



HAL
open science

Vapnik-Chervonenkis Dimension of Axis-Parallel Cuts

Servane Gey

► **To cite this version:**

| Servane Gey. Vapnik-Chervonenkis Dimension of Axis-Parallel Cuts. 2012. hal-00675553v1

HAL Id: hal-00675553

<https://hal.science/hal-00675553v1>

Preprint submitted on 1 Mar 2012 (v1), last revised 19 Oct 2016 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vapnik-Chervonenkis Dimension of Axis-Parallel Cuts

Servane Gey*

March 1, 2012

Abstract

The Vapnik-Chervonenkis dimension (VC dimension) of the set of half-spaces of \mathbb{R}^d with frontiers parallel to the axes is computed exactly. It is shown that this VC dimension is smaller than the intuitive value of d . An additional approximation using the Stirling's formula is given. This result may be used to evaluate the performance of classifiers or regressors based on dyadic partitioning of \mathbb{R}^d for instance. Algorithms using axis-parallel cuts to partition \mathbb{R}^d are often used to reduce the computational time of such estimators when d is large.

Keywords: Vapnik-Chervonenkis dimension, axis-parallel cuts. *MSC 2010 Classification:* 62G99 62H99

1 Introduction

The Vapnik-Chervonenkis dimension (denoted by VC dimension) of a set of subsets, or equivalently of a function space, has been introduced by Vapnik and Chervonenkis [9, 10] to measure the complexity of the corresponding set. It plays an important role in many applications. In particular, the VC dimension appears commonly in the statistical learning area when evaluating the performance of classifiers or regressors.

For example, Vapnik's theory in the classification framework is now widely known (see [3] for instance): let (X, Y) be a couple of variables taking values in $\mathbb{R}^d \times \{0, 1\}$, and let \mathcal{L} be a sample of n independent replications of (X, Y) . If \hat{f} is a classifier minimizing the average misclassification rate of \mathcal{L} on a set

*Servane.Gey@parisdescartes.fr, Laboratoire MAP5 - UMR 8145, Université Paris Descartes, 75270 Paris Cedex 06, France

of classifiers having finite VC dimension V , then, if no assumption is made on the distribution P of (X, Y) , the performance of \hat{f} is evaluated as follows:

$$\mathbb{E}_{\mathcal{L}} \left[P \left(\hat{f}(X) \neq Y \right) \right] \leq C_1 \text{bias}^2(\hat{f}) + C_2 \sqrt{\frac{V}{n}}, \quad (1)$$

where $\mathbb{E}_{\mathcal{L}}$ denotes the expectation with respect to the sample distribution, $\text{bias}(\hat{f})$ denotes the bias of the classifier \hat{f} , and C_1 and C_2 are absolute constants.

Functional estimates defined on partitions of \mathbb{R}^d are often used to estimate relationships between two variables $X \in \mathbb{R}^d$ and $Y \in \{0; 1\}$ or $Y \in \mathbb{R}$ (such as histograms, piecewise polynomials, or splines for example). In many cases, the VC dimension of the set of subsets used to construct the partition appears inside risk bounds when evaluating the performance of such estimators. For example, if the set used is the set of all half-spaces of \mathbb{R}^d , often its VC dimension $d + 1$ has to be taken into account.

When d is large, it is often computationally easier to construct partitions using axis-parallel cuts. For example, some theoretical developments on dyadic partitions of \mathbb{R}^2 are given in [4, 1], and the VC dimension of axis-parallel cuts appears more particularly in the results obtained on the performance of classification and regression binary decision trees (CART) introduced by Breiman *et. al* [2] in 1984, and theoretically studied in [8, 7, 5, 6].

In the present paper, we compute the VC dimension of axis-parallel cuts after a slight reminder about VC dimension. A good approximation of the thus computed VC dimension is also given via the Stirling's formula.

2 Reminder about VC Dimension

The VC dimension of a set \mathcal{A} of subsets of some measurable space \mathcal{X} is based on counting the number of intersects of \mathcal{A} with a finite set of fixed points in \mathcal{X} . This count is closely related to the shatter coefficient.

Definition 1 (Shatter Coefficient). *Let \mathcal{A} be a set of subsets of some measurable space \mathcal{X} . For $n \geq 1$, the shatter coefficient of \mathcal{A} of order n is defined as*

$$S_{\mathcal{A}}(n) = \max_{(x_1, \dots, x_n) \in \mathcal{X}^n} |\{ \{x_1, \dots, x_n\} \cap A ; A \in \mathcal{A} \}|.$$

Example : $\mathcal{A} = \{] - \infty; x] ; x \in \mathbb{R}\}$. Then, for all $n \geq 1$, $S_{\mathcal{A}}(n) = n + 1$.

Hence, the VC dimension is defined as follows.

