ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 167-172

Prediction of Mental Development of Preterm Newborns at Birth Time using LS-SVM

L. Ameye[1], **C. Lu**[1], **L. Lukas**[1],
J. De Brabanter[1], **J.A.K. Suykens**[1], **S. Van Huffel**[1],
H. Daniels[2], **G. Naulaers**[2], **H. Devlieger**[2]

[1] K.U. Leuven, Dept. of Electrical Engineering ESAT-SISTA,
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
[2] K.U. Leuven, Dept. of Paediatrics, University Hospitals Leuven,
Herestraat 49, B-3000 Leuven, Belgium
{lieveke.ameye,chuan.lu,llukas,brabant,johan.suykens,sabine.vanhuffel}@esat.kuleuven.ac.be

Abstract. Premature babies have a higher risk for mental development disturbances (the lower the post-menstrual age, the higher the risk) and therefore it is important to predict at birth time the probability for a normal mental development and analyse the main risk factors. A data analysis study has been performed in order to classify premature babies according to their Mental Development Index. Using a data set with 170 observations and 26 explanatory variables we developed prediction models using Logistic Regression (LR) and Least Squares Support Vector Machines (LS-SVM) and compared their performances by means of ROC-curves.

Keywords. Classification, Mental Development, Logistic Regression, Least Squares Support Vector Machines.

1 Introduction

Premature babies have a higher risk for mental development disturbances (the lower the post-menstrual age, the higher the risk) and therefore it is important to predict at birth time the probability for a normal mental development [2, 8].

The neonatology division of the University Hospitals Leuven (Belgium) has measured the Mental Development Index of 170 premature babies born in the period 1996 - 1998. This index is measured 7 months after the baby is born and is a number between 50 and 150. If the index is larger than 110 then the baby has a normal mental development, otherwise the baby has an abnormal mental development. For all the premature babies 26 explanatory variables were measured at birth time, e.g. the weight at birth, the sex, the postmenstrual age, whether the baby has problems concerning the central nervous system or not, and others. From the 26 factors, there are 19 binary, 1 categorical and 6 continuous factors.

A reliable test for discrimination at birth time between normal and abnormal babies would be of considerable help for the clinicians. The aim is to construct a model, based on the factors measured at birth time in order to predict whether the baby will have a normal or abnormal mental development at 7 months, i.e. if the index will be higher than 110 or not.

In this paper two methods are applied to construct a predictive classification model: Logistic Regression (LR) and Least Squares Support Vector Machines (LS-SVM). LR [1, 4] is a linear model that many statisticians are using as a standard model to classify observations and is frequently applied in a biomedical context. The drawback of this method is that linear techniques are not capable to capture possible nonlinear dependencies. Since the introduction of Support Vector Machines (SVM) [7], there has been a growing interest in kernel-based methods and many successes have been reported. In this paper we focus on the use of Least Squares SVM classifiers (LS-SVM) [5]. In [6] it has been shown that LS-SVM classifiers consistently perform very well on a large variety of problems in comparison with a large amount of different methods. LS-SVMs lead to linear systems instead of QP problems and sparseness and robustness might be imposed where needed. In comparison with classical neural networks (LS)-SVM perform well in high dimensional feature spaces.

This paper is organized as follows. In Section 2 we briefly explain the LR model. In Section 3 we give a short introduction to LS-SVM. Section 4 presents the application of LR and LS-SVM on this classification problem and their performances are compared by means of their Receiver Operating Characteristic (ROC) curves [3].

2 Logistic Regression

Suppose we have N observations and the response variable Y can only take two possible values, denoted by -1 and 1. For example, $Y = 1$ if the baby has a normal mental development and $Y = -1$ otherwise. In that case, logistic regression [1, 4] can be used to explore the relation between the response variable Y and several explanatory variables $X_1, X_2, ..., X_n$ with n the number of inputs.

