Separating balls with partly random hyperplanes with a view to partly random neural networks

Olov Schavemaker, Utrecht University, the Netherlands

***Abstract*—We derive exact expressions for the probabilities that partly random hyperplanes separate two Euclidean balls. The probability that a fully random hyperplane separates two balls turns out to be significantly smaller than the corresponding probabilities for hyperplanes which are not fully random in certain cases. Our results motivate studying partially random neural networks and provide a first step in this direction.**

I. Introduction

In this short paper we deduce exact expressions for the probabilities that three types of partly random hyperplanes separate two Euclidean balls. Letting $\mathsf{H}[w;b]$ be the zero locus of $u \mapsto (w \mid u) - b$, we consider the following three types of partly random hyperplanes: $\mathsf{H}[w;\mathfrak{b}], \mathsf{H}[\mathfrak{w};b]$, and $\mathsf{H}[\mathfrak{w};\mathfrak{b}]$, where w, b are deterministic and $\mathfrak{w}, \mathfrak{b}$ are random variates.

The hyperplane parametrization $\mathsf{H}[w;b]$ is redolent of neural networks by design; a "neuron" in a neural network consists of a nonlinear activation function being applied to a linear function of the form $u \mapsto (w \mid u) - b$.

This paper was inspired by [2], where the authors showed that random neural networks can make two fairly arbitrary subsets of Euclidean space linearly separable by covering those two sets with small enough balls. In [2] they consider an architecture with two layers. Their result relies on letting the first layer be wide enough for it to separate all the balls covering one set from all of the balls covering the other set with high probability. This boils down to drawing enough fully random hyperplanes to separate all the balls covering one set from all the balls covering the other set with high probability. Clearly the probability one random hyperplane separates one pair of balls is of interest if one wants to get insight into how many random hyperplanes are needed to separate all pairs of balls with high probability.

Since sufficiently nice manifolds are known to admit nice covering number bounds [6, Thm 4.2], [7, Lemma B.1], [8, §5] the result in [2] and possible improvements thereof could be of great interest to those interested in classifying manifolds; a commonly assumed situation.

Random neural networks might sound void of practical significance at first as one generally wants to learn the weights and biases. Well, learning the weights and biases can be slow, sensitive to hyperparameters, and final values may be suboptimal [3].

One way to avoid these undesirabilities is to forgo learning some of the weights or/and biases; a popular architecture with this design philosophy is the so-called Random Vector Functional Link (RVFL) network [3]. In the case of RVFL networks, it has been known for quite a while that they are performant despite the random weights and biases [4], [5].

An RVFL consists of two layers, the first one being fully random and the second one being learned. As far as the author knows, partially random layers have not yet been studied despite having the potential to be a best of both worlds golden mean. In §VI we shall discuss how our theorems hint at this indeed being the case.

In §II we clarify the setting wherein we operate throughout the paper, including notation and assumptions. §§III–V are each dedicated to a probability that a partly random hyperplane separates two balls; $\mathsf{H}[w;\mathfrak{b}], \mathsf{H}[\mathfrak{w};b]$, and $\mathsf{H}[\mathfrak{w};\mathfrak{b}]$ in that order. Lastly, we discuss our results and their possible implications in §VI.

II. Setting

Since we are interested in separating balls in $\mathbb{R}^n$ with hyperplanes, we introduce the following notation.

The Euclidean inner product and norm on $\mathbb{R}^n$ will be denoted by $(\cdot \mid \cdot)$ and $|\cdot|$ resp.

Let $\mathsf{B}[c,r] = \{u \in \mathbb{R}^n : |c-u| < r\}$, where $c \in \mathbb{R}^n$ and $r > 0$. Let $\mathsf{H}[w;b]$ be the zero locus of $u \mapsto (w \mid u) - b$, where $b \in \mathbb{R}$ and $w \in \mathbb{S}^{n-1} = \{u \in \mathbb{R}^n : |u| = 1\}$.

