Journal of Mathematics Research; Vol. 5, No. 1; 2013
ISSN 1916-9795 E-ISSN 1916-9809
Published by Canadian Center of Science and Education

Note on the Rademacher-Walsh Polynomial Basis Functions

Hamse Y. Mussa[1], Jonathan D. Tyzack[1] & Robert C. Glen[1]

[1] Unilever Centre for Molecular Science Informatics, Department of Chemistry, Lensfield Road, Cambridge, United Kingdom

Correspondence: H. Y. Mussa, Unilever Centre for Molecular Science Informatics, Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW, United Kingdom. Tel: 44-1223-763-854. E-mail: hym21@cam.ac.uk

Received: October 22, 2012 Accepted: January 30, 2013 Online Published: February 20, 2013

doi:10.5539/jmr.v5n1p114 URL: http://dx.doi.org/10.5539/jmr.v5n1p114

Unilever Centre for Molecular Science Informatics, Department of Chemistry

Abstract

Over the years, one of the methods of choice to estimate probability density functions for a given random variable (defined on binary input space) has been the expansion of the estimation function in Rademacher-Walsh Polynomial basis functions. For a set of L features (often considered as an "L-dimensional binary vector"), the Rademacher-Walsh Polynomial approach requires 2^L basis functions. This can quickly become computationally complicated and notationally clumsy to handle whenever the value of L is large. In current pattern recognition applications it is often the case that the value of L can be 100 or more.

In this paper we show that the expansion of the probability density function estimation in Rademacher-Walsh Polynomial basis functions is equivalent to the expansion of the estimation function in a set of Dirac kernel functions. The latter approach is not only able to eloquently allay the computational bottle–neck and notational awkwardness mentioned above, but may also be naturally neater and more "elegant" than the Rademacher-Walsh Polynomial basis function approach even when this latter approach is computationally feasible.

Keywords: Rademacher-Walsh, Dirac kernel function, binary input spaces

1. Introduction

When $\mathbf{x}$ is an "L-dimensional binary vector" whose components can take binary values (0 or 1), the probability density function, $p(\mathbf{x})$, for $\mathbf{x}$ can be approximated by using a set of basis functions. It is often the case that $p(\mathbf{x})$ is estimated through 2^L Rademacher-Walsh Polynomial basis functions φ_i (Note 1) (Duda & Hart, 1973; Hand, 1981) as

$$p(\mathbf{x}) = \sum_{i=0}^{2^L-1} \alpha_i \varphi_i(\mathbf{x}) \tag{1}$$

where

$$\alpha_i = \frac{1}{2^L} \sum_{\mathbf{x}\in\mathcal{B}}^{N} p(\mathbf{x})\varphi_i(\mathbf{x}) \tag{2}$$

and N refers to the number of available samples, $\{\mathbf{x}_j\}_{j=1}^N$, drawn from the underlying probability distribution being estimated. The coefficients α_i can be viewed as moments (Duda & Hart, 1973), which can be estimated as

$$\hat{\alpha}_i = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{2^L} \varphi_i(\mathbf{x}_j) \tag{3}$$

Throughout the paper it is assumed that the L-dimensional random variables reside in a binary input space $\mathcal{B}$ with $\mathcal{B} = \{0, 1\}^L$ and the descriptions and notations given in (Duda & Hart, 1973) are adopted.

If the available N samples are distinct instances and $N = 2^L$, the estimated coefficients $\hat{\alpha}_i$ are exact (Tou & Gonzalez, 1974). However, exact or not, the expansion in Equation 1 requires 2^L Rademacher-Walsh Polynomial basis functions, which can make the estimation notationally clumsy and computationally complicated whenever the value of L is large (Duda & Hart, 1973).

In passing we point out that one can employ a subset of the 2^L Rademacher-Walsh Polynomials in the expansion, but this may result in an estimate of $p(\mathbf{x})$-*i.e.*,$\hat{p}(\mathbf{x})$-that can have negative values (Hand, 1981, pp. 106).

