International Journal of Hybrid Information Technology
Vol.8, No.11 (2015), pp.235-248
http://dx.doi.org/10.14257/ijhit.2015.8.11.20

Optimization of the Smoothing Parameter of the Variable Kernel Estimator used in Bayes Classifier - Application to Microarray Data Analysis

Yissam Lakhdar[1] and El Hassan Sbai[2]

[1]*Department of Physics, Faculty of Science, University Moulay Ismail Meknes, Morocco*
[2]*Ecole Supérieure de Technologie, University Moulay Ismail BP 3103 Route d'Agouray, 50006 Toulal, Meknes, Morocco*
[1]*lakhdar_yissam@yahoo.fr*, [2]*sbaielhassan@yahoo.fr*

Abstract

The estimation of probability density function (pdf) by the nonparametric kernel methods requires a reliable estimate of the bandwidth. There have been several studies on how to efficiently estimate this parameter. In this work, we propose a new optimization method of the smoothing parameter of the variable kernel estimator (VKE) based on the statistical properties of the probability distributions of random variables. In this setting, we show how to use the maximum entropy principle for estimating the smoothing parameter. The optimized estimator is after used in building the Bayesian classifier. In the same setting, the estimated probability density function is called optimal in the sense of having a minimum error rate of classifying data. Finally, a practical implementation with the aid of a dataset of DNA microarray is used to illustrate the behavior of the optimization technique.

Keywords: *variable kernel estimator, smoothing parameter, probability density function, maximum entropy principle, microarray data, bayes classifier*

1. Introduction

Data classification is intended to form groups or classes of homogeneous objects in order to structure a set of data. The development of methods that attempt to classify the data has led to distinguish two types of classification strategies: supervised and unsupervised. Supervised classification seeks to classify a new data, based on the characteristics of a set of pre-classified data in the classes known in advance while the clustering has been designed to group that data without a priori knowledge of the group structure [1-4]. In this work, we are interested in learning methods based on statistical techniques that introduce the concept of non-parametric estimation of the probability density function (pdf). We concentrate on the nonparametric kernel estimator particularly the variable kernel estimator (VKE). This exploratory statistical approach consists to structure the information contained in the data. However, many problems affect the quality of these methods such as the optimal choice of the bandwidth of the estimator. While it is advantageous to apply small bandwidth to densely populated region of the data space, larger bandwidths may be appropriate for sparsely populated regions. As a result, many studies have proposed techniques for selecting the smoothing parameter such as the family of cross validation methods and the family of plug-in methods. In this sense, this paper presents a method for optimizing the smoothing parameter of the variable kernel estimator using the maximum entropy principle [5-8]. The optimized estimator is after used to construct the Bayes classifier. In the proposed criterion, the estimated pdf is called optimal in the sense of having maximal classification accuracy.

ISSN: 1738-9968 IJHIT

Cancer classification is one of the challenging studies for research in the last century. In the literature, several researchers have attempted to provide detailed technical on the classification of different tumor types. These techniques have limitations in their analysis. In recent years, research studies have attempted to address these limitations.

The statistical and machine learning area is not new to cancer research. Today different classification approaches based on gene expression analysis have been proposed [9, 10].

We note that some researchers believe that the genes selection is an important step which must be applied before the classification step. Also, for most of those proposed methods on gene classification, the authors are only concerned with the accuracy of the classification and did not pay much attention to the running time. However, there are major challenges regarding the nature and the structure of the gene expression which we should note the highly large size of the data (thousands to tens of thousands of genes) versus the small number of patients. It is noteworthy that we have evaluated the performance of our cancer classification method in one aspect which is the classification accuracy.

The paper is organized as follows. In the second section, we review the variable kernel estimation method which combines the properties of both techniques; the k-nearest neighbor approach and the Parzen approach to obtain a method that works well in various situations. In the third section, we present a new approach to optimize the smoothing parameter of the variable kernel estimator, based on the maximum entropy principle (MEP). Then, we describe the use of the variable kernel estimator in the construction of the Bayes classifier. In the fourth and last section, we illustrate the efficiency of our algorithm to optimize the bandwidth of variable kernel estimator using a set of DNA microarray gene expression patterns for a certain number of patients. Finally, we provide a conclusion and future prospects.

2. Variable Kernel Estimator

The mathematical expression of the Bayes classifier takes into consideration the probability densities that are not precisely known. This explains why it is necessary to estimate these densities from the available data. In this section, we review the variable kernel estimator (VKE) of the probability density function from a data set of size n in an Euclidean space of dimension d.

The variable kernel estimator is an estimator which combines the k-nearest neighbors estimator and the Parzen estimator where the scale parameter of the bumps placed on the data points are allowed to vary from data point to the other [11-13].

