Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.6, 2014

Comparing Prediction Accuracy for Supervised Techniques in Gene Expression Data

Setu Chandra Kar
B.Sc. (Honours) and M. S. Department of Statistics, Shahjalal University of Science & Technology
Sylhet-3114, Bangladesh

Abstract

Classification is one of the most important tasks for different application such as text categorization, tone recognition, image classification, micro-array gene expression, proteins structure predictions, data classification etc. Microarray based gene expression profiling has been emerged as an efficient technique for cancer classification, as well as for diagnosis, prognosis, and treatment purposes. The classification of different tumor types is of great significance in cancer diagnosis and drug innovation. One challenging area in the studies of gene expression data is the classification of different types of tumors into correct classes. Diagonal discriminant analysis, regularized discriminant analysis, support vector machines and k-nearest neighbor have been suggested as among the best methods for small sample size situations. The methods are applied to datasets from four recently published cancer gene expression studies. Four publicly available microarray data sets are Leukemia, Lymphoma, SRBCT & Prostate. The performance of the classification technique has been evaluated according to the percentage of misclassification through hold-out cross validation.

1 Introduction: Microarray technology is a new tool that can automate the diagnostic task and improve the accuracy of the traditional diagnostic techniques. Today, the analysis of gene expression data is one of the major topics in health informatics. With microarrays, it is possible to examine the expression of thousands of genes at once. Accurate prediction of different tumor types has great value in providing better treatment and toxicity minimization on the patients. The gene expression profiles that are obtained from particular microarray experiments have been widely used for cancer classification to build an effective model. This model can differentiate normal or different cancerous states by using selected informative genes (Chuang, 2011). However, the curse of dimensionality, the small number of samples, and the level of irrelevant and noise genes make the classification task of a test sample more challenging (Ghorai, 2010). In the past several decades, many dimension reduction techniques have been proposed. Roughly, these methods follow two approaches, feature selection and feature extraction (Li et al., 2005). Several gene selection methods have been developed to select these predictive genes, such as t-statistics, information gain, towing rule, the ratio of between-groups to within-groups sum of squares (BSS/WSS), Principal Component Analysis (PCA), and Genetic Algorithm (GA). Different classification methods from statistical and machine learning area have been applied to cancer classification such as Fisher Linear Discrimination Analysis (FLDA), Diagonal Discriminant Analysis (DLDA), Regularized Discriminant Analysis (RLDA), Maximum Likelihood Discriminant Rules, Classification Tree, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and the aggregating classifiers.

Following are the objectives of our study:

- To examine the classification performance of the selected supervised algorithms using delete-d method;
- To reveal the effect of gene selection on different classification methods;
- To assess the performance of different classification methods after dimension reduction.

2 Overview of molecular biology

Every cell in an organism contains a full set of chromosomes and identical genes. At a given point of time, only a subset of these genes is active. These genes define certain unique properties of a cell type. The information contained in the DNA is transcribed into messenger RNA (mRNA) molecules, which are then translated into proteins, which perform most of the important functions of the cell. The cells are the smallest units of life. The self producing, structural and functional unit matter of protoplasm which is surrounded by a living membrane is known as cell. They provide structure for the body and contain the body's hereditary material.

A DNA or Deoxyribonucleic acid molecule is a double-stranded polymer composed of four basic molecular units called nucleotides. It comprises of five carbon Deoxyribose sugar, inorganic phosphate and one of the four nitrogen bases. The nitrogen bases are adenine (A), guanine (G), cytosine(C) and thymine (T). Adenine and Guanine are called purin base (double ringed) and Thymine and Cytosine are called the pyrimedine base (single ringed). Ribonucleic Acid (RNA) is similar to DNA in the fact that it is constructed from nucleotides. However, instead of thymine (T), an alternative base uracil (U) is found in RNA. The self-reproducing thread like structure which is made of nucleic acid and protein from nuclear reticulum and carries the hereditary characteristics to the successive generations, playing a vital role in mutation, variation and evolution is known as chromosome.

