# On the Linear Separability of Random Points in the $d$-dimensional Spherical Layer and in the $d$-dimensional Cube

S. V. Sidorov
*Institute of Information Technologies,*
*Mathematics and Mechanics*
*Lobachevsky State University of Nizhni Novgorod*
Nizhni Novgorod, Russia
sergey.sidorov@itmm.unn.ru

N. Yu. Zolotykh
*Institute of Information Technologies,*
*Mathematics and Mechanics*
*Lobachevsky State University of Nizhni Novgorod*
Nizhni Novgorod, Russia
nikolai.zolotykh@itmm.unn.ru

***Abstract*—The authors of [6] propose a method for correcting errors of artificial intelligence systems by separating erroneous cases with the Fisher linear discriminant. It turned out that if the dimension is large this approach works well even for an exponential (of the dimension) number of samples. In this paper, we specify the limits of applicability of this approach by estimating the number of points that are linearly separable with a probability close to $1$ in two particular cases: when the points drawn randomly, independently and uniformly from a $d$-dimensional spherical layer and from the $d$-dimensional cube. Our bounds for these two cases improve some bounds obtained in [6].**

***Index Terms*—random points, $1$-convex set, linear separability, Fisher separability, Fisher linear discriminant**

## I. Introduction

Currently, artificial intelligence systems for data mining consume huge and fast-growing amounts of heterogeneous data. During processing, numerous errors can appear. One of the most important tasks is developing methods to correct the errors without damaging existing skills. In [4]–[6], the following methodology is being developed. The correction can be performed using binary classifiers, which separate the situations with errors from correctly solved tasks.

Surprisingly (this is one of the manifestations of the blessing of dimensionality), if the dimension of the data is large, then this correction can be solved using simple Fisher linear discriminant, even if the data are exponentially large with respect to dimension. In particular, in [6], it was shown that a random $n$-element set in $\mathbb{R}^d$ is linearly separable with probability $p > 1 - \alpha$, if $n < ae^{bd}$. The exact form of the exponent function depends on the probability distribution that determines how the random set is drawn, and on the constant $\alpha$ $(0 < \alpha < 1)$.

In this paper we give more exact bounds for the cardinality of the set for two particular distributions: when the points are drawn randomly, independently and uniformly from a $d$-dimensional spherical layer and when they are drawn from the $d$-dimensional cube. Our results improve some bounds obtained in [6].

We note that there are many algorithms for constructing a functional separating a point from a set of points (Fisher linear discriminant, SVM algorithm, the Rosenblatt perceptron algorithm etc.). Fisher linear discriminant is computationally cheap, simple, and robust [4].

A set of points $\{X_1, \dots, X_n\} \subset \mathbb{R}^d$ is called $1$-*convex* [1] or *linear separable* [6] if the set of vertices of their convex hull, $\mathrm{conv}(X_1, \dots, X_n)$, coincides with $\{X_1, \dots, X_n\}$ . The set $\{X_1, \dots, X_n\}$ is called *Fisher separable* if for some $\beta$, where $0 \le \beta < 1$, the inequality $(X_i, X_j) \le \beta \cdot (X_i, X_i)$ holds for all $i$, $j$, such that $i \ne j$ [4], [5].

Fisher separability implies linear separability but not vice versa (even if the set is centered and normalized to unit variance).

Let $B_d = \{X \in \mathbb{R}^d : \ \|X\| \le 1\}$ be the $d$-dimensional unit ball centered at the origin ($\|X\|$ means Euclidean norm), $rB_d$ is the $d$-dimensional ball of radius $r < 1$ centered at the origin, and $Q_d = \{X = (x_1, \dots, x_d) \in \mathbb{R}^d : \ 0 \le x_j \le 1 \ (j = 1, \dots, d)\}$ is the $d$-dimensional unit cube.

First, let $M_n = \{X_1, \dots, X_n\} \subset B_d \setminus rB_d$ be the set of points chosen randomly, independently, according to the uniform distribution on the spherical layer $B_d \setminus rB_d$.

In [6] it is shown that for all $r$, $\alpha$, $n$, where $0 < r < 1$, $0 < \alpha < 1$, if

$$n < \left(\frac{r}{\sqrt{1-r^2}}\right)^d \left(\sqrt{1 + \frac{2\alpha(1-r^2)^{d/2}}{r^{2d}}} - 1\right), \qquad (1)$$

then $M_n$ is Fisher separable with a probability greater than $1 - \alpha$. (The authors of [6] formulate their result for linearly separable sets of points, but in fact in the proof they used that the sets are only Fisher separable.)