The notations $\mathsf{B}[c,r]$ and $\mathsf{H}[w;b]$ parametrize all balls and hyperplanes resp., albeit not quite uniquely, as $\mathsf{H}[w;b] = \mathsf{H}[-w;-b]$.

Definition. We say $\mathsf{H}[w;b]$ <u>separates</u> $\mathsf{B}[c,r]$ and $\mathsf{B}[x,p]$ if $(w \mid u) - b \gtrless 0$ for all $u \in \mathsf{B}[c,r]$ whereas $(w \mid u) - b \lessgtr 0$ for all $u \in \mathsf{B}[x,p]$.

The balls $\mathsf{B}[c,r]$ and $\mathsf{B}[x,p]$ clearly cannot be separated if $\mathsf{B}[c,r] \cap \mathsf{B}[x,p] \neq \varnothing$. Since we are interested in partially random hyperplanes, we assume $|c-x| = p + r + \delta$ for some $\delta > 0$. If $\delta = 0$, only one hyperplane separates the balls $\mathsf{B}[c,r]$ and $\mathsf{B}[x,p]$; namely, $\{c-x\}^{\perp} + v$, where $\{v\} = \partial\mathsf{B}[c,r] \cap \partial\mathsf{B}[x,p]$.

When considering a hyperplane $\mathsf{H}[w;b]$, we shall refer to w, b as the <u>weight</u> and <u>bias</u> resp., in keeping with neural networks. We account random weights uniformly on $\mathbb{S}^{n-1}$ and random biases uniformly on $[-k,k]$ with $k > 0$.

Henceforth we shall always consider $\mathsf{B}[c,r]$ and $\mathsf{B}[x,p]$ so that $|c-x| = p+r+\delta$ for some $\delta > 0$. We withal assume that $n \geqslant 2$ and that $\mathfrak{w}, \mathfrak{b}$ are uniformly distributed on $\mathbb{S}^{n-1}$ and $[-k,k]$ resp., where $k \geqslant |c| \vee |x| > 0$ ($\Leftarrow c \neq x$).

III. Random biases

The first and easiest probability we are interested in is

$$\mathbb{P}_{\mathfrak{b}}\{\exists w \in \mathbb{S}^{n-1} : \mathsf{H}[w;\mathfrak{b}] \text{ separates } \mathsf{B}[c,r] \text{ and } \mathsf{B}[x,p]\}$$

To compact notation, we will shorten this to

$$\mathbb{P}_{\mathfrak{b}}\{\exists w : \mathsf{H}[w;\mathfrak{b}] \text{ separates}\}$$

Before we evaluate this probability, however, we return to the aforementioned $\mathsf{H}[w;b] = \mathsf{H}[-w;-b]$, which suggests that it may be more natural to consider $w \in \mathbb{S}^{n-1}/\{\pm\} \cong \mathbb{P}^{n-1}$ instead, where $\mathbb{P}^{n-1}$ comprises lines in $\mathbb{R}^n$ through the origin. Indeed, the orthogonal complement of a hyperplane is a line through the origin.

We therefore introduce a second parametrization of hyperplanes, namely $\mathsf{H}[a,v] = \{a\}^{\perp} + v$, where $a \in \mathbb{P}^{n-1}$ and $v \in a$ is the unique point on the line a wherethrough the hyperplane passes. This parametrization is unique and we have $\mathsf{H}[w;(w \mid v)] = \mathsf{H}[w\mathbb{R}, v]$ whenever $w \in \mathbb{S}^{n-1}$ and $v \in \mathbb{R}w$.

Theorem III. Let $\mathfrak{b}$ be as in §II. Then

$$\begin{aligned}\mathbb{P}_{\mathfrak{b}}\{\exists w : \mathsf{H}[w;\mathfrak{b}] \text{ separates}\} &= \\ \mathbb{P}_{\mathfrak{b}}\{\mathsf{H}[\omega\mathbb{R}, \omega\mathfrak{b}] \text{ separates}\} &= \delta/(2k)\end{aligned}$$

where $\mathbb{S}^{n-1} \ni \omega$ is parallel to the line through c and x, which we will call a.