The functional unit of chromosome situated in a particular locus in which recombination may occur and can take part in mutation is called gene. Therefore, any single base-pair of DNA is called gene. A gene is the basic physical and functional unit of heredity. cDNA is single-stranded DNA that is complementary to a certain sequence of messenger RNA. It is usually formed in a laboratory by the action of the enzyme reverse transcriptase on a messenger RNA template. Complementary DNA is a popular tool for molecular hybridization or cloning studies. Gene expression is the process of transforming genetic information into proteins which then serve biological function. In a first step, the DNA template of a gene is transcribed into mRNA using an enzyme called RNA polymerase. The second step includes translating the mRNA sequence into a protein. This two-step process is also known as the fundamental dogma of molecular biology.

2.1 Microarray

The microarray technology was developed at the beginning of the last decade by Mark Schena and co-workers at Stanford University and has now become a widely used research tool. A microarray is a tool for measuring the expression levels of genes (mRNA) in a tissue sample. This is done by hybridizing the sample to the spots on the array and then measuring the amount of hybridized material, where each spot is associated with a unique gene. There are many types of microarrays, e.g. Affymetrix, Agilent, spotted cDNA (complementary DNA) arrays and spotted oligo arrays. Although they are constructed in different ways, the general idea behind them is the same.

Microarrays are microscope slides that contain an ordered series of samples (DNA, RNA, protein, tissue). The type of microarray depends upon the material placed onto the slide: DNA, DNA microarray; RNA, RNA microarray; protein, protein microarray; tissue, tissue microarray. Since the samples are arranged in an ordered fashion, data obtained from the microarray can be traced back to any of the samples. This means that genes on the microarray are addressable. The number of ordered samples on a microarray can number into the hundreds of thousands. The typical microarray contains several thousands of addressable genes.

3 Related Works:

3.1 KNN: KNN is a distance-based approach for classification. In order to classify an observation X in the test set, KNN takes the following steps: (i) select an integer K (i.e., by cross-validation) and find the K closest observations in the training set; (ii) classify the observation by majority vote, that is, choose the class that is most common among those K neighbors (Dudoit et al., 2002; Speed, 2003). The KNN method is the simplest, yet useful approach to general pattern classification. Its error rate has been proven to be asymptotically at most twice that of the Bayesian error rate (R.O. Duda, 2002). However, its performance deteriorates dramatically when the input data set has a relatively low local relevance (J.H. Friedman, 1994).

3.2 DLDA: Diagonal Linear Discriminant Analysis (Dudoit, 2002) is a simplification of classical LDA, which applies the common diagonal covariance matrix to all classes. It is computationally more efficient than other LDA-based algorithms. Interestingly, the "weighted voting scheme" for binary classification proposed by T.R.Golub (1999) can be shown to be a variant of DLDA.

3.3 RDA: Classical Linear Discriminant Analysis (LDA) is not applicable for small sample size problems due to the singularity of the scatter matrices involved. Regularized LDA (RLDA) provides a simple strategy to overcome the singularity problem by applying a regularization term, which is commonly estimated via cross-validation from a set of candidates (Ye, J., 2006). RDA is better able to extract the relevant discriminatory information from training data than the other classifiers tested, thus obtaining a lower error rate (Pima & Aladjem, 2004). The RDA combines strengths of linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). It solves the small sample size and ill-posed problems suffered from QDA and LDA through a regularization technique (Lee et al., 2010).

3.4 SVM: The SVM has been shown to give superb performance in binary classification tasks. Intuitively, SVM aims at searching for a hyperplane that separates the two classes of data with largest margin (the margin is the distance between the hyperplane and the point closest to it). For multiclass SVM, there are many decomposition techniques that can adapt SVM to identify non-binary class divisions such as one-versus-the rest, pair wise comparison, and error-correcting output coding. SVM was used by Park and Cho (2003) and Li et al. (2004).

4 Microarray Datasets

Four publicly available microarray data sets are used for this study, with sample sizes ranging from 38 to 102 and numbers of genes ranging from 2,308 to 6033. Gene expression values for all the datasets are available from the Bioconductor libraries.

A. Leukemia: This dataset contains gene expression (3051 genes and 38 tumor mRNA samples) levels of n =38 patients either suffering from acute lymphoblastic leukemia (ALL, 27 cases) or acute myeloid leukemia (AML, 11 cases) where ALL and AML classes are coded in 1 and 2 respectively. It was obtained from Affymetrix

oligonucleotide microarrays. Pre-processing was done as described in Dudoit et al. (2002). Following the protocol in Dudoit et al. (2002), preprocessed them by thresholding, filtering, a logarithmic transformation and standardization, so that the data finally comprise the expression values of p=3051 genes. The data are described in Golub et al. (1999) and can be freely downloaded from http://www-genome.wi.mit.edu/MPR/.

