

Comparison between SVM and Logistic Regression: Which One is Better to Discriminate?

Comparación entre SVM y regresión logística: ¿cuál es más
recomendable para discriminar?

DIEGO ALEJANDRO SALAZAR^{1,a}, JORGE IVÁN VÉLEZ^{2,b},
JUAN CARLOS SALAZAR^{1,2,c}

¹ESCUELA DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DE COLOMBIA, MEDELLÍN, COLOMBIA

²GRUPO DE INVESTIGACIÓN EN ESTADÍSTICA, UNIVERSIDAD NACIONAL DE COLOMBIA,
MEDELLÍN, COLOMBIA

Abstract

The classification of individuals is a common problem in applied statistics. If X is a data set corresponding to a sample from an specific population in which observations belong to g different categories, the goal of classification methods is to determine to which of them a *new* observation will belong to. When $g = 2$, logistic regression (LR) is one of the most widely used classification methods. More recently, Support Vector Machines (SVM) has become an important alternative. In this paper, the fundamentals of LR and SVM are described, and the question of which one is better to discriminate is addressed using statistical simulation. An application with real data from a microarray experiment is presented as illustration.

Key words: Classification, Genetics, Logistic regression, Simulation, Support vector machines.

Resumen

La clasificación de individuos es un problema muy común en el trabajo estadístico aplicado. Si X es un conjunto de datos de una población en la que sus elementos pertenecen a g clases, el objetivo de los métodos de clasificación es determinar a cuál de ellas pertenecerá una *nueva* observación. Cuando $g = 2$, uno de los métodos más utilizados es la regresión logística. Recientemente, las Máquinas de Soporte Vectorial se han convertido en una alternativa importante. En este trabajo se exponen los principios básicos de ambos métodos y se da respuesta a la pregunta de cuál es más recomendable

^aMSc student. E-mail: diasalazarbl@unal.edu.co

^bResearcher. E-mail: jorgeivanvelez@gmail.com

^cAssociate professor. E-mail: jcsalaza@unal.edu.co

para discriminar, vía simulación. Finalmente, se presenta una aplicación con datos provenientes de un experimento con microarreglos.

Palabras clave: clasificación, genética, máquinas de soporte vectorial, regresión logística, simulación.

1. Introduction

In applied statistics, it is common that observations belong to one of two mutually exclusive categories, e.g., presence or absence of a disease. By using a (random) sample from a particular population, classification methods allow researchers to discriminate *new* observations, i.e. assign the group to which this new observation belongs based on discriminant function (Fisher 1936, Anderson 1984) after the assumptions on which it relies on are validated. However, in practice, these assumptions cannot always be validated and, as a consequence, veracity of results is doubtful. Moreover, the implications of wrongly classifying a new observation can be disastrous.

To relax the theoretical assumptions of classical statistical methods, several alternatives have been proposed (Cornfield 1962, Cox 1966, Day & Kerridge 1967, Hosmer & Lemeshow 1989), including logistic regression (LR), one of the most widely used techniques for classification purposes today. More recently, new methodologies based on iterative calculations (algorithms) have emerged, e.g., neural networks (NN) and machine learning. However, pure computational approaches have been seen as “black boxes” in which data sets are throw in and solutions are obtained, without knowing exactly what happens inside. This, in turn, limits their interpretation.

Support Vector Machine (SVM) (Cortes & Vapnik 1995) is a classification and regression method that combines computational algorithms with theoretical results; these two characteristics gave it good reputation and have promoted its use in different areas. Since its appearance, SVM has been compared with other classification methods using real data (Lee, Park & Song 2005, Verplancke, Van Looy, Benoit, Vansteelandt, Depuydt, De Turck & Decruyenaere 2008, Shou, Hsiao & Huang 2009, Westreich, Lessler & Jonsson 2010) and several findings have been reported. In particular, *(i)* SVM required less variables than LR to achieve an equivalent misclassification rate (MCR) (Verplancke et al. 2008), *(ii)* SVM, LR and NN have similar MCRs to diagnose malignant tumors using imaging data (Shou et al. 2009), and *(iii)* NN were much better than LR with sparse binary data (Asparoukhova & Krzanowskib 2001).

In this paper we compare, by statistical simulation, the MCRs for SVM and LR when the data comes from a population in which individuals can be classified in one of two mutually exclusive categories. We consider different scenarios in which the training data set and other functional parameters are controlled. This control allowed us to generate data sets with specific characteristics and further decide whether SVM or LR should be used in that particular situation (Salazar 2012).

2. SVM and Logistic Regression

2.1. SVM for Two Groups

Moguerza & Muñoz (2006) and Tibshirani & Friedman (2008) consider a classification problem in which the discriminant function is nonlinear (Figure 1a), and there exists a kernel function Φ to a *characteristic space* on which the data is linearly separable (Figure 1b). On this new space, each data point corresponds to an abstract point on a p -dimensional space, being p the number of variables in the data set.