Proof. Starting with the second equality, we can introduce coordinates onto a by considering $\Omega : a \ni v \mapsto (\omega \mid v)$. Let the intervals (s,t) and (y,z) be the images of $\mathsf{B}[c,r] \cap a$ & $\mathsf{B}[x,p] \cap a$ under Ω resp.; i.e., s is the coordinate of one of the endpoints of the line segment $\mathsf{B}[c,r] \cap a$, etc.

WLOG assuming that $t < y$, we must have $y - t = \delta$. Now, $\Omega(\omega\mathfrak{b}) = \mathfrak{b}$ so

$$\mathbb{P}_{\mathfrak{b}}\{\mathsf{H}[\omega\mathbb{R}, \omega\mathfrak{b}] \text{ separates}\} = \mathbb{P}_{\mathfrak{b}}\{\mathfrak{b} \in [t,y]\} = \delta/(2k)$$

if $[t,y] \subseteq [-k,k]$; indeed, since $t, y \in (c,x)$,

$$|t| \vee |y| \leqslant |c| \vee |x| \leqslant k$$

As for the first equality, it is geometrically optimal to have $\mathsf{H}[w;\mathfrak{b}] \perp a$, which is achieved by letting $w = \omega$. Indeed, each separating hyperplane must intersect a with coordinate in the interval $[t,y]$ and every coordinate is achieved by at most one realization of $\mathfrak{b}$. Only $\mathsf{H}[\omega\mathbb{R}, \omega\mathfrak{b}]$ can achieve the coordinates t and y, and since we have already shown it in fact achieves the entire interval, $w = \omega$ is optimal.

Since $(\omega \mid \omega\mathfrak{b}) = \mathfrak{b}|\omega|^2 = \mathfrak{b}$ the first equality follows. □

Henceforth a will always denote the line through c and x.

IV. Random weights

In this section we consider

$$\mathbb{P}_{\mathfrak{w}}\{\exists b \in [-k,k] : \mathsf{H}[\mathfrak{w};b] \text{ separates}\}$$

But first we recall two facts. First off, $\mathbb{P}^{n-1}$ is metrized by the angle between two lines $\measuredangle : \mathbb{P}^{n-1} \times \mathbb{P}^{n-1} \to [0, \frac{\pi}{2}]$. A closed metric ball on $\mathbb{P}^{n-1}$ is thus of the form

$$\mathsf{C}[\varepsilon, \alpha] = \{\ell \in \mathbb{P}^{n-1} : \measuredangle(\ell, \varepsilon) \leqslant \alpha\}$$

which is a (double) cone with axis ε.

Secondly, the <u>regularized beta function</u> is given by

$$I(\kappa; y, z) = \frac{1}{B(y,z)} \int_0^{\kappa} s^{y-1}(1-s)^{z-1}\, ds$$

where $B(y,z)$ is the familiar beta function, $y, z > 0$, and $\kappa \in [0,1]$.

Theorem IV. Let $\mathfrak{w}$ be as in §II. Then

$$\begin{aligned}\mathbb{P}_{\mathfrak{w}}\{\exists b : \mathsf{H}[\mathfrak{w};b] \text{ separates}\} &= I(\mathcal{Q}; \tfrac{n-1}{2}, \tfrac{1}{2}) \\ &= \mathbb{P}_{\mathfrak{w}}\{\mathsf{H}[\mathfrak{w}\mathbb{R}, (\mathfrak{w} \mid v)\mathfrak{w}] \text{ separates}\}\end{aligned}$$

where $\mathcal{Q} = 1 - (\frac{p+r}{p+r+\delta})^2$ and $v = \frac{p}{p+r}c + \frac{r}{p+r}x$.

Remark. The point v is the vertex of the (unique) double cone touching both balls. The line a is its axis.

