Approximation of Functions in Barron's Space Using Sums of Separable Functions

OHIO UNIVERSITY

Indupama Herath and Dr. Martin J. Mohlenkamp

Department of Mathematics , Ohio University.

ABSTRACT

High dimensional functions occur in many machine learning examples. Approximating them is a challenging problem due to many methods that are available suffer from the curse of dimensionality. Andrew R. Barron in 1993 proved dimension independent approximation bounds for functions with certain Fourier conditions. He showed that single layer neural networks with sigmoidal activation functions can achieve approximation error of order $\mathcal{O}(1/n)$, where n is the number of nodes in the hidden layer, for functions in Barron's space. Sums of separable functions is an alternative method based on constructing a multivariate function as a product of univariate functions. In this study we show that functions in Barron's space can be approximated with the same approximation rate using sums of separable functions with complex exponentials. Approximation power of both methods in approximating functions in this space is further discussed.

1. Background

Consider approximating a real valued continuous function f on a d dimensional rectangular grid. Usually the number of parameters n needed to approximate f is given as $\mathcal{O}(\epsilon^{-d})$. The exponential dependence of n on d is known as the ***curse of dimensionality*** [1] and is unavoidable in general.

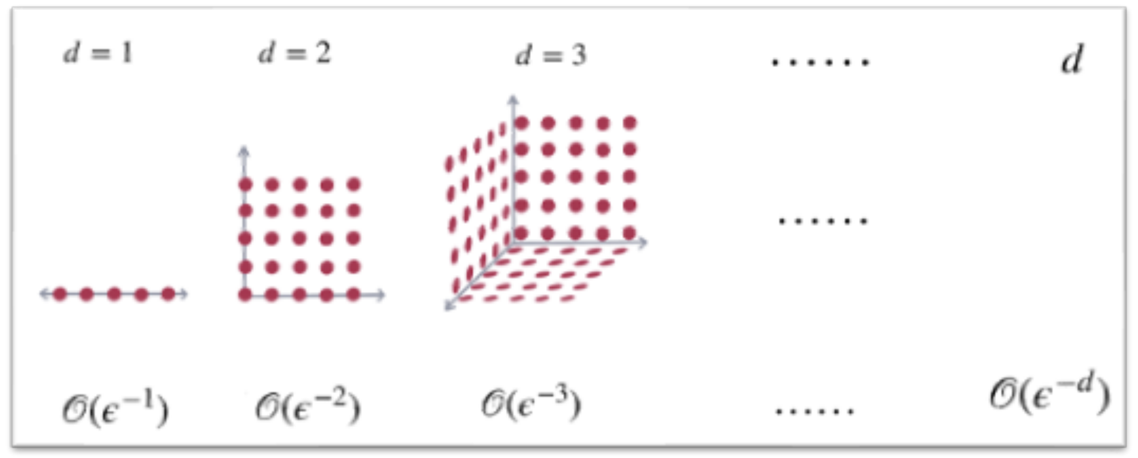

Figure 2: Figure illustrating the curse of dimensionality

However, by imposing strong assumptions on the function class one can reduce or possibly avoid the curse.

In [2], it is showed that for functions in Barron's space single layer neural networks can obtain integrated squared error of order $\mathcal{O}(1/n)$, where n is the number of hidden nodes. In this study we show that sums of separable functions can also achieve same dimension independent approximation error bounds for Barron's functions.

1.1 Barron's Function Space

The Fourier distribution of a function f on $\mathbb{R}^d$ is the complex valued measure given by $\tilde{F}(d\omega) = e^{i\theta(\omega)}F(d\omega)$,where $\theta(\omega)$ is the phase and $F(d\omega)$ is the magnitude distribution, and satisfies

$$f(x) = f(0) + \int (e^{i\omega\cdot x} - 1)\tilde{F}(d\omega) \text{ —————————(1)}$$

For $C > 0$ and $|\omega| = (\omega\cdot\omega)^{1/2}$ define

$$\Gamma_C = \{f : \mathbf{R}^d \to \mathbf{R} \mid C_f = \int_{\mathbf{R}^d} |\omega| F(d\omega) \le C\}.$$