We will show (see Theorem 1) that if we want to guarantee only the linear separability, then the upper bound (1) can be increased to $n < \sqrt{\alpha 2^d(1 - r^d)}$. Statement 1 compares these two bounds.

The work is supported by the Ministry of Education and Science of Russian Federation (project 14.Y26.31.0022).

Also, in [6] a product distribution in the $Q_d$ is considered. Let the coordinates of a random point $X = (x_1, \dots, x_d) \in Q_d$ be independent random variables with variances $\sigma_i^2 > \sigma_0^2 > 0$ $(i = 1, \dots, d)$. In [6] it is shown that for all $\alpha$ and $n$, where $0 < \alpha < 1$, if

$$n < \sqrt{\frac{\alpha e^{0.5 d \sigma_0^4}}{3}}, \tag{2}$$

then $M_n$ is Fisher separable with a probability greater than $1 - \alpha$.

If all random variables $x_1, \dots, x_d$ has uniform distribution on the segment $[0, 1]$ then $\sigma_0^2 = \frac{1}{12}$. So the inequality (2) takes the form

$$n < \sqrt{\frac{\alpha e^{d/288}}{3}}. \tag{3}$$

(As above, the authors of [6] formulate their result for the linearly separable case, but in fact they used only the Fisher separability.)

We will show (see Theorem 2) that if we want to guarantee only the linear separability, then the bound (3) can be increased to $n < \sqrt{\frac{\alpha c^d}{d+1}}$, where $c = 1.18858$.

Note that using the Fisher separability in [4]–[6] allows one to write out the explicit equations for the linear discriminant in terms of inner products. Our bounds for linear separability are better, but these are only existence theorems. We pay for this improvement by the absence of an explicit form of the discriminant function. Apparently, this is a natural trade-off.

## II. Random points in the spherical layer

The following theorem gives an improved estimate for the number of points $n$ guaranteeing the linear separability of a random $n$-element set $M_n$ in $B_d \backslash r B_d$ with probability at least $1 - \alpha$. The proof uses an approach borrowed from [1]. Denote by $A_n$ the event that $M_n$ is linear separable.

**Theorem 1.** *Let* $0 \le r < 1$, $0 < \alpha < 1$,

$$n < \sqrt{\alpha 2^d (1 - r^d)}. \tag{4}$$

*Then* $P(A_n) > 1 - \alpha$.

*Proof.* Denote the event that $X_i \notin \operatorname{conv}(M_n \setminus \{X_i\})$ by $C_i$ $(i = 1, \dots, n)$. Clearly $A_n = C_1 \cap \ldots \cap C_n$ and $P(A_n) = P(C_1 \cap \ldots \cap C_n) = 1 - P(\overline{C}_1 \cup \ldots \cup \overline{C}_n) \ge 1 - \sum_{i=1}^{n} P(\overline{C}_i)$. Let us find the upper bound for the probability of the event $\overline{C}_i$. This event means the point $X_i$ belongs to the convex hull of the remaining points, i.e. $X_i \in \operatorname{conv}(M_n \backslash \{X_i\})$. Since the points in $M_n$ have the uniform distribution, then the probability of $\overline{C}_i$ is

$$P(\overline{C}_i) = \frac{\operatorname{Vol}\Big(\operatorname{conv}(M_n \backslash \{X_i\})\Big) - \operatorname{Vol}\Big(\operatorname{conv}(M_n \backslash \{X_i\}) \cap r B_d\Big)}{\operatorname{Vol}(B_d) - \operatorname{Vol}(r B_d)}$$

$(i = 1, \dots, n)$.

Hence

$$P(\overline{C}_i) \le \frac{\operatorname{Vol}\Big(\operatorname{conv}(M_n \setminus \{X_i\})\Big)}{\gamma_d (1 - r^d)},$$

where $\operatorname{Vol}(r B_d) = \gamma_d r^d$ is the volume of a ball of radius $r$.

