Partitioning Points by Parallel Planes

Martin Anthony*

November 2002
CDAM Research Report LSE-CDAM-2002-10

Abstract

A new upper bound is given on the number of ways in which a set of N points in $\mathbb{R}^n$ can be partitioned by k parallel hyperplanes. This bound improves upon a result of Olafsson and Abu-Mostafa [*IEEE Trans. Pattern Analysis and Machine Intelligence* 10(2), 1988: 277-281]; it agrees with the known (tight) result for the case $k = 1$; and it is, for fixed k and n, tight to within a constant. A previously published claimed improvement to the bound of Olafsson and Abu-Mostafa is shown to be incorrect.

1 Introduction

The partitioning of points in n-space by a single hyperplane has been well-studied. (Here, by a hyperplane I mean an $(n-1)$-flat; it need not contain the origin.) The question of how many such partitions are possible arises naturally in the theory of pattern classification and machine learning [13, 9, 3, 8, 18], and is an interesting problem in its own right. It is known [6] that the number of ways in which N points in $\mathbb{R}^n$ can be partitioned into two blocks separated by an $(n-1)$-dimensional hyperplane is at most

$$\sum_{i=0}^{n} \binom{N-1}{i}.$$

*Department of Mathematics and Centre for Discrete and Applicable Mathematics, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK, `m.anthony@lse.ac.uk`

Furthermore, this is tight, with the maximum number of partitions being achieved when the points are in *general position*, meaning that no r-flat for $r < n$ contains $r+2$ of the points. This proof of this result uses the classical fact [6, 16] that the number of connected components into which $\mathbb{R}^n$ can be divided by N hyperplanes, each passing through the origin, is at most $C(N,n) = 2\sum_{i=0}^{n-1}\binom{N-1}{i}$ (with equality if the normals are in general position).

Partitioning by a single hyperplane is fairly limited, and attention has been given to more complex partitioning methods that arise as simple generalizations, such as separation by surfaces with polynomial equations [7, 1], for intance.

In the context of pattern classification and learning, one proposed way of obtaining more powerful partitioning methods is to use some number k of parallel hyperplanes. The following question arises: what is the maximum number $P(N,k,n)$ of ways in which N points in $\mathbb{R}^n$ can be partitioned by k parallel hyperplanes (none of which contains any of the points)?

Answering this question, or obtaining good bound on the answer, has consequences for the 'capacity' of what have been called *multilevel threshold functions* [15] or *multilevel threshold elements* [17, 5], generalizations of the threshold functions and threshold elements so central to the theory of artificial neural networks. (These consequences are discussed in [2].)

2 Previous work

Olafsson and Abu-Mostafa [15] gave an upper bound on $P(N,k,n)$, correcting a claimed upper bound of Takiyama [17]. Their result is as follows.

Theorem 2.1 *The maximum possible number of ways in which N points in $\mathbb{R}^n$ can be partitioned by k parallel hyperplanes is bounded as follows:*

$$P(N,k,n) \leq \sum_{l=0}^{k}\binom{N-1}{l}\sum_{i=0}^{n-1}\binom{\binom{N}{2}-1}{i}.$$

Olafsson and Abu-Mostafa observed that if the k parallel hyperplanes have normal vector w, then the partition induced by these planes depends on two things: first, the ordering of the points when projected onto the line from the origin in direction w and, secondly, on the location of the k planes with respect to this ordering. They bounded the number of possible orderings when the points are projected onto a line, and noted that there are then at most $\sum_{l=0}^{k}\binom{N-1}{l}$ choices for the positioning of the planes with respect to the points. They arrived at Theorem 2.1 by deriving an upper bound of

$$C\left(\binom{N}{2}, n\right) = 2\sum_{i=0}^{n-1}\binom{\binom{N}{2}-1}{i}$$

on the number of orderings. (Their original bound has a factor 2 attached to it as a consequence, but this can be dropped here because, while this paper is concerned simply with the number of partitions, they were interested in the number of ways the points could be partitioned into $\{0,1\}$-labelled blocks, with adjacent blocks labelled differently.)

Ngom *et al.* [14] claimed to have proved that the number of partitions achievable with k parallel planes, in which none of the $(k+1)$ blocks is empty, is no more than

$$L(N,k,n) = \binom{N-2}{k-1}P(N,1,n) = \binom{N-2}{k-1}\sum_{i=0}^{n}\binom{N-1}{i}.$$