The k-nearest neighbors (kNN) estimator is defined as follows [13]:

$$\hat{f}_{knn}(x) = \frac{k/n}{V_k(x)} = \frac{k/n}{c_d\, r_k(x)} \quad (1)$$

Let k be a positive integer, $r_k(x)$ is the distance from x to the kth nearest point and $V_k(x)$ is the volume of a sphere of radius $r_k(x)$ and c_d is the volume of the unit sphere in d dimensions.

The degree of smoothing of this estimator is controlled by the parameter k, chosen to be very smaller than the sample size.

The Parzen-Rosenblatt estimator with kernel K_d is defined by the following expression [14, 15]:

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K_d\left(\frac{D(x,X_i)}{h}\right) \quad (2)$$

where h is the smoothing parameter also called Bandwidth and $D(x, X_i)$ is the distance between sample x and X_i.

The VKE is constructed similary to the classical kernel estimator. It is defined by [13]:

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} \frac{1}{(r_{i,k})^d} K_d\left(\frac{D(x,X_i)}{hr_{i,k}}\right) \quad (3)$$

with $r_{i,k}$ is a distance between a data point X_i and the kth nearest point of the other $n-1$ data points.

We note that:

- if we fix k, the overall degree of smoothing depends on the discretization step h;
- the window width does not depend on the point x where we want to estimate the density but depends only on the distances between the data points.

It is worth noting that in this manuscript we used the Euclidean and the Mahalanobis distances.

3. Bandwidth Selection Method

In this work, we optimize the smoothing parameter of the variable kernel estimator by a criterion introduced by Jayens Edwin called the maximum entropy principle [6, 7].

3.1. Entropy

Claude Shannon introduced the concept of the entropy associated with a discrete random variable X as a basic concept in information theory [8]. In fact, Shannon introduced the quantity $-\log(p(X_i))$ as a measure of the information brought by the realization x_i knowing the probability distribution p. Shannon entropy $S(p)$, thus, appears as the average missing information which is the weighted sum of the quantity of information provided by the realization of this random variable.

Let the distribution of probabilities $p = \{p_1, p_2, \dots p_n\}$ associated with the realizations of X. The Shannon entropy is calculated using formula:

$$S(p) = -\sum_{i=1}^{n} p(X_i) \log(p(X_i)) \quad (4)$$

Entropy measures the uncertainty associated with a random variable. Therefore the realization of a rare event provides more information about the phenomenon that the realization of a frequent event.

3.2. Maximum Entropy Principle

The maximum entropy principle (MEP) means that the probability to choose among a set of possible laws in statistical modeling should be one that maximizes the Shannon entropy. Therefore, the best estimate of the probability density function is one that has a maximal entropy [5-7, 16, 17]. The optimal probability density function can be defined as following:

$$S(\hat{f}^*) = \max\{S(\hat{f}, h)\} \quad (5)$$

with $S(\hat{f}, h)$ is the Shannon entropy defined as:

$$S(\hat{f}, h) = -\sum_{i=1}^{n} \hat{f}(X_i, h) \log\left(\hat{f}(X_i, h)\right) \quad (6)$$

subject to the constraints $\sum_{i=1}^{n} \hat{f}(X_i, h) = 1$ with n is the number of observations and $\hat{f}$ is the estimated pdf.

3.3. Criterion based on the Maximum Entropy Principle

In this work, we propose a new approach to optimize the bandwidth of the variable kernel estimator, it is noteworthy that the choice of the bandwidth is critical while the type of the kernel has little effect on the result. For this we normalize the estimated values of the pdf and search the best estimate of the underlying pdf to the distribution of observations corresponding to the one that maximizes the entropy [5-7, 12, 16, 18-20]. The obtained estimator is called optimized variable kernel estimator (OVKE). The proposed algorithm is summarized as following:

Input:

X: The input data;

Parameters: k: Nearest neighbours, h: Smoothing Parameter;
Initialization: Choose the initial value of bandwidth to zero ($h = 0$);
Gene Selection
While (The entropy S is not maximum)
Begin

- Incrementation of h with a small step (step = 0.01);
- Estimation of the pdf with the relation (3);
- Normalization of the estimated density function $\hat{f}$ which will be noted $\hat{g}(X_i, h)$;
- Calculation the entropy of the normalized density $\hat{g}(X_i, h)$ with the relation (6);

end While

3.4. Choice of Parameter K of Variable Kernel Estimator

Another major issue in our variable kernel estimator is how to choose the optimum value of the parameter *k* (number of nearest neighbors). The choice of *k* is critical in the estimation of the density and depends on data. A small value of *k* increases the influence of noise on the classification result while a high value reduces the effect of noise. The best choice of *k* may be selected by various techniques such as cross-validation techniques and heuristic techniques. In this article, the validation indices (CH and DB) [21, 22] are used as methods to determine the value of k.