Let $B \subseteq \mathbb{R}^d$ be a bounded set including $x = 0$. Then

$$\Gamma_B = \{f : B \to \mathbf{R} \mid \text{(1) holds for } \mathbf{x} \in B \text{ for some } \tilde{F} \text{ for which } C_f < \infty\}.$$

For $C > 0$ and $|\omega|_B = \sup_{\mathbf{x}\in B} |\omega\cdot\mathbf{x}|$ define

$$\Gamma_{C,B} = \{f \in \Gamma_B \mid \tilde{F} \text{ representing } f \text{ on } B,\ C_f = \int |\omega|_B F(d\omega) \le C\}.$$

1.2 Single Layer Neural Networks

Artificial Neural Networks is a method inspired by the learning in human brain. They have applications in wide variety of machine learning problems such as multivariate regression, classification, pattern recognition, speech recognition, and computer vision.

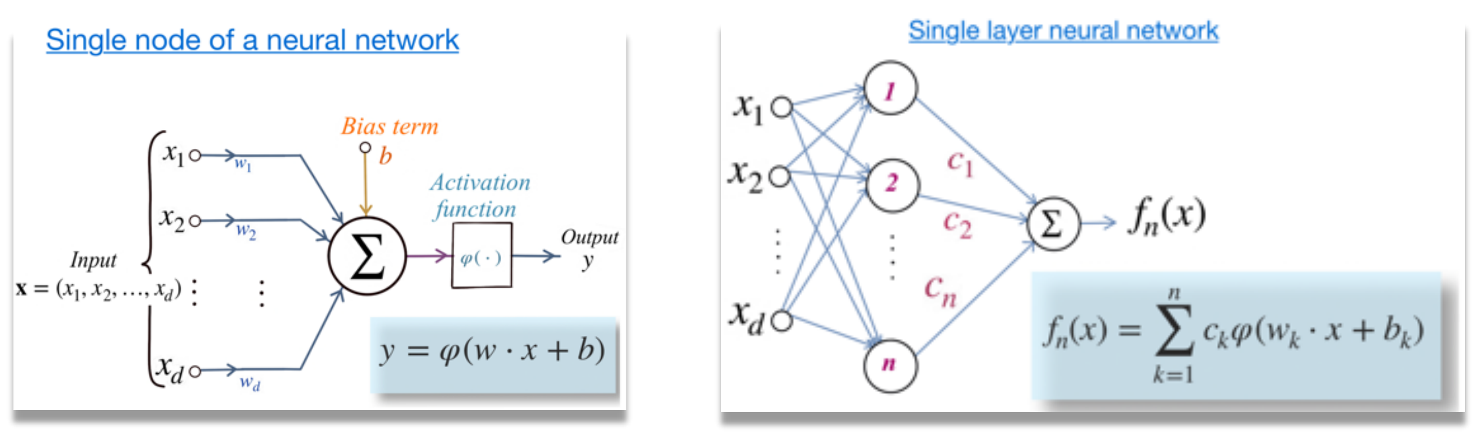

Figure 1: (Left) a model of an artificial neuron(node), (right) a model of a single layer neural network of n hidden layers

1.3 Sums of Separable Functions

This function approximation method is based on the classical approach of constructing a multivariate function as a product of univariate functions as described below.

$$f(\mathbf{x}) \approx g(\mathbf{x}) = \sum_{l=1}^{r} s_l \prod_{i=1}^{d} g_i^l(x_i),$$

where s_l are (optional) normalization coefficients, g_i^l are unknown univariate component functions and r is the separation rank. The univariate functions may be constrained to a space or required to have a particular form, but are not restricted to come from a particular basis set.

1.4 Single Layer Neural Networks for Barron's Functions

Given below is the main theorem in [1] that gives a bound for the integrated squared error for approximation by linear combination of a sigmoidal functions.

Theorem 1 (Theorem 1 of [2])

For every function $f \in \Gamma_{C,B}$ every sigmoidal function ϕ, every probability measure μ, and every $n \ge 1$, there exists a linear combination of sigmoidal functions $f_n(\mathbf{x})$ of the form

$$f_n(\mathbf{x}) = c_0 + \sum_{k=1}^{n} c_k\phi(a_k\cdot\mathbf{x} + b_k), \text{ where } a_k \in \mathbb{R}^d \text{ and } b_k, c_k \in \mathbb{R} \text{------(2)}$$

such that $\int_B (f(\mathbf{x}) - f_n(\mathbf{x}))^2\mu(d\mathbf{x}) \le \frac{(2C)^2}{n}$.