Putting Equation 3 into Equation 1 yields (Meisel, 1972)

$$\begin{aligned}\hat{p}(\mathbf{x}) &= \sum_{i=0}^{2^L-1} \frac{1}{N} \sum_{j=1}^{N} \frac{1}{2^L} \varphi_i(\mathbf{x}_j)\varphi_i(\mathbf{x}) \\ &= \frac{1}{N} \sum_{j=1}^{N} \sum_{i=0}^{2^L-1} \frac{\varphi_i(\mathbf{x}_j)}{\sqrt{2^L}} \frac{\varphi_i(\mathbf{x})}{\sqrt{2^L}} \\ &= \frac{1}{N} \sum_{j=1}^{N} K(\mathbf{x}_j, \mathbf{x}) \qquad (4)\end{aligned}$$

where

$$K(\mathbf{x}_j, \mathbf{x}) = \sum_{i=0}^{2^L-1} \frac{\varphi_i(\mathbf{x}_j)}{\sqrt{2^L}} \frac{\varphi_i(\mathbf{x})}{\sqrt{2^L}} \qquad (5)$$

Clearly, for all practical purposes, $L << \infty$; besides $\varphi_i(\mathbf{x}_j)$ and $\varphi_i(\mathbf{x})$ can only take values 1, or -1 as illustrated by Note 1. Hence

$$\sum_{i=0}^{2^L-1} \| \varphi_i(\mathbf{x}) \|^2 < \infty, \ \forall \mathbf{x} \in \mathcal{B}$$

which means that $K(\mathbf{x}_j, \mathbf{x})$ is a valid positive definite kernel function (Aronszajn, 1950; Shawe-Taylor & Cristianini, 2004) in $\mathcal{B} \times \mathcal{B}$.

In Equation 4, the estimation of $p(\mathbf{x})$ at $\mathbf{x}$ can be instructively viewed as an average of how similar $\mathbf{x}$ is to the given N samples $\mathbf{x}_j$, where $K(\mathbf{x}_j, \mathbf{x})$ is the similarity function (*cf.* the popular Parzen Window approach (Parzen, 1962)). For the expression in Equation 4 to have any practical use, knowledge of the closed form of the kernel function $K(\mathbf{x}_j, \mathbf{x})$ is essential. In the following section we formulate a theorem stating that $K(\mathbf{x}_j, \mathbf{x})$ in Equation 5 is a Dirac kernel function (Jacob & Vert, 2008). In this section we also develop the tools necessary for proving this theorem. The full proof of the theorem is presented in Section 3 followed by our concluding remarks in the final section.

For notational simplicity, in this work $\mathbf{x}$ denotes both the random vector and the values it may assume.

2. Method

This section introduces a theorem which constitutes the core of this paper-that is, $K(\mathbf{x}_j, \mathbf{x})$ is a Dirac kernel function. Also in this section tools (e.g., definitions, lemmas, propositions and remarks) that are essential for proving the theorem are presented.

Theorem 1 *If* $\mathbf{x}$ *and* $\mathbf{x}_j \in \mathcal{B}$ *with* $\mathcal{B} = \{0, 1\}^L$, *and* $\varphi_i(\cdot)$ *are Rademacher-Walsh Polynomial basis functions defined on* $\mathcal{B}$*, then*

$$K(\mathbf{x}_j, \mathbf{x}) = \sum_{i=0}^{2^L-1} \frac{\varphi_i(\mathbf{x}_j)}{\sqrt{2^L}} \frac{\varphi_i(\mathbf{x})}{\sqrt{2^L}} = \begin{cases} 1 & \mathbf{x}_j = \mathbf{x} \\ 0 & \mathbf{x}_j \neq \mathbf{x} \end{cases} \qquad (6)$$

i.e., $K(\boldsymbol{x}_j, \boldsymbol{x})$ *is a Dirac kernel function.*

Where $\mathbf{x}_j = \mathbf{x}$ means that $x_{j1} = x_1, x_{j2} = x_2, ..., x_{jL} = x_L$, with x_{jl} and x_l referring to the binary-valued l^{th} elements of $\mathbf{x}_j$ and $\mathbf{x}$, respectively. (Equation 6 can be viewed as showing that the basis functions $\frac{\varphi_i}{\sqrt{2^L}}$ satisfy the "orthonormality" relation in the arguments $\mathbf{x}_i$ and $\mathbf{x}$.)