It should be noted that we will estimate the probability density function for each class using the optimized variable kernel estimator. Therefore, we will require to choose two values of k, k_{c_1} and k_{c_2}, because the datasets used in our article are divided into two classes.

4. Experimental Results

We illustrate our optimization approach of smoothing parameter of the variable kernel estimator using two real datasets both of which are composed of a matrix of gene expression obtained from the DNA microarrays: Leukemia and Colon datasets. Both distributions are divided into two classes. The leukemia dataset is divided into two subsets, acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). We use these 72 data samples in order to compare our approach with previously published results. This distribution of data is divided into two subsets, the training set and the test set. All training data consists of 38 bone marrow samples with 27 patients with ALL and 11 patients with AML. In addition to this, 34 test samples are provided with 20 cases of ALL and 14 cases of AML. All data samples have 7129 genes, corresponding to some gene expression values extracted from the microarray image [23]. The Colon dataset consists of 62 tissues which include 22 normal and 40 colon cancer tissues. Each sample contains 2000 gene expression levels, some genes are non-human genes. We split randomly the data into 31 tissues for training and 31 tissues for testing.

We note that the OVKE is used to construct the Bayes classifier which is defined as following [24]:

$$P(C = C_i | X = X_j) = (\tilde{f}_{C_i}(X_j)\tilde{p}_{C_i}) / \sum_{i=1}^{m} \tilde{f}_{C_i}(X_j)\tilde{p}_{C_i} \quad (7)$$

where $\tilde{p}_{C_i}$ is the prior probability of the class C_i and $\tilde{f}_{C_i}(X_j)$ is the estimated probability using the optimized variable kernel estimator at the point X_j.

The most probable class is assigned to X_j using the assumption of the maximum a posteriori (MAP):

$$C_{aff} = C_{MAP} = \text{argmax}_{C_i \in C} P(C = C_i | X = X_j) \quad (8)$$

It should be noted that our methodology consists of two phases. The first phase uses variable selection techniques because the microarray data are characterized by high dimensionality and small sample size, the number of genes may range from 2000 to

60,000 per respect to a limited number of patients. However, the expression data is highly redundant and most genes are considered less significant; it requires providing innovative pre-processing techniques to reduce the dimension of the microarray data and select the most relevant and most informative genes for improving the efficiency of the classification process. The hybrid selection filters used are the information gain (IG), the reliefF filter and the minimum redundancy maximum relevance filter (mRMR) [25, 26]. The result of this pretreatment process is used in the second phase to find the optimal value of the smoothing parameter h_{opt} of the variable kernel estimator.

4.1. Leukemia Dataset

The results of selected genes through our pre-treatment process are presented in table 1. For each gene is given its place in our initial phase of selection, its identification number (ID) in the leukemia dataset, and its index [23, 27-36].

Table 1. Genes Ranked for Leukemia Dataset

Rank	Gene ID	Index
1	X95735	4847
2	X17042	4196
3	M23197	1834
4	L09209s	6041
5	U46499	3252
6	M27891	1882
7	M16038	1745
8	M22960	1829
9	M63138	2121
10	M55150	2020
11	M62762	2111
12	U50136	3320
13	X61587	4366
14	M32304	6005
15	X52056	4229
16	D49950	461
17	X59417	4328
18	M31211	6281
19	M92287	2354
20	X64072	6185
21	L09717	1260
22	M31523	6855

The representation of 1882 and 6855 genes is shown in Figure 1, The optimal values of the bandwidth for our genes 1882 and 6855 are $h_{opt} = 15.59$ for $\tilde{f}_{C_1}$ and $h_{opt} = 1.12$ for $\tilde{f}_{C_2}$, see Figures 2 and 3.