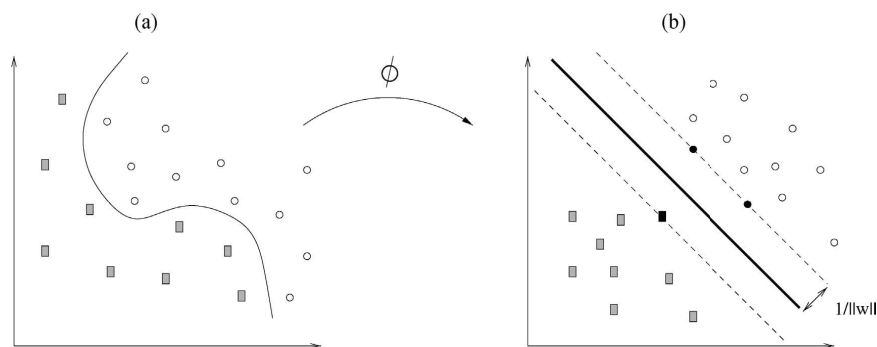


FIGURE 1: An illustration of a SVM model for two groups modified from Moguerza & Muñoz (2006). Panel (a) shows the data and a non-linear discriminant function; (b) how the data becomes separable after a kernel function Φ is applied.

When Φ is applied to the original data, a new data $\{(\Phi(\mathbf{x}_i), y_i)\}_{i=1}^n$ is obtained; $y_i = \{-1, 1\}$ indicates the two possible classes (categories), and any equidistant hyperplane to the closest point of each class on the new space is denoted by $\mathbf{w}^T \Phi(\mathbf{x}) + b = 0$. Under the separability assumption (Cover 1965), it is possible to find \mathbf{w} and b such that $|\mathbf{w}^T \Phi(\mathbf{x}) + b| = 1$ for all points closer to the hyperplane. Thus,

$$\mathbf{w}^T \Phi(\mathbf{x}) + b \begin{cases} \geq 1, & \text{if } y_i = 1 \\ \leq -1, & \text{if } y_i = -1 \end{cases} \quad i = 1, \dots, n, \quad (1)$$

such that the distance (margin) from the closest point of each class to the hyperplane is $1/\|\mathbf{w}\|$ and the distance between the two groups is $2/\|\mathbf{w}\|$. Maximizing the margin implies to solve

$$\min_{\mathbf{w}, b} \|\mathbf{w}\|^2 \quad \text{subject to} \quad y_i(\mathbf{w}^T \Phi(\mathbf{x}) + b) \geq 1 \quad i = 1, \dots, n \quad (2)$$

Let \mathbf{w}^* and b^* the solution of (2) that defines the hyperplane

$$D^*(\mathbf{x}) = (\mathbf{w}^*)^T \Phi(\mathbf{x}) + b^* = 0$$

on the *characteristic space*. All values of $\Phi(x_i)$ satisfying the equality in (2) are called *support vectors*. From the infinite number of hyperplanes separating the

data, SVM gives the optimal margin hyperplane, i.e., the one on which the classes are more distant.

Once the optimal margin hyperplane has been found, it is projected on the data's original space to obtain a discriminant function. For example, Figure 2(a) shows a data set in \mathbb{R}^2 in which two groups, linearly separable, are characterized by white and black dots that are not linearly separable. In Figure 2(b), the data is transformed to \mathbb{R}^3 where it is separable by a plane and, when it is projected back to the original space, a circular discriminant function is obtained.

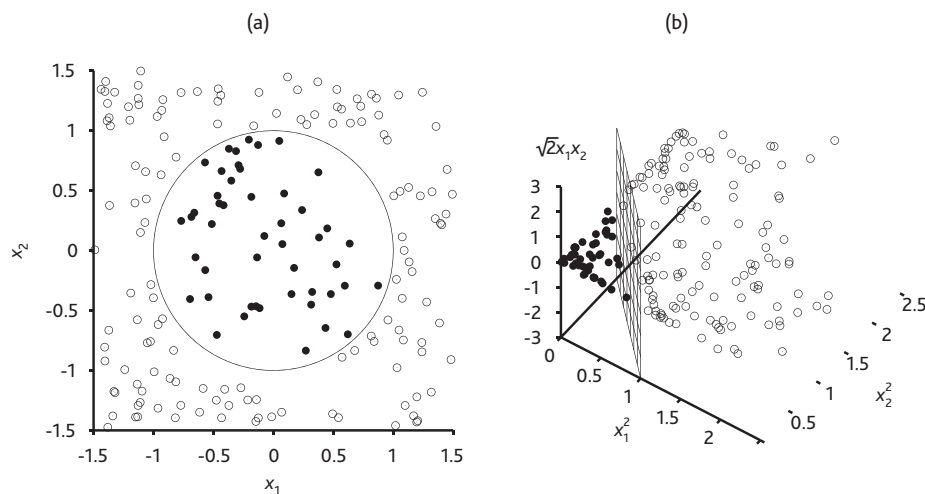


FIGURE 2: An SVM example in which (a) the two-dimensional training data set (black circles represent cases) becomes a linear decision boundary in three dimensions (b). Modified from Verplancke et al. (2008).