As described in the Introduction, the set $\{\varphi_i(\mathbf{x})\}_{i=0}^{2^L-1}$ is obtained by systematically forming products of $(2x_l - 1)$ none at a time, one at a time, two at a time, three at a time, etc., where $l = 1, 2, ..., L$. By the same token the set $\{\varphi_i(\mathbf{x}_j)\varphi_i(\mathbf{x})\}_{i=0}^{2^L-1}$ is obtained by forming products of the distinct terms $(2x_{jl} - 1)(2x_l - 1)$ none at a time, one at a time, two at a time, three at a time, and so on:

$$\begin{aligned}
\varphi_0(\mathbf{x}_j)\varphi_0(\mathbf{x}) &= 1\\
\varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x}) &= (2x_{j1}-1)(2x_1-1)\\
\varphi_2(\mathbf{x}_j)\varphi_2(\mathbf{x}) &= (2x_{j2}-1)(2x_2-1)\\
\varphi_3(\mathbf{x}_j)\varphi_3(\mathbf{x}) &= (2x_{j3}-1)(2x_3-1)\\
&.\\
&.\\
&.\\
\varphi_L(\mathbf{x}_j)\varphi_L(\mathbf{x}) &= (2x_{jL}-1)(2x_L-1)\\
\varphi_{L+1}(\mathbf{x}_j)\varphi_{L+1}(\mathbf{x}) &= [\varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x})][\varphi_2(\mathbf{x}_j)\varphi_2(\mathbf{x})]\\
\varphi_{L+2}(\mathbf{x}_j)\varphi_{L+2}(\mathbf{x}) &= [\varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x})][\varphi_3(\mathbf{x}_j)\varphi_3(\mathbf{x})]\\
\varphi_{L+3}(\mathbf{x}_j)\varphi_{L+3}(\mathbf{x}) &= [(\varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x})][\varphi_4(\mathbf{x}_j)\varphi_4(\mathbf{x})]\\
&.\\
&.\\
&.\\
\varphi_{2^L-1}(\mathbf{x}_j)\varphi_{2^L-1}(\mathbf{x}) &= [\varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x})][\varphi_2(\mathbf{x}_j)\varphi_2(\mathbf{x})]...[\varphi_L(\mathbf{x}_j)\varphi_L(\mathbf{x})]
\end{aligned} \tag{7}$$

Remark 1 By definition $\varphi_0(\mathbf{x}_j)\varphi_0(\mathbf{x}) = 1$; and self-evidently $\varphi_i(\mathbf{x}_j)\varphi_i(\mathbf{x})$ (where $1 \le i \le L$) can only take the values of 1 or -1. This also means that $\varphi_i(\mathbf{x}_j)\varphi_i(\mathbf{x})$ can only take the values of 1 or -1, where $L+1 \le i \le 2^L - 1$.

Before we embark on proving Theorem 1, we show that the following lemma (Lemma 1) holds.

Lemma 1 *Let $a_1, a_2, ..., a_L$ be L distinguishable real variables which can take the values of 1 and -1, and assume that combinatorial compositions can be considered as products, i.e., $a_k a_j$ means $a_k \times a_j$, where $k, j = 1, 2, ..., L$. The sum of their possible combinatorial compositions z_i, with $i = 0, 1, ..., 2^L - 1$, then gives:*

$$\sum_{i=0}^{2^L-1} z_i = \begin{cases} 2^L & if\ a_1, a_2, ..., a_L = 1 \\ 0 & if\ not \end{cases} \tag{8}$$

Proof. The possible combinations are the L variables chosen: none at a time; 1 variable, a_i, at a time; 2 variables, $a_i a_j$, at a time; three variables,$a_i a_j a_k$, at a time;,...,; or L variables, $a_1 a_2 ... a_L$, at a time.

There are three possible scenarios:

Scenario (1): All the L variables are positive, i.e., $a_k = +1$, where $k = 1, 2, .., L$.

Let $z_0 = +1$ (when none is chosen); $z_1 = a_1 = +1, z_2 = a_2 = +1, ..., z_L = a_L = +1; z_{L+1} = a_1 a_2 = +1, z_{L+2} = a_1 a_3 = +1, ..., z_{L+\frac{L(L-1)}{2}} = a_{L-1} a_L = +1$; ...; and $z_{2^L-1} = a_1 a_2 ... a_L = 1$.

Clearly, the number of times that none of the variable is chosen is $\binom{L}{0}$, which can also written as ${}^L C_0$. The number of combinatorial terms containing one variable is ${}^L C_1 = \binom{L}{1}$. Similarly the number combinatorial terms consisting of two, three, four, ..., and L variables are ${}^L C_2 = \binom{L}{2}$, ${}^L C_3 = \binom{L}{3}$, ${}^L C_4 = \binom{L}{4}$, ..., and ${}^L C_L = \binom{L}{L}$, respectively.