Let $\mathcal{P} = \mathcal{P}(X_1, X_2, ..., X_n)$ denote the probability that the mental development is abnormal, given the explanatory variables $X_1, X_2, ...X_n$. The logistic regression model can then be written as follows:

$$\log \frac{\mathcal{P}}{1-\mathcal{P}} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n, \tag{1}$$

where $\mathcal{P}/(1-\mathcal{P})$ is called the *odds* that the baby has a normal mental development given $X_1, X_2, ..., X_n$. This equation models the log of the odds or the *logit* as a linear function of the explanatory variables; β_0 is the intercept parameter and $\beta_1, ...\beta_n$ are the slope parameters. The logistic regression model can also be rewritten as follows:

$$\mathcal{P} = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n)}. \tag{2}$$

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 167-172

Note that the logit can range from minus to plus infinity, while $\mathcal{P}$ falls between 0 and 1. The odds ratio $\exp(\beta_i)$ is the ratio of odds for a one-unit increase in X_i. This means that the odds increase multiplicatively by $\exp(\beta_i)$ for every one-unit increase in X_i.

Stepwise LR [4] is applied to detect the main risk factors. Stepwise selection is the process of adding or removing variables from a model until the final model is relatively the best of the competing models for the data. At each step in the process, the significance levels of the variables are computed and compared to specified significance level criteria for entry to or removal from the model. These significance levels are p-values with respect to a chi-square distribution with one degree of freedom (χ_1^2). A crucial aspect of stepwise LR is the choice of a significance level to judge the variables. When the stepwise selection process is terminated, the variables that are not significant at 5% significance level are removed. This is done in several stages, that is one variable at a time is removed, and at each stage stepwise regression is again performed. The drawback of the LR model is that it assumes that the output is a linear combination of the explanatory factors and therefore is not always capable to fit a good model to the data.

3 Least Squares Support Vector Machines

Consider a training set of N data points $\{z_k = (x_k, y_k)\}_{k=1}^N$, with input data $x_k \in \mathbb{R}^n$ and output data $y_k \in \{-1, 1\}$ in the simple classification case of two classes. For the case of two classes, one assumes $w^T \varphi(x_k) + b \geq +1$ if $y_k = +1$ and $w^T \varphi(x_k) + b \leq -1$ if $y_k = -1$ which is equivalent to $y_k[w^T \varphi(x_k) + b] \geq 1$ for $k = 1, ..., N$ where w is the primal parameter vector, b is the bias (a real constant), and $\varphi(.)$ is a nonlinear function which maps the input space into a higher dimensional space. The parameter vector cannot be calculated explicitly, because the nonlinear mapping is not explicitly known. This problem can be solved by introducing kernel functions.

The support vector method approach [5, 7] aims at constructing a classifier of the form:

$$y(x) = \text{sign}\left[\sum_{k=1}^{N} \alpha_k \, y_k \, K(x, x_k) + b\right] \tag{3}$$

where α_k are support values and b is a real constant and are the solution to a linear system when LS-SVMs [5, 6] are used as defined below. For standard SVMs [7] a quadratic programming problem must be solved. $K(\cdot, \cdot)$ is a positive definite kernel, often chosen as $K(x, x_k) = \exp\{-\|x - x_k\|_2^2/\sigma^2\}$ (RBF kernel) or $K(x, x_k) = x_k^T x$ (linear SVM).

LS-SVM classifiers are obtained as solution to the following optimization problem:

$$\min_{w,b,e} \mathcal{J}_{LS}(w, b, e) = \frac{1}{2} w^T w + \gamma \, \frac{1}{2} \sum_{k=1}^{N} e_k^2 \tag{4}$$

subject to the equality constraints

$$y_k \, [w^T \varphi(x_k) + b] = 1 - e_k, \; k = 1, ..., N.$$

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 167-172

The first term in the optimization problem is a regularisation term and the second term controls the minimization of the error. One defines the Lagrangian