B. Lymphoma: The lymphoma dataset consists of 42 samples of diffuse large B-cell lymphoma (DLBCL), 9 samples of follicular lymphoma (FL), and 11 samples of chronic lymphocytic leukemia (CLL). DBLCL, FL, and CLL classes are coded in 1, 2, and 3, respectively. The total sample size is n=62, and the expression of p =4026 well-measured genes, preferentially expressed in lymphoid cells or with known immunological or oncological importance is documented. Matrix of gene expression data and arrays were normalized, imputed, log transformed, and standardized to zero mean and unit variance across genes as described in Dettling (2004) and Dettling and Beuhlmann (2002). More information on these data can be found in Alizadeh et al. (2000) and can be freely downloaded from http://llmpp.nih.gov/lymphoma

C. SRBCT: This gene expression data (2308 genes for 83 samples) obtained from the microarray experiments of Small Round Blue Cell Tumors (SRBCT) of childhood cancer study. This data set contains 83 samples with 2308 genes: 29 cases of Ewing sarcoma (EWS), coded 1, 11 cases of Burkitt lymphoma (BL), coded 2, 18 cases of neuroblastoma (NB), coded 3, 25 cases of rhabdomyosarcoma (RMS), coded 4. A total of 63 training samples and 25 test samples are provided in Khan et al. (2001) and was obtained from cDNA microarrays. Five of the test set are non-SRBCT and are not considered here. The training sample indexes correspond to 1:65 and the test sample indexes (without non-SRBCT sample) correspond to 66:83. Each tissue sample is associated with a thoroughly preprocessed expression profile of p= 2,308 genes, already standardized to zero mean and unit variance across genes. Data can be freely downloaded from http://www.thep.lu.se/pub/Preprints/01/lu_tp_01_06_supp.html.

D. Prostate: The prostate dataset consists of 52 prostate tumor and 50 normal samples and obtained using the Affymetrix technology. Normal and tumor classes are coded in 1 and 2, respectively. Matrix of gene expression data and arrays were normalized, log transformed, and standardized to zero mean and unit variance across genes as described in Dettling (2004) and Dettling and Beuhlmann (2002). More information on these data can be found in Chung and Keles (2010) and can be freely downloaded from http://www-genome.wi.mit.edu/cancer.

5 Experimental Results and Discussions:

We have compared the performances of four supervised learning machines on experimental datasets of gene expression profiling. The classification algorithms were taken from those freshly proposed and applied in scientific literature for the interpretation of cancer gene expression profiling data obtained with DNA array. They are: K-Nearest Neighbor method (KNN), Diagonal Linear Discriminant Analysis (DLDA), Regularized Discriminant Analysis (RDA) & Support Vector Machines (SVM) with different kernel functions

5.1 Results: We have examined the performances of the selected supervised technologies after the application of methodology for the reduction of data dimension. In particular, we have considered the principal component analysis (PC). We evaluate their performances when different numbers of PCs are used in the classification. For each data set, experiments are carried out with 10, 20, 30 and all components. Each experiment is repeated 500 times, and the average test error rates and their standard variances over the 500 experiments are stated.

Number of selected gene components= 10

Method/ Dataset		Leukemia	Lymphoma	SRBCT	Prostate
KNN		**0.035(0.057)**	0.017(0.023)	0.161(0.071)	0.235(0.071)
DLDA		0.062(0.069)	0.016(0.023)	0.067(0.049)	0.178(0.058)
RDA		0.084(0.073)	0.030(0.043)	**0.050(0.054)**	0.181(0.057)
Kernels for SVM	Linear	0.038(0.057)	0.012(0.021)	0.076(0.055)	**0.168(0.056)**
	Polynomial	0.180(0.070)	0.092(0.056)	0.270(0.087)	0.394(0.060)
	Radial	0.096(0.077)	0.044(0.038)	0.097(0.056)	0.181(0.060)
	Sigmoid	0.083(0.072)	**0.007(0.018)**	0.076(0.051)	0.171(0.058)