In other words

$$\sum_{i=0}^{2^L-1} z_i = \sum_{\varrho=0}^{L} {}^L C_\varrho = \sum_{\varrho=0}^{L} \binom{L}{\varrho} \tag{9}$$

But from the binomial expansion theorem, $\sum_{\varrho=0}^{L} \binom{L}{\varrho} = 2^L$, which means

$$\sum_{i=0}^{2^L-1} z_i = 2^L \tag{10}$$

The use of both ${}^{L}C_{\varrho}$ and $\binom{L}{\varrho}$ may seem somewhat superfluous, but the reason for this will become clear in the following discussions.

Scenario (2): All the L variables take the value of -1, i.e., a_k = -1, where k is as defined before.

Let $z_0 = +1$; $z_1 = a_1 = (-1)^1, z_2 = a_2 = (-1)^1, \dots, z_L = a_L = (-1)^1; z_{L+1} = a_1a_2 = (-1)^2, z_{L+2} = a_1a_3 = (-1)^2, \dots, z_{L+\frac{L(L-1)}{2}} = a_{L-1}a_L = (-1)^2; \dots$; and $z_{2^L-1} = a_1a_2...a_L = (-1)^L$. By the same token (as we reasoned above): ${}^{L}C_0 = 1 = (-1)^0\binom{L}{0}; {}^{L}C_1 = (-1)^1\binom{L}{1}; {}^{L}C_2 = (-1)^2\binom{L}{2}; \dots$, *i.e.*, ${}^{L}C_{\varrho} = (-1)^{\varrho}\binom{L}{\varrho}$ with $\varrho = 0, 1, 2, \dots, L$. This means

$$\sum_{i=0}^{2^L-1} z_i = \sum_{\varrho=0}^{L} {}^{L}C_{\varrho} = \sum_{\varrho=0}^{L}(-1)^{\varrho}\binom{L}{\varrho}. \tag{11}$$

$\sum_{\varrho=0}^{L}(-1)^{\varrho}\binom{L}{\varrho}$ can be expressed as $\sum_{\varrho=0}^{L}\binom{L}{\varrho}1^{L-\varrho}(-1)^{\varrho}$, where $1 = (1)^{L-\varrho}$. But $\sum_{\varrho=0}^{L}\binom{L}{\varrho}1^{L-\varrho}(-1)^{\varrho}$ is basically the binomial expansion of $(1-1)^L$. Thus,

$$\sum_{i=0}^{2^L-1} z_i = \sum_{\varrho=0}^{L} {}^{L}C_{\varrho} = \sum_{\varrho=0}^{L}\binom{L}{\varrho}1^{L-\varrho}(-1)^{\varrho} = (1-1)^L = 0 \tag{12}$$

Scenario (3): Some of the L variables assume the values of -1 while the remaining variables take the value of 1.

For no specific reason and loss of generality, let us consider that *m* and *k* denote the number of variables that take the values -1 and 1, respectively, where $L = m + k$.

In this scenario one is required to demonstrate that

$$\sum_{i=0}^{2^L-1} z_i = \sum_{\varrho=0}^{m+k} {}^{m+k}C_{\varrho} = 0 \tag{13}$$

Fortunately, Equation 13 can be readily proved by use of induction providing ${}^{m+k}C_{\varrho}$ is expressed in terms of ${}^{m}C_{\varrho}$'s. In this case it is germane to recall the following important identities where where *r*, *j*, *n* are non-negative integers and $r \le n$ (Riley et al., 2007):

I: ${}^{n}C_r = {}^{n-1}C_r + {}^{n-1}C_{r-1}$,

II: ${}^{n+r}C_{n+r} = {}^{n+r-1}C_{n+r-1}$, and

III: ${}^{n+j}C_0 = {}^{n+j-1}C_0$,

with *k = 1*, *i.e.*, $\sum_{\varrho=0}^{m+k} {}^{m+k}C_{\varrho}$ becomes $\sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho}$, which can be expressed as

$$\sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} = {}^{m+1}C_0 + \sum_{\varrho=1}^{m} {}^{m+1}C_{\varrho} + {}^{m+1}C_{m+1}$$