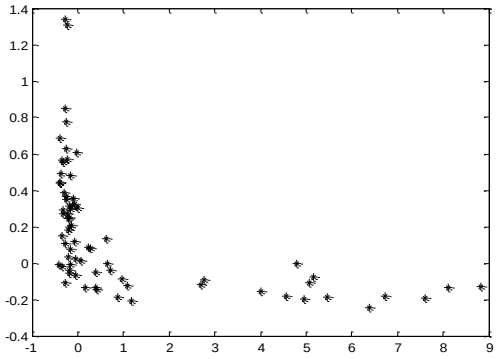

Figure 1. The Data Sample with the Two Genes 1882 and 6855

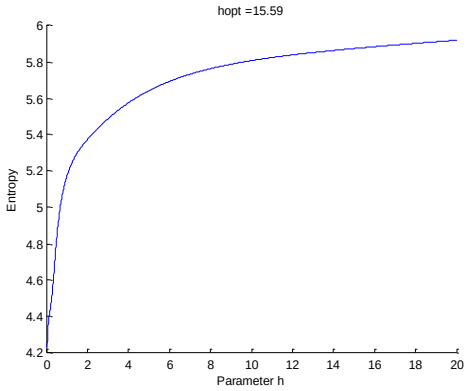

Figure 2. The Evolution of Entropy with $\mathbf{k_{C_1}}$ = 3 using VKE

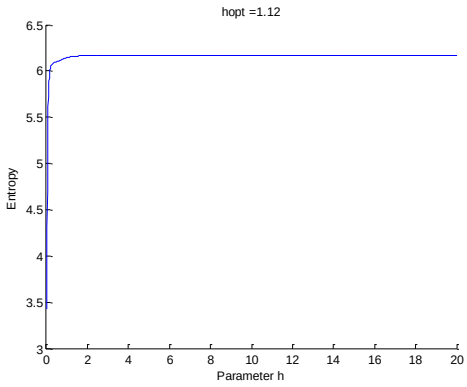

Figure 3. The Evolution of Entropy with $\mathbf{k_{C_2}}$ = 8 using VKE

They show the evolution of the entropy of the estimated density function by the variable kernel estimator (VKE) which is a function of the smoothing parameter h. The entropy is calculated after applying the normalization technique of the estimated values from the probability density function. It is noted by using with the normalization that the value of the entropy increases regularly and it then remained stable. Indeed, from a certain parameter h, $h_{opt} = 15.59$ for $\tilde{f}_{C_1}$ and $h_{opt} = 1.12$ for $\tilde{f}_{C_2}$, called optimal smoothing parameter, all relevant information is extracted.

It is worth mentioning that we validate our classification results by the validation indices CH and DB for different values of parameter k since the choice of this parameter is crucial and it has an influence on the classification accuracy. The figures 4 and 5 show

the evolution of the CH and DB indices with the parameter k. It is also important to point out that we have chosen two values of k namely k_{C_1} and k_{C_2} because the datasets are distributed into two classes. Figure 6 illustrates the classification result with OVKE using Bayes classifier.

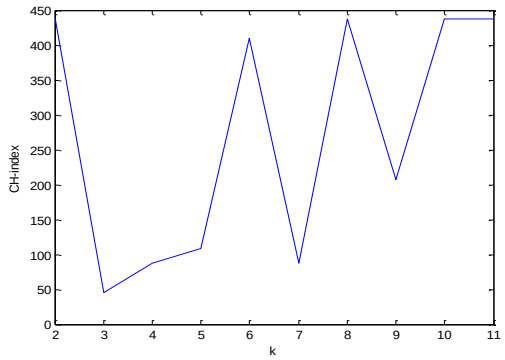

Figure 4. The Evolution of the CH-index with the Parameter $k_{C_1} = 3$ using VKE

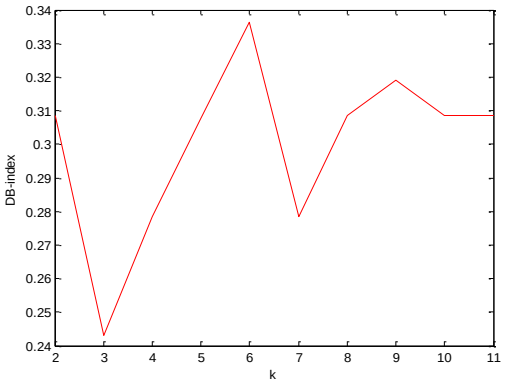

Figure 5. The Evolution of the DB-index with the Parameter $k_{C_1} = 3$ using VKE

The classification rates (CR) of our optimization process of the smoothing parameter determined by the maximum entropy principle (see tables 2 and 3), are compared with the classification rate obtained with the Parzen, k-NN, weighted voting method (WVM), associative artificial neural network (AANN) and other works [23, 27, 28, 31-35, 37].