Proof. Let ϕ be the angle between the generatrices of the aforementioned double cone and a. Then $\phi \in (0, \frac{\pi}{2})$ and $\sin\phi = \frac{p+r}{p+r+\delta}$ because

$$\begin{aligned}p + r + \delta = |c - x| &= |c - v| + |v - x| \\ &= \frac{r}{\sin\phi} + \frac{p}{\sin\phi}\end{aligned} \tag{1}$$

using that tangent lines to circles are perpendicular to the radii at the points of tangency. WLOG let $v = 0$, so that $a \in \mathbb{P}^{n-1}$ for convenience.

The idea of the proof is to show that having the randomly oriented hyperplane pass through the vertex of the double cone is optimal. Separating the balls is then equivalent to separating the two nappes of the double cone.

Let $\alpha \in (0, \frac{\pi}{2})$. We claim that

$$\begin{aligned}\{\omega \in \mathbb{P}^{n-1} : \measuredangle(\ell, \omega) < \tfrac{\pi}{2} \text{ for all } \ell \in \mathsf{C}[a,\alpha]\} \\ = \mathsf{C}[a, \tfrac{\pi}{2} - \alpha]\end{aligned}$$

Indeed, if $\ell \in \mathsf{C}[a,\alpha]$ and $\omega \in \mathsf{C}[a, \frac{\pi}{2} - \alpha]$, then

$$\measuredangle(\ell, \omega) \leqslant \measuredangle(\ell, a) + \measuredangle(a, \omega) \leqslant \alpha + (\tfrac{\pi}{2} - \alpha) = \tfrac{\pi}{2}$$

Conversely, if $\omega \notin \mathsf{C}[a, \frac{\pi}{2} - \alpha]$, then $\measuredangle(\omega, \nu) < \alpha$, with ν being the line through v perpendicular to a in the plane a spans with ω. Letting ℓ be the 90° rotation of ω about v within the plane it spans with a, plainly $\ell \in \mathsf{C}[a,\alpha]$ since $\measuredangle(\ell, a) = \measuredangle(\omega, \nu)$ and $\measuredangle(\ell, \omega) = \frac{\pi}{2}$.

Ergo, $\mathbb{P}_{\mathfrak{w}}\{v + \{\mathfrak{w}\}^{\perp} \text{ separates}\}$ equals the probability of separating the two nappes, which may be expressed as

$$\begin{aligned}&\mathbb{P}_{\mathfrak{w}}\{\measuredangle(\ell, \mathfrak{w}\mathbb{R}) < \tfrac{\pi}{2} \text{ for all } \ell \in \mathsf{C}[a, \phi]\}\\ &= \mathbb{P}_{\mathfrak{w}}\{\mathfrak{w}\mathbb{R} \in \mathsf{C}[a, \tfrac{\pi}{2} - \phi]\}\end{aligned} \tag{2}$$

because $\mathbb{P}^{n-1} \cong \mathbb{S}^{n-1}/\{\pm\}$. However, [1, (1)] yields that

$$\mathbb{P}_{\mathfrak{w}}\{\mathfrak{w} \in \mathsf{C}[a, \tfrac{\pi}{2} - \phi]\} = 2 \times \tfrac{1}{2} I(\cos^2\phi; \tfrac{n-1}{2}, \tfrac{1}{2})$$

Using that $\cos^2\phi = 1 - \sin^2\phi$, we have thus shown that

$$\mathbb{P}_{\mathfrak{w}}\{v + \{\mathfrak{w}\}^{\perp} \text{ separates}\} = I(1 - (\tfrac{p+r}{p+r+\delta})^2; \tfrac{n-1}{2}, \tfrac{1}{2})$$

All that remains to be shown is therefore that

$$\begin{aligned}\mathbb{P}_{\mathfrak{w}}\{\exists b : \mathsf{H}[\mathfrak{w}; b] \text{ separates}\} &= \mathbb{P}_{\mathfrak{w}}\{v + \{\mathfrak{w}\}^{\perp} \text{ separates}\}\\ &= \mathbb{P}_{\mathfrak{w}}\{\mathsf{H}[\mathfrak{w}\mathbb{R}, (\mathfrak{w} \mid v)\mathfrak{w}] \text{ separates}\}\end{aligned}$$