Making use of Identity I, the ${}^{m+1}C_{\varrho}$ on the RHS of the equation above becomes ${}^{m}C_{\varrho} + {}^{m}C_{\varrho-1}$, *i.e.*, the equation can be rewritten as

$$\begin{aligned}\sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} &= {}^{m+1}C_0 + \sum_{\varrho=1}^{m} {}^{m}C_{\varrho} + {}^{m}C_{\varrho-1} + {}^{m+1}C_{m+1} \\ &= {}^{m+1}C_0 + \sum_{\varrho=1}^{m} {}^{m}C_{\varrho} + \sum_{\varrho=1}^{m} {}^{m}C_{\varrho-1} + {}^{m+1}C_{m+1}\end{aligned}$$

This equation can be modified further by applying Identities III and II to the first and last terms on the RHS of the last equation, respectively, resulting in

$$\sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} = {}^{m}C_0 + \sum_{\varrho=1}^{m} {}^{m}C_{\varrho} + \sum_{\varrho=1}^{m} {}^{m}C_{\varrho-1} + {}^{m}C_m$$

By expressing $\sum_{\varrho=1}^{m} {}^{m}C_{\varrho-1}$ as ${}^{m}C_0 + {}^{m}C_1 + ... + {}^{m}C_{m-1}$, evidently the equation above can be written as

$$\begin{aligned}\sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} &= {}^{m}C_0 + \sum_{\varrho=1}^{m} {}^{m}C_{\varrho} + {}^{m}C_0 + {}^{m}C_1 + ... + {}^{m}C_{m-1} + {}^{m}C_m \\ &= \sum_{\varrho=0}^{m} {}^{m}C_{\varrho} + {}^{m}C_0 + {}^{m}C_1 + ... + {}^{m}C_{m-1} + {}^{m}C_m \\ &= \sum_{\varrho=0}^{m} {}^{m}C_{\varrho} + \sum_{\varrho=0}^{m} {}^{m}C_{\varrho}\end{aligned}$$

In Scenario (2) we have demonstrated that in the case that all the variables (denoted here by m) take the value of -1, ${}^{m}C_{\varrho} = (-1)^{\varrho}\binom{m}{\varrho}$. This means

$$\begin{aligned}\sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} &= \sum_{\varrho=0}^{m} {}^{m}C_{\varrho} + \sum_{\varrho=0}^{m} {}^{m}C_{\varrho} \\ &= \sum_{\varrho=0}^{m}(-1)^{\varrho}\binom{m}{\varrho} + \sum_{\varrho=0}^{m}(-1)^{\varrho}\binom{m}{\varrho} \\ &= 2\sum_{\varrho=0}^{m}(-1)^{\varrho}\binom{m}{\varrho} \qquad (14)\end{aligned}$$

In the case of k =2, $\sum_{\varrho=0}^{m+k} {}^{m+k}C_{\varrho}$ becomes $\sum_{\varrho=0}^{m+2} {}^{m+2}C_{\varrho}$, which can be expressed as

$$\sum_{\varrho=0}^{m+2} {}^{m+2}C_{\varrho} = {}^{m+2}C_0 + \sum_{\varrho=1}^{m+1} {}^{m+2}C_{\varrho} + {}^{m+2}C_{m+2}$$

Applying Identity I to ${}^{m+2}C_{\varrho}$ in the middle term on the RHS of the equation above, we obtain

$$\begin{aligned}\sum_{\varrho=0}^{m+2} {}^{m+2}C_{\varrho} &= {}^{m+2}C_0 + \sum_{\varrho=1}^{m+1} {}^{m+1}C_{\varrho} + {}^{m+1}C_{\varrho-1} + {}^{m+2}C_{m+2} \\ &= {}^{m+2}C_0 + \sum_{\varrho=1}^{m+1} {}^{m+1}C_{\varrho} + \sum_{\varrho=1}^{m+1} {}^{m+1}C_{\varrho-1} + {}^{m+2}C_{m+2}\end{aligned}$$

By following the same line of reasoning as employed in the case of k=1, we can modify the equation above further. Applying Identities III and II to the first and last terms on the RHS of the equation above, respectively, gives

$$\begin{aligned}\sum_{\varrho=0}^{m+2} {}^{m+2}C_{\varrho} &= {}^{m+1}C_0 + \sum_{\varrho=1}^{m+1} {}^{m+1}C_{\varrho} + \sum_{\varrho=1}^{m+1} {}^{m+1}C_{\varrho-1} + {}^{m+1}C_{m+1} \\ &= \sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} + \sum_{\varrho=1}^{m+1} {}^{m+1}C_{\varrho-1} + {}^{m+1}C_{m+1} \\ &= \sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} + {}^{m+1}C_0 + {}^{m+1}C_1 + ... + {}^{m+1}C_m + {}^{m+1}C_{m+1} \\ &= \sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} + \sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} \\ &= 2\sum_{\varrho=0}^{m+1} {}^{m+1}C_{\varrho} \qquad (15)\end{aligned}$$

