One-Class Classification for Microarray Datasets with Feature Selection

Beatriz Pérez-Sánchez(✉), Oscar Fontenla-Romero,
and Noelia Sánchez-Maroño

Laboratory for Research and Development in Artificial Intelligence (LIDIA),
Department of Computer Science, Faculty of Informatics,
University of A Coruña, Campus de Elviña, S/n, 15071 A Coruña, Spain
bperezs@udc.es

Abstract. Microarray data classification is a critical challenge for computational techniques due to its inherent characteristics, mainly small sample size and high dimension of the input space. For this type of data two-class classification techniques have been widely applied while one-class learning is considered as a promising approach. In this paper, we study the suitability of employing the one-class classification for microarray datasets while the role played by feature selection is analyzed. The superiority of this approach is demonstrated by comparison with the classical approach, with two classes, on different benchmark data sets.

1 Introduction

In the last years there has been a boom in the acquisition of biomedical data. The application of innovative technologies in the field of molecular genetics - such as deoxyribonucleic acid (DNA) microarrays - provides a global view of the cell enabling the measurement of simultaneous expression of tens of thousands of genes. Thus, microarrays allow for creating data sets that represent a system which may be of interest from a biological or clinical point of view. Moreover, due to the intrinsic characteristics of these data, in recent years they also have become a challenge for the scientific community in terms of machine learning and bioinformatics. Microarray data present a large dimensionality of the feature space (often reaching several thousands of genes) compared to the small number of samples available (usually less than a hundred). The high dimensionality of the feature space degrades the performance of the classifier and increases its computational complexity. As a consequence, the application of conventional statistical and machine learning techniques for classification purposes is heavily limited.

Several studies have shown that most genes measured in a DNA microarray experiment are not relevant in the accurate classification of different classes of the problem [1]. To overcome the problem known as *curse of dimensionality* or Hughes effect [2] that appears when with a fixed number of training samples, the predictive power of the learner reduces as the feature dimensionality increases, dimensionality reduction techniques plays a crucial role. Among dimensionality

L. Iliadis and C. Jayne (Eds.): EANN 2015, CCIS 517, pp. 325–334, 2015.
DOI: 10.1007/978-3-319-23983-5_30

reduction methods, feature selection identifies and discards irrelevant features from the training data as a previous step of the classification stage. Thus, the learning algorithm can focus on those useful aspects of data improving its performance. There are usually three varieties of feature selection methods: wrappers, filters and embedded methods. Wrapper models involve optimizing a predictor as part of the selection process. Filter methods rely on the general characteristics of the training data to select feature independent of any predictor. And finally, embedded techniques generally use machine learning models for classification and then an optimal subset of features is built by the classifier algorithm. Both wrapper and embedded methods have an important computational cost due to the interaction with the classifier. In the case of microarray data classification the most employed methods fall into the filter category [3]. In [4] the authors review the most up-to-date feature selection methods developed in this field (filters and one embedded method) and a comparative among them is introduced.

Machine learning has been widely applied for handling microarray data sets being supervised machine learning a promising approach. Several surveys about machine learning in microarray were published over time [5–7]. To date, two-class classification techniques have mainly been applied and Support Vector Machines are among the most popular classifiers used for this task. In fact, there is a clear tendency in the literature to use SVM versus other classifiers. Many microarray data are used for cancer diagnosis. In this context, some datasets present unbalanced classes as healthy patients normally predominate. For that reason, the one-class classification (OCC) paradigm also has been employed to treat this kind of data but its use is not so extended as the two-class perspective. In [8] the author proposes to use a one-class approach to classify microarrays because of these models rely only on objects coming from single class distribution. In this case two approaches can be considered since training of the classifier can be focused on the majority class or on the minority one. Despite of having less information to distinguish between classes, one-class models can easily learn the specific properties of a given data set and are robust to intrinsic difficulties of the data.

Thus, the aim of this paper is to compare the behavior of the two-class classification versus the one-class paradigm when dealing with microarray data sets and analyze the role played by feature selection.