Table 5.1: *The average misclassification rates(%) and standard deviation based on 500 random partitions into training sets and test set by using delete-d method (where d=0.33) for the first 10 PCs.*

Number of selected gene components=20

Method/ Dataset		Leukemia	Lymphoma	SRBCT	Prostate
KNN		**0.034(0.057)**	0.021(0.026)	0.171(0.077)	0.203(0.069)
DLDA		0.109(0.080)	0.021(0.028)	0.056(0.045)	0.150(0.063)
RDA		0.073(0.073)	0.020(0.027)	**0.049(0.045)**	0.125(0.060)
Kernels for SVM	Linear	0.107(0.087)	**0.014(0.028)**	0.102(0.058)	0.142(0.056)
	Polynomial	0.307(0.007)	0.282(0.040)	0.458(0.084)	0.436(0.052)
	Radial	0.303(0.019)	0.087(0.050)	0.154(0.078)	**0.115(0.041)**
	Sigmoid	0.187(0.084)	0.033(0.039)	0.089(0.057)	0.136(0.057)

Table 5.2: *The average misclassification rates(%) and standard deviation based on 500 random partitions into training sets and test set by using delete-d method (where d=0.33) for the first 20 PCs.*

Number of selected gene components=30

Method/ Dataset		Leukemia	Lymphoma	SRBCT	Prostate
KNN		0.060(0.067)	0.026(0.033)	0.203(0.076)	0.178(0.066)
DLDA		0.241(0.077)	0.026(0.031)	0.068(0.049)	0.174(0.061)
RDA		**0.051(.069)**	0.023(0.028)	**0.052(0.051)**	0.136(0.062)
Kernels for SVM	Linear	0.254(0.082)	**0.020(0.036)**	0.167(0.080)	0.191(0.062)
	Polynomial	0.308(0.000)	0.331(0.011)	0.581(0.050)	0.476(0.017)
	Radial	0.308(0.000)	0.123(0.049)	0.307(0.081)	**0.122(0.046)**
	Sigmoid	0.301(0.026)	0.087(0.052)	0.176(0.081)	0.214(0.067)

Table 5.3: *The average misclassification rates(%) and standard deviation based on 500 random partitions into training sets and test set by using delete-d method (where d=0.33) for the first 30 PCs.*

Number of selected all gene components

Method/ Dataset		Leukemia	Lymphoma	SRBCT	Prostate
KNN		**0.034(0.062)**	0.030(0.034)	0.161(0.078)	0.193(0.073)
DLDA		0.192(0.071)	0.087(0.060)	0.301(0.096)	0.345(0.080)
RDA		0.047(0.064)	**0.019(0.024)**	**0.039(0.042)**	**0.100(0.047)**
Kernels for SVM	Linear	0.264(0.053)	0.332(0.008)	0.675(0.094)	0.548(0.101)
	Polynomial	0.308(0.000)	0.333(0.003)	0.646(0.028)	0.475(0.031)
	Radial	0.308(0.000)	0.333(0.000)	0.636(0.000)	0.484(0.010)
	Sigmoid	0.293(0.035)	0.333(0.000)	0.656(0.039)	0.521(0.079)

Table 5.4: *The average misclassification rates(%) and standard deviation based on 500 random partitions into training sets and test set by using delete-d method (where d=0.33) for all PCs.*