By the virtue of Equation 14, $2\sum_{\varrho=0}^{m+1} {}^{m+1}C_\varrho$ can be written as

$$2\sum_{\varrho=0}^{m+1} {}^{m+1}C_\varrho = 2\Big[2\sum_{\varrho=0}^{m}(-1)^\varrho\binom{m}{\varrho}\Big] = 2^2\sum_{\varrho=0}^{m}(-1)^\varrho\binom{m}{\varrho} \tag{16}$$

For $k = k$, one just needs to repeat the process above k times, which gives

$$\sum_{\varrho=0}^{m+k} {}^{m+k}C_\varrho = 2^k\sum_{\varrho=0}^{m}(-1)^\varrho\binom{m}{\varrho}$$

In Scenario (2) it was shown that $\sum_{\varrho=0}^{m}(-1)^\varrho\binom{m}{\varrho} = 0$ when all m variables take the value -1. (Note that in Scenario (2) $L = m$).

Hence

$$\sum_{\varrho=0}^{m+k} {}^{m+k}C_\varrho = 2^k\Big[\sum_{\varrho=0}^{m}(-1)^\varrho\binom{m}{\varrho} = 0\Big] = 0$$

which is basically the RHS of Equation 13.

$$\sum_{i=0}^{2^L-1} z_i = \sum_{\varrho=0}^{m+k} {}^{m+k}C_\varrho = 2^k\sum_{\varrho=0}^{m}(-1)^\varrho\binom{m}{\varrho} = 0 \tag{17}$$

This finalizes the proof of Lemma 1.

3. Results

In the preceding section, we have attempted to develop the essential tools for proving the proposed theorem, Theorem 1. In this section the full proof of the theorem is given.

Proof of Theorem 1.

Recall that (by the virtue of Remark 1) $\varphi_0(\mathbf{x}_j)\varphi_0(\mathbf{x}) = 1$ and the terms $\varphi_i(\mathbf{x}_j)\varphi_i(\mathbf{x})$ take the values +1 or -1, where $i = 1, 2, ..., L$.

Now, if we consider $\varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x})$, $\varphi_2(\mathbf{x}_j)\varphi_2(\mathbf{x})$, ..., and $\varphi_L(\mathbf{x}_j)\varphi_L(\mathbf{x})$ as the real L variables in Lemma 1, then (see Equation 7)

$$z_0 = \varphi_0(\mathbf{x}_j)\varphi_0(\mathbf{x}),$$

$$z_1 = \varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x}),$$

$$.$$

$$.$$

$$.$$

$$z_{2^L-1} = \varphi_{2^L-1}(\mathbf{x}_j)\varphi_{2^L-1}(\mathbf{x}) = [\varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x})][\varphi_2(\mathbf{x}_j)\varphi_2(\mathbf{x})]...[\varphi_L(\mathbf{x}_j)\varphi_L(\mathbf{x})].$$

Then by the virtue of Lemma 1,

$$\sum_{i=0}^{2^L-1} z_i = \sum_{i=0}^{2^L-1}\varphi_i(\mathbf{x}_j)\varphi_i(\mathbf{x}) = \begin{cases} 2^L & \text{if } \varphi_1(\mathbf{x}_j)\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}_j)\varphi_2(\mathbf{x}), ..., \varphi_L(\mathbf{x}_j)\varphi_L(\mathbf{x}) = 1 \\ 0 & if\ not \end{cases} \tag{18}$$

(Recall that only if $\mathbf{x} = \mathbf{x}_j$, the elements of the set $\{\varphi_i(\mathbf{x}_j)\varphi_i(\mathbf{x})\}_{i=1}^{L}$ take the value of +1.)