The validation of the results by the CH and the DB indices gives the following results. For 6041 and 3252 genes, the application of B-OVKE algorithm with Euclidean and the Mahalanobis distances give the same classification rate, which is equal to 97.23%. The latter value may be considered the best relative to the other classification methods. However, the use of B-OParzen algorithm with the Euclidean distance give a good classification rate that is equal to 93.06%, while the B-Parzen and B-k-NN algorithms provide a similar result, which is equal to 91%. This value is lower than that found with other algorithms. In the case of 6855 and 4847 genes, we get 97.23% performance by applying the B-OVKE technique. The accuracy rate is better than the rate of B-k-NN classifiers and B-OParzen. The classification rate of genes 1882 and 6855 using a B-OVKE is 91.67% with the Euclidean distance and 94.45% with the Mahalanobis distance respectively. These values are less efficient than B-OParzen method which is equal to 97.23%. Likewise for 6041 and 6855 genes, we get a classification rate equal to 91.67%

by using the B-OVKE method with the Euclidean distance and 93.06% with the Mahalanobis distance. We conclude that the classification rate obtained by the OVKE varies between 91.67 % and 97.23% for the test samples. The achieved results, presented in Table 2 and 3, show the usefulness of the proposed methodology and we note that the classification performances are improved.

Table 2. The Comparison of Classification Methods for the Leukemia Data

Method (Work)	CR (Test)	Method (Work)	CR (Test)
SVM	**94.12%**	**Tang**	**100.00%**
WVM	**85.29%**	**Cho**	**95.90%**
AANN	**65.00%**	**Ding**	**100.00%**
Wang	**95.80%**	**Marohnic**	**100.00%**

Table 3. Comparative Results on the Data Sample

G 1	G 2	Algorithms	Dist	CR (Test) in %	CR (Training) in %
1882	6855	B-k-NN[a]	ED[c]	91	100
		B-OParzen[b]	ED	97.23	94.45
		B-OVKE	ED	91.67	90.28
			MD[d]	94.45	94.45
6041	3252	B-k-NN[a]	ED	91	95
		B-Parzen[a]	ED	91	97
		B-OParzen[b]	ED	93.06	90.28
		B-OVKE	ED	97.23	90.28
			MD	97.23	95.84
6041	6855	B-k-NN[a]	ED	94	97
		B-Parzen[a]	ED	91	100
		B-OParzen[b]	ED	91.67	90.28
		B-OVKE	ED	91.67	91.67
			MD	93.06	91.67
6855	4847	B-k-NN[a]	ED	91	100
		B-Parzen[a]	ED	91	100
		B-OParzen[b]	ED	90.28	93.06
		B-OVKE	ED	97.23	95.84
			MD	97.23	98.62

[a] Peters and Valafar method [11],
[b] The optimized Parzen Estimator by MEP,
[c] Euclidean distance,
[d] Mahalanobis distance.

4.2. Colon Dataset

Tables 4, 5 and 6 present the filtered genes by the selection process using the information gain, the mRMR and the ReliefF filters respectively.

Table 4. Genes Ranked with IG for Colon Dataset

Rank	Gene ID	Index
1	H86060	12
2	H80240	8
3	R16255	248
4	L11369	650
5	H80240	25
6	M96839	523
7	U14973	5
8	H80240	25
9	L34774	551
10	H55933	1
11	M99626	146
12	T63508	18
13	H85596	97

Table 5. Genes Ranked with mRMR for Colon Dataset

Rank	Gene ID	Index
1	T47377	1325
2	X54942	1730
3	M19045	317
4	R36977	1042
5	M22382	513
6	M26383	1671
7	H40095	780
8	X63629	1582
9	T86473	964
10	U26312	1406
11	R08183	1002
12	U17899	1770
13	X56597	1900

Table 6. Genes Ranked with reliefF for Colon Dataset

Rank	Gene ID	Index
1	M80815	1058
2	M63239	1985
3	H71150	673
4	M76378	765
5	R80427	1644
6	M76378	245
7	M76378	267
8	L07648	1873
9	T55117	1087
10	M63391	249
11	R39209	1328
12	Z50753	377
13	H20543	1098

The representation of 780 and 377 genes is shown in Figure 7, while Figures 8 and 9 show the evolution of the entropy as a function of the bandwidth, we also remark that entropy increases progressively up to a maximum value that is equal to our genes $h_{opt} = 1.29$ for $\tilde{f}_{C_1}$ and $h_{opt} = 2.06$ for $\tilde{f}_{C_2}$ and it remains stable thereafter. Figure 12 illustrates the classification results using the Bayes classifier with the optimized variable kernel estimator (OVKE). These results are obtained for $k_{C_1} = 11$ and $k_{C_2} = 17$ with OVKE. The figures 10 and 11 show the evolution of the CH and DB indices with the parameter k.

The classification rates (CR) obtained with the OVKE are shown in Table 7. The application of Bayes classifier with optimized variable kernel estimator by the maximum entropy principle give better classification rate when compared with other research work. The Tables 7 and 8 show the following results using the validation indices.