2.2. Logistic Regression

Let Y be a random variable such that

$$Y = \begin{cases} 1, & \text{if the condition is present} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

and $\mathbf{x} = (x_1, x_2, \dots, x_p)$ be covariates of interest. Define

$$\pi(\mathbf{x}) = E(Y|x_1, \dots, x_p)$$

as the probability that one observation \mathbf{x} belongs to one of the two groups. The Logistic Regression model is given by Hosmer & Lemeshow (1989):

$$\pi(\mathbf{x}) = \frac{\exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p\}} \quad (4)$$

Applying the transformation

$$\text{logit}(y) = \log(y/(1 - y)) \quad (5)$$

on (4) yields to a linear model in the parameters. If $\hat{\beta}$ be the maximum likelihood estimation of $\beta = (\beta_0, \beta_1, \dots, \beta_p)$, then the probability that a *new* observation $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_p^*)$ belongs to one of the two groups is

$$\hat{\pi}(x^*) = \frac{\exp\{\hat{\beta}_0 + \hat{\beta}_1 x_1^* + \dots + \hat{\beta}_p x_p^*\}}{1 + \exp\{\hat{\beta}_0 + \hat{\beta}_1 x_1^* + \dots + \hat{\beta}_p x_p^*\}} \quad (6)$$

such that a *new* observation \mathbf{x}^* will be classified in the group for which (6) is higher.

3. Simulation and Comparison Strategies

Let $g = \{1, 2\}$ be the group to which an observation belongs to. In our simulation approach, different probability distributions were considered for simulating the *training* and *validation* data sets (see Table 1). As an example, consider the Poisson case in which we generate n_1 and n_2 observations from a Poisson distribution with parameter $\lambda = 1$ (first group, $g = 1$) and $\lambda = d$ (second group, $g = 2$), respectively, with d taking the values 3, 5, 8, and 10. In real-world applications, the Poisson data being generated could be seen as white blood cell counts. Note that the greater the value of d , the greater the separation between groups.

TABLE 1: Probability distributions in our simulation study.

Distribution	$g = 1$	$g = 2$	d
Poisson	Poisson(1)	Poisson(d)	{3, 5, 8, 10}
Exponential	Exp(1)	Exp(d)	{3, 5, 8, 10}
Normal	$N(0, 1)$	$N(d, 1)$	{0.5, 1, 2, 2.5}
Cauchy-Normal	Cauchy(0, 1)	$N(d, 1)$	{1, 2, 4, 5}
Normal-Poisson	$N(0, 1)$	Poisson(d)	{1, 2, 4, 5}
Bivariate Normal ^a	$N_2(\mathbf{0}, \Sigma_1)$	$N_2(d, \Sigma_1)$	^b

^a Σ_1 is a 2×2 correlation matrix whose off-diagonal elements are $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$

^b d is a bivariate vector with elements $(d_1, d_2) = \{(0, 0), (1, 0), (1, 1.5), (2.5, 0)\}$

Our simulation and comparison strategies involve the following steps:

1. Define a probability distribution to work with.
2. Draw n_g individuals (see Hernández & Correa 2009) to form the D , the *training* data set.
3. On D , fit the LR and SVM models.
4. Draw new observations as in 1 to form D^* , the *validation* data set.
5. On D^* , evaluate the models fitted in 2. Determine their accuracy by estimating the misclassification rate (MCR)¹ calculated as $(g_{1,2} + g_{2,1}) / (n_1 + n_2)$,

¹These tables are available from the authors under request.

where $g_{i,j}$ is the number of individuals belonging to group i being classified in group j , $i, j = 1, 2$.

6. Repeat 3 and 4, $B = 5000$ times and calculate the average MCR.

Steps 2-6 were programmed in R (R Development Core Team 2011) considering several probability distributions (Table 1). Of note, either or both the expected value, variance or correlation parameter were controlled by the simulation parameter d . As samples sizes (i) $n_1 = n_2 = 20, 50, 100$ and (ii) $n_1 \neq n_2$ were used.

In LR, models were fitted using the `glm()` function from R and individuals were assigned to the group g for which the probability was higher.

SVM models including (i) linear, (ii), polynomial, (iii) radial and (iv) tangential kernels were fitted and tuned using the `e1071` facilities (Dimitriadou, Hornik, Leisch, Meyer, & Weingessel 2011). When tuning these models, the parameters γ , which controls the complexity of the classification function build by the SVM, and C , which controls the penalty paid by the SVM for missclassifying a training point (Karatzoglou, Meyer & Hornik 2006, pp. 3), were determined using the `tune.svm()` function in `e1071`.

4. Results

4.1. Univariate Distributions

Results for the Normal, Poisson and Exponential distributions are reported in Figure 3².

In the Normal case, the MCR for the polynomial SVM model is higher (poor performance). On the other hand, the performances of LR and linear, radial and tangential SVM models are equivalent. When the sample sizes differ, the MCR of the tangential and polynomial kernel is lower than when the groups have the same number of individuals. However, the former presents lower MCRs.

When the observations from both groups come from a Poisson distribution and the number of individuals by group is the same, the polynomial SVM kernel performs poorer compared with other methods, which are good competitors to LR. Additionally, the performance of the tangential kernel is not as good as it is for the LR and radial and linear kernels. LR is preferable to SVM methods when the sample sizes are not equal.

In the Exponential case, except for the polynomial kernel, SVM models perform equally well than LR when both groups have the same number of individuals. Conversely, LR performs better than SVM methods when the sample sizes are not the same. As in the Normal and Poisson distributions, the polynomial SVM is not recommended.

