16]. A finite Gaussian mixture model as implemented by the programme AutoClass [6] was estimated from $P$=1,987, $N$=19-dimensional unlabelled data points. The trained model was used to assign each feature (gene) to one of the resultant clusters. Five iterations of nominal LIKNON identified the majority of genes assigned to a small number of discriminative clusters. Furthermore, these genes constituted most of the important distinguishing genes defined by the original authors [1].

## 7 Discussion

Nominal LIKNON implements a mathematical technique for finding a sparse hyperplane. When applied to two-class high-dimensional real-world molecular profiling data, it identifies a small number of relevant features and creates a classifer that generalises well. As discussed elsewehere [4, 7], many subsets of relevant features are likely to exist. Although nominal LIKNON specifies but one set of discriminatory features, this "low-hanging fruit" approach does suggest genes of interest to experimentalists. Iterating the procedure provides a rapid mechanism for highlighting additional sets of relevant features that yield good classifiers. Since nominal LIKNON is a single-pass method, one disadvantage is that the learned parameters cannot be adjusted (improved) as would be possible with a more typical train/test methodology.

## 8 Future directions

Computational biology and chemistry are generating high-dimensional data so sparse solutions for classification and regression problems are of widespread importance. A general purpose toolbox containing specific implementations of particular statistical techniques would be of considerable practical utility. Future plans include developing a suite of software modules to aid in performing tasks such as the following. A. Create high-dimensional input data. (i) Direct generation by high-throughput experimental technologies. (ii) Systematic formulation and extraction of large numbers of features from data that may be in the form of strings, images, and so on (*a priori*, features "relevant" for one problem may be "irrelevant" for another). B. Enunciate sparse solutions for classification and regression problems in high-dimensions. C. Construct and assess models. (i) Learn a variety of models by a grid search through the space of adjustable parameters. (ii) Evaluate the generalisation error of each model. D. Combine best models to create a final decision function. E. Propose hypotheses for domain expert.


## 9 Acknowledgements

This work was supported by NSF grant IIS-9988642, the Director, Office of Energy Research, Office of Health and Environmental Research, Division of the U.S. Department of Energy under Contract No. DE-AC03-76F00098 and an LBNL/LDRD through U.S. Department of Energy Contract No. DE-AC03-76SF00098.

## References


1. S.V. Allander, N.N. Nupponen, M. Ringner, G. Hostetter, G.W. Maher, N. Goldberger, Y. Chen, Carpten J., A.G. Elkahloun, and P.S. Meltzer. Gastrointestinal Stromal Tumors with KIT mutations exhibit a remarkably homogeneous gene expression profile. *Cancer Research*, 61:8624–8628, 2001.
2. K. Bennett and A. Demiriz. Semi-supervised support vector machines. In *Neural and Information Processing Systems*, volume 11. MIT Press, Cambridge MA, 1999.
3. A. Bhattacharjee, W.G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E.J. Mark, E.S. Lander, W. Wong, B.E. Johnson, T.R. Golub, D.J. Sugarbaker, and M. Meyerson. Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci.*, 98:13790–13795, 2001.
4. C. Bhattacharyya, L.R. Grate, A. Rizki, D.C. Radisky, F.J. Molina, M.I. Jordan, M.J. Bissell, and I.S. Mian. Simultaneous relevant feature identification and classification in high-dimensional spaces: application to molecular profiling data. *Submitted, Signal Processing*, 2002.
5. M.P. Brown, W.N. Grundy, D. Lin, N. Cristianini, C.W. Sugnet, T.S. Furey, M. Ares, Jr, and D. Haussler. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci.*, 97:262–267, 2000.
6. P. Cheeseman and J. Stutz. Bayesian Classification (AutoClass): Theory and Results. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 153–180. AAAI Press/MIT Press, 1995. The software is available at the URL `http://www.gnu.org/directory/autoclass.html`.
7. M.L. Chow, E.J. Moler, and I.S. Mian. Identifying marker genes in transcription profile data using a mixture of feature relevance experts. *Physiological Genomics*, 5:99–111, 2001.
8. N. Cristianini and J. Shawe-Taylor. *Support Vector Machines and other kernel-based learning methods.* Cambridge University Press, Cambridge, England, 2000.
9. S.M. Dhanasekaran, T.R. Barrette, R. Ghosh, D. Shah, S. Varambally, K. Kurachi, K.J. Pienta, M.J. Rubin, and A.M. Chinnaiyan. Delineation of prognostic biomarkers in prostate cancer. *Nature*, 432, 2001.
10. D.L. Donoho and X. Huo. Uncertainty principles and idea atomic decomposition. Technical Report, Statistics Department, Stanford University, 1999.
11. R. Fletcher. *Practical Methods in Optimization.* John Wiley & Sons, New York, 2000.
12. T. Furey, N. Cristianini, N. Duffy, D. Bednarski, M. Schummer, and D. Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16:906–914, 2000.
13. M.E. Garber, O.G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyana-Gengelbach, M. van de Rijn, G.D. Rosen, C.M. Perou, R.I. Whyte, R.B. Altman, P.O. Brown, D. Botstein, and I. Petersen. Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci.*, 98:13784–13789, 2001.
14. T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfeld, and E.S. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537, 1999. The data are available at the URL `waldo.wi.mit.edu/MPR/data_sets.html`.