However, this is incorrect. They argue as follows, in a manner similar to that of Olafsson and Abu-Mostafa. For each single-plane partition (of which there are at most $P(N,1,n)$), order the N points according to their distance from the hyperplane. Then, to construct a k-parallel plane partition, one adds $k-1$ more parallel hyperplanes. The partition obtained is determined by the choice of the position of these parallel planes, and there are $\binom{N-2}{k-1}$ choices (if no block is to be empty). But this argument does not work. It *is* the case that a given hyperplane realizing a particular single-plane partition can give rise to at most $\binom{N-2}{k-1}$ partitions by a set of k parallel planes of which it is one. But there are many different hyperplanes realizing a particular single-plane partition, and these may give rise to different corresponding sets of k-parallel plane partitions. For a specific example, suppose that $A=(0,0)$, $B=(0,1)$, $C=(1,0)$ and $D=(1,1)$ in $\mathbb{R}^2$ and consider the single-plane partition with blocks $\{A\}$ and $\{B,C,D\}$. This is realizable by the hyperplane (line) H_1 with equation $x+2y=1/2$, and also by the hyperplane H_2 with equation $2x+y=1/2$. Each of H_1 and H_2 can, together with the introduction of an additional parallel plane, realize two partitions resulting from two parallel planes, with nonempty blocks. Explicitly, with H_1 and a further parallel plane, the partitions

$$\{A\}|\{C\}|\{BD\},\ \ \{A\}|\{B,C\}|\{D\}$$

can be obtained, and with H_2 and a further parallel plane, the partitions

$$\{A\}|\{B\}|\{CD\},\ \ \{A\}|\{B,C\}|\{D\}$$

can be obtained. So there are in fact at least three—and not at most two—distinct partitions (with non-empty blocks) consistent with the given single-hyperplane partition.

It is not possible, by some other means, to obtain the upper bound claimed by Ngom *et al.* [14], because it is does not hold, as can easily be shown by extending the argument just given, as follows.

Theorem 2.2 *Let* $\mathbf{o}$ *denote the origin in* $\{0,1\}^n$ *and* $\mathbf{e}_1, \dots, \mathbf{e}_n \in \{0,1\}^n$ *the unit basis vectors (with exactly one co-ordinate equal to* 1*). Then all partitions of these* $n+1$ *points into three non-empty blocks can be achieved using two parallel hyperplanes.*

Proof: Suppose that $E|F|G$ is any given partition into non-empty sets of $X = \{\mathbf{o}, \mathbf{e}_1, \dots, \mathbf{e}_n\}$. Suppose $\mathbf{o} \in E$. Define $w \in \mathbb{R}^n$ as follows: for $i = 1, 2, \dots, n$, $w_i = 0$ if $\mathbf{e}_i \in E$, $w_i = 1$ if $\mathbf{e}_i \in F$ and $w_i = 2$ if $\mathbf{e}_i \in G$. Consider the two hyperplanes H_1, H_2 with equations

$$H_1 : w^T x = 1/2, \quad H_2 : w^T x = 3/2.$$

Then it is easily seen, since $w^T \mathbf{e}_i > 3/2$ if and only if $\mathbf{e}_i \in G$ and $1/2 < w^T \mathbf{e}_i < 3/2$ if and only if $\mathbf{e}_i \in F$, that the parallel planes H_1 and H_2 realize the given partition.

It follows that, for all n, $P(n+1, 2, n)$ is at least $s(n+1, 3)$, where $s(n+1, 3)$, the Stirling number of the second kind, is the number of partitions into three non-empty blocks of $n+1$ objects, given explicitly by

$$s(n+1, 3) = \frac{1}{2}\left(3^n - 2^{n+1} + 1\right).$$

However, $L(n+1, 2, n) = 2^n(n-1)$, so the claimed bound of Ngom *et al.* is incorrect by a considerable margin. (It fails first, in this case, when $n = 7$, for $s(7+1, 3) = 966 > 762 = L(8, 2, 7)$.) Theorem 2.2 generalizes easily to the case of $k \geq 3$ hyperplanes by a very similar argument. It can then see that $P(n+1, k, n)$ is at least $(k+1)^n/(k+1)!$, while $L(n+1, k, n) = 2^n \binom{n-1}{k-1} \leq 2^n n^{k-1}/(k-1)!$ is very much smaller.

3 A refinement of a previous upper bound

To bound the number of possible orderings of N points when projected onto a line, Olafsson and Abu-Mostafa observed that an upper bound is given by the number of regions into which the planes with normals $x_i - x_j$ (for all x_i, x_j among the N points) partition $\mathbb{R}^n$. However, the problem of counting the number of orderings was previously considered by Cover [7]. Following Cover, a permutation π of $\{1, 2, \ldots, N\}$, is said to be a *linearly inducible ordering* of $X = \{x_1, x_2, \ldots, x_n\}$ if there exists $w \in \mathbb{R}^n$ such that

$$w^T x_{\pi(1)} > w^T x_{\pi(2)} > \cdots > w^T x_{\pi(n)},$$

meaning that π describes the order of the points when they are projected onto the line with direction w. Cover proved that if the N points of X are in general position then the number of possible orderings is exactly

$$Q(N, n) = 2 + 2 \sum_{i=1}^{n-1} R(N, i),$$

where, for $1 \leq i \leq n - 1$,

$$R(N, i) = \sum_{2 \leq y_1 < \cdots < y_i \leq N-1} y_1 y_2 \ldots y_i$$

is the sum of all the $\binom{N-2}{i}$ products of i numbers between 2 and $n - 1$.