It remains to estimate $\operatorname{Vol}(\operatorname{conv}(M_n \setminus \{X_i\}))$. Let

$$V(k, d) = \sup \Big\{ \operatorname{Vol}\Big(\operatorname{conv}(Y_1, \dots, Y_k)\Big) : \ Y_1, \dots, Y_k \in B_d \Big\}$$

and

$$V_r(k, d) = \sup \Big\{ \operatorname{Vol}\Big(\operatorname{conv}(Y_1, \dots, Y_k)\Big) : \ Y_1, \dots, Y_k \in B_d \setminus r B_d \Big\}.$$

Clearly, $V_r(k, d) \le V(k, d)$ for $0 < r < 1$. It is known (see e.g. [3]) that

$$V(k, d) \le \frac{k \gamma_d}{2^d}.$$

Since

$$\operatorname{Vol}\Big(\operatorname{conv}(M_n \setminus \{X_i\})\Big) \le \le V_r(n-1, d) \le V(n-1, d) \le \frac{(n-1)\gamma_d}{2^d},$$

then

$$P(\overline{C}_i) \le \frac{n-1}{2^d (1 - r^d)} \qquad (i = 1, \dots, n).$$

Hence

$$P(A_n) \ge 1 - \sum_{i=1}^{n} P(\overline{C}_i) \ge 1 - \frac{n(n-1)}{2^d (1 - r^d)} \ge 1 - \frac{n^2}{2^d (1 - r^d)}.$$

Thus if $n$ satisfies the condition $\frac{n^2}{2^d (1 - r^d)} < \alpha$, i.e. $n < \sqrt{\alpha 2^d (1 - r^d)}$, then the inequality $P(A_n) > 1 - \alpha$ holds. $\square$

Let us compare the obtained bound (4) with the bound (1) proposed in [6].

**Statement 1.** *Let* $g = \left(\frac{r}{\sqrt{1-r^2}}\right)^d \left(\sqrt{1 + \frac{2\alpha (1-r^2)^{d/2}}{r^{2d}}} - 1\right)$, $0 < r < 1$, $0 < \alpha < 1$, $d \in \mathbb{N}$. *If $r$ and $\alpha$ are fixed then the following asymptotic estimates hold:*

1. $g \sim \frac{\alpha}{r^d}$, *if* $\sqrt{\frac{\sqrt{5}-1}{2}} < r < 1$.
2. $g \sim \frac{2\alpha}{r^d(\sqrt{1+2\alpha}+1)} = \frac{\sqrt{1+2\alpha}-1}{r^d} = (\sqrt{1+2\alpha} - 1)(\frac{\sqrt{5}+1}{2})^{d/2}$, *if* $r = \sqrt{\frac{\sqrt{5}-1}{2}}$.
3. $g \sim \frac{\sqrt{2\alpha}}{(1-r^2)^{d/4}}$, *if* $0 < r < \sqrt{\frac{\sqrt{5}-1}{2}}$.

*Proof.* We have

$$g = \frac{\left(\frac{r}{\sqrt{1-r^2}}\right)^d \frac{2\alpha(1-r^2)^{d/2}}{r^{2d}}}{\sqrt{1 + \frac{2\alpha(1-r^2)^{d/2}}{r^{2d}}} + 1} = \frac{2\alpha}{r^d \left( \sqrt{1 + 2\alpha \left(\frac{\sqrt{1-r^2}}{r^2}\right)^d} + 1 \right)}.$$

If $0 < \frac{\sqrt{1-r^2}}{r^2} < 1$ then $g \sim \frac{\alpha}{r^d}$.
If $\frac{\sqrt{1-r^2}}{r^2} = 1$ then $g \sim \frac{2\alpha}{r^d(\sqrt{1+2\alpha}+1)} = \frac{\sqrt{1+2\alpha}-1}{r^d}$.
If $\frac{\sqrt{1-r^2}}{r^2} > 1$ then $g \sim \frac{2\alpha}{r^d \sqrt{2\alpha\left(\frac{\sqrt{1-r^2}}{r^2}\right)^d}} = \frac{\sqrt{2\alpha}}{(1-r^2)^{d/4}}$.

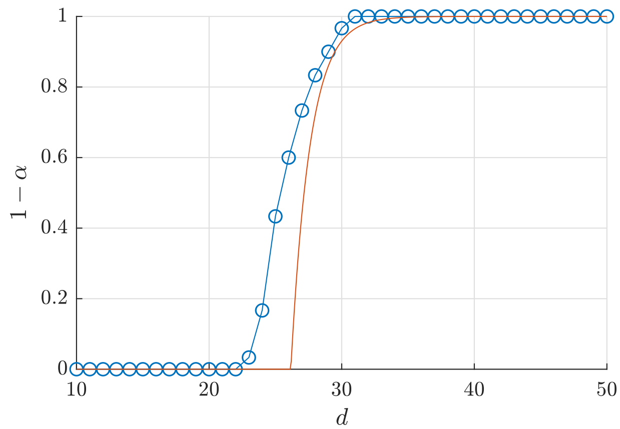


Fig. 1. The probability (frequency) that the set of $n = 10000$ random points in the layer $B_d \setminus 0.5B_d$ is linear separable. The red line shows the theoretical bound obtained from Theorem 1. The blue circles correspond to the empirical frequencies obtained in 30 trials for each dimension $d$.

