# A LINK BETWEEN DEGREES OF FREEDOM AND VC-DIMENSION

DAN SIMON

ABSTRACT. It has been said that intuitively, the VC dimension measures the number of degrees of freedom of a set of hypotheses. However, this is commonly treated more as a heuristic than as any type of formal result; as a result, this connection is generally not stated in papers, instead being mentioned in passing in classes such as [4]. Furthermore, there are cases where the number of degrees of freedom is clearly less than the VC dimension. We provide a topological definition of "degrees of freedom" that matches up with intuition in many common cases, and present the hypothesis that this is a lower bound for the VC dimension. We provide examples of cases where the intuitive number of degrees of freedom appears to be more than the VC dimension, and show that surprisingly these seem to be the cases where our topological definition disagrees with intuition. We prove our hypothesis in the case where the number of degrees of freedom is less than 3 (that is, 0, 1, or 2) and provide some progress toward a general proof. We also give some rather broad and potentially useful consequences that would easily follow.

## 1. SOME BASIC TOPOLOGY

It turns out that the conjecture made in this paper has much more of a relationship with topology than with other fields, like combinatorics, which are more often associated with VC-dimension. Thus, from now on, we often use basic topological concepts such as paths, compactness, the interior of a set, etc. One can read any basic topology book for an overview of these concepts and facts here used about them. However, here is a brief explanation of some of these concepts, which will hopefully let readers who do not know much topology get an idea of what is meant when topological terms are used. Readers who know topology can likely skip this section without loss.

The interior of a set is the union of all the open balls contained entirely within a set, or equivalently, the largest open set contained entirely within a set. For example, the interior of the closed disk of a certain radius is the open disk of the same radius. The interior of any set is open.

An open cover of a set is a set of open sets whose union is the entire set. A compact set is a set of which any open cover has a finite subset which is also an open cover. It is common knowledge that the sphere is compact.

A path from $a$ to $b$ in a topological space $S$ is a continuous map $f$ from $[0, 1]$ to $S$ such that $f(0) = a$ and $f(1) = b$.

## 2. MOTIVATION AND CONJECTURE

It is said that the VC-dimension measures the number of degrees of freedom of a set of hypotheses. But mathematically, a set of hypotheses $\mathcal{H}$ is generally said to have $n$ degrees of freedom when it can be parametrized by $n$ real numbers, that is, when there is a 'natural' map from $\mathbb{R}^n$ (the $n$ parameters) to $\mathcal{H}$ (the set of parametrized hypotheses).

But what makes a map 'natural'? At the very least, we want to rule out constant maps: maps where every point in $\mathbb{R}^n$ maps to the same hypothesis. It is easy to see that hypothesis sets generated by such constant maps will only have a single hypothesis, and thus not be able to shatter any nonempty subset of points. So we forbid such constant maps. In fact, we forbid any two points in $\mathbb{R}^n$ from mapping to the same hypothesis, that is, we require the map to be injective.

However, this is still not enough since we are not using any of the structure of $\mathbb{R}^n$. Without structure, each $\mathbb{R}^n$ is the same; an uncountable collection of points. Thus, if a hypothesis set $\mathcal{H}$ has 1 degree of freedom due to a map from $\mathbb{R}$ to $\mathcal{H}$, we can compose this map with a bijective map from $\mathbb{R}^2$ to $\mathbb{R}$ to get a map from $\mathbb{R}^2$ to $\mathcal{H}$, showing that $\mathcal{H}$ has 2 degrees of freedom. This is very undesirable, so we use the minimum structure we can get away with, which is topology. We can in fact combine this with injectivity in a nice way, which is given below.

