Foetal Weight Estimation by Support Vector Regression

Fernando Sereno*, J.P. Marques de Sá†, Ana Matos†, João Bernardes ‡‡

* FEUP – Faculdade de Engenharia da Universidade do Porto, Portugal
E-mail: fsereno@fe.up.pt
† HSJ – Hospital de S. João, Dep. Ginecologia e Obstetrícia, Porto, Portugal
‡‡ FMUP - Faculdade de Medicina da Universidade do Porto, Portugal
INEB - Instituto de Engenharia Biomédica, Porto, Portugal

Abstract

Foetal weight estimation based on echographic measurements has paramount importance. This paper reports some results using data taken from a dataset of four Portuguese hospitals that participated in the collection of clinical and echographic data (414 cases) during 1998-99. Firstly, it revises some theoretical concepts from Statistical Learning Theory. Then, reports some results using Support Vector regression to predict foetal weight in lower and higher bands. Finally, it concludes that the error given by the SVR methods are better than traditional formulas and can be improved further.

1 Introduction

Foetal weight estimation based on echographic measurements has paramount importance in delivery risk assessment [1,2].

The research objective is to know to what extent SV machines can improve over the 15% error of FW estimation performed by prediction formulas in current day clinical use [3].

Four Portuguese hospitals participated in the collection of clinical and echographic data (414 cases) during 1998-99, according to a protocol. Each case consists of foetal weight (FW) at birth, and five echographic measurements, taken one week before birth. These are: biparietal diameter, cephalic circumference, abdominal circumference, femur length and umbilical artery resistance index.

2 Statistical inference

Parametric statistics aims to create simple statistical methods of inference that can be used to solve real-life problems and is based in the assumption that the investigator knows the problem, the function to be found up to a finite number of parameters.

Using information about the statistical law and the maximum likelihood method applied to the data one finds the target function and estimates its parameters, which is the essence of classical Fisherean inference.

When one does not have reliable a priori information about the statistical law underlying the problem or about the class of functions and the conditions under which one can get better approximations with an increasing number of examples, one is in the general inference approach, a development that was started by Glivenko, Cantelli and Kolmogorov.

In the last 40 years of research this approach culminated in inductive methods, a different type of inference which is more general and more powerful than parametric inference [4,5].

3 Minimizing the risk functional from empirical data

The basic problem is to formulate a constructive criterion for choosing from parametric sets of functions one function that minimizes the mathematical expectation

\[ R(\alpha) = \int Q(z,\alpha) dF(z), \quad \alpha \in \Lambda, \] (1)

where \( Q(z,\alpha) \) is called the loss function, \( z \) is a variable that represents random independent observations \( z_1, \ldots, z_i \), obtained according to unknown distribution \( F(z) \), \( \alpha \) is a parameter from a
set $\Lambda$, arbitrary, it can be a set of scalar quantities, a set of vectors or a set of abstract elements, and the integral is Lebesgue-Stieltjes for a bounded nonnegative function [4].

4 The problem of regression estimation

Estimating the stochastic dependence based on empirical data pairs $(y_1, x_1), \ldots, (y_\ell, x_\ell)$, taken randomly and independently from a joint distribution function $F(x,y)$, means estimating the conditional distribution function $F(y|x)$. This is often an ill-posed problem [4,5,6], that can however be determined by the mathematical expectation

$$r(x) = \int ydF(y|x)$$

called regression function. It can be shown that this function can be estimated for sets of functions $f(x,\alpha)$ in the metric $L^2(P)$, by minimizing the functional

$$R(\alpha) = \int (y - f(x,\alpha))^2 dF(x,y)$$

5 Principle of Empirical Risk Minimization

One cannot minimize the functional (3) directly since one ignores the probability distribution function $F(x,y)$ that defines the risk. Instead one can use the classical induction principle, that consists in minimizing an empirical risk functional, for example (4). There are probabilistic bounds on the distance between empirical and expected risks involving the number of examples $\ell$ and the capacity $h$ of the function space, a quantity measuring the “complexity” of the space.

The solution of learning $f(x,\alpha)$ is found by solving for each constant $A_m$, related to hypothesis spaces, an optimization problem:

$$\sum_{i=1}^\ell Q(y_i, f(x_i)) + \lambda ||f||^2$$

subject to

$$A_m \leq A_m, \quad m \in \Lambda$$

and choosing among the solutions found for each $A_m$, the one with the best trade off between empirical risk and capacity [4,5,7].

The regularization parameter $\lambda$ penalizes functions with high capacity. In Support Vector Regression (SVR) we used the loss function :

$$Q(y_i, f(x_i)) = |y_i - f(x_i)|_\varepsilon$$

where the function $|.|_\varepsilon$, is called $\varepsilon$-insensitive loss.

The function given has the general form:

$$f(x) = \sum_{i=1}^\ell c_i K(x, x_i)$$

The data points $x_i$ associated with nonzero $c_i$ are called support vectors and they represent the most informative data points and compress the information contained in the training set.
7 Experimental results

The SVR training algorithm [8] has been tested on two subsets of our Foetal weight (FW) data set, each one corresponding to the inferior and superior tails of the FW distribution function, as shown in figures 1 and 2. The central and most frequent cases will not belong to our sub-sets. The experiment consists in the performance determination in a test separate set.

The predicted FW ($FW_{pred}$) was computed from two echographic features abdominal circumference (AC) and femur length (FL), in two different portions of the distribution function, with almost the same number of examples. The polynomial kernels used in this experiment were of order $\leq 7$. The $\varepsilon$-insensitive loss function used values of $\varepsilon >0.05$. The number of support vectors returned by our algorithm was $SV_{inf}=90.5\%$ and $SV_{sup}=96.5\%$, respectively in the inferior and superior tails of the FW distribution function.

Finally, the error rates we got were $E_{inf}=11.2\%$ $E_{sup}=10.0\%$, in the inferior and superior tails of the FW distribution function, respectively.

8 Conclusions

SVR is equivalent to maximizing the margin between training examples and the regression function. It is an alternative to other neural networks with training methods that optimize cost functions such as the mean square error, therefore it can be applied to FW estimation.

SVR is motivated by the statistical learning theory, which characterizes the performance of SVR learning using bounds on their ability to predict future data.

The training consists in solving a constrained quadratic optimization problem [4,5,9]. Among others, this implies that there is a unique optimal solution for each choice of the SVR parameters. This is unlike other learning machines, such as standard Neural Networks trained using backpropagation.
Acknowledgement

The authors would like to thank Steve Gunn and the Image Speech & Intelligent Systems Group, University of Southampton, United Kingdom, for letting us experiment the Matlab software developed for Support Vector Machines for Classification and Regression.

References