Multiplying on both sides of Equation 18 by $\frac{1}{\sqrt{2^L}}\frac{1}{\sqrt{2^L}}$ yields

$$\sum_{i=0}^{2^L-1} \frac{\varphi_i(\mathbf{x}_j)}{\sqrt{2^L}} \frac{\varphi_i(\mathbf{x})}{\sqrt{2^L}} = \begin{cases} 1 & if\ \mathbf{x}_j = \mathbf{x} \\ 0 & if\ \mathbf{x}_j \neq \mathbf{x} \end{cases} \tag{19}$$

which is Equation 6 and this completes the proof of Theorem 1.

4. Summary

For a long time, expansion in Rademacher-Walsh Polynomial basis functions has been the method of choice to estimate the probability density function for given random variables/features defined on binary input space. For a set of L features, the Rademacher-Walsh Polynomial approach requires 2^L basis functions, which can quickly become notationally clumsy and computationally difficult to handle whenever the value of L is large. In realistic pattern recognition applications, the value of L can be 100 or more.

In this paper we have demonstrated that, on binary descriptor space, the expansion of the probability density estimation function in Rademacher-Walsh Polynomial basis functions is equivalent to the expansion of the estimation function in a set of Dirac kernel functions. The probability density estimation based on the Dirac kernel function scheme certainly alleviates both the computational bottle-necks and notational complexity associated with the Rademacher-Walsh Polynomial basis function approach as discussed in the preceding sections.

Therefore it is hoped that the statistical and machine learning communities find the proposed theorem and its proof highly useful when it comes to estimating probability density functions on binary input spaces.

Acknowledgements

The authors would like to thank Unilever for financial support.

References

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc., 68*, 337-404. Retrieved from http://www.ams.org/journals/tran/1950-068-03/S0002-9947-1950-0051437-7/

Duda, R, O., & Hart, P. E. J. (1973). *Pattern and Scene Analysis* (1st ed., Chapter 4). New York, US: John Wiley & Sons.

Hand, D. J. (1981). *Discrimination and Classification* (1st ed., pp. 106). Chichester, UK: John Wiley & Sons.

Jacob, L., & Vert, J. (2008). Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics, 24*, 2149-2156. http://dx.doi.org/10.1093/bioinformatics/btn409

Meisel, W. S. (1972). *Computer-Oriented Approaches to Pattern Recognition* (1st ed., pp. 106). London, UK: Academic Press.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics, 33*, 1065-1076. http://dx.doi.org/10.1214/aoms/1177704472

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel Methods for Pattern Analysis* (1st ed., pp. 60-66). Cambridge, UK: Cambridge University Press.

Tou, J. R., & Gonzalez, R. C. (1974). *Pattern Recognition Principles* (1st ed., pp. 152-153). New York, US: Addison-Wesley.

Riley, K. F., Hobson, M. P., & Bence, S. J. (2007). *Mathematical Methods for Physics and Engineering* (3rd ed., p. 26). Cambridge, UK: Cambridge University Press.

Notes

Note 1. According to Duda and Hart (1973) this basis function set consists of a set of polynomials that can be generated by systematically forming the products of the distinct terms $2x_i - 1$ taken none at a time, one at a time, two at a time, three at a time, and so on, as follows:

$$\varphi_i(\mathbf{x}) = \begin{cases} 1 & i = 0 \\ 2x_1 - 1 & i = 1 \\ . & . \\ . & . \\ . & . \\ 2x_L - 1 & i = L \\ (2x_1 - 1)(2x_2 - 1) & i = L + 1 \\ . & . \\ . & . \\ . & . \\ (2x_{L-1} - 1)(2x_L - 1) & i = L + 1 + \frac{L(L-1)}{2} \\ (2x_1 - 1)(2x_2 - 1)(2x_3 - 1) & i = L + 2 + \frac{L(L-1)}{2} \\ . & . \\ . & . \\ . & . \\ (2x_1 - 1)(2x_2 - 1)...(2x_L - 1) & i = 2^L - 1. \end{cases}$$

where $\mathbf{x} = (x_1, x_2, ..., x_L)$. The set forms a complete basis set satisfying an orthogonality relation in their order – *i.e.*, $\varphi_i(\mathbf{x})$ and $\varphi_k(\mathbf{x})$ – with respect to the weighting function $w(\mathbf{x}) = 1$,

$$\sum_{\mathbf{x}} \varphi_i(\mathbf{x})\varphi_k(\mathbf{x}) = \begin{cases} 2^L & i = k \\ 0 & i \neq k \end{cases}$$

where the summation is taken over all 2^L values of the binary vectors.