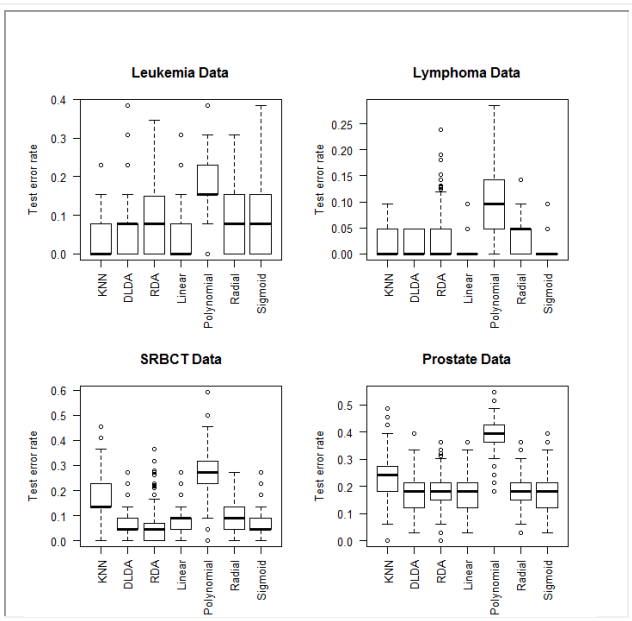


Figure 5.1: *Boxplots of error rates on four datasets for the first 10 PCs.*

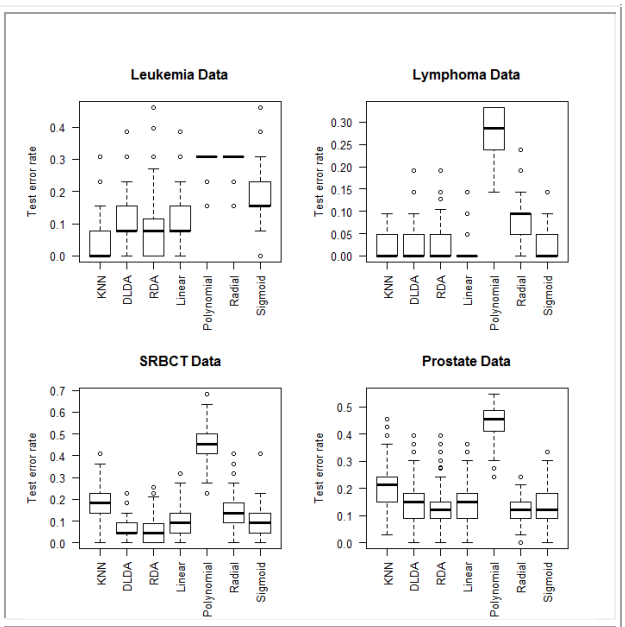


Figure 5.2: *Boxplots of error rates on four datasets for the first 20 PCs.*

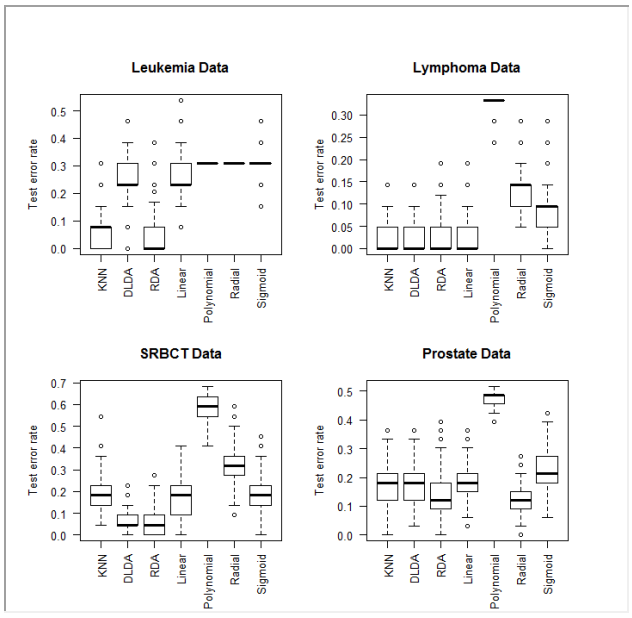


Figure 5.3: *Boxplots of error rates on four datasets for the first 30 PCs.*

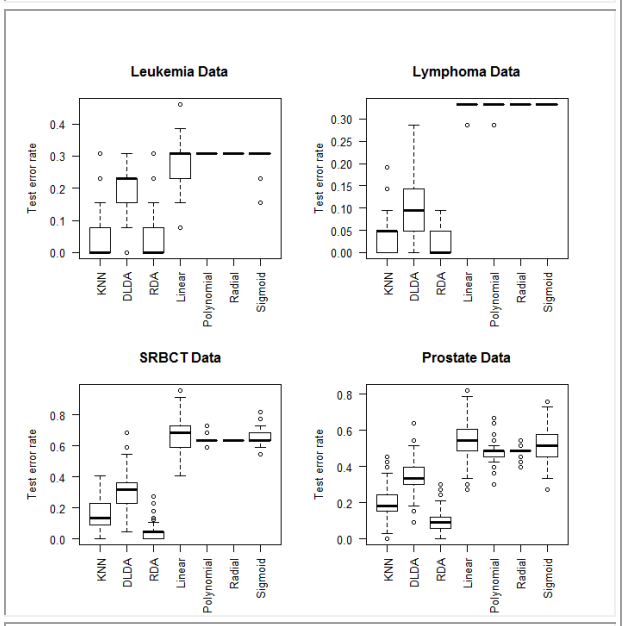


Figure 5.4: *Boxplots of error rates on four datasets for all PCs.*

In table 5.1 & Figure 5.1 we compare the misclassification values as well as standard deviation obtained by the statistical methodologies when applied with the help of the PC transformation as an additional tool for dimension reduction. Here, we have considered the first 10 PCs. As can be seen, for leukemia data the performance of the KNN & SVM with linear kernel are good enough. But on an average it is better to use KNN (3.5%) for this datasets because it obtains the lowest misclassification rate. For lymphoma data, on an average it seems to be good to consider SVM with linear (0.7%) kernel function because it has lower misclassification rate than other classifiers. For SRBCT data, RDA (5%) achieves the lowest misclassification rate. And for prostate data, on an

average it is better to consider SVM with linear (16.8%) kernel function because it obtains the lowest misclassification rate.