As the intersection of $v + \{\mathfrak{w}\}^{\perp}$ and $\mathfrak{w}\mathbb{R}$ is $\mathfrak{w}(\mathfrak{w} \mid v)$, it follows that

$$v + \{\mathfrak{w}\}^{\perp} = \mathsf{H}[\mathfrak{w}\mathbb{R}, (\mathfrak{w} \mid v)\mathfrak{w}] = \mathsf{H}[\mathfrak{w}; (\mathfrak{w} \mid v)]$$

That $b = (\mathfrak{w} \mid v)$, or equivalently, $v \in \mathsf{H}[\mathfrak{w}; b]$, is geometrically optimal may be seen as follows: if q, the intersection of a and $\mathsf{H}[\mathfrak{w}; b]$, is WLOG closer to $\mathsf{B}[c, r]$ than v is, then the cone touching $\mathsf{B}[c, r]$ with vertex q has wider aperture than $\mathsf{C}[a, \phi]$, so the corresponding separating probability would be smaller because (2) is decreasing in ϕ.

Lastly, $b = (\mathfrak{w} \mid v) \Rightarrow |b| \leqslant |\mathfrak{w}|.|v| \leqslant |c| \vee |x| \leqslant k$. □

Henceforth $\mathcal{Q}, v$ will always be as in Theorem IV.

V. Fully random case

Lastly, we compute $\mathbb{P}_{\mathfrak{w},\mathfrak{b}}\{\mathsf{H}[\mathfrak{w}; \mathfrak{b}] \text{ separates}\}$.

Theorem V. Let $\mathfrak{w}, \mathfrak{b}$ be as in §II. Then

$$\begin{aligned}&\mathbb{P}_{\mathfrak{w},\mathfrak{b}}\{\mathsf{H}[\mathfrak{w}; \mathfrak{b}] \text{ separates}\} =\\ &\frac{|c-x|}{2k}\left(\frac{\mathcal{Q}^{(n-1)/2}}{\frac{n-1}{2}B(\frac{n-1}{2}, \frac{1}{2})} - \tfrac{p+r}{p+r+\delta} I(\mathcal{Q}; \tfrac{n-1}{2}, \tfrac{1}{2})\right)\end{aligned}$$

Proof. WLOG let $v = 0$ again. By Adam's law,

$$\mathbb{P}_{\mathfrak{w},\mathfrak{b}}\{\mathsf{H}[\mathfrak{w}; \mathfrak{b}] \text{ separates}\} = \mathop{\mathbb{E}}_{\mathfrak{w}}(\mathbb{P}_{\mathfrak{b}}\{\mathsf{H}[\mathfrak{w}; \mathfrak{b}] \text{ separates} \mid \mathfrak{w}\})$$

Let $\alpha = \cos^{-1}|(\mathfrak{w} \mid \omega)|$, with ω as in Theorem III. Then $\alpha = \measuredangle(a, \mathbb{R}\mathfrak{w})$. It follows from (2) that $\mathsf{H}[\mathfrak{w}; (\mathfrak{w} \mid v)]$, a fortiori $\mathsf{H}[\mathfrak{w}; \mathfrak{b}]$, separates only if $\alpha \leqslant \frac{\pi}{2} - \phi$, where

$$\phi = \sin^{-1} \tfrac{p+r}{p+r+\delta} \tag{3}$$

Smaller α correspond to larger ranges of separating biases; e.g., if $\alpha = \frac{\pi}{2} - \phi$, only $\mathfrak{b} = (\mathfrak{w} \mid v)$ separates.