2 One-class Classifiers

In a multi-class classification problem, data from several categories are available and the decision boundary is chosen thanks to objects from each class. The main goal is to categorize an unknown object as belongs to one specific class from the more broad set of classes, two in the simplest case of binary classification. Nevertheless, occasionally the classification task does not consist just in categorizing an object into one specific class but deciding if this object fits to a particular class or not. OCC paradigm shows a favorable perspective to solve these kinds of problems since in OCC only instances from one of the classes are available or considered. The objects from this category will be called the *target* objects while all other are the *outlier* ones.

OCC problems are usual in the real world, there are different situations where normal examples are widely accessible but outliers are expensive or even unfeasible to collect. For example, for identifying failures in a machine, data about the regular working of the machine can be easily obtained. However, obtaining examples about failures is expensive and sometimes impossible since some of them would not have taken place. Another scenario refers to the diagnosis of a disease, as in the previously commented case of cancer. Although several methods to solve the one-class classification problem have been proposed we have selected the Support Vector Data Description [9]. This algorithm is one of the most up-to-date one-class classifiers applied to handle the type of data of interest. This method is briefly introduced as follows.

2.1 Support Vector Data Description.

A one-class classification technique based on Support Vector Machines (SVMs) [10] was proposed by Tax [9] and it is known as Support Vector Data Description (SVDD). This approach establishes a closed boundary around the target data by means of a small hypersphere which encloses all target data. As SVDD is based on SVMs, its decision boundary is described by a certain target data called support vectors. The hypersphere is characterized by a center $\mathbf{c}$ and a radius R ($R > 0$) around the dataset which has minimal volume [11] as it can be seen in Fig. 1. The error function to minimize is given below,

$$\min_{\mathbf{c},R} R^2$$

where

$$\|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2, i = 1, 2, \ldots, n.$$

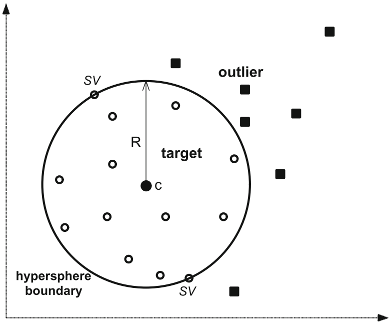

Fig. 1. The hypersphere boundary includes all target data. The objects which are on the boundary are known as the support vectors (SV).

As SVDD is a variant of the SVM, also it has to solve a quadratic optimization problem during its training. Therefore a dual formulation, in terms of inner products, can be derived. Replacing the normal inner products $(\mathbf{x}_i \cdot \mathbf{x}_j)$ by a kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ the flexibility of the model is increased. Data are mapped to a high dimensional feature space without much extra computational cost. The most commonly used kernel function is Gaussian kernel. The distance from the sphere center to a test object $\mathbf{x}$ is calculated by means of the following equation

$$d_{SVDD}(\mathbf{x}, X_t) = K(\mathbf{x}, \mathbf{x}) - 2\sum_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$$

where the parameters α_i are obtained by quadratic optimization. To obtain a more efficient description there is the possibility of including outliers examples provided when these are available into the training procedure. In this case, the distance from $\mathbf{x}_i$ to the center $\mathbf{c}$ should not be strictly smaller than R^2, but larger distances should be penalized. Therefore, slack variables ξ_i are introduced and the new minimization problem is,

$$\min_{\mathbf{c},R} R^2 + C_i \sum_i \xi_i$$

where

$$\|\mathbf{x}_i - \mathbf{c}\|^2 \leq R^2 + \xi_i, \ \xi_i \geq 0, \ \forall i$$

taking into account that the parameter C controls the trade-off between the volume and the errors. In this way a more robust classifier can be obtained.

3 Experimental Setup

In this section, we test the suitability of the one-class learning approach for microarray datasets and compare the results with those obtained by the two-class approach, specifically the SVM algorithm. It is worth mentioning that the OCC is addressed by using both minority and majority class as target concept. Furthermore, we study the role played by feature selection as a previous step to the classification stage. Next, we establish certain considerations which have been taken into account in the experimental study.