Gould [10] subsequently expressed $Q(N, n)$ in terms of the Stirling numbers of the first kind, $S(r, s)$, where (with Gould's definitions) $S(r, s)$ is defined as the coefficient of x^s in $\prod_{j=1}^{r}(1 + jx)$. He showed that

$$Q(N, n) = 2 \sum_{j=0}^{\lfloor (n-1)/2 \rfloor} S(N - 1, n - 1 - 2j).$$

The following upper bound is therefore obtained.

Theorem 3.1 *The maximum possible number of ways in which N points in $\mathbb{R}^n$ can be partitioned by k parallel hyperplanes is bounded as follows:*

$$P(N, k, n) \leq 2 \sum_{l=0}^{k} \binom{N-1}{l} \sum_{j=0}^{\lfloor (n-1)/2 \rfloor} S(N - 1, n - 1 - 2j),$$

where $S(r, s)$ is the Stirling number of the first kind, the coefficient of x^s in $\prod_{j=1}^{r}(1 + jx)$.

4 A new upper bound

Theorem 3.1 is an improvement of Theorem 2.1, based on the same idea. But it is possible to obtain another bound using a different technique that has its roots in the proof of the $k = 1$ case given by Cover [6], and which can be traced back to Schläfli [16]. (Generalizations of this technique have recently proven useful for more complex partitioning methods; see [11, 4, 12, 3].)

Theorem 4.1 *The maximum possible number of ways in which N points in $\mathbb{R}^n$ can be partitioned by k parallel hyperplanes is bounded as follows:*

$$P(N, k, n) \leq \sum_{i=0}^{n+k-1} \binom{Nk-1}{i}.$$

Proof: Let N points, $x_1, x_2, \ldots, x_N \in \mathbb{R}^n$ be given. Now, each partition by k parallel hyperplanes can be described by a permissible parameter vector $\vec{p} = (w_1, w_2, \ldots, w_n, \theta_1, \theta_2, \ldots, \theta_k)$, where, to say the vector is permissible means that $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$. The proof hinges on considering the partitioning of the *parameter space* $\mathbb{R}^{n+k}$ by certain hyperplanes corresponding to the given points $x_1, x_2, \ldots, x_N$. For $1 \leq i \leq N$ and $1 \leq j \leq k$, suppose that $x_i = (x_1^i, x_2^i, \ldots, x_n^i)$ and define H_j^i to be the hyperplane with equation

$$\sum_{r=1}^{n} x_r^i w_r - \theta_j = 0,$$

which passes through the origin in $\mathbb{R}^{n+k}$. Note that, here, the variables are $w_1, \ldots, w_n$ and $\theta_1, \ldots, \theta_k$: the hyperplane is in parameter space, and is one of k corresponding to x_i. In this way, Nk hyperplanes are obtained. These Nk hyperplanes divide parameter space $\mathbb{R}^{n+k}$ into a number of regions, or cells. Suppose that two permissible parameter vectors

$$\vec{p} = (p_1, p_2, \ldots, p_n, \phi_1, \ldots, \phi_k), \quad \vec{q} = (q_1, q_2, \ldots, q_n, \psi_1, \ldots, \psi_k)$$

belong to the same cell in this decomposition. Then this means that for every $1 \leq i \leq N$ and $1 \leq j \leq k$, $\vec{p}$ and $\vec{q}$ lie on the same side of hyperplane H_j^i; that is,

$$\sum_{r=1}^{n} x_r^i p_i - \phi_j > 0 \Longleftrightarrow \sum_{r=1}^{n} x_r^i q_i - \psi_j > 0.$$

But this means that the two sets of parallel hyperplanes corresponding to $\vec{p}$ and $\vec{q}$ induce the same partition of X. By the classical result on the number of regions created by a set of hyperplanes passing through the origin [6, 16], the number of cells into which Nk planes can divide $\mathbb{R}^{n+k}$ is at most $C(Nk, n+k)$. Now, the fact that parameter vectors $\vec{p}$ and $-\vec{p}$ induce the same partition, and belong to distinct cells, means that the number of distinct ways in which N points in $\mathbb{R}^n$ can be partitioned by k parallel hyperplanes is at most $C(Nk, n+k)/2$, which is as required.

