Classification by Polynomial Surfaces

Martin Anthony
Department of Statistical and Mathematical Sciences
London School of Economics and Political Science
Houghton Street, London WC2A 2AE
anthony@vax.lse.ac.uk

Abstract

Linear threshold functions (for real and Boolean inputs) have received much attention, for they are the component parts of many artificial neural networks. Linear threshold functions are exactly those functions such that the positive and negative examples are separated by a hyperplane. One extension of this notion is to allow separators to be surfaces whose equations are polynomials of at most a given degree (linear separation being the degree-1 case). We investigate the representational and expressive power of polynomial separators. Restricting to the Boolean domain, by using an upper bound on the number of functions defined on $\{0,1\}^n$ by polynomial separators having at most a given degree, we show, as conjectured by Wang and Williams [26], that for almost every Boolean function, one needs a polynomial surface of degree at least $\lfloor n/2 \rfloor$ in order to separate the negative examples from the positive examples. Further, we show that, for odd n, at most half of all Boolean functions are realisable by a separating surface of degree $\lfloor n/2 \rfloor$. We then compute the Vapnik-Chervonenkis dimension of the class of functions realised by polynomial separating surfaces of at most a given degree, both for the case of Boolean inputs and real inputs. In the case of linear separators, the VC dimensions coincide for these two cases, but for surfaces of higher degree, there is a strict divergence. We then use these results on the VC dimension to quantify the sample size required for valid generalisation in Valiant's probably approximately correct framework [24, 6].

1 Introduction

A $\{0,1\}$-valued function f on $\mathbf{R}^n$ is said to be a *linear threshold function* if there are real numbers $w_1, w_2, \ldots, w_n$ and θ such that $f(x) = 1$ if and only if the inner product $\langle w, x\rangle = \sum_{i=0}^{n} w_i x_i$ is greater than θ. The coefficients w_i $(1 \leq i \leq n)$ are known as the *weights*, the vector w is known as the *weight-vector*, and θ is called the *threshold*. When $\theta = 0$, we say that t is a *homogeneous* linear threshold function. Linear threshold functions have been studied extensively: see, for example, [19, 13, 17, 15, 3]. Observe that a function is a linear threshold function if and only if the positive examples can be separated from the negative examples by a separating hyperplane. (In the case of a homogeneous linear threshold function, this hyperplane contains the origin.) One may consider different kinds of *separators*, generalising the notion of a linear threshold function. As in [7, 20, 26] (for example), one may consider separators which are quadratic, cubic, or in general multilinear. A surface in $\mathbf{R}^n$ is said to be a *polynomial discriminator* of order m if it can be described by a multinomial equation of degree m in the variables $x_1, x_2, \ldots, x_n$. For example, the set of discriminators of order 2 in $\mathbf{R}^3$ consists of all surfaces whose equations are of the form

$$w_1x_1 + w_2x_2 + w_3x_3 + w_4x_1^2 + w_5x_2^2 + w_6x_3^2 + w_7x_1x_2 + w_8x_1x_3 + w_9x_2x_3 = w_{10},$$

for some contants $w_i, (1 \leq i \leq 10)$, where at least one of the terms of order 2 has a non-zero coefficient. (This last condition ensures that the order of a discriminator is well-defined.)

Linear and polynomial discriminators have been studied in the the context of pattern classification (see, for example, [9, 8, 20]), where the aim is to classify a given set of test data points into two categories correctly, this classification being used as a (hopefully valid) means of classifying further points. Although we shall discuss only the method of separation by polynomial surfaces in this paper, it should be mentioned that other methods have usefully been applied when the desired classification of the data points cannot be achieved by linear separators; see, for example [16, 27]. The representation and approximation of boolean functions by polynomials has useful applications in a number of areas of computer science; a discussion of some of the problems studied and the techniques used may be found in the survey of Saks [23]. In addition, polynomial discriminators have recently been employed in signal processing [22]. It is therefore an important problem to determine the 'power' of classification achievable by such discriminators.

Here, we investigate the representational and expressive power of polynomial separators. Restricting to the Boolean domain, and bounding the number of functions defined on $\{0,1\}^n$ by polynomial separators having at most a given degree, we show that for almost every Boolean function, one needs a polynomial surface of degree at least $\lfloor n/2 \rfloor$ in order to separate the negative examples from the positive examples. We then compute the Vapnik-Chervonenkis dimension of the class of functions realised by polynomial separating surfaces of at most a given degree; we do this both for the case of Boolean inputs and real inputs. In the case of linear separators, the VC dimensions coincide for these two cases, but for surfaces of higher degree, there is a strict divergence. Having computed the VC dimensions, we obtain an indication of how large a set of test data should be used for valid further classification of previously unseen points (within the criteria of the 'probably approximately correct' model of machine learning [24, 6, 1]).