$$\mathcal{L}(w, b, e; \alpha) = \mathcal{J}_{LS} - \sum_{k=1}^{N} \alpha_k \{y_k[w^T \varphi(x_k) + b] - 1 + e_k\} \tag{5}$$

where α_k are Lagrange multipliers. The Kuhn-Tucker conditions for optimality $\frac{\partial \mathcal{L}}{\partial w} = 0$, $\frac{\partial \mathcal{L}}{\partial b} = 0$, $\frac{\partial \mathcal{L}}{\partial e_k} = 0$, $\frac{\partial \mathcal{L}}{\partial \alpha_k} = 0$ provide a set of linear equations $w = \sum_{k=1}^{N} \alpha_k y_k \varphi(x_k)$, $\sum_{k=1}^{N} \alpha_k y_k = 0$, $\alpha_k = \gamma e_k$, $y_k[w^T \varphi(x_k) + b] - 1 + e_k = 0$, for $k = 1, ..., N$. Elimination of w and e gives

$$\left[\begin{array}{c|c} 0 & Y^T \\ \hline Y & \Omega + \gamma^{-1} I \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline 1_v \end{array}\right] \tag{6}$$

with $Y = [y_1; ...; y_N]$, $1_v = [1; ...; 1]$, $e = [e_1; ...; e_N]$, $\alpha = [\alpha_1; ...; \alpha_N]$. Mercer's theorem is applied to the matrix Ω with

$$\begin{aligned} \Omega_{kl} &= y_k y_l \, \varphi(x_k)^T \varphi(x_l) \\ &= y_k y_l \, K(x_k, x_l). \end{aligned} \tag{7}$$

Hence one chooses a positive definite kernel $K(\cdot, \cdot)$ that satisfies Mercer condition. Note that the LS-SVM predictive model depends on an adjustable hyperparameter vector $\theta = [\sigma; \gamma]$. 10-fold cross-validation will be used here as a method to choose the value of the hyperparameter vector that optimizes the performance of the LS-SVM model.

4 Classification of the mental development index using LR and LS-SVM

4.1 Logistic Regression

Important predictors can be selected from stepwise multiple regression analysis [1, 4]. A crucial aspect of using stepwise LR is the choice of a significance level to judge the importance of the variables. The significance level for entry in the model has been chosen equal to 0.15 and the significance level for removing a variable equals 0.20. When the stepwise selection process is terminated, the variables that are not significant at 5% significance level are removed one by one.

The selected main risk factors for mental development disturbances of premature babies are:

- a low weight at birth
- a low apgarscore 1 minute after birth
- problems concerning the central nervous system
- congenetic problems

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 167-172

- iatrogenic problems

No interaction terms or transformed factors are found to be significant.

To check the performance on a independent set of observations the whole data set of 170 observations is randomly split up in a training set (70% of the data) to get the parameter values of the model and a test set (the remaining 30%) to test the performance of the model on new data. The LR model using the above 5 predictors gives in 73% of the cases in the test set a correct classification. The area under the ROC-curve (AUC) equals 0.80.

4.2 Least Squares Support Vector Machines

The LS-SVM predictive model [5, 6] depends on an adjustable hyperparameter vector $\theta = [\sigma; \gamma]$. 10-fold cross-validation is used as a method to choose the value of the hyperparameter vector that optimizes the performance of the LS-SVM model with RBF kernel. When using all the 26 explanatory variables the LS-SVM model gives a performance of 67% correct classification in the test set (AUC=0.69).

A much better performance can be obtained using a selected set of inputs. The LS-SVM model using the 5 predictors selected by the multiple regression model gives a performance of 92% correct classifications in the test set. The area under the ROC curve equals 0.93.