²Color versions of all figures presented from now on are available from the authors under request.

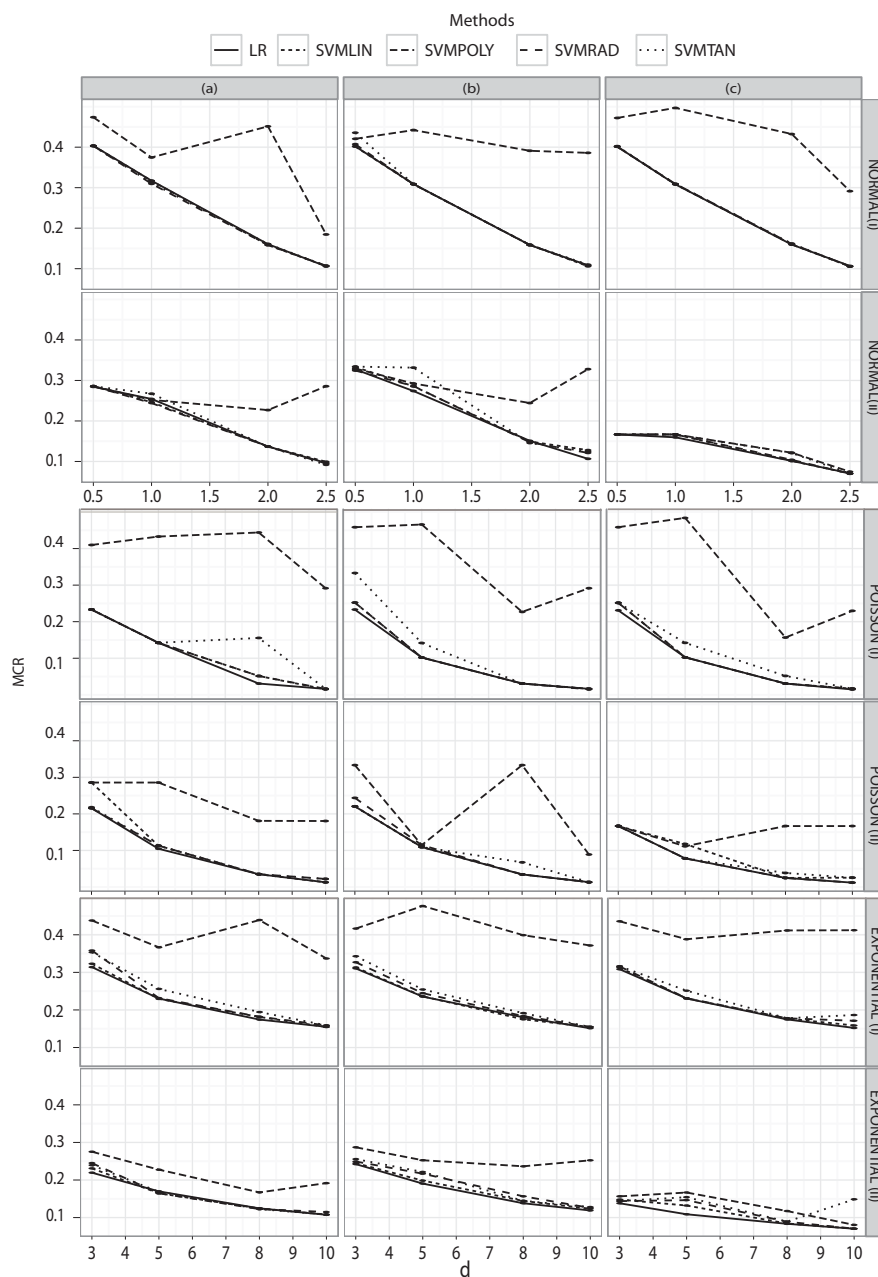


FIGURE 3: MCR as a function of d for the LR and SVM models when the observations come from the Normal, Poisson and Exponential distributions. Sample sizes in (i) are equal to (a) 20, (b) 50 and (c) 100 individuals per group. In row (ii), (a) $n_1 = 20, n_2 = 50$, (b) $n_1 = 50, n_2 = 100$, (c) $n_1 = 20, n_2 = 100$ individuals. See Table 1 for more details.

4.2. Mixture of Distributions

In Figure 4, the MCR for the Cauchy-Normal and Normal-Poisson mixtures is presented. Regardless the groups' sample sizes, SVM models perform better than LR in a Cauchy-Normal mixture. Interestingly, the polynomial kernel performs poorer when the number of individuals in both groups is the same (upper panel), but its performance improves when they are different (lower panel).