Here is how we will define "degrees of freedom" for the purpose of this talk:

**Definition 1.** *A hypothesis set $\mathcal{H}$, the elements of which are subsets of a space $\mathcal{X}$ from which we do not require any structure, has $n$ degrees of freedom if and only if there is a map $f$ (not necessarily surjective, and we will see why later) from $\mathbb{R}^n$ to $\mathcal{H}$ such that for any $p_1$ and $p_2$ in $\mathbb{R}^n$, there is an element $x \in \mathcal{X}$, a ball $B_1$ around one of $p_1$ and $p_2$, and a ball $B_2$ around the other, such that for $q \in B_1$, $x \in f(q)$ and for $q \in B_2$, $x \notin f(q)$.*

In other words, for any two hypotheses that are values of the map from $\mathbb{R}^n$ to $\mathcal{H}$, there is not only a point in $\mathcal{X}$ at which they differ, but a point in $\mathcal{X}$ at which all hypotheses sufficiently close to one differ from all hypotheses sufficiently close to the other. This ties together injectivity, the topology of $\mathbb{R}^n$ (we refer to balls, but it is easy to see that open neighborhoods would have been equivalent), and even the desire that hypotheses generated from

close points in $\mathbb{R}^n$ should be similar (since we say that all hypotheses resulting from points in certain balls have the same value at certain points).

The definition above is complicated, but as we will see in the consequences section, if a set of hypotheses seems to have $n$ real degrees of freedom, this definition probably implies that it does, and it is often very easy to see.

The conjecture is now simply:

**Conjecture 2.** *If a set has $n$ degrees of freedom, its VC-dimension is at least $n$.*

We need 'at least' here. Unit circles in the plane have 2 degrees of freedom, since they are determined by the $x$-coordinate and $y$-coordinate of their center. However, it is easy to see that an equilateral triangle is shattered by the set of unit circles, and thus this set of hypotheses has VC-dimension at least 3. With this in mind, note that a subset of a hypothesis set has VC-dimension less than or equal to that of the original set, so if a subset has VC-dimension at least $n$ so does the original set. Thus, if we know that a subset of a hypothesis set has $n$ degrees of freedom, then the conjecture claims that its VC-dimension, and thus the VC-dimension of the original set, is at least $n$. Thus, for the purpose of our conjecture, it makes sense to say that if a subset of a set has $n$ degrees of freedom, so does the whole set. But we get this without changing anything if we do not require the map from $\mathbb{R}^n$ to $\mathcal{H}$ to be surjective, in which case the subset of hypothesis that is the image of this map also has $n$ degress of freedom. That is why we did not require surjectivity. Nevertheless, it is reasonable to only consider the hypotheses in its image, and we will often do this.

Before continuing, let us examine an interesting case. The set of points of $\mathbb{R}^2$ intuitively has 2 degrees of freedom. However, if these points are considered as hypotheses only containing one point, any two points have no hypothesis containing both, and thus are not shattered. So the VC dimension is less than 2. This seems to contradict our conjecture. However, we can see that our definition does not say that this set of hypotheses has 2 degrees of freedom. For that to happen, for any two hypotheses $H_1$ and $H_2$, there would need to be a point $p$ in $\mathbb{R}^2$ at which all hypotheses sufficiently close to one differ from all hypotheses sufficiently close to the other. So either all hypotheses sufficiently close to $H_1$ contain $p$, or all hypotheses sufficiently close to $H_2$ contain $p$. In either case, an infinite number of hypotheses (the image of a ball in $\mathbb{R}^2$ under an injective map is infinite) contain $p$. But only 1 hypothesis does, so our definition says that set set of single-point hypotheses does not have 2 degrees of freedom. On one hand, this does conflict with intuition, but if it were not the case the conjecture made above would be false, so we cannot complain too much.

## 3. An important alternate way of viewing the problem

We can look at points in the sample space as subsets of the set of hypotheses, where the subset corresponding to a point is the subset of hypotheses containing it. We know that a set of points $S$ is shattered if there is a hypothesis containing any subset $S'$ of these points and none of the points in $S \setminus S'$. When we look at this in terms of subsets of $\mathbb{R}^n$, we see that it is equivalent to saying that for any subset $S'$ of the set $S$ of subsets of $\mathbb{R}^n$ corresponding to these points there is a point $p$ in $\mathbb{R}^n$ (a hypothesis) in each element of $S'$ but not in any elements of $S \setminus S'$. (We are slightly abusing notation here by referring to points on $\mathbb{R}^n$ as if they are hypotheses, even though they just map to hypotheses, and we will continue to abuse notation in this way.) Thinking of the problem in this way as a problem about $\mathbb{R}^n$, and thinking of the sample space as a set of subsets of $\mathbb{R}^n$, is often easier than doing the "dual" and thinking about hypotheses as subsets of the sample space. Thus, from now on the sample space refers to a set of subsets of $\mathbb{R}^n$.