Throughout the next paragraph we consider $\mathfrak{w}$ fixed so that $\alpha \in (0, \frac{\pi}{2} - \phi]$. Note that $\alpha > 0$ almost surely. We also let

$$U = \bigcup\{\mathsf{H}[\mathfrak{w}; b] : b \in [-k, k] \text{ so that } \mathsf{H}[\mathfrak{w}; b] \text{ separates}\}$$

Using that $k \geqslant |c| \vee |x| > 0$, a computation similar to (1) shows that the length of line segment $a \cap U$ is

$$|c-x| - \frac{p+r}{\sin(\frac{\pi}{2} - \alpha)} = \left(1 - \frac{\sin\phi}{\sin(\frac{\pi}{2} - \alpha)}\right)|c-x|$$

Indeed, letting P be the plane $\mathfrak{w}$ spans with a, the angle between a and the line $\mathsf{H}[\mathfrak{w}; \mathfrak{b}] \cap \mathsf{P}$ is $\frac{\pi}{2} - \alpha$. Projecting onto $\mathbb{R}\mathfrak{w}$ yields that the length of $\mathbb{R}\mathfrak{w} \cap U$ is

$$\left(1 - \frac{\sin\phi}{\sin(\frac{\pi}{2} - \alpha)}\right)|c-x|\cos\alpha = (\cos\alpha - \sin\phi)|c-x|$$

Ergo, using Iverson brackets, $\mathbb{P}_{\mathfrak{b}}\{\mathsf{H}[\mathfrak{w}; \mathfrak{b}] \text{ separates} \mid \mathfrak{w}\} =$

$$\frac{|c-x|}{2k}(\cos\alpha - \sin\phi)[\cos\alpha \geqslant \sin\phi]$$

Plugging this back into the expectation, we get

$$\frac{|c-x|}{2k}\mathop{\mathbb{E}}_{\mathfrak{w}}\Big(\Big(|(\mathfrak{w} \mid \omega)| - \sin\phi\Big)\Big[|(\mathfrak{w} \mid \omega)| \geqslant \sin\phi\Big]\Big)$$

By the well-known Funk–Hecke formula [9] the above expectation equals

$$\begin{aligned}&\frac{1}{B(\frac{n-1}{2}, \frac{1}{2})}\int_{-1}^{+1}\Big(|s| - \sin\phi\Big)\Big[|s| \geqslant \sin\phi\Big](1-s^2)^{(n-3)/2}\,ds\\ &= \frac{2}{B(\frac{n-1}{2}, \frac{1}{2})}\int_{\sin\phi}^{1}(s - \sin\phi)(1-s^2)^{(n-3)/2}\,ds\end{aligned}$$

The above integral evaluates to

$$\frac{\mathcal{Q}^{(n-1)/2}}{\frac{n-1}{2}B(\frac{n-1}{2}, \frac{1}{2})} - I(\mathcal{Q}; \tfrac{n-1}{2}, \tfrac{1}{2})\sin\phi$$

where we recall that $\mathcal{Q} = 1 - \sin^2\phi$. Putting everything together now yields the desideratum. □

It behooves us to verify that the resultant probability in Theorem V is indeed a probability, since it is no longer immediately clear.

Lemma. If $\alpha \in (0, \frac{\pi}{2})$, then

$$\begin{aligned}&I(\cos^2\alpha; \tfrac{n-1}{2}, \tfrac{1}{2})\sin\alpha \leqslant\\ &\frac{\cos^{n-1}\alpha}{\frac{n-1}{2}B(\frac{n-1}{2}, \frac{1}{2})} \leqslant I(\cos^2\alpha; \tfrac{n-1}{2}, \tfrac{1}{2})\end{aligned}$$

Proof. Let $\kappa = \cos^2\alpha$. Plainly $s \in [0, \kappa] \Rightarrow$

$$1 \leqslant (1-s)^{-1/2} \leqslant (1 - \cos^2\alpha)^{-1/2} = \csc\alpha$$

Integration now yields that

$$\begin{aligned}\int_0^{\kappa} s^{(n-3)/2}\,ds &\leqslant \int_0^{\kappa} s^{(n-3)/2}(1-s)^{-1/2}\,ds\\ &\leqslant \csc\alpha\int_0^{\kappa} s^{(n-3)/2}\,ds\end{aligned}$$

The rest is trivial. □

Applying the Lemma with $\alpha = \phi$ as in (3) evinces that the resultant probability in Theorem V is indeed a probability, since $|c - x| \leqslant |c| + |x| \leqslant 2k$ as well.