- The input dataset was previously normalized to have zero mean and a standard deviation of 1.
- In order to obtain statistically significant results, 30 simulations were run with the cross-validation technique to tune the parameters of each method.
- For the implementation of classifiers two different toolboxs for Matlab was used. The data description toolbox, DDtools library[12], for SVDD and the Statistics and Machine Learning toolbox for SVM.

- Regarding the parameters associated with the SVDD classifier, the fraction of the target set which will be rejected was established to $0,01$, and the width parameter in the radial basis function kernel was selected by means of cross-validation technique. In case of SVM, the kernel function was getting by cross-validation and the percentage of the target object that can be considered as outliers was established as 0.
- With reference to the feature selection methods, all techniques are available in the well-known Weka tool [13], except for mRMR filter, whose implementation is available for *Matlab*.
- To evaluate the goodness of the selected set of genes in terms of accuracy of the classifier it is necessary to have an independent test set with data which have not been seen by neither the feature selection method nor the classifier. The selected data sets come originally distributed into training and test sets, so the training set was employed to perform the feature selection process and posterior classification while the test set was used to evaluate the appropriateness of the selection and the posterior classification.
- Finally, a statistical study was conducted to determine whether the results are statistically different. First at all, the normality conditions of each distribution are checked by means of Kolmogorow Smirnov test. As in any case, normal conditions are verified then the non parametric Kruskal-Wallis test was applied.

The remainder of this section is devoted to the analysis of the characteristics of the selected datasets and also to introduce the feature selection methods and the evaluation measures.

3.1 Datasets Characteristics

For this experimental study, we have considered two widely used binary microarray dataset which are available for download at [14,15]. Both data sets come originally separated in training and test thus, Table 1 summarizes the main characteristics of both partitions. For each set, we introduce its number of examples (# Ex.), attributes (# Atts.) and some information for majority and minority classes (number of examples/percentage of examples). Moreover, we provide the imbalance ratio (IR) defined as, the number of outlier examples that are divided by the number of normal examples. A value of 1 indicates balance whereas a large value denotes a high imbalance. As it can be seen in Table 1 both datasets present more imbalance in the test set especially in prostate data set. This may be caused by the problem known as *dataset shift* [16]. It occurs when testing (unseen) data experience a phenomenon that leads to a change in the distribution of a single feature, a combination of features, or the class boundaries. As a result, the common assumption that the training and testing data follow the same distributions is often violated in real world application and scenarios. In this regard, prostate dataset poses a big challenge for machine learning methods since the test dataset was extracted from a different experiment. It is possible that some classifiers, whose features are selected according to the training set, assign all samples to the majority class.

Table 1. Description of the train and test binary datasets used in the experimental study.

		Train				Test			
Dataset	# Atts.	# Ex.	Min. class	Maj. class	IR	# Ex.	Min. class	Maj. class	IR
Breast	24.481	78	34/43,59%	44/56,41%	1,29	19	7/36,84%	12/63,16%	1,71
Prostate	12.600	102	50/49,02%	52/50,98%	1,04	34	9/26,47%	25/73,53%	2,78

3.2 Feature Selection Methods

In this study we choose seven classical feature selection methods widely used by the researchers in this field. Moreover, for the sake of comparison, these methods are those used in [4], where a thorough study can be found. Such methods are: Correlation-based Feature Selection (CFS) [17], Fast Correlation-Based Filter (FCBF) [18], INTERACT algorithm [19], Information Gain (IG) [20], ReliefF [21], minimum Redundancy Maximum Relevance (mRMR) [22] and Suppor Vector Machine based on Recursive Feature Elimination (SVM-RFE) [23]. All these methods are in the filter category except SVM-REF, the most famous embedded method to specifically deal with gene selection for cancer classification. The three first feature selection methods (CFS, FCBF and INTERACT) return a subset of features, the number of selected ones for each dataset is shown in Table 2. Remaining techniques (IG, ReliefF, mRMR and SVM-REF) provide an ordered ranking of the features, for these one we show the performance when the top 10 and top 50 features are retained.

Table 2. Total number of features and selected number by subset methods.