Table 5.2 and Figure 5.2 display the average misclassification rate & standard deviation and boxplots of test error rate of the classifiers on four datasets, respectively for the first 20 PCs.

KNN (3.4%) performs better than the other technique on leukemia data when 20 PC's were used. For lymphoma data, on an average it appears to be good to consider SVM with linear (1.4%) kernel function because it obtains the lowest misclassification rate among all the algorithms. For SRBCT data, RDA (4.9%) achieves lower misclassification rate than other classifiers. For prostate data, on an average SVM with radial (11.5%) kernel function gains the lowest misclassification rate.

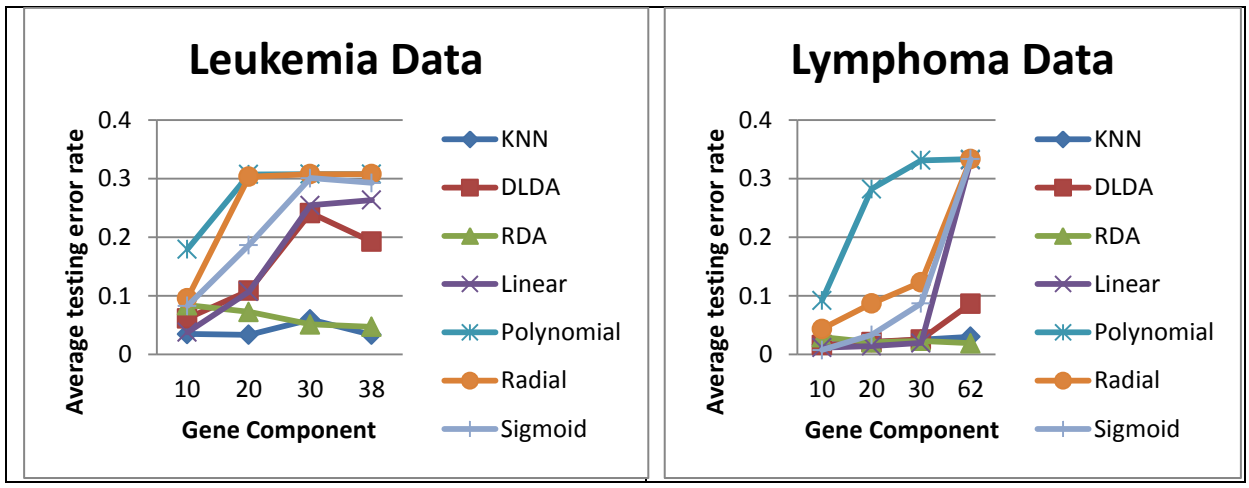
Table 5.3 and Figure 5.3 show the average misclassification rate & standard deviation and boxplots of test error rate for the four classifiers on four datasets for the first 30 PCs.

For leukemia & SRBCT data, it is observed that RDA acquires the lowest misclassification rate i.e. 5.1% & 5.2%, respectively. For lymphoma data, on an average SVM with linear (2.0%) kernel function achieves lower misclassification rate than other algorithms. It appears that SVM with radial (12.2%) kernel function obtains the lowest misclassification rate among all the techniques.