15. T. Graepel, B. Herbrich, R. Schölkopf, A.J. Smola, P. Bartlett, K. Müller, K. Obermayer, and R.C. Williamson. Classification on proximity data with lp-machines. In *Ninth International Conference on Artificial Neural Networks*, volume 470, pages 304–309. IEE, London, 1999.
16. L.R. Grate, C. Bhattacharyya, M.I. Jordan, and I.S. Mian. Integrated analysis of transcript profiling and protein sequence data. *In press, Mechanisms of Ageing and Development*, 2002.
17. T. Hastie, R. Tibshirani, , and Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag, New York, 2000.
18. 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, and J. Trent. Gene-expression profiles in hereditary breast cancer. *New England Journal of Medicine*, 344:539–548, 2001.
19. J. Khan, J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, Antonescu C.R., Peterson C., and P.S. Meltzer. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 7:673–679, 2001.
20. G. Lanckerit, L. El Ghaoui, C. Bhattacharyya, and M.I. Jordan. Minimax probability machine. *Advances in Neural Processing systems*, 14, 2001.
21. L.A. Liotta, E.C. Kohn, and E.F. Perticoin. Clinical proteomics. personalized molecular medicine. *JAMA*, 14:2211–2214, 2001.
22. E.J. Moler, M.L. Chow, and I.S. Mian. Analysis of molecular profile data using generative and discriminative methods. *Physiological Genomics*, 4:109–126, 2000.
23. D.A. Notterman, U. Alon, A.J. Sierk, and A.J. Levine. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Research*, 61:3124–3130, 2001.
24. E.F. Petricoin III, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B Mills, C. Simone, D.A. Fishman, E.C. Kohn, and L.A. Liotta. Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet*, 359:572–577, 2002.
25. S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J.P. Mesirov, T. Poggio, W. Gerald, M. Loda, E.S. Lander, and T.R. Golub. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci.*, 98:15149–15154, 2001. The data are available from `http://www-genome.wi.mit.edu/mpr/GCM.html`.
26. A. Smola, T.T. Friess, and B. Schölkopf. Semiparametric support vector and linear programming machines. In *Neural and Information Processing Systems*, volume 11. MIT Press, Cambridge MA, 1999.
27. T. Sorlie, C.M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, T. Thorsen, H. Quist, J.C. Matese, P.O. Brown, D. Botstein, P.E. Lonning, and A.-L. Borresen-Dale. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.*, 98:10869–10874, 2001.
28. A.I. Su, J.B. Welsh, L.M. Sapinoso, S.G. Kern, P. Dimitrov, H. Lapp, P.G. Schultz, S.M. Powell, C.A. Moskaluk, H.F. Frierson Jr, and G.M. Hampton. Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Research*, 61:7388–7393, 2001.

29. L.J. van 't Veer, H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, van der Kooy K., M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards, and S.H. Friend. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530–536, 2002.
30. V. Vapnik. *Statistical Learning Theory.* Wiley, New York, 1998.
31. J.B. Welsh, L.M. Sapinoso, A.I. Su, S.G. Kern, J. Wang-Rodriguez, C.A. Moskaluk, J.F. Frierson Jr, and G.M. Hampton. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Research*, 61:5974–5978, 2001.
32. J. Weston, Mukherjee S., O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik. Feature Selection for SVMs. In *Advances in Neural Information Processing Systems*, volume 13, 2000.