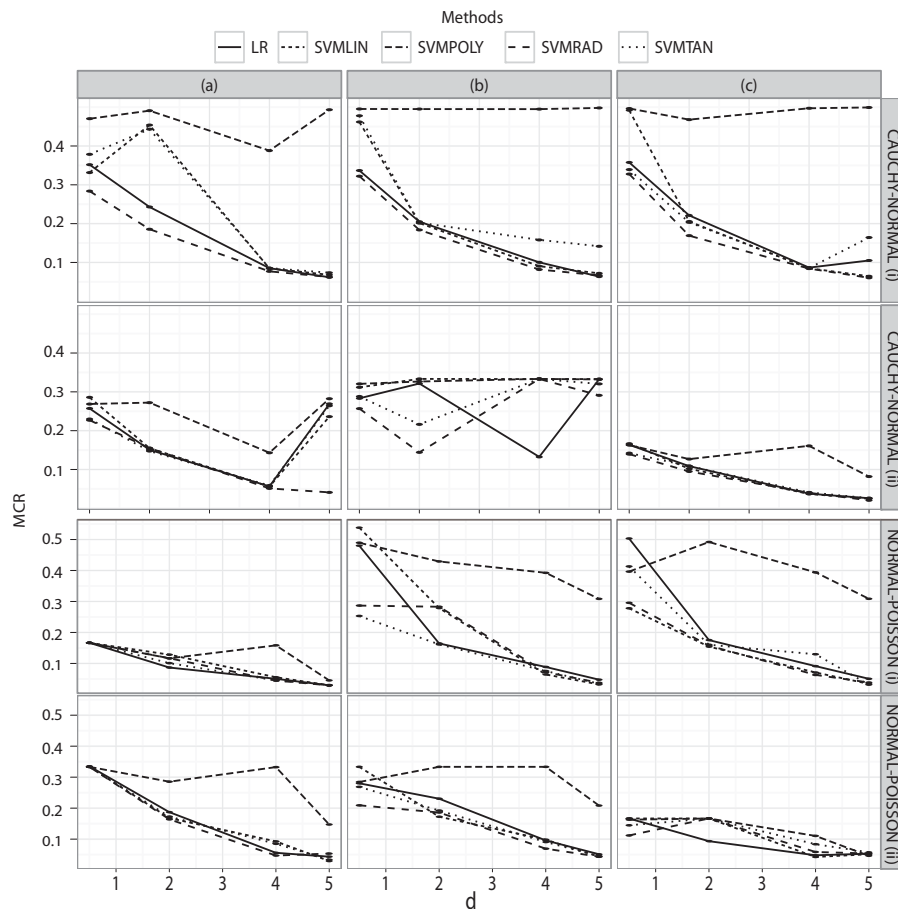


FIGURE 4: MCR as a function of d for the LR and SVM models when the observations come from a Cauchy-Normal and a Normal-Poisson mixture distributions. Conventions as in Figure 3. See Table 1 for more details.

In the Normal-Poisson mixture, the MCRs for SVM are lower than those for LR, especially when d is low, i.e., the expected value of both groups is similar. When $n_1 = n_2$ (upper panel), the linear and radial SVM models present lower MCRs than LR when the sample sizes increase.

Results for the Bivariate Normal distribution are presented in Figure 5. For all methods, the MCR decreases when ρ increases and the sample size is the same

for both groups. However, if the number of individuals per group is different and d is low, the MCR for LR is similar regardless ρ . Under this scenario, the radial and tangential SVM models perform as good as LR. Conversely, the linear kernel shows a poor performance.