Table 5.4 represents the average misclassification rate & standard deviation over the four datasets. Figure 5.4 shows the boxplots of error rate for all the PCs.

In this case it is observed that the performance of SVM with all the kernel functions has become of poorer quality on four datasets when all gene components were used. KNN (3.4%) obtains the lowest misclassification rate for leukemia data. For lymphoma, SRBCT & prostate datasets, RDA has lower misclassification rate than the other classifiers i.e. 1.9%, 3.9% & 10%, respectively.

5.2 Discussions: The aim of our exertion was the comparative evaluation of four most used statistical analyses (nearest-neighbor, diagonalized linear discriminant analysis, regularized discriminant analysis & support vector machine with different kernel functions) by testing them on published datasets and measuring their misclassification rate given by hold-out cross validation. Now we briefly addressed the effects of variable selection on the relative performance of the classifiers.



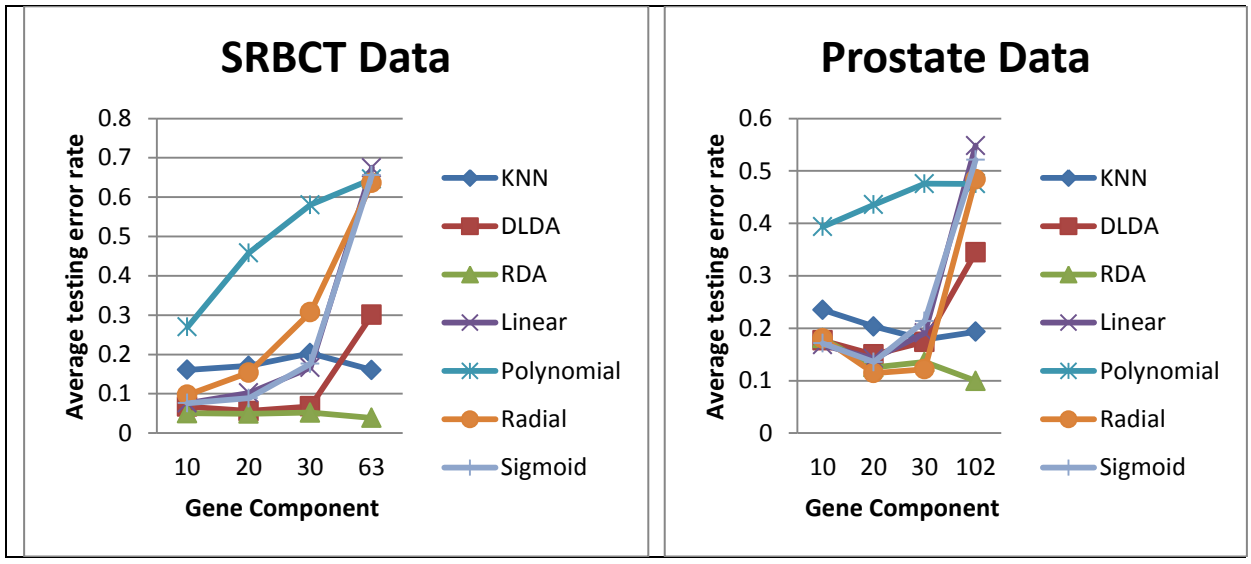


Figure 5.5: *Misclassification trend of the selected statistical technologies on leukemia, lymphoma, SRBCT and prostate gene expression dataset.*

Here, we have found that KNN (1.7%) reaches its minimum value of misclassification when 10 gene components were used among four datasets. DLDA (1.6%) achieves its minimum value of misclassification when 10 gene components were used among four datasets. RDA (1.9%) has received the lowest level of misclassification when 62 gene components were used among four datasets. SVM with linear (1.2%), polynomial (9.2%), radial (4.4%) & sigmoid (0.7%) kernel functions reach their minimum value of misclassification when 10 gene components were used.

6 Conclusions:

In this paper, we have shown the comparative results of SVM using different kernels, KNN, DLDA and RDA on four different datasets. Here, we have attempted to explore the best choice among KNN, DLDA, RDA and SVM with different kernel functions. The selected kernels are linear, polynomial, radial basis function (RBF) and sigmoid kernels. Then, focusing on PC transformation, we have compared the classification techniques for varying number of selected features and explore the most suitable method in each situation.