VI. Discussion

In §I we suggested that partially random neural network layers could be a golden mean between fully learned and fully random layers. We shall therefore discuss how the separating probability for fully random hyperplanes compares to those for random weights and random biases.

Comparing the resultant probabilities in Theorems IV and V is easy thanks to our Lemma:

$$\begin{aligned}\frac{|c-x|}{2k}\left(\frac{\mathcal{Q}^{(n-1)/2}}{\frac{n-1}{2}B(\frac{n-1}{2},\frac{1}{2})} - \tfrac{p+r}{p+r+\delta}I(\mathcal{Q};\tfrac{n-1}{2},\tfrac{1}{2})\right)\\ \leqslant \frac{\mathcal{Q}^{(n-1)/2}}{\frac{n-1}{2}B(\frac{n-1}{2},\frac{1}{2})} - \tfrac{p+r}{p+r+\delta}I(\mathcal{Q};\tfrac{n-1}{2},\tfrac{1}{2})\\ \leqslant \frac{\mathcal{Q}^{(n-1)/2}}{\frac{n-1}{2}B(\frac{n-1}{2},\frac{1}{2})} \leqslant I(\mathcal{Q};\tfrac{n-1}{2},\tfrac{1}{2})\end{aligned}$$

which makes sense: choosing the bias optimally should improve the separating probability.

Of note is also that $\mathcal{Q} \in (0,1) \Rightarrow$

$$\begin{aligned}\frac{\mathcal{Q}^{(n-1)/2}}{\frac{n-1}{2}B(\frac{n-1}{2},\frac{1}{2})} &< \frac{\Gamma(\frac{n}{2})}{\frac{n-1}{2}\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}\\ &= \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})\sqrt{\pi}} \sim \sqrt{2/(n\pi)}\end{aligned}$$

so $\mathbb{P}\{\mathsf{H}[\mathfrak{w};\mathfrak{b}] \text{ separates}\} \preccurlyeq 1/\sqrt{n}$ where "$\sim$" & "$\preccurlyeq$" are as in [10, pp. 2f.]; i.e., fully random hyperplanes separate balls really poorly if the ambient dimension is large regardless of p, r, δ.

Note that this is not the case for hyperplanes with random biases (and optimal weights). There the separating probability can get as close to 1 as desired if $\frac{p+r}{p+r+\delta}$ is sufficiently small (compared to the ambient dimension).

Comparing the resultant probabilities in Theorems III and V is a bit harder. Recall that $\mathbb{P}\{\mathsf{H}[\mathfrak{w};\mathfrak{b}] \text{ separates} \mid \mathfrak{w}\} =$

$$\begin{aligned}\frac{\cos\alpha}{2k}\left(|c-x| - \frac{p+r}{\sin(\frac{\pi}{2}-\alpha)}\right)[\cos\alpha \geqslant \sin\phi]\\ \leqslant \frac{1}{2k}\Big(|c-x| - p - r\Big) = \delta/(2k)\end{aligned}$$

so $\mathbb{P}\{\mathsf{H}[\mathfrak{w};\mathfrak{b}] \text{ separates}\} = \mathbb{E}(\mathbb{P}\{\mathsf{H}[\mathfrak{w};\mathfrak{b}] \text{ separates} \mid \mathfrak{w}\}) \leqslant \mathbb{E}(\delta/(2k)) = \delta/(2k) = \mathbb{P}\{\exists w : \mathsf{H}[w,\mathfrak{b}] \text{ separates}\}$, as is to be expected.