5 Discussion and conclusions

An existing upper bound on the number of ways in which a set of points can be partitioned by parallel hyperplanes has been improved. A previously claimed improvement has been shown to be incorrect. The new bound of Theorem 4.1 agrees in the case $k = 1$ with the well-known (tight) bound mentioned in the Introduction. It can also be shown to be quite tight, as follows. Olafsson and Abu-Mostafa [15] (in establishing that the bound claimed by Takiyama [17] failed as an upper bound, but did provide a lower bound) proved that N points of $\mathbb{R}^n$ in general position can be partitioned in at least $\sum_{i=0}^{n+k-1} \binom{N-1}{i}$ ways by k parallel planes. This quantity therefore provides a lower bound on $P(N, k, n)$. If n and k are fixed, then, as a function of N, this lower bound is $\Omega(N^{n+k-1})$, and the upper bound of Theorem 4.1 is of order N^{n+k-1}. Thus, for fixed n and k, the new upper bound of Theorem 4.1 is tight to within a constant.

The original upper bound, Theorem 2.1, of Olafsson and Abu-Mostafa is, for fixed n and k, $O(N^{2n+k})$. As noted, their lower bound is $\Omega(N^{n+k-1})$. Olafsson and Abu-Mostafa [15] claimed that the dependence upon n in their upper bound appeared to be, in a sense, asymptotically correct, asserting that "Expressing the result asymptotically as $N^{\alpha n+k}$, we find that [...] $1 \leq \alpha < 2$" and that "we are led to conclude that α is indeed greater than 1, and apparently approaches 2 as N, n, and k approach infinity, with n and k growing logarithmically in N." However, the bound of Theorem 4.1 now suggests otherwise: this has been commented on already for fixed n and k and follows, for n, k growing slowly with respect to N, from the observation that

$$\sum_{i=0}^{n+k-1} \binom{Nk-1}{i} < N^{n+k} k^{n+k},$$

which is $N^{(n+k)(1+o(1))}$ as $n, k, N \to \infty$ with n, k logarithmic in N.

As a final remark, it can be shown that the upper bound given in Theorem 3.1, while it is an improvement of Theorem 2.1 because it is based on a tight bound on the number of linearly inducible orders, is still, as a function of N, $\Omega(N^{2n+k})$ for fixed n and k. In this sense, therefore, the bound given by Theorem 4.1 is better.

Acknowledgements

I am grateful to Graham Brightwell for useful comments, particularly concerning asymptotics. I started to think about this problem while visiting RUTCOR, Rutgers University, and I thank Peter Hammer and colleagues for their hospitality.

References

[1] M. Anthony. Classification by polynomial surfaces. *Discrete Applied Mathematics*, 61, 1995: 91–103.

[2] M. Anthony. Analysis of data with threshold decision lists. In preparation.

[3] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

[4] S. Ben-David and M. Lindenbaum. Localization vs. identification of semi-algebraic sets. In *Proceedings of the* 6*th Annual Workshop on Computational Learning Theory*, ACM Press, New York, NY, 1993: 327–336.

[5] V. Bohossian and J. Bruck. Multiple threshold neural logic. In *Advances in Neural Information Processing, Volume 10: NIPS'1997*, Michael Jordan, Michael Kearns, Sara Solla (eds), MIT Press, 1998.

[6] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Trans. Electronic Computers* 14, 1965: 326–334.

[7] T. M. Cover. The number of linearly inducible orderings of points in d-space. *SIAM Journal on Applied Mathematics*, 15(2), 1967: 434–439.

[8] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK, 2000.

[9] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*, John Wiley, 1973.

[10] H. W. Gould. A note on the number of linearly inducible orderings of points in d-space. *SIAM Journal on Applied Mathematics*, 26(3), 1974: 528–530.

[11] P.W. Goldberg and M. R. Jerrum. Bounding the Vapnik-Chervonenkis dimension of concept classes parameterized by real numbers. *Machine Learning* 18(2/3), 1995: 131–148.

[12] M. Karpinski and A. MacIntyre. Polynomial bounds for VC dimension of sigmoidal and general Pfaffian neural networks. *Journal of Computer and System Sciences*, **54**, 1997: 169–176.

[13] M. Minsky and S. Papert. *Perceptrons*. MIT Press, Cambridge, MA., 1969. (Expanded edition 1988.)

[14] A. Ngom, I. Stojmenović and J. Žunić. On the number of multilinear partitions and the computing capacity of multiple-valued multiple-threshold perceptrons. In *Twenty-Ninth IEEE International Symposium on Multiple-Valued Logic*, 1999. IEEE Press.

[15] S. Olafsson and Y. S. Abu-Mostafa. The capacity of multilevel threshold functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10 (2), 1988: 277–281.

[16] L. Schläfli. *Gesammelte Mathematische Abhandlungen I*, Birkhäuser, Basel, 1950.

[17] R. Takiyama. The separating capacity of a multi-threshold element. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7, 1985: 112–116.

[18] V. N. Vapnik: *Statistical Learning Theory*, Wiley, 1998.