Logistic discriminant analysis and probabilistic neural networks in estimating erosion risk

Anna Bartkowiak[1] and Niki Evelpidou[2]

[1] Department of Computer Science, Univ. of Wrocław, 51-151 Poland
[2] Remote Sensing Lab., Geology Dept., Univ. of Athens, 15-784, Greece

Abstract. Logistic discriminant function $f(\mathbf{x};\boldsymbol{\theta})$ may be viewed as the probability a posteriori, that the observation $\mathbf{x}$ belongs do class C1, which implies in our elaboration a high erosion risk, as opposed to class C2, denoting low erosion risk.
It is known that when the two classes are completely separated (which happens in our case), the maximum likelihood estimates of $\boldsymbol{\theta}$ do not exist.
We show, that when working with typical neural networks (perceptrons) we may get reasonable solutions separating the two classes, although the solutions are not unique and have some other deficiencies.

Keywords. Prediction of erosion risk, discriminant analysis, probabilistic neural networks

1 Problem and results

We consider 3 dimensional data containing variables relevant for determining erosion risk. We are given some data vectors which may be considered as typical representants of high and low erosion risk (the description of the data may be found in Gournellos et al, 2002).

Our task is to establish an index indicating – for a given territorial unit – its erosion risk.

Our first approach was to use for this purpose the logistic discriminant function $f(\mathbf{x};\boldsymbol{\theta})$. It yields the probability a posteriori, that a given observation $\mathbf{x}$ belongs to class C1, which in our case means the class of high erosion risk. This probability (belonging to the interval [0,1]) could serve as the indication of the erosion risk.

However, when estimating the parameters of the assumed logistic discriminant function, we have encountered the following obstacle: for our data the two classes C1 and C2 are completely separated, thus for such a situation the maximum likelihood estimates of the parameters $\boldsymbol{\theta}$ of the logistic discriminant function do not exist (see., e.g. Lesaffre and Albert, 1989; Christman and Rousseeuw, 2000).

Despite that result we may obtain a continuum of parameters $\boldsymbol{\theta}$ such, that $f(\mathbf{x};\boldsymbol{\theta})$ separates perfectly the two classes. The problem with these parameters is, that they give only values practically equal 1.0 or 0.0, thus do not contribute anything to a more clear differentiation between the analyzed units.

The logistic function plays an important role in neural networks, where it is termed *activation* function. Bishop (1995) has shown that a properly trained neural network with a logistic activation function yields at its output just the desired probabilities a posteriori.

Using the Netlab software for neural networks (Nabney, 2001), we have performed several trials for our data (perceptron with one hidden neuron, random initialization of the parameters).

A typical output is shown in the figure below.

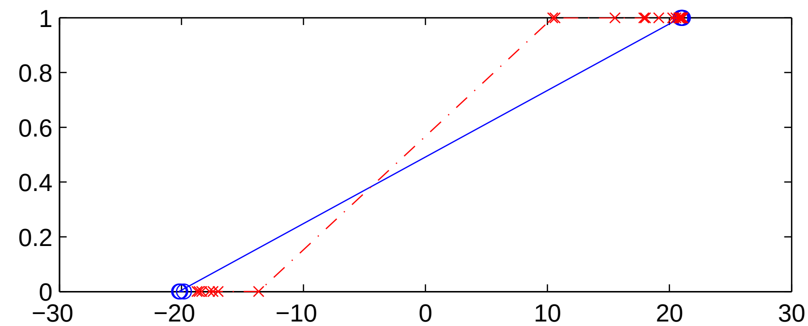

Fig. 1. Values of the *probability a posteriori* evaluated by neural network (perceptron), truly the values of the logistic discriminant function put against its input scores. Circles denote items with known low and high erosion risk used as training sample. Crosses denote items with known 'medium' risk – they were not used for training the network. The sharp differentiation of the output: either *one* either *zero* is clearly visible.

We stated the following:

– Depending on the initialization, we obtained differentiated estimates of θ, however yielding practically the same results,
– The magnitude of the parameters was dependent on the algorithm used for optimization: when using the regularized conjugate gradient method ('scg') the parameter values were much smaller as when using typical quasi-Newton (variable metric) method.
– A sharp differentiation of the output – either *one* either *zero* – as shown in the figure above.

The mentioned inconveniences may be alleviated, when using addition information from representatives of the *medium risk class*. We have worked along two paths: (1) – using a regression-like output in the network, (2) – using a non-parametric approach constructing densities by radial basis or kernel functions. The second approach needs additional information on the width of the kernels used. This can be estimated by a data-driven method using information from the medium class.

References

Bishop Ch.M. (1990). *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford UK. Reprinted twice 1996.

Christman A. and Rousseeuw P.J. (2000). Measuring overlap in logistic regression. In: *Compstat 2000 Proceedings in Computational Statistics, Short Communications and Posters*, 19–20. Statistics Netherlands.

Gournellos T., Evelpidou N. and Vassipolous A. (2002). Developing an erosion risk map using soft computing. Submitted.

Lesaffre E., Albert A. (1989). Partial separation in logistic discrimination. *J. R. Statist. Soc.* B, **51**, No. 1, 109–116.

Nabney I. (2002). *Netlab: Algorithms for Pattern Recognition.* Springer