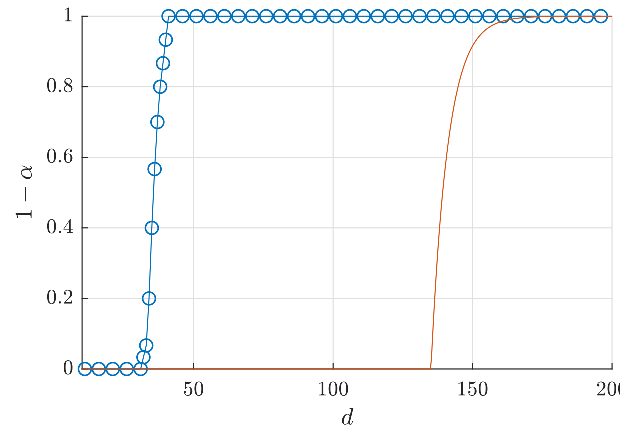


Fig. 2. The probability (frequency) that the set of $n = 10000$ random points in the unit cube is linear separable. The red line shows the theoretical bound obtained from Theorem 2. The blue circles correspond to the empirical frequencies obtained in 30 trials for each dimension $d$.

The equality $\frac{\sqrt{1-r^2}}{r^2} = 1$ holds if $r^4 + r^2 - 1 = 0$, that is $r^2 = \frac{\sqrt{5}-1}{2}$, $r = \sqrt{\frac{\sqrt{5}-1}{2}}$. The inequality $0 < \frac{\sqrt{1-r^2}}{r^2} < 1$ holds if $r^4 + r^2 - 1 > 0$, that is for $\sqrt{\frac{\sqrt{5}-1}{2}} < r < 1$. The inequality $\frac{\sqrt{1-r^2}}{r^2} > 1$ holds if $r^4 + r^2 - 1 < 0$, that is for $0 < r < \sqrt{\frac{\sqrt{5}-1}{2}}$. □

**Corollary 1.** *Let* $f = \sqrt{\alpha 2^d (1-r^d)}$, $g = \left(\frac{r}{\sqrt{1-r^2}}\right)^d \left(\sqrt{1 + \frac{2\alpha(1-r^2)^{d/2}}{r^{2d}}} - 1\right)$, $0 < r < 1$, $0 < \alpha < 1$, $d \in \mathbb{N}$. *If* $r$ *and* $\alpha$ *are fixed then the following asymptotic estimates of the quotient* $\frac{f}{g}$ *hold:*

1. $\frac{f}{g} \sim \frac{1}{\sqrt{\alpha}}(r\sqrt{2})^d \to \infty$, *if* $\sqrt{\frac{\sqrt{5}-1}{2}} < r < 1$.
2. $\frac{f}{g} \sim \frac{\sqrt{1+2\alpha}+1}{2\sqrt{\alpha}}(\sqrt{5}-1)^{\frac{d}{2}} \to \infty$, *if* $r = \sqrt{\frac{\sqrt{5}-1}{2}}$.
3. $\frac{f}{g} \sim \frac{1}{\sqrt{2}}(2\sqrt{1-r^2})^{d/2} \to \infty$, *if* $0 < r < \sqrt{\frac{\sqrt{5}-1}{2}}$.

*Proof.* Obviously $f \sim \sqrt{\alpha 2^d}$ for $0 \le r < 1$.

If $\sqrt{\frac{\sqrt{5}-1}{2}} < r < 1$, then $\frac{f}{g} \sim \frac{\sqrt{\alpha 2^d}}{\alpha / r^d} = \frac{1}{\sqrt{\alpha}}(r\sqrt{2})^d \to \infty$ for $d \to \infty$, since $r\sqrt{2} > 1$ for $r > \sqrt{\frac{\sqrt{5}-1}{2}}$.

If $r = \sqrt{\frac{\sqrt{5}-1}{2}}$, then $\frac{f}{g} \sim \frac{r^d(\sqrt{1+2\alpha}+1)\sqrt{\alpha 2^d}}{2\alpha} = \frac{\sqrt{1+2\alpha}+1}{2\sqrt{\alpha}}(r\sqrt{2})^d = \frac{\sqrt{1+2\alpha}+1}{2\sqrt{\alpha}}(\sqrt{5}-1)^{\frac{d}{2}} \to \infty$ for $d \to \infty$.