Of course, our condition of degrees of freedom is translated to the statement that for any two different points in $\mathbb{R}^n$, there is a subset $S$ in the sample space containing a neighborhood of one but disjoint with a neighborhood of the other.

## 4. Proofs in Special Cases ($n \leq 2$, where $n$ is the number of degrees of freedom)

$n = 0$: This is clear since there are always $0$ points that are shattered.

$n = 1$: Pick any two hypotheses (say, those corresponding to $0$ and $1$) and consider a point on which they differ. This point is shattered.

$n = 2$: Note that all we need here is a circle $C$ that maps to the hypothesis set, not all of $\mathbb{R}^2$. Choose two different points $p_1$ and $p_2$ on $C$. Choose a subset $S$ in the sample space which is in all hypotheses in a neighborhood of one of $p_1$ and $p_2$, and not in any hypotheses in a neighborhood of the other. Without loss of generality (by switching $p_1$ and $p_2$ if needed), we can assume that $s$ is in all hypotheses in a neighborhood of $p_1$ and in none in a neighborhood of $p_2$.

Much of this paragraph is basically just a formalization of the idea that we can find two points on $C$ on the boundary of $S$, and then take $T$ to be a subset in the sample space containing a neighborhood of one of these points but not the other. There are two non-intersecting paths on $C$ from $p_1$ to $p_2$, one clockwise and the other counterclockwise. Topologically, these paths can be represented as continuous functions $f_1$ and $f_2$ from $[0,1]$ to $C$ such that $f_1(0) = f_2(0) = p_1$, $f_1(1) = f_2(1) = p_2$, and $f_1$ and $f_2$ have no other common values. Consider the sets $S_1$ and $S_2$, where $S_i = \{x \in [0,1] : f_i(x) \in S\}$. If $x$ is sufficiently close to $1$, $f_i(x)$ is close to $p_2$, so $f_i(x) \notin S$

and $x \notin S_i$. Similarly, if $x$ is sufficiently close to 0, $f_i(x)$ is close to $p_1$, so $f_i(x) \in S$ and $x \in S_i$. Thus $\sup S_i$ is neither 0 (since small enough numbers are in $S_i$) nor 1 (since large enough numbers are not in $S_i$). So $f_1(\sup S_1) \neq f_2(\sup S_2)$. There is a subset $T$ in the sample space such that $T$ contains a neighborhood of $f_1(\sup S_1)$ or $f_2(\sup S_2)$ and is disjoint with a neighborhood of the other.

I claim $S$ and $T$ are shattered (which suffices to get what we want, which is a proof of VC-dimension 2). This is true since for small enough $\epsilon$, $f_1(\sup S_1 + \epsilon)$ is still near enough to $f_1(\sup S_1)$ to be in $T$ or not in $T$ with the same status as $f_1(\sup S_1)$, but $f_2(\sup S_2 + \epsilon)$ is near enough to $f_2(\sup S_2)$ to have the same status as $f_2(\sup S_2)$ concerning being in $T$, and thus the status opposite to that of $f_1(\sup S_1 + \epsilon)$. However, neither $f_1(\sup S_1 + \epsilon)$ nor $f_2(\sup S_2 + \epsilon)$ is in $S$ (since otherwise $\sup S_1$ or $\sup S_2$ would be higher) so there is a point in neither $S$ nor $T$ and a point in $T$ but not $S$. Now all we have to do to prove shattered-ness is to find a point in $S$ but not $T$ and a point in $S$ and $T$. We can do this by noticing that for any $\epsilon > 0$, there is an $\epsilon' < \epsilon$ such that $\sup S_i - \epsilon' \in S_i$, and thus by decreasing $\epsilon$ we can get points $f_i(\sup S_i - \epsilon')$ in $S$ and arbitrarily close to $f_i(\sup S_i)$. By getting these points close enough, the one close to $f_1(\sup S_1)$ will be in $T$ or not in $T$ in accordance with whether $f_1(\sup S_1)$ in in $T$, and the one close to $f_2(\sup S_2)$ will be the opposite with respect to being in $T$. This gives up two points in $S$, one of which is in $T$ and the other of which is not, as required.