	Features			
	Total	CFS	FCBF	INT
Breast	24.481	130	99	102
Prostate	12.600	89	77	73

3.3 Evaluation Measures

Most of performance measures for a binary class problem are built over the classical confusion matrix from which four measures can be directly obtained. TP and TN denote the number of positive and negative cases correctly classified, while FP and FN refer to the number of misclassified positive and negative examples, respectively. *Accuracy*, defined as $Acc = (TP + TN) / (TP + FN + TN + FP)$, is the most common metric for assessing the performance of learning systems. Moreover, the *true negative rate* or *specificity* $Sp = TN/ (TN + FP)$, is the percentage of correctly classified negative examples (e.g. the rate of healthy patients who are correctly classify as not having cancer). Analogously, the *true positive rate*, also called *recall* or *sensitivity*, $Se = TP/ (TP + FN)$ is the percentage of correctly classified positive instances (e.g. the rate of cancer patients who are correctly identified as having cancer). A perfect predictor would be described as 100% sensitive and 100% specific. Regardless of the class (majority or minority) used as target class in the OCC approach, it should be mentioned that sensitivity and specificity measures are always calculated on the same criteria. We consider as positive the cancer samples and as negative the healthy ones.

4 Experimental Results

In this section we present the experimental results achieved on the Breast and Prostate datasets previously introduced. Tables 3 and 4 show the performance obtained by SVM and SVDD classifiers over both test datasets. In case of SVDD classifier we introduce the results reached by using both classes (minority and majority) as the target concept in training process. Each column represents one of the three performance measures while rows indicate the feature selection methods. Note that, for the sake of comparison, last row shows the results achieved using the whole set of features, i.e., no feature selection is applied. To facilitate the analysis of the results, in both tables the results corresponding to the best values (statistically speaking) of each performance measures for each dataset are marked in boldface type.

Firstly we focus on Breast dataset whose results are shown in Table 3. At first glance it seems that for all cases the results obtained by SVDD classifier class are better than those achieved by the SVM and statistical tests confirm this assumption. Only for FCBF and INT filters the SVM obtain a higher value in the specificity measure, however in both cases SVDD achieves the best value of accuracy and specificity and also balanced values for sensitivity and specificity. Furthermore, it is worth mentioning the importance of using feature selection methods (see last row in Table 3) because they help prevent overfitting.

Consider now the Prostate dataset whose results are shown in Table 4. As it can be observed, the results obtained by SVM and SVDD on this dataset follows along the same line as the previous one. The one-class approach overcomes the results obtained by SVM, showing important differences in all cases. As it was stated earlier, Prostate dataset suffers the dataset shift problem since the test distribution differs significantly from the train distribution. In this situation it is possible that classifiers assign the vast majority of samples to one of the classes such it shows for SVM. However, SVDD seems not to suffer this problem and very good results are reached. Although the results obtained when no feature selection is applied are good, it should take into account the important needs both computational and time to manage the original datasets.

Finally, another point to be borne in mind is that SVDD presents an important advantage respect to SVM. Although in provided tables the results are not statistical different, SVDD allows us to use minority or majority class as the target class in the training process and remain the best results depending on the specific application. The ideal situation would be obtain a classifier 100% sensitive and 100% specific but this fact is not easy. Therefore, a trade-off becomes

Table 3. Results for **SVM** and **SVDD** classifiers on Breast dataset.