If $0 < r < \sqrt{\frac{\sqrt{5}-1}{2}}$, then $\frac{f}{g} \sim \frac{\sqrt{\alpha 2^d}(1-r^2)^{d/4}}{\sqrt{2\alpha}} = \frac{1}{\sqrt{2}}(2\sqrt{1-r^2})^{d/2} \to \infty$ for $d \to \infty$, since $2\sqrt{1-r^2} > 1$ for $0 < r < \sqrt{\frac{\sqrt{5}-1}{2}}$. □

The bound (4) is quite accurate as is illustrated with Fig. 1.

## III. Random points in the cube

Let us consider a random set of points $M_n = \{X_1, \dots, X_n\}$ chosen from the cube $Q_d$. As above, $A_n$ denotes the event that $M_n$ is linearly separable.

The following theorem gives an improved bound for the number of points $n$ guaranteeing the linear separability of $M_n \subset Q_d$ with probability at least $1 - \alpha$.

**Theorem 2.** *Let* $0 < \alpha < 1$ *and* $M_n = \{X_1, \dots, X_n\} \subset Q_d$ *be points choosing randomly, independently and according to the uniform distribution on the* $d$*-dimensional unit cube* $Q_d$*. If*

$$n < \sqrt{\frac{\alpha c^d}{d+1}}, \qquad c = 1.18858, \tag{5}$$

*then* $P(A_n) > 1 - \alpha$.

*Proof.* In [2] it is proved that the upper bound for the maximal volume of the convex hull of $k$ points placed in $Q_d$ is $\frac{(d+1)k}{c^d}$. Thus $\mathrm{Vol}\Big(\mathrm{conv}(Y_1, \dots, Y_k)\Big) < \frac{(d+1)k}{c^d}$.

As in the proof of Theorem 1, we can obtain the inequality

$$P(A_n) > 1 - \frac{n(d+1)(n-1)}{c^d} > 1 - \frac{n^2(d+1)}{c^d}.$$

Hence $P(A_n) > 1 - \alpha$ if $n < \sqrt{\frac{\alpha c^d}{d+1}}$. □

Let us compare the obtained bound (5) with the bound (3). We have

$$\frac{\sqrt{\frac{\alpha c^d}{d+1}}}{\sqrt{\frac{\alpha e^{d/288}}{3}}} = \sqrt{\frac{3}{d+1}\left(\frac{c}{e^{\frac{1}{288}}}\right)^d} \to \infty$$

for $d \to \infty$, since $\frac{c}{e^{\frac{1}{288}}} \approx 1.18446$.

The bound (5) could be rather conservative as is illustrated with Fig. 2.

## IV. Conclusion

In the paper, we clarify the limits of applicability of the method proposed in [6] for correcting errors of large artificial intelligence systems by separating erroneous cases by means of the Fisher linear discriminant. First, we show that if $n < \sqrt{\alpha 2^d(1-r^d)}$, then the set of points drawn randomly, independently and uniformly from the spherical layer $B_d \setminus rB_d$ is linearly separable with a probability greater than $1-\alpha$. Second, we show that if $n < \sqrt{\alpha c^d/(d+1)}$, where $c = 1.18858$, then the set of points drawn randomly, independently and uniformly from the cube $Q_d$ is linearly separable with a probability greater than $1-\alpha$. These results refine some results obtained in [6].

## Acknowledgements

Authors are grateful to Prof. A. N. Gorban for useful discussions.

## References

[1] *I. Bárány, Z. Füredi* On the shape of the convex hull of random points. Probab. Th. Rel. Fields 77, 231–240 (1988)

[2] *M. E. Dyer, Z. Füredi, C. McDiarmid* Random points in the $n$-cube. DIMACS Series in Discrete Mathematics and Theoretical Computer Science 1, 33–38 (1990)

[3] *G. Elekes* A geometric inequality and the complexity of computing volume. Discrete. Comput. Geom. 1, 289–292 (1986)

[4] *A. N. Gorban, A. Golubkov, B. Grechuk, E. M. Mirkes, I. Y. Tyukin* Correction of AI systems by linear discriminants: Probabilistic foundations. Information Sciences 466, 303–322 (2018)

[5] *A. N. Gorban, B. Grechuk, I. Y. Tyukin* Augmented artificial intelligence: a conceptual framework. https://arxiv.org/abs/1802.02172v3 (2018)

[6] *A. N. Gorban, I. Y. Tyukin* Stochastic separation theorems. Neural Networks 94, 255–259 (2017)