## 5. Some progress toward a general proof

Here we note that the set of pairs of different points in $\mathbb{R}^n$ has a subset containing pairs of antipodal (opposite from each other) points in $S^{n-1}$, where, as in mathematical notation, $S^{n-1}$ is the unit sphere in $\mathbb{R}^n$. So it suffices to prove

**Conjecture 3.** *If for any two antipodal points in $S^{n-1}$, there is a subset $S$ in the sample space containing a neighborhood of one but disjoint with a neighborhood of the other, then the VC-dimension is at least $n$.*

(All we are saying is that if we can find a proof with a weaker assumption, then we can prove the theorem in its original form.)

But for each subset $S$ in the sample space, let $g(S)$ be $(S \cap A(S)^{co}) \cup A(S \cap A(S)^{co})$ where $A(S)$ is the set of antipodes of points in $S$, and the $o$ superscript marks interior. A point is in $g(S)$ if and only if either it is in the interior of $S$ and its antipode is in the interior of the complement of $S$, or its antipode is in the interior of $S$ and its antipode's antipode (which is itself) is n the interior of the complement of $S$. But this is equivalent to there being either a neighborhood in $S$ around it and a neighborhood in $S^c$

around its antipode, or a neighborhood in $S^c$ around it and a neighborhood in $S$ around its antipode. So claiming that for any two antipodal points in $S^{n-1}$, there is a subset $S$ in the sample space containing a neighborhood of one but disjoint with a neighborhood of the other is equivalent to claiming that for any point in $S^{n-1}$, either there is a neighborhood in $S$ around it and a neighborhood in $S^c$ around its antipode, or there is a neighborhood in $S^c$ around it and a neighborhood in $S$ around its antipode, and this is equivalent to saying that every point in $S^{n-1}$ is in some $g(S)$, that is, $S^{n-1}$ is covered by the sets $g(S)$.

But each $g(S)$ is open and $S^{n-1}$ is compact, so we only need a finite number of sets from the cover to satisfy the hypothesis. So it suffices to prove that if the sample space is finite and for any two antipodal points in $S^{n-1}$, there is a subset $S$ in the sample space containing a neighborhood of one but disjoint with a neighborhood of the other, then the VC-dimension is at least $n$, since we can reduce all other cases to the finite case.

Furthermore, for any $S$ in the sample space we can divide $S^{n-1}$ into $g(S) \cap S$ (which is open), $g(s) \cap S^c$ (which is open), and $g(S)^c$ (which is closed). By combining all of these three-fold partitions we can divide $S^{n-1}$ into at most $3^n$ pieces, where $n$ is the number of elements in the sample space. For any nonempty piece $P$, we can pick a point $p \in P$. If we change each $S$ in the sample space so that it contains $P$ if $p \in S$ and is disjoint with $P$ otherwise, this change preserves the truth of the assumption and cannot increase the VC dimension.