	Acc			Se			Sp		
	SVM	OCCmin	OCCmaj	SVM	OCCmin	OCCmaj	SVM	OCCmin	OCCmaj
CFS	0,5295	**0,6116**	**0,6167**	0,3352	**0,4756**	**0,4772**	0,6428	**0,8448**	**0,8476**
FCBF	0,6109	**0,6937**	**0,6940**	0,0714	**0,7400**	**0,7067**	**0,9256**	0,6667	0,6724
INT	0,5846	**0,7091**	**0,7116**	0,1400	**0,7294**	**0,7400**	**0,8439**	0,6743	0,6629
IG-10	0,5284	**0,6733**	**0,6726**	0,4886	**0,6411**	**0,6450**	0,5517	**0,7286**	**0,7200**
IG-50	0,5435	**0,7329**	**0,7442**	0,3762	**0,7956**	**0.8128**	0,6411	0,6257	0,6267
RelieF-10	0,4958	**0,7905**	**0,7898**	0,5533	**0,7500**	**0,7500**	0,4622	**0,8600**	**0,8581**
RelieF-50	0,4705	**0,7351**	**0,7337**	0,5267	**0,6811**	**0,6739**	0,4378	**0,8276**	**0,8362**
SVM-REF-10	0,4944	**0,8849**	**0,8881**	0,5790	**0,8944**	**0,9006**	0,4450	**0,8686**	**0,8667**
SVM-REF-50	0,4821	**0,8351**	**0,8403**	0,6638	**0,8333**	**0,8450**	0,3751	**0,8381**	**0,8324**
mRMR-10	0,4944	**0,7614**	**0,7467**	0,4981	**0,7711**	**0,7556**	0,4922	**0,7448**	**0,7314**
mRMR-50	0,5151	**0,7586**	**0,7617**	0,4105	**0,8011**	**0,8028**	0,5761	**0,6857**	**0,6914**
no FS	0,4979	**0,6432**	**0,6358**	**0,5505**	0,4806	0,4705	0,4672	**0,9219**	**0,9190**

OCCmin corresponds to the test results obtained by training with minority class
OCCmaj corresponds to the test results obtained by training with majority class

Table 4. Results for **SVM** and **SVDD** classifiers on Prostate dataset.

	Acc			Se			Sp		
	SVM	OCCmin	OCCmaj	SVM	OCCmin	OCCmaj	SVM	OCCmin	OCCmaj
CFS	0,5909	**0,9747**	**0,9745**	0,3029	**0,9656**	**0,9653**	0,6947	**1,0000**	**1,0000**
FCBF	0,6216	**0,9245**	**0,9225**	0,1681	**0,9011**	**0,8989**	0,7848	**0,9896**	**0,9881**
INT	0,6549	**0,9590**	**0,9571**	0,1259	**0,9443**	**0,9416**	0,8453	**1,0000**	**1,0000**
IG-10	0,5976	**0,9508**	**0,93925**	0,2844	**0,9416**	**0,9269**	0,7104	**0,9763**	**0,9733**
IG-50	0,6470	**0,9933**	**0,9904**	0,2185	**0,9909**	**0,9869**	0,8013	**1,0000**	**1,0000**
RelieF-10	0,6141	**0,9309**	**0,9284**	0,2563	**0,9061**	**0,9067**	0,7429	**1,0000**	**1,0000**
RelieF-50	0,6937	**0,9571**	**0,9584**	0,1437	**0,9416**	**0,9437**	0,8917	**1,0000**	**0,9992**
SVM-REF-10	0,6369	**0,8696**	**0,8718**	0,2733	**0,8589**	**0,8592**	0,7677	**0,8993**	**0,9067**
SVM-REF-50	0,6278	**0,9598**	**0,9602**	0,2229	**0,9453**	**0,9459**	0,7736	**1,0000**	**1,0000**
mRMR-10	0,6384	**0,9486**	**0,9594**	0,2193	**0,9389**	**0,9496**	0,7893	**0,9756**	**0,9867**
mRMR-50	0,6153	**0,9363**	**0,9369**	0,2178	**0,9133**	**0,9141**	0,7584	**1,0000**	**1,0000**
no FS	0,5400	**0,9012**	**0,8976**	0,3348	**0,8656**	**0,8608**	0,6139	**1,0000**	**1,0000**

OCCmin corresponds to the test results obtained by training with minority class
OCCmaj corresponds to the test results obtained by training with majority class

a good option. For example, in case of cancer diagnosis purpose should take into account that a low value of Sensitivity (cancer patients who are correctly identified as suffering the disease) is more critic than a low value of specificity (healthy patients who are correctly classify as not suffering the disease).