Like the resultant probability for random weights, the resultant probability for random biases is unlike the resultant probability for fully random hyperplanes in the sense that $\delta/(2k) \not\preccurlyeq 1/\sqrt{n}$.

All in all, we have seen that hyperplanes with random biases (and optimal weights) or random weights (and optimal biases) are significantly better at separating small balls in a high dimensional space than fully random hyperplanes. But small balls in high dimensional space is exactly the situation we talked about in §I. This suggests that partially random neural networks may be quite a bit better at classifying low dimensional manifolds than their fully random counterparts. A detailed exploration of this topic is beyond the scope of this paper, however.

The author is currently cosupervising Guido Wagenvoorde (master student at Utrecht University) who is currently, i.a., numerically comparing RVFL, a neural net of the form

$$\begin{aligned}\mathbb{R}^n \ni x \mapsto \sum_{j=1}^m a_j\rho((w_j \mid x) + b_j) = (a \mid \varrho) \qquad (4)\\ a = \{a_j\}_{j=1}^m, \varrho = \Big\{\rho((w_j \mid x) + b_j)\Big\}_{j=1}^m\end{aligned}$$

with random w_j's and b_j's, with RVFL with learned biases (random w_j's only) in his thesis. Many popular activation functions ρ, like ReLU & (hard) tanh, are positive on positive inputs and nonpositive on nonpositive inputs, thus in which cell of the hyperplane tessellation induced by the $\mathsf{H}[w_j;b_j]$'s x lies determines in which octant ϱ lies whenever ρ is positive for positive inputs and nonpositive for nonpositive inputs, evincing the linkage between (4) and hyperplane tessellations.

In light of the linkage between (4) and hyperplane tessellations, RVFL thus corresponds to fully random hyperplanes and RVFL with learned biases corresponds to hyperplanes with random weights (and optimal biases). Extending our results to partly random hyperplane tessellations would be a nice follow-up direction, seeing as (random) hyperplane tessellation is an active area of research interesting in its own right [11].

Acknowledgements

The author wants to thank Guido Wagenvoorde for making plots especially for this paper, and apologize for not being able to fit them within the allotted four pages. I'm sorry.

The author would also like to thank Palina Salanevich for her feedback on the initial draft of this paper.

References

[1] *Concise Formulas for the Area and Volume of a Hyperspherical Cap* by Li. Asian J. Math. Stat. 4 (1): 66–70, 2011.

[2] *The Separation Capacity of Random Neural Networks* by Dirksen et al. J. Mach. Learn. Res. 23 (2023) 1–47.

[3] *Random vector functional link network: Recent developments, applications, and future directions* by Malik et al. Applied Soft Computing 143 (2023) 110377.

[4] *Feed Forward Neural Networks With Random Weights* by Schmidt et al. International Conference on Pattern Recognition, IEEE, 1992.

[5] *Random Activation Weight Neural Net (RAWN) for Fast Non-iterative Training* by Te Braake and Van Straten.
Engng Applic. Artif. Intell. Vol. 8, No. 1, pp. 71–80, 1995.

[6] *On Fast Johnson-Lindenstrauss Embeddings of Compact Submanifolds of RN with Boundary* by Iwen et al.
Discrete Comput. Geom. (2024) 71:498–555.

[7] *Stable Manifold Embeddings With Structured Random Matrices* by Yap et al. IEEE Journal of Selected Topics in Signal Processing, Vol. 7, No. 4, August 2013.

[8] *Finding the Homology of Submanifolds with High Confidence from Random Samples* by Niyogi et al.
Discrete Comput Geom (2008) 39: 419–441.

[9] *The Funk-Hecke formula* by Chafaï.
`https://djalil.chafai.net/blog/2021/05/22/the-funk-hecke-formula/`

[10] *Orders of Infinity* by Hardy. Cambridge University Press, 1910.

[11] *Sharp Estimates on Random Hyperplane Tessellations* by Dirksen et al. SIAM J. Math. Data Sci., Vol. 4, No. 4, pp. 1396–1419.