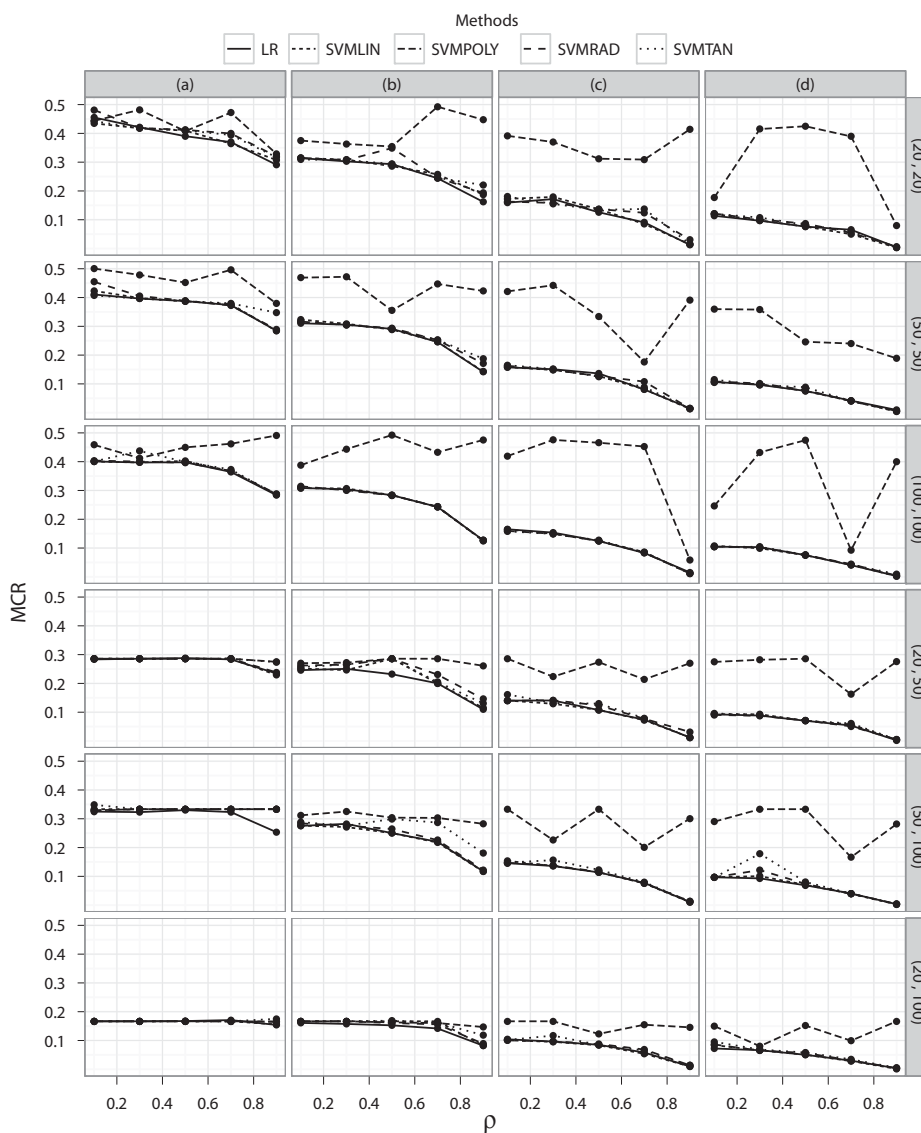


FIGURE 5: MCR as a function of ρ for the Bivariate Normal distribution when the mean vector is (a) (0,0), (b) (1,0), (c) (1, 1.5) and (d) (2.5, 0). Rows correspond to combinations of n_1 and n_2 of the form (n_1, n_2) . Here (20, 50) corresponds to $n_1 = 20$ and $n_2 = 50$.

5. Application

Mootha, Lindgren, Eriksson, Subramanian, Sihag, Lehar, Puigserver, Carlsson, Ridderstrele, Laurila, Houstis, Daly, Patterson, Mesirov, Golub, Tamayo, Spiegelman, Lander, Hirschhorn, Altshuler & Groop (2003) introduce an analytical strategy for detecting modest but coordinate changes in the expression of groups of functionally related genes, and illustrate it with DNA microarrays measuring gene expression levels in 43 age-matched skeletal muscle biopsy samples from males, 17 with normal glucose tolerance (NGT), 8 with impaired glucose tolerance (IGT) and 18 with type 2 diabetes (T2D). As a result, they identified a set of genes involved in oxidative phosphorylation.

TABLE 2: Statistics for the top 10 differentially expressed genes. No correction by multiple testing was applied.

Gene	t -statistic	$\bar{x}_{\text{NGT}} - \bar{x}_{\text{T2D}}$	P -value
G557	3.8788	0.1632	0.0005
G591	-3.6406	-0.1008	0.0009
G226	3.0621	0.1285	0.0044
G718	-3.0566	-0.1093	0.0044
G45	-2.8978	-0.1275	0.0066
G137	2.8432	0.1255	0.0076
G737	-2.6544	-0.1947	0.0121
G587	-2.5774	-0.2654	0.0146
G232	-2.5607	-0.3213	0.0152
G185	-2.5368	-0.2752	0.0161

For analysis, expression levels were processed as follows. First, a subset of 1000 genes was randomly selected from the original data. Second, the expression levels in samples from NGT (controls, group 1) and T2D individuals (cases, group 2) were compared using a two-sample t -test as implemented in `genefilter` (Gentleman, Carey, Huber & Hahne 2011). Third, only the expression levels for the top 30 differentially expressed (DE) genes were subsequently used to fit the LR and SVM models.

Summary statistics for the top 10 genes found to be DE are presented in Table 2; genes G557, G226 and G137 are down-regulated, i.e., their expression levels are lower in T2D than in NGT samples. Figure 6 depicts a scatterplot for the top 5 genes by disease status. In there, some (expected) correlation structures are observed; these correlations might constitute a potential problem for any classification method.

LR and SVM models were fitted using the disease status as dependent variable and the expression levels of k genes as covariates. Our findings are reported in Figure 7. For predicting the disease status in this data set, (i) SVM models required less variables (genes); (ii) all methods performed similarly when $k < 5$, but the radial SVM model is more consistent, and (iii) the polynomial and tangential SMVs are definitely not an option. These results may provide important insights in the diagnosis of genetic diseases using this type of models.