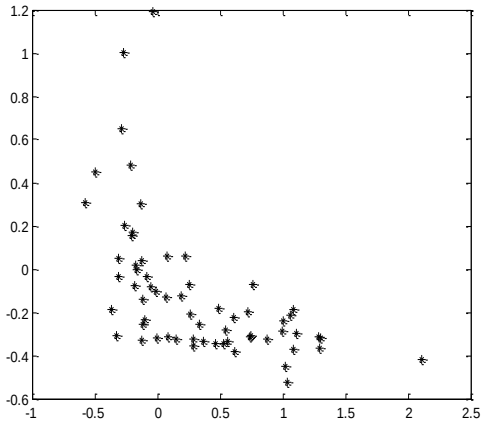

Figure 7. The Data Sample with the Two Genes 780 and 377

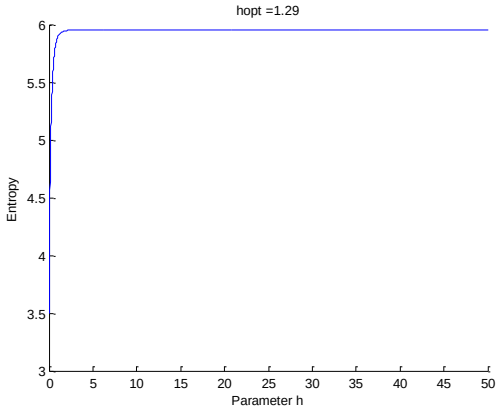

Figure 8. The Evolution of Entropy with $\mathbf{k_{C_1}}$ = 11 using VKE

Table 7. Comparative Results on the Data Sample

G 1	G 2	Algorithms	Dist	CR (Test) in %	CR (Trai-ning) in %
780	765	B-OParzen	ED	85.49	93.55
		B-OVKE	ED	85.49	93.55
249	377	B-OParzen	ED	90.33	88.71
		B-OVKE	ED	91.93	93.55
780	377	B-OParzen	ED	88.71	85.49
		B-OVKE	ED	90.33	88.71
765	1582	B-OParzen	ED	88.71	93.55
		B-OVKE	ED	85.49	93.55
765	249	B-OParzen	ED	83.88	87.10
		B-OVKE	ED	88.71	93.55
249	1582	B-OParzen	ED	90.33	88.71
		B-OVKE	ED	91.93	93.55
780	1582	B-OParzen	ED	83.88	80.65
		B-OVKE	ED	87.10	80.65

Table 8. The Comparison of Classification Methods for the Colon Data

Work	CR (Test)
Hernandez [38]	84.6%
Zhang [39]	90.3%
Li [40]	83.5%
Wang [41]	93.5%
Li [42]	93.6%
Cho & Won [29]	87.7%

The classification rate obtained by implementation of the OVKE method to test data sets is ranging from 85.49% to 91.93%. In the case of 780 and 1582 genes we get 87.10% performance by applying the B-OVKE with the Euclidean distance, this rate is better than the rates obtained by B-OParzen method. The classification rate of genes 780 and 377 using a B-OVKE is 90.33%. For 780 and 765 genes, we find that the application of both algorithms B-OVKE and the B-OParzen with Euclidean distance give the same classification rate equal to 85.49%. For 249 and 377 genes, the implementation of OVKE algorithm gives better classification rate equal to 91.94% by comparing it with the B-OParzen algorithm which is equal to 90.33%. Likewise for 249 and 1582 genes, the B-OVKE algorithms give a upper result equal to 91.94%. The same result is observed for 765 and 249 genes, we find that the B-OVKE algorithm gives a better performance rate compared with B-OParzen method. The classification rate of genes 765 and 1582 using a B-OVKE method is less powerful than the B-OParzen method.

5. Conclusion

In this work, we proposed a new method for optimizing the smoothing parameter of the variable kernel estimator based on the maximum entropy principle. In the majority of cases, this technique gives a minimum error rate in the process of data classification. We evaluate the performance of our method on two datasets in the oncology area. We note that we have evaluated the performance of our classification method in one aspect which is the classification accuracy. The experimental results presented in this paper clearly demonstrate the interest and robustness of the maximum entropy principle. This criterion allows us to obtain optimal bandwidth of the variable kernel estimator.