The claim directly above is true since for any $S$ in the sample space, if $P \in g(S) \cap S$ or $P \in g(S) \cap S^c$, $S$ already is the same throughout $P$ (including on $p$) and so there is no change, and if $P \in g(S)^c$ a change in $S$ on it cannot remove points from $g(S)$, so the sets $g(S)$ will still cover $S^{n-1}$. On the other hand, consider the possible subsets of the sample space consisting of elements including a given point in $S^{n-1}$ (after we change each $S$ on $P$). Call the set of such subsets $Q$. They are all either subsets that occur for points outside of $P$ (which occurred before) or the subset of of the sample space consisting of elements containing $P$ (which also occurred before). The VC-dimension is at least $n$ if there is a subset $S$ of $n$ points in the sample space that are shattered, which occurs if and only if for each subset $S'$ of $S$ there is a point in those and no others, which itself occurs if and only if for each subset $S'$ of $S$ there is an element $q$ of $Q$ such that $q \cap S = S'$, since such a $q$ comes from a point in $S^{n-1}$ for which the subset of the sample space that contains it includes every element in $S'$ and none in $S \setminus S'$. So the VC-dimension is at least $n$ if and only if there exists a subset $S$ of the sample space with $n$ elements such that for each subset $S'$ of $S$ there is an element $q$ of $Q$ such that $q \cap S = S'$. Thus the VC-dimension

cannot go up without new elements in $Q$, and since our change to each $S$ on $P$ did not add new elements to $Q$ it cannot increase the VC dimension.

So, if for each nonempty piece $P$ we pick a point $p \in P$ and change each $S$ in the sample space so that it contains $P$ if $p \in S$ and is disjoint with $P$ otherwise, the hypothesis of Conjecture 3 (that $S^{n-1}$ is covered by the sets $g(S)$) is still true and if the conclusion is true now, the VC dimension could only have been higher or equal originally, so the conclusion was true originally. Thus we only have to prove Conjecture 3 in the case where for each $S$ and each piece $P$, $S$ either contains or is disjoint with $P$. Note that changing $S$ may have made $g(S)$ larger and thus changed the boundaries of the partitions. However, each part is still a finite intersection of open and closed sets, and each $S$ in the sample space is a finite union of parts, so each $S$ in the sample space is a finite union of finite intersections of open and closed sets, and thus measurable. In particular, this tells us that we only have to consider the case where each $S$ in the sample is measurable.

## 6. CONSEQUENCES

The applications below are for hypotheses which can be considered as a subset of $\mathbb{R}^n$ (where each sample point's location in $\mathbb{R}^n$ is determined by real-valued features). However, no structure on the sample space is needed to use Conjecture 2.

Note that as a general principle, since we only use the topological structure of $\mathbb{R}^n$ and $\mathbb{R}^n$ is topologically equivalent to an open ball, it suffices to find a map from an open ball to hypotheses.

In all the cases given below, the actual VC-dimension of the hypothesis set in question has been found. As can be seen, the lower bound Conjecture 2 gives is often the correct answer.

6.1. **Spheres.** An $n$-dimensional sphere (to be precise, we are here talking about an $n$-dimensional ball, not a phere) has $n + 1$ degrees of freedom: 1 for the radius and $n$ for the center. The open ball we can use as the domain of our parametrization map is centered at $(1, 0, ..., 0)$ and of radius $\frac{1}{2}$, and it is easy to see that if the first coordinate gives the radius and the rest are the coordinates of the center, all the hypotheses mapped to are distinct spheres. Furthermore, any two of these distinct spheres have a point which one of the spheres (and all spheres with similar enough parameters) contains and which the other (and, again, all spheres with similar enough parameters) does not contain. (The interior of one will have points in the interior of the complement of the other, and they vary continuously. Basically, no two spheres are anywhere near the same.) But this is exactly what Conjecture 2 requires, so assuming it we see that the VC-dimension of the set of spheres is $n + 1$. This is indeed the value given by [3].

Note that in general, under any sane parametrization, any two shapes will have a point which they differ on and which they agree with some neighborhood of themselves on. This can, as said above, be seen by noting that they disagree somewhere other than their boundaries (this is not true for such hypotheses as points or lines, but it is almost always otherwise true, and it is true in all the examples we consider), so there is a point $p$ in the interior of one and not in the interior of the other. The shapes vary continuously (this should generally be clear from the parametrization), so any shapes sufficiently "nearby" the originals (in the parametrization) will have the same behavior as the originals with respect to containing $p$. This logic is sufficiently generally applicable that it works with all our consequences in this section, and thus we do not again mention it explicitly.