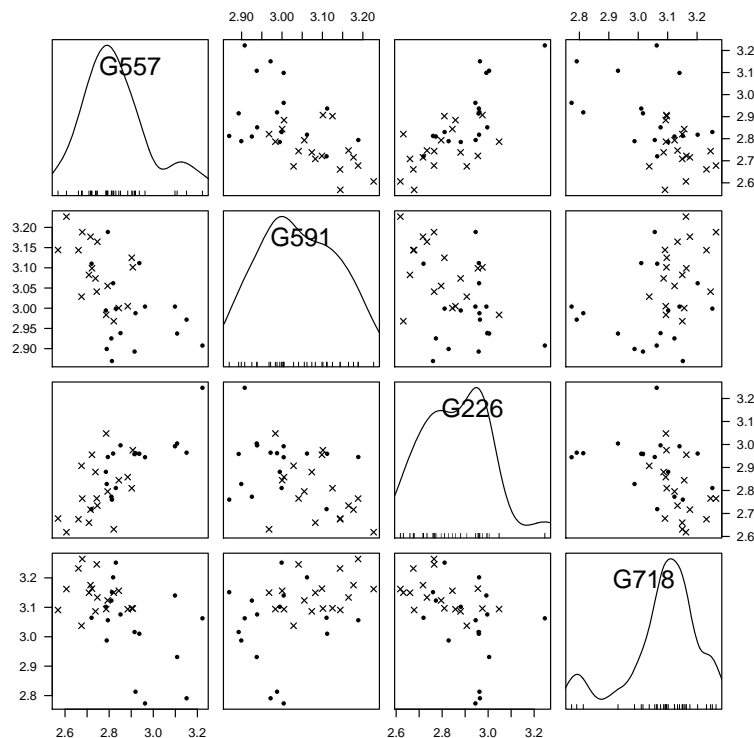


FIGURE 6: Scatterplot matrix for some of the genes presented in Table 2. Filled dots correspond to NGT samples (controls). In the diagonal panel, density plots are shown.

Although in our application we only used a subset of the genes available in the microarray experiment, it illustrates how a SVM model can be used to predict the (disease) status of a patient using his/her genetic information. Furthermore, we evaluated the possibility of including “one-gene-at-the-time” and determine the MCR of the (full) SVM model as more genetic profiles were added. Using a similar strategy and by combining SVM with other classification methods such as genetic algorithms, several authors have been able to build accurate predictive models that, in the near future, could be used to diagnose patients in the clinical setting. Some examples include the work by David & Lerner (2005) in genetic syndrome diagnosis, and Furey, Cristianini, Duffy, Bednarski, Schummer & Haussler (2000), Peng, Xum, Bruce Ling, Peng, Du & Chen (2003), and Li, Jiang, Li, Moser, Guo, Du, Wang, Topol, Wang & Rao (2005) in cancer.

SVM models have shown to be highly accurate when cancer diagnosis is of interest and either microarray expression data (Furey et al. 2000, Noble 2006) or tumor marker detection (TMD) results for colorectal, gastric and lung cancer (Wang & Huan 2011) are available. For instance, Furey et al. (2000) used 6817 gene expression measurements and fitted a SVM model that achieved near-perfect classification accuracy on the ALL/AML data set (Golub, Slonim, Tamayo, Huard,

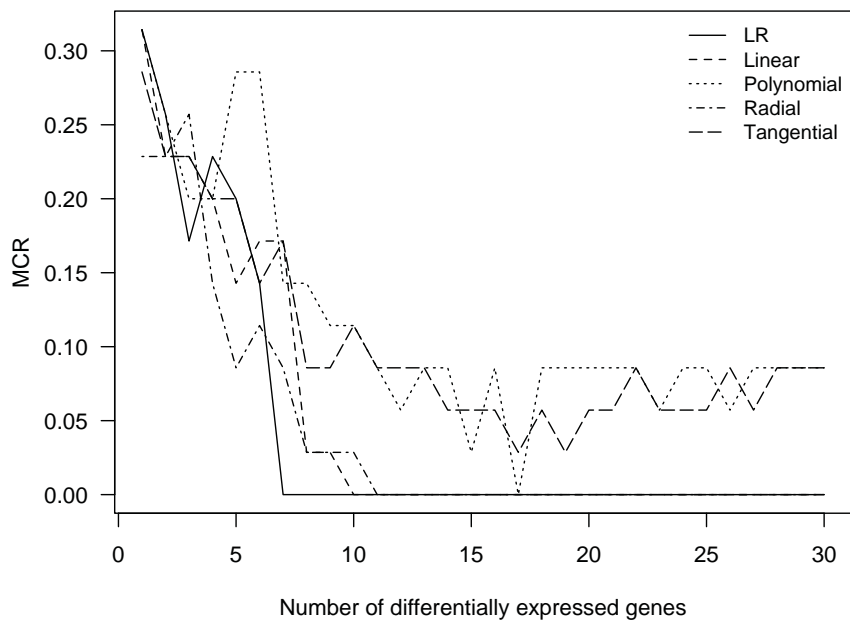


FIGURE 7: MCR as a function of the number of differentially expressed genes.

Gaasenbeek, Mesirov, Collier, Loh, Downing, Caligiuri, Bloomfield & Lander 1999). For TMD, Wang & Huan (2011) created, trained, optimized and validated SVM models that resulted to be highly accurate compared to others, indicating a potential application of the method as a diagnostic model in cancer. Similarly, Peng et al. (2003) combined genetic algorithms and paired SVM for multiclass cancer identification to narrow a set of genes to a very compact cancer-related predictive gene set; this method outperformed others previously published.

6. Conclusions

We have presented a framework to compare, by statistical simulation, the performance of several classification methods when individuals belong to one of two mutually exclusive categories. As a test case, we compared SVM and LR.

When it is of interest to predict the group to which a *new* observation belongs to based on a single variable, SVM models are a feasible alternative to LR. However, as shown for the Poisson, Exponential and Normal distributions, the polynomial SVM model is not recommended since its MCR is higher.

In the case of multivariate and mixture of distributions, SVM performs better than LR when high correlation structures are observed in the data (as shown in Figure 6). Furthermore, SVM methods required less variables than LR to achieve a better (or equivalent) MCR. This latter result is consistent with Verplancke et al. (2008).