References

[1] C. Cortes and V. Vapnik, “Support-Vector Networks”, Machine Learning, vol. 20, **(1995)**, pp. 273-297.
[2] E. W. Forgy, “Cluster analysis of multivariate data: efficiency versus interpretability of classifications”. Biometrics, vol. 21, **(1965)**, pp. 768–769.
[3] J. MacQueen, “Some methods for classification and analysis of multivariate observations”. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, California, Statistics, vol. 1, **(1967)**, pp. 281-297.
[4] J. Pearl, “Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning”. Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, CA, **(1985)** August 15-17, pp. 329-334.
[5] M. Bouhafid, H. Qjidaa and L. Radouane,”Optimal discretization of the data space in unsupervised data clustering”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **(1996)**.
[6] M. Jan, C. Van and T. M. Cover, “Maximum entropy and conditional probability”, IEEE Trans. On Information Theory, vol-IT-27, no. 4, **(1981)** July, pp. 483-489.
[7] E. T. Jaynes, “On the rational of maximum entropy methods”. Proceeding of the IEEE, vol. 70, **(1982)**, pp. 939–952.
[8] C. E. Shannon and W. Weaver, “The mathematical theory of communication”. University Of Illinois Press, Urbana, IL, **(1949)**.
[9] C. F. Aliferis, D. Hardin and P. P. Massion, “Machine learning models for lung cancer classification using array comparative genomic hybridization”, Proceedings of the American Medical Informatics Association 2002 Symposium, San Antonio, AMIA, USA, **(2002)** November 9-13, pp. 7-11.

[10] A. Ben-Dor, R. Shamir and Z. Yakhini, "Clustering gene expression patterns", Journal of Computational Biology, vol. 6, **(1999)**, pp. 281-297.
[11] A. P. Christopher and V. Faramarz, "Comparison of three nonparametric density estimation techniques using Bayes' classifiers applied to microarray data analysis", Proceeding of the International Conference on Mathematics and Engineering Techniques in Medecine and Biological Sciences, CSREA Press, Las Vegas, Nevada, **(2003)** June 23-26, pp. 119-125.
[12] Y. Lakhdar and E. Sbai, "Optimization of the Smoothing Parameter of Variable Kernel Estimator", International Conference on Communications, Computing and Control Applications, Marseilles-France, **(2012)** December 06-08.
[13] B. W. Silverman, "Density estimation for statistics and data analysis", Chapman and Hall, London, **(1986)**.
[14] E. Parzen, "On estimation of a probability density function and modes". Annals Mathematical Statistics, vol. 33, **(1962)**, pp. 1065-1076.
[15] M. Rosenblatt, "Remarks On some nonparametric estimates of a density function", Annals Mathematical Statistics, vol. 27, **(1956)**, pp. 832-837.
[16] E. S. John and R. W. Johson, "Axiomatic, derivation of the principle of maximum entropy and the principle minimum cross entropy". IEEE Transactions on Information Theory, vol-IT-26, no. 1, **(1980)** January, pp. 26-37.
[17] E. Sbai, "Classification automatique par les statistiques d'ordre, la réduction des données et l'estimation de la fonction de densité de probabilité par la méthode des moments", Thèse de Doctorat d'Etat, Faculty of Science, University Moulay Ismail, Meknes, Morocco, **(2001)**.
[18] Y. Lakhdar and E. Sbai, "Optimisation du paramètre de lissage des estimateurs à noyau adaptatifs", Colloque International TELECOM'2013 & 8ème JFMMA, Marrakech-Maroc, **(2013)** March 13-15.
[19] Y. Lakhdar and E. Sbai, "The Optimized Adaptive Density Estimation Technique Applied to Microarray Data Analysis", International Conference on Multimedia computing and Systems (ICMCS'14), Marrakech-Maroc, **(2014)** April 14-16.
[20] Y. Lakhdar and E. Sbai, "Optimization of the smoothing parameter of the adaptive kernel estimator used in Bayes classifier - Application to microarray data analysis", International Journal of Software Engineering and Its Applications, vol. 9, no. 3, **(2015)**, pp. 79-94.
[21] R. Calinsky and J. Harabsz, "A dendrite method for cluster analysis". Communications in Statistics, vol. 3, **(1974)**, pp. 1-27.
[22] D. Davies and D. Bouldin, "A cluster separation measure", IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 1, **(1979)**, pp. 224–227.
[23] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield and E. S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring", Science, vol. 286, **(1999)**, pp. 531-537.
[24] A. Pérez, P. Larrañaga and I. Inza, "Bayesian classifiers based on kernel density estimation: Flexible classifiers". International Journal of Approximate Reasoning, vol. 50, no. 2, **(2009)**, pp. 341-362.
[25] K. Kira and L. Rendell, "A practical approach to feature selection", In Proceedings of the 9th International Conference on Machine Learning, Scotland, UK, **(1992)** July 1-3, pp. 249-256.
[26] H. Peng, F. Long and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, **(2005)**, pp. 1226–1238.
[27] S. Bicciato, M. Pandin, G. Didonè and C. Di Bello, "Analysis of an associative memory neural network for pattern identification in gene expression data", In Proceedings of the ACM SIGKDD Workshop on Data Mining in Bioinformatics, San Francisco, USA, **(2001)** August 26-29, pp. 22-30.
[28] S. B. Cho and H. H. Won, "Machine learning in DNA microarray analysis for cancer classification", In Proceedings of the First Asia-Pacific Bioinformatics Conference, Adelaide, Australia, **(2003)**, pp. 189-198.
[29] S. B. Cho and H. H. Won, "Cancer classification using ensemble of neural networks with multiple significant gene subsets", Applied Intelligence, vol. 26, no. 3, **(2007)**, pp. 243–250.
[30] W. Chu, Z. Ghahramani, F. Falciani and D L. Wild, "Biomarker discovery in microarray gene expression data with Gaussian processes", Bioinformatics, vol. 21, **(2005)**, pp. 3385-3393.
[31] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data", J. Bioinformatics and Computational Biology, vol. 3, **(2005)**, pp. 185-205.
[32] V. Marohnic, Z. Debeljak and N. Bogunovic, "Mutual information based reduction of data mining dimensionality in gene expression analysis", In Proceedings of the 26th International Conference on Information Technology Interfaces, vol. 1, pp. 249-254, **(2004)** July 7-10, Cavtat, Croatia.
[33] Y. Tang, *et al.*, "Development of two-stage SVMRFE gene selection strategy for microarray expression data analysis", IEEE/ACM Transactions Comput. Biol. Bioinform., vol. 4, **(2007)**, pp. 365-381.
[34] Y. C. Tang, Y.-Q. Zhang and Z. Huang, "FCM-SVM-RFE gene feature selection algorithm for leukemia classification from microarray gene expression data", In Proceedings of the 14th IEEE International Conference on Fuzzy Systems, pp. 97-101, **(2005)** May 22-25, Reno, USA.