6.2. **Axis-aligned Hyper-rectangles.** An $n$-dimensional axis-aligned hyper-rectangle has $2n$ degrees of freedom. We can parametrize this by letting coordinates $2m+1$ and $2m+2$ of the parametrization vector be the lower and upper bounds for the $m$th coordinate in the axis-aligned hyper-rectangle, where of course coordinate $2m+1$ has to be less than coordinate $2m+2$. The open ball we can use as the domain of our parametrization map is centered at $(0, 1, 0, 1, ..., 0, 1)$ and of radius $\frac{1}{2}$, and it is easy to see that the radius is small enough that all hypotheses generated are indeed axis-aligned hyper-rectangles, rather than just being empty. So the VC-dimension of the set of $n$-dimensional axis-aligned hyper-rectangles is at least $2n$. [2] (in Example 2.2) tells us it is exactly $2n$.

6.3. **Polygons.** An $n$-gon has $2n$ degrees of freedom: $2$ for each of the $n$ vertices. We can represent the coordinates of vertex $n$ by two coordinates in the parametrization. It is intuitively clear that there is an $\epsilon > 0$ such that if we start with an equilateral, regular, $n$-gon and change the location of no vertex by more than $\epsilon$, we still have an $n$-gon and it is still clear which vertex comes from which. But this shows that we can map an open set around a point (the points within $\epsilon$ of it for each coordinate) in $2n$-dimensional space to distinct $n$-gons. Thus $n$-gons have VC-dimension at least $2n$. Note that [2] (in Example 3.2.2) provides the answer $2n + 1$ for the VC-dimension of the set of $n$-gons (furthermore, it easily provides $2n + 1$ points that are shattered), which is higher than the bound we got but is at least within one. Still, this is another case where Conjecture 2 is not an equality.

6.4. **Ellipsoids.** An $n$-dimensional ellipsoid can be defined by the equation

$$c_{n+1}(x_1 - c_1)^2 + ... + c_{2n}(x_2 - c_n)^2 + c_{2n+1}(x_1 - c_1)(x_2 - c_2)+$$
$$c_{2n+2}(x_1 - c_1)(x_3 - c_3) + ... + c_{\frac{n^2+3n}{2}}(x_{n-1} - c_{n-1})(x_n - c_n) = 1,$$

which has $\frac{n^2+3n}{2}$ parameters (the first $n$ for translation, the second $n$ for the square terms, and the remaining $\frac{n^2-n}{2}$ for the products of different coefficients), each a degree of freedom. A sphere, an example of an ellipsoid, can have its parameters slightly changed and remain an ellipsoid, so the domain for our parametrization can be sets of parameters within an open ball of the sphere's parameters. Thus the VC-dimension of the set of ellipsoids is at least $\frac{n^2+3n}{2}$. [1] proved this (and also showed that $\frac{n^2+3n}{2}$ is a upper bound) just five years ago, so unlike some of the results above it is rather recent.

## 7. FINAL NOTES

We have seen that Conjecture 2 seems to give many interesting lower bounds rather quickly and easily (although it does not give upper bounds). Thus, it seems to be a useful idea. However, this depends on Conjecture 2 being true, and I do not see any way to easily reach a proof of it from where I am right now (although I have not been able to find a counter-example, either). If there could be a proof of this conjecture, though, I expect that it would provide a lot of new and interesting VC-dimension bounds.

## REFERENCES

[1] Y. Akama and K. Irie. VC dimension of ellipsoids, 2011.
[2] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis Dimension. *Journal of the ACM*, 36:929–965, 1989.
[3] R. M. Dudley. Balls in $\mathbb{R}^k$ Do Not Cut All Subsets of $k + 2$ Points. *Advances in Mathematics*, 31:306–308, 1979.
[4] M. Mohri. Foundations of Machine Learning, Fall 2016.