When dimension of the datasets were reduced by PC transformation, we have observed that the performance of the KNN & RDA are consistent for different number of PCs. But SVM with different kernel functions is negatively influenced by PC transformation. It is monitored that all the techniques except RDA have attained minimum value of misclassification when first 10 gene components were used.

Through this survey, we can conclude that classification using gene expression data has a promising future in providing a more systematical and unbiased approach in differentiating different tumor types. However, there is still a great amount of work that can be done in order to achieve the goal of optimum prediction.

Acknowledgements

I would like to express my greatest appreciation to my thesis advisor Professor Dr. Md. Shahidul Islam for his guidance and encouragement. This paper would have not been possible without his guidance and support. I feel very fortunate that I have got the opportunity to work under him, and to have discussions with him, which invariably turn out to make me learn something I would have struggled to know otherwise.

Bibliography:

Alizadeh, A., Eisen, M. B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J .Jr., Lu ,L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., & Staudt, L.M. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, *Nature*, Vol. 403, pp. 503–511.

Chuang, L. Y., Yang, C. H., Wu, K. C. and Yang, C. H. (2011). A hybrid feature selection method for DNA microarray data. Computers in Biology and Medicine, vol. 41, no. 4, pp. 228–237.

Chung, D. & Keles, S. (2010). Sparse partial least squares classification for high dimensional data. Statistical Applications in Genetics and Molecular Biology, Vol. 9, Article 17.

Dettling, M. & Beuhlmann, P. (2002). Supervised clustering of genes, Genome Biology, Vol. 3, pp. research0069.1–0069.15.

Dettling, M (2004). BagBoosting for tumor classification with gene expression data, Bioinformatics, Vol. 20, pp. 3583–3593.

Duda, R. O., Hart, P. E. & Stork, D. G. (2000). Pattern Classification (Second Edition). A Wiley-Interscience Publication.

Dudoit, S., Fridlyand, J. & Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of the American Statistical Association, Vol. 97, No. 457, p. 77–87.

Friedman, J. H. (1994). Flexible metric nearest neighbor classification. Technical report, Dept. of Statistics, Stanford University.

Ghorai, S., Mukherjee, A., Sengupta, S. and Dutta, P. (2010). Multicategory cancer classification from gene expression data by multiclass nppc ensemble. International Conference on Systems in Medicine and Biology (ICSMB), pp. 4–48.

Golub, G. & Loan, C. F. Van. (1996). Matrix Computations. The Johns Hopkins University Press, 3rd edition.

Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., and Caligiuri, M. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science, Vol. 286, 531–537.

Khan, J. and Wei, J. S. and Ringner, M. and Saal, L. H. and Ladanyi, M. and Westermann, F. and Berthold, F. and Schwab, M. and Antonescu, C. R. and Peterson, C. and Meltzer, P. S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nature Medecine, 7, 673–679.

Lee, C.-C.; Huang, S.-S. & Shih, C.-Y. (2010). Facial Affect Recognition Using Regularized Discriminant Analysis-Based Algorithms, EURASIP *J. Adv. Sig. Proc.* 2010 .

Li, T., Zhang, C., and Ogihara, M. (2004). A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. Bioinformatics.

Li, H., Zhang, K. & Jiang, T. (2005). Robust and Accurate Cancer Classification with Gene Expression Profiling. Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference, pages 320-321.

Park, C. and Cho, S.-B. (2003). Genetic search for optimal ensemble of feature classifier pairs in dna gene expression profiles. In Neural Networks, volume 3, pages 1702– 1707. Proceedings of the International Joint Conference on, IEEE.

Pima, I. & Aladjem, M. (2004). Regularized discriminant analysis for face recognition. Pattern Recognition **37** (9) , 1945-1948.

Speed, T., Ed. (2003). Statistical Analysis of Gene Expression Microarray Data. Interdisciplinary Statistics Series, Chapman & Hall/CRC.

Ye, J., Xiong, T., Li, Q., Janardan, R., Bi, J., Cherkassky, V., Kambhamettu, C. (2006). Efficient Model Selection for Regularized Linear Discriminant Analysis.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Recent conferences: http://www.iiste.org/conference/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar