Approximating Pseudo-Boolean Functions on Non-Uniform Domains*

R. F. Lax[1]**, Guoli Ding**[1]**, Peter P. Chen**[2]**,** and **J. Chen**[2]
[1] Department of Mathematics, LSU, Baton Rouge, LA 70803
[2] Department of Computer Science, 298 Coates Hall, LSU, Baton Rouge, LA 70803
lax@math.lsu.edu, ding@math.lsu.edu, chen@bit.csc.lsu.edu, jianhua@bit.csc.lsu.edu,

Abstract

In Machine Learning (ML) and Evolutionary Computation (EC), it is often beneficial to approximate a complicated function by a simpler one, such as a linear or quadratic function, for computational efficiency or feasibility reasons (cf. [Jin, 2005]). A complicated function (the target function in ML or the fitness function in EC) may require an exponential amount of computation to learn/evaluate, and thus approximations by simpler functions are needed. We consider the problem of approximating pseudo-Boolean functions by simpler (e.g., linear) functions when the instance space is associated with a probability distribution. We consider $\{0,1\}^n$ as a sample space with a (possibly non-uniform) probability measure on it, thus making pseudo-Boolean functions into random variables. This is also in the spirit of the PAC learning framework of Valiant [Valiant, 1984] where the instance space has a probability distribution on it. The best approximation to a target function f is then defined as the function g (from all possible approximating functions of the simpler form) that minimizes the expected distance to f. In an example, we use methods from linear algebra to find, in this more general setting, the best approximation to a given pseudo-Boolean function by a linear function.

1 Introduction

A pseudo-Boolean function of n variables is a function from $\{0,1\}^n$ to the real numbers. Such functions are used in 0-1 optimization problems, cooperative game theory, multicriteria decision making, and as fitness functions. It is not hard to see that such a function $f(x_1,\ldots,x_n)$ has a unique expression as a multilinear polynomial

$$f(x_1,\ldots,x_n) = \sum_{T\subseteq N}\left[a_T \prod_{i\in T} x_i\right],$$

where $N=\{1,\ldots,n\}$ and the a_T are real numbers. By the *degree* of a pseudo-Boolean function, we mean the degree of its multilinear polynomial representation.

Several authors have considered the problem of finding the best pseudo-Boolean function of degree $\leq k$ approximating a given pseudo-Boolean function f, where "best" means a least squares criterion. Hammer and Holzman [Hammer and Holzman, 1992] derived a system of equations for finding such a best degree $\leq k$ approximation, and gave explicit solutions when $k=1$ and $k=2$. They proved that such an approximation is characterized as the unique function of degree $\leq k$ that agrees with f in all average m-th order derivatives for $m=0,1,\ldots,k$, in analogy with the Taylor polynomials from calculus. Grabisch, Marichal, and Roubens [Grabisch *et al.*, 2000] solve the system of equations derived by Hammer and Holzman, and give explicit formulas for the coefficients of the best degree $\leq k$ function. Zhang and Rowe [Zhang and Rowe, 2004] use linear algebra to find the best approximation that lies in a linear subspace of the space of pseudo-Boolean functions; for example, these methods can be used to find the best approximation of degree $\leq k$.

Here, instead of simply viewing the domain of a pseudo-Boolean function as the set $\{0,1\}^n$, we consider $\{0,1\}^n$ as a discrete sample space and introduce a probability measure on this space. Thus, a pseudo-Boolean function will be a random variable on this sample space. (Viewing $\{0,1\}$ simply as a set corresponds to viewing all of its points as equally likely outcomes.) Given a pseudo-Boolean random variable f, we then use methods from linear algebra to find the best approximation to f that lies in a linear subspace, taking into account the weighting of the elements of $\{0,1\}^n$. Such a best approximation will then be close to f at the "most likely" n-tuples, and may not be so close to f at the "least likely" n-tuples.

2 Best Approximation on a Non-Uniform Domain

We will identify the integers $0,1,\ldots,2^n-1$ with the elements in B^n via binary representation. Let $p(i), i=0,1,\ldots,2^n-1$, be a probability measure on B^n. Let $\mathcal{F}$ denote the space of all pseudo-Boolean functions in n variables. Then $\mathcal{F}$ has the structure of a real vector space. Define

*Research partially supported by NSF grant ITR-0326387 and AFOSR grants F49620-03-1-0238,F49620-03-1-0239, and F49620-03-1-0241

an inner product $\langle\ ,\ \rangle_p$ on $\mathcal{F}$ by

$$\langle f, g\rangle_p = \sum_{i=0}^{2^n-1} f(i)g(i)p(i).$$

We note that $\langle f, g\rangle_p$ is the expected value of the random variable fg. Put $\| f \|_p = \sqrt{\langle f, f\rangle_p}$.

Now let $\mathcal{L}$ be a vector subspace of $\mathcal{F}$ of dimension m. For example, $\mathcal{L}$ might be the space of all pseudo-Boolean functions of degree at most k, for some fixed k. We recall how to use an orthonormal basis of $\mathcal{L}$ to find the best approximation to a given element of $\mathcal{F}$ (cf. [Hoffman and Kunze, 1971]).

Let $v_1, \ldots, v_m$ be a basis for $\mathcal{L}$. We can find an orthonormal basis $u_1, \ldots, u_m$ for $\mathcal{L}$ by applying the Gram-Schmidt algorithm. This orthonormal basis satisfies the property $\langle u_r, u_s\rangle_p = \delta_{rs}$ for $r, s = 1, \ldots, m$, where δ_{rs} equals 0 if $r \neq s$ and equals 1 if $r = s$. The orthonormal basis can be obtained as follows: Take $u_1 = (1/\| v_1 \|_p)v_1$. If $u_1, \ldots, u_{r-1}$ have been obtained, then put $w_r = v_r - \sum_{j=1}^{r-1}\langle v_r, u_j\rangle_p u_j$, and take $u_r = (1/\| w_r \|_p)w_r$.

Given $f \in \mathcal{F}$, the "best approximation" to f by functions in $\mathcal{L}$ is that function $g \in \mathcal{L}$ that minimizes

$$\| f - g \|_p = \sqrt{\sum_{i=0}^{2^n-1} (f(i) - g(i))^2 p(i)}.$$

Notice that if we take the uniform distribution on B^n, so that $p(i) = (1/2)^n$ for all i, then the best approximation to f in $\mathcal{L}$ is the function $g \in \mathcal{L}$ that also minimizes $\sum_{i=0}^{2^n-1}(f(i) - g(i))^2$. This is the usual "least squares" condition used in [Hammer and Holzman, 1992], [Grabisch *et al.*, 2000], [Zhang and Rowe, 2004], and in this case one may simply use the usual Euclidean inner product in $\mathbb{R}^{2^n}$. In our more general setting, it follows from section 8.2 of [Hoffman and Kunze, 1971] that the best approximation to f by functions in $\mathcal{L}$ is the unique function $g = \sum_{j=1}^{m}\langle f, u_j\rangle_p u_j$.

3 Example

To illustrate these ideas, we look at an example considered by [Zhang and Rowe, 2004]. Take $n = 3$ and $f(x_1, x_2, x_3) = 5x_1 + 13x_3 + 9x_1x_2 - 4x_1x_3 - 4x_2x_3 + 4x_1x_2x_3$. We wish to approximate f by the best linear function, so we let $\mathcal{L}$ be the space spanned by the functions $v_1 = 1, v_2 = x_1, v_3 = x_2, v_4 = x_3$. If we take the uniform distribution on B^3, so that $p(i) = 1/8$ for $i = 0, 1, \ldots, 7$, then by applying the Gram-Schmidt algorithm we get the following orthonormal basis for $\mathcal{L}$ with respect to the inner product $\langle\ ,\ \rangle_p$:

$$u_1 = 1, u_2 = 2x_1 - 1, u_3 = 2x_2 - 1, u_4 = 2x_3 - 1.$$

(More generally, one can show that, for any n, an orthonormal basis for the space of pseudo-Boolean functions of degree at most 1 with respect to the uniform distribution is $1, 2x_1 - 1, \ldots, 2x_n - 1$.) Then the best linear approximation to f is $g(x_1, x_2, x_3) = \sum_{j=1}^{4}\langle f, u_j\rangle_p u_j =$

$$\begin{aligned} &= \frac{39}{4}\cdot 1 + \frac{17}{4}(2x_1 - 1) + \frac{7}{4}(2x_2 - 1) + 5(2x_3 - 1) \\ &= -\frac{5}{4} + \frac{17}{2}x_1 + \frac{7}{2}x_2 + 10x_3, \end{aligned}$$

in agreement with Example 4.1 of [Zhang and Rowe, 2004]. Here, $\| f - g \|_p \approx 2.88$.

Now we take a different probability measure on B^3. Supposing that a "1" is twice as likely as a "0" we define a probability measure $\tilde{p}$ on B^3 by $\tilde{p}(0) = 1/27, \tilde{p}(1) = 2/27, \tilde{p}(2) = 2/27, \tilde{p}(3) = 4/27, \tilde{p}(4) = 2/27, \tilde{p}(5) = 4/27, \tilde{p}(6) = 4/27, \tilde{p}(7) = 8/27$. An orthonormal basis for $\mathcal{L}$ with respect to the inner product $\langle\ ,\ \rangle_{\tilde{p}}$ is then

$$\tilde{u}_1 = 1, \tilde{u}_2 = \frac{3x_1 - 2}{\sqrt{2}}, \tilde{u}_3 = \frac{3x_2 - 2}{\sqrt{2}}, \tilde{u}_4 = \frac{3x_3 - 2}{\sqrt{2}}.$$

Then the best linear approximation to f is now $\tilde{g}(x_1, x_2, x_3) = \sum_{j=1}^{4}\langle f, \tilde{u}_j\rangle_{\tilde{p}}\tilde{u}_j =$

$$\begin{aligned} &= \frac{368}{27}\cdot 1 + \frac{91\sqrt{2}}{27}\tilde{u}_2 + \frac{46\sqrt{2}}{27}\tilde{u}_3 + \frac{85\sqrt{2}}{27}\tilde{u}_4 \\ &= (1/27)(-76 + 273x_1 + 138x_2 + 255x_3). \end{aligned}$$

Here, $\| f - \tilde{g} \|_{\tilde{p}} \approx 2.55$. For comparison, the distance now between the linear function g we found above and the function f is $\| f - g \|_{\tilde{p}} \approx 2.79$.

4 Conclusion

Instead of considering $B^n = \{0, 1\}^n$ simply as a set, we allow it to be viewed as a sample space wth a probability measure p. Then pseudo-Boolean functions are random variables on this sample space. Given a complicated pseudo-Boolean function, it is natural to want to approximate it by a simpler function, for example a linear or quadratic function. As an example, we found the best linear approximation to a given pseudo-Boolean function in three variables with respect to two different probability measures on B^3. Further research is needed to find an effective method of computing the best approximation on a non-uniform domain when the number of variables is large.

References

[Grabisch *et al.*, 2000] Michel Grabisch, Jean-Luc Marichal, and Marc Roubens. Equivalent representations of set functions. *Mathematics of Operations Research*, 25(2):157–178, May 2000.

[Hammer and Holzman, 1992] P. L. Hammer and R. Holzman. Approximations of pseudo-boolean functions; applications to game theory. *ZOR–Methods and Models of Operations Research*, 36:3–21, 1992.

[Hoffman and Kunze, 1971] Kenneth Hoffman and Ray Kunze. *Linear Algebra, 2nd edition*. Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

[Jin, 2005] Yaochu Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft Computing*, 9(1):3–12, 2005.

[Valiant, 1984] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[Zhang and Rowe, 2004] Hang Zhang and Jonathan E. Rowe. Best approximations of fitness functions of binary strings. *Natural Computing*, 3(1):113–124, 2004.