The coefficients of the linear combination in (2) may be restricted to satisfy $\sum_{k=1}^{n} |c_k| \le 2C$, and $c_0 = f(0)$.

The proof of this theorem is based on the following lemma which gives an approximation scheme using convex combinations in a Hilbert space.

Lemma 1 (Lemma 1 of [2])

If $\bar{f}(\mathbf{x}) = f(\mathbf{x}) - f(0)$ is in the closure of the convex hull of a set G in a Hilbert space, with $\|g\| \le b$ for each $g \in G$, then for every $n \ge 1$, and every $c' > b^2 - \|\bar{f}\|^2$, there is an f_n in the convex hull of n points in G such that

$$\|\bar{f} - f_n\|^2 \le \frac{c'}{n}.$$

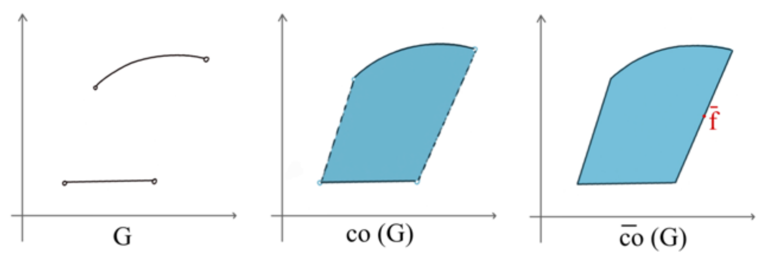

Figure 2: Figure illustrating lemma 1 in 2 dimensional Hilbert space

2. Results

2.1 Sums of Separable Functions for Barron's Functions

The following theorem is the sums of separable function version of the Theorem 1 of [1].

Theorem 2 (Main result)

For every function $f \in \Gamma_{C,B}$, every probability measure μ, and every $n \ge 1$, there exists a linear combination of complex valued exponential functions $f_n(\mathbf{x})$ of the form

$$f_n(\mathbf{x}) = s_0 + \sum_{k=1}^{n} s_k\left(\prod_{j=1}^{d} e^{i\omega_{kj}x_j} - 1\right), \text{ such that ------(3)}$$

$$\int_B |f(\mathbf{x}) - f_n(\mathbf{x})|^2\mu(d\mathbf{x}) \le \frac{C^2}{n}.$$

The coefficients of the linear combination in (3) may be restricted to satisfy $\sum_{k=1}^{n} |s_k||\omega_l|_B \le C$, and $s_0 = f(0)$.

The proof of this theorem is based on lemma 1 and the following theorem.

Theorem 3

For $\omega \in \Omega = \{\omega \in \mathbb{R}^d \mid \omega \ne 0\}$ let $G_{\exp}$ be defined as

$$G_{\exp} = \left\{ g : B \to \mathbb{C} \mid g(\mathbf{x}) = \gamma\left(\prod_{j=1}^{d} e^{i\omega_j x_j} - 1\right), |\gamma| \le \frac{C}{|\omega|_B} \right\}.$$

For every function $f \in \Gamma_{C,B}$, the function $f(\mathbf{x}) - f(0)$ is in the closure of convex hull of $G_{\exp}$, where the closure is taken in $L_2(\mu, B)$.

Remark

- In the neural network literature, [2] plays an important role as it obtained an approximation rate independent of input dimension d for approximating functions using single layer neural networks. Our study showing the same approximation bounds for the Barron's function class is therefore a significant result for the sums of separable function method.
- In the sums of separable approach, the univariate functions are given by exponential functions and the representation has the rank $n+1$.
- Dimension independent bounds can be obtained by applying strong assumptions on the smoothness of the function class as done in [2]. The issue with such function classes is that, even though they can be approximated with relatively smaller number of parameters (number of parameters not showing exponential dependence in d) the function class may not be useful in practical applications because the strong assumptions can make the function class shrink.

References

[1] Richard Bellman. Adaptive Control Processes: a Guided Tour. Princeton Univ. Press, Princeton, New Jersey, 1961.

[2] A. R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. IEEE Transactions on Information Theory, 39(3):930–945, May 1993. doi:10.1109/18.256500