The performance of the classical LR model and the LS-SVM using the inputs selected by stepwise LR can be compared by their ROC-curves using the following test statistic [3]:

$$z = \frac{AUC_{LR} - AUC_{LS-SVM}}{(SE_{AUC_{LR}})^2 + (SE_{AUC_{LS-SVM}})^2 - 2 * r * SE_{AUC_{LR}} * SE_{AUC_{LS-SVM}}} \tag{8}$$

with AUC_{LR} the area under the ROC curve of the LR model, AUC_{LS-SVM} the area under the ROC curve of the LS-SVM model using the selected set of inputs, SE the standard error on the AUC and r the estimated correlation between AUC_{LR} and AUC_{LS-SVM}. Using this z-statistic the obtained difference between the two ROC-curves (Figure 1) is denoted as significant (p-value $= 0.03$).

5 Conclusions

A data analysis study has been performed in order to classify premature babies according to their Mental Development Index. Using a data set with 170 observations and 26 explanatory variables we developed prediction models using Logistic Regression (LR) and Least Squares Support Vector Machines (LS-SVM) and compared their performances by means of ROC-curves. It is concluded that LS-SVMs with an area under the ROC-curve (AUC) of 93% outperform significantly LR models (AUC=80%) in order to predict the mental development at birth time.

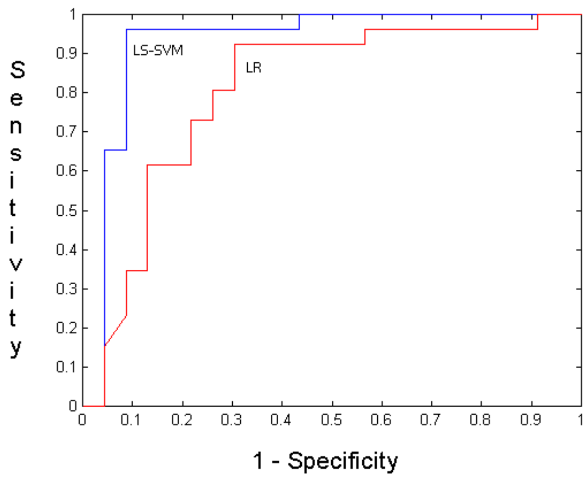

Figure 1: *ROC curves of the selected LR and LS-SVM classifiers. The Area Under the Curve (AUC) on test data is 0.80 for the LR model and 0.93 for the LS-SVM model with RBF kernel.*

Acknowledgements

This research work was carried out at the ESAT laboratory and the Interdisciplinary Center of Neural Networks ICNN of the Katholieke Universiteit Leuven, in the framework of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture (IUAP P4-02 & IUAP P4-24), the Concerted Action Project MEFISTO of the Flemish Community, the FWO project G.0407.02 and the IDO/99/03 project. JS is a postdoctoral researcher with the National Fund for Scientific Research FWO - Flanders.

References

[1] Agresti A., *Introduction to Categorical Data Analysis*, Wiley Series in Probability and Statistics, 1996.

[2] De Cock P., *Development perspectives in children with birth weight below 2000g* (in Dutch), Leuven University Press, 1994.

[3] Hanley A.J., McNeil B.J., "The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve," *Radiology*, 143(1):29-36, 1982.

[4] Hosmer D.W. Jr., Lemeshow S., *Applied Logistic Regression*, John Wiley & Sons, Inc., 1989.

[5] Suykens J.A.K., Vandewalle J., "Least Squares support vector machine classifiers," *Neural Processing Letters*, 9(3):293-300, 1999.

[6] Van Gestel T., Suykens J., Baesens B., Viaene S., Vanthienen J., Dedene G., De Moor B., Vandewalle J., "Benchmarking Least Squares Support Vector Machine Classifiers," ESAT-SISTA Technical Report 00-37 ESAT-SISTA, K.U.Leuven Belgium, 2000 (downloadable from www.esat.kuleuven.ac.be/sista/members/suykens.html).

[7] Vapnik V., *The nature of statistical learning theory*, Springer-Verlag, 1995.

[8] Zernikow B. et al., "Artificial neural network for risk assessment in preterm neonates," *Arch Dis Child Fetal Neonatal Ed*, 79:F129-F134, 1998.