5 Conclusions

Microarray data classification is a difficult challenge for learning systems due to its intrinsic characteristics. Machine learning has predominantly been employed for this kind of data and two-class classification techniques have been widely applied. Recent research indicates that the one-class approach is suitable to handle this kind of data because of it relies only on object coming from single class distribution. Despite of having less information to distinguish between classes, one-class models can easily learn the specific properties of a given dataset and are more robust to intrinsic difficulties of the data. In this paper, we demonstrate the suitability of applying one-class learning to handle microarray datasets. We made an experimental study to analyze and compare the behavior of one and two class classifiers, SVDD and SVM respectively, on two microarray datasets. At the same time the effect of applying feature selection techniques is considered, denoting its importance to reduce overfitting. The experimental results allow us to prove the superiority of the one-class classification. Therefore, we can confirm that one-class approach is a good technique to handle this kind of data offering a fine global performance and a good trade-off between sensitivity and specificity measures. Moreover, it offers the possibility of selecting one of the two available class as target concept in the learning process and remain the best results depending on the specific application. As lines of future research, we will conduct a study that includes other feature selection methods, since the tendency is toward focusing on new combinations (such as hybrid or ensemble methods). Moreover, we will incorporate both new microarray datasets as other one-class classification methods.

References

1. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286**(5439), 531–537 (1999)
2. Hughes, G.: On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory **14**(1), 55–63 (1968)
3. Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A.: Feature Extraction: Foundations and Applications. Studies in Fuzziness and Soft Computing. Springer-Verlag New York, Inc. (2006)
4. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J.M., Herrera, F.: A review of microarray datasets and applied feature selection methods. Information Sciences **282**, 111–135 (2014)

5. Valafar, F.: Pattern recognition techniques in microarray data analysis: a survey. Annals of the NewYork Academy of Sciences **980**, 41–64 (2002)
6. Larrañaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J.A., Armañanzas, R., Santafé, G., Pérez, A., Robles, V.: Machine learning in bioinformatics. Briefings in Bioinformatics **7**(1), 86–112 (2006)
7. Yip, W.K., Amin, S.B., Li, C.: A survey of classification techniques for microarray data analysis. In: Handbook of Statistical Bioinformatics. Springer Handbooks of Computational Statistics, pp. 193–223 (2011)
8. Krawczyk, B.: Combining one-class support vector machines for microarray classification. In: Proc. Federated Conference on Computer Science and Information Systems (FedCSIS 2013), pp. 83–89 (2013)
9. Tax, D.M.J., Duin, R.P.W.: Support vector domain description. Pattern Recognition Letters **20**(11), 1191–1199 (1999)
10. Vapnik, V.: Statistical Learning Theory. Wiley (1998)
11. Tax, D.M.J., Duin, R.P.W.: Support vector data description. Machine Learning **54**, 45–66 (2004)
12. Tax, D.M.J.: DDtools, the data description toolbox for matlab. Delft University of Technology (2005)
13. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD Explorations Newsletter **11**(1), 10–18 (2009)
14. Kent Ridge Bio-Medical Dataset. http://datam.i2r.a-star.edu.sg/datasets/krbd (online; accessed January 2015)
15. Microarray Cancers, Plymouth University. http://www.tech.plym.ac.uk/spmc/links/bioinformatics/microarray/microarray_cancers.html (online; accessed January 2015)
16. Moreno-Torres, J.G., Raeder, T., Alaiz-RodríGuez, R., Chawla, N.V., Herrera, F.: A Unifying View on Dataset Shift in Classification. Pattern Recognition **45**(1), 521–530 (2012)
17. Hall, M.: Correlation-Based Feature Selection for Machine Learning, PhD. Thesis (1999)
18. Yu, L., Liu, H.: Feature selection for high-dimensional data: A fast correlation-based filter solution, pp. 856–863 (2003)
19. Zhao, Z., Liu, H.: Searching for interacting features. In: Proceedings of the International Joint Conference on Artifical Intelligence, pp. 1156–1161 (2007)
20. Hall, M., Smith, L.: Practical feature subset selection for machine learning. Computer Science **98**, 181–191 (1998)
21. Kononenko, I.: Estimating attributes: analysis and extensions of RELIEF. In: Bergadano, Francesco, De Raedt, Luc (eds.) ECML 1994. LNCS, vol. 784, pp. 171–182. Springer, Heidelberg (1994)
22. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**, 1226–1238 (2005)
23. Guyon, I., Weston, J., Barnhill, S., Vapnik, V., Cristianini, N.: Gene selection for cancer classification using support vector machines. Machine Learning **46**(1–3), 389–422 (2002)