[35] Z. Wang, V. Palade and Y. Xu, "Neuro-fuzzy ensemble approach for microarray cancer gene expression data analysis", In Proceedings of the Second International Symposium on Evolving Fuzzy Systems, pp. 241-246, **(2006)** September 7-9, Lake District, UK.
[36] X. Zhou, *et al.*, "Gene selection using logistic regressions based on Aic, Bic and MDL criteria", New Math. Nat. Comput., vol. 1, **(2005)**, pp. 129-145.
[37] T. S. Furey, N. Duy, N. Cristianini, D. Bednarski, M. Schummer and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data", Bioinformatics, vol. 16, **(2000)**, pp. 906-914.
[38] J. C. Hernandez, B. Duval and J. Hao, "A genetic embedded approach for gene selection and classification", EVOBIO, vol. 4447, **(2007)**, pp. 90–101.
[39] L. Zhang, Z. Li and H. Chen, "An effective gene selection method based on relevance analysis and discernibility matrix", PAKDD, vol. 4426, **(2007)**, pp. 1088–1095.
[40] G. Li, X. Zeng, J. Yang and M. Yang, "Partial least squares based dimension reduction with gene selection for tumor classification", IEEE 7th International Symposium on Bioinformatics and Bioengineering, pp. 1439–1444, **(2007)** October 14-17.
[41] S. Wang, H. Chen, S. Li and D. Zhang, "Feature extraction from tumor gene expression profiles using DCTt and DFT", EPIA Workshops, pp. 485–496, **(2007)** December 3-7.
[42] S. Li, X. Wu and X. Hu, "Gene selection using genetic algorithm and support vectors machines", Soft Computing, vol. 12, no. 7, **(2008)**, pp. 693–698.

Authors

Yissam Lakhdar, he got a master degree on "computer sciences, computer graphics and imagery" from faculty of sciences, Dhar El Mehraz, university Sidi Mohamed Ben Abdallah, Fez, Morocco, in 2011. Since October 2011, he has been working on his Phd thesis at the university of My Ismaïl, Meknes, Morocco, with his research team "Control Systems and Information Processing (CSIP)" at higher school of technology of Meknes. His scientific interests include data analysis, artificial intelligence, cluster analysis, image processing and machine learning.

Sbai El Hassan, received the Doctorate degree (Doctorat de 3ème cycle) from the university of Sidi Mohammed Ben Abdellah, Fez, Morocco in 1992, and the Doctorate of sciences (Doctorat d'Etat) from the university of My Ismaïl, Meknes, Morocco in 2001. Since 1993, he is a Professor in automatics and informatics at higher school of technology of Meknes. His main research interest are cluster analysis, pattern recognition, machine learning and computer vision.