Further work includes the evaluation of the MCR of SVM and LR methods for other probability distributions, different variance-covariance matrices among groups, and high-dimensional (non) correlated data with less variables than observations, e.g., genetic data with up to 5 million genotypes and ~ 1000 cases and controls.

Acknowledgements

DAS and JCS were supported by Programa Nacional de Investigación en Genómica, Bioinformática y Estadística para Estudios en Enfermedades Neurosiquiátricas. Fase I: Enfermedad de Alzheimer, código 9373, Grupo de Neurociencias de la Universidad Nacional de Colombia, Sede Bogotá. We thank two anonymous reviewers for their helpful comments and suggestions.

[Recibido: septiembre de 2011 — Aceptado: febrero de 2012]

References

- Anderson, T. (1984), *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York.
- Asparoukhova, K. & Krzanowskib, J. (2001), ‘A comparison of discriminant procedures for binary variables’, *Computational Statistics & Data Analysis* **38**, 139–160.
- Cornfield, J. (1962), ‘Joint dependence of the risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis’, *Proceedings of the Federal American Society of Experimental Biology* **21**, 58–61.
- Cortes, C. & Vapnik, V. (1995), ‘Support-vector networks’, *Machine Learning* **20**(3), 273–297.
- Cover, T. M. (1965), ‘Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition’, *IEEE Transactions on Electronic Computers* **14**, 326–334.
- Cox, D. (1966), *Some Procedures Associated with the Logistic Qualitative Response Curve*, John Wiley & Sons, New York.
- David, A. & Lerner, B. (2005), ‘Support vector machine-based image classification for genetic syndrome diagnosis’, *Pattern Recognition Letters* **26**, 1029–1038.
- Day, N. & Kerridge, D. (1967), ‘A general maximum likelihood discriminant’, *Biometrics* **23**, 313–323.

- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. (2011), *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*. R package version 1.5-27.
*<http://CRAN.R-project.org/package=e1071>
- Fisher, R. (1936), 'The use of multiple measurements in taxonomic problems', *Annual Eugenics* **7**, 179–188.
- Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M. & Hausler, D. (2000), 'Support vector machine classification and validation of cancer tissue samples using microarray expression data', *Bioinformatics* **16**(10), 906–914.
- Gentleman, R., Carey, V., Huber, W. & Hahne, F. (2011), *Genefilter: Methods for filtering genes from microarray experiments*. R package version 1.34.0.
- Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M., Bloomfield, C. & Lander, E. (1999), 'Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring', *Science* **286**, 531–537.
- Hernández, F. & Correa, J. C. (2009), 'Comparación entre tres técnicas de clasificación', *Revista Colombiana de Estadística* **32**(2), 247–265.
- Hosmer, D. & Lemeshow, S. (1989), *Applied Logistic Regression*, John Wiley & Sons, New York.
- Karatzoglou, A., Meyer, D. & Hornik, K. (2006), 'Support vector machines in R', *Journal of Statistical Software* **15**(8), 267–73.
- Lee, J. B., Park, M. & Song, H. S. (2005), 'An extensive comparison of recent classification tools applied to microarray data', *Computational Statistics & Data Analysis* **48**, 869–885.
- Li, L., Jiang, W., Li, X., Moser, K. L., Guo, Z., Du, L., Wang, Q., Topol, E. J., Wang, Q. & Rao, S. (2005), 'A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset', *Genomics* **85**(1), 16–23.
- Moguerza, J. & Muñoz, A. (2006), 'Vector machines with applications', *Statistical Science* **21**(3), 322–336.
- Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrele, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D. & Groop, L. C. (2003), 'Pgc-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes', *Nature Genetics* **34**(3), 267–73.

- Noble, W. (2006), 'What is a support vector machine?', *Nature Biotechnology* **24**(12), 1565–1567.
- Peng, S., Xum, Q., Bruce Ling, X., Peng, X., Du, W. & Chen, L. (2003), 'Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines', *FEBS Letters* **555**, 358 – 362.
- R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*<http://www.R-project.org/>
- Salazar, D. (2012), Comparación de Máquinas de Soporte vectorial vs. Regresión Logística: cuál es más recomendable para discriminar?, Tesis de Maestría, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.
- Shou, T., Hsiao, Y. & Huang, Y. (2009), 'Comparative analysis of logistic regression, support vector machine and artificial neural network for the differential diagnosis of benign and malignant solid breast tumors by the use of three-dimensional power doppler', *Korean Journal of Radiology* **10**, 464–471.
- Tibshirani, R. & Friedman, J. (2008), *The Elements of Statistical Learning*, Springer, California.
- Verplancke, T., Van Looy, S., Benoit, D., Vansteelandt, S., Depuydt, P., De Turck, F. & Decruyenaere, J. (2008), 'Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies', *BMC Medical Informatics and Decision Making* **8**, 56–64.
- Wang, G. & Huan, G. (2011), 'Application of support vector machine in cancer diagnosis', *Med. Oncol.* **28**(1), 613–618.
- Westreich, D., Lessler, J. & Jonsson, M. (2010), 'Propensity score estimation: Neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression', *Journal of Clinical Epidemiology* **63**, 826–833.