Definition 2 (Vapnik-Chervonenkis Dimension). Let \mathcal{A} be a set of subsets of some measurable space \mathcal{X} . Then $(x_1, \dots, x_n) \in \mathcal{X}^n$ will be said to be shattered by \mathcal{A} if all subsets of $\{x_1; \dots; x_n\}$ are covered by \mathcal{A} , that is if $|\{\{x_1, \dots, x_n\} \cap A ; A \in \mathcal{A}\}| = 2^n$.

The Vapnik-Chervonenkis dimension $VC(\mathcal{A})$ of \mathcal{A} is then defined as the maximal integer n such that there exists n points in \mathcal{X} shattered by \mathcal{A} , i.e.

$$VC(\mathcal{A}) = \max \{n ; S_{\mathcal{A}}(n) = 2^n\}.$$

If no such n exists, then $VC(\mathcal{A}) = +\infty$.

Thus, it is easily seen that the larger $VC(\mathcal{A})$, the more complex \mathcal{A} .

Examples

- 1) $\mathcal{A} = \{] - \infty; x] ; x \in \mathbb{R}\}$. Since, for all $n \geq 1$, $S_{\mathcal{A}}(n) = n + 1$, then $VC(\mathcal{A}) = 1$.
- 2) If \mathcal{A} is the set of all half-spaces in \mathbb{R}^d , then $VC(\mathcal{A}) = d + 1$.

3 VC Dimension of axis-parallel cuts

We give a formula to compute the VC dimension of axis-parallel cuts in \mathbb{R}^d . Since the obtained formula is not always easy to handle, an approximation is also given.

Lemma 1. Let

$$\mathcal{A}_d = \left\{ \{x \in \mathbb{R}^d ; x^i \leq a\} ; i = 1, \dots, d, a \in \mathbb{R} \right\}.$$

Then

$$VC(\mathcal{A}_d) = \max \left\{ n ; \binom{n}{\lfloor n/2 \rfloor} \leq d \right\},$$

where $\lfloor n/2 \rfloor$ denotes the integer part of $n/2$.

Furthermore, the following approximation of $VC(\mathcal{A}_d)$ is available for all $d \geq 3$:

$$\frac{\log d}{\log 2} - 1.18 \leq VC(\mathcal{A}_d) \leq d.$$

Intuitively, since \mathcal{A}_d is a subset of half-spaces of \mathbb{R}^d , it could be natural to think that its VC dimension is of order d because the VC dimension of all half-spaces of \mathbb{R}^d is equal to $d + 1$. Thus, Lemma 1 shows that, for some d , $VC(\mathcal{A}_d)$ is far from d .

Remark: a simple calculation gives $VC(\mathcal{A}_1) = 1$ and $VC(\mathcal{A}_2) = 2$.

Proof. The idea is that, to have n points (x_1, \dots, x_n) shattered by \mathcal{A}_d , all the subsets of $\{x_1, \dots, x_n\}$ should be covered by \mathcal{A}_d . But, if there exists $p \leq n$ such that there is more than $d + 1$ subsets of $\{x_1, \dots, x_n\}$ having p elements, then \mathcal{A}_d will miss at least $\binom{n}{p} - d$ subsets: let $n \geq 1$ and (x_1, \dots, x_n) be n points in \mathbb{R}^d . Suppose that n is such that $\binom{n}{\lfloor n/2 \rfloor} > d$. This means that there are at least $d + 1$ subsets of $\{x_1, \dots, x_n\}$ of size $\lfloor n/2 \rfloor$. For each coordinate $i = 1, \dots, d$, let us denote by $x_{i(\cdot)}$ the ordered statistic computed from the i^{th} coordinate of (x_1, \dots, x_n) , that is, for all $i = 1, \dots, d$,

$$x_{i(1)}^i \leq x_{i(2)}^i \leq \dots \leq x_{i(n)}^i.$$

Let $p = \lfloor n/2 \rfloor$ and let

$$\begin{aligned} \mathcal{B}_p &= \{ \{x_{i(1)}; \dots; x_{i(p)}\} ; i = 1, \dots, d \text{ and } |\{x_{i(1)}; \dots; x_{i(p)}\}| = p \}, \\ \mathcal{B}_p^c &= \{ B \subset \{x_1, \dots, x_n\}; |B| = p \text{ and } B \notin \mathcal{B}_p \}. \end{aligned}$$

Hence \mathcal{B}_p is covered by \mathcal{A}_d (by simply taking $A = \{x^i \leq (x_{i(p)}^i + x_{i(p+1)}^i)/2\}$ for each coordinate), and we have that:

$$|\mathcal{B}_p| \leq d \quad \text{and} \quad |\mathcal{B}_p^c| \geq \binom{n}{p} - d > 0.$$

Let $B \in \mathcal{B}_p^c$ and $A = \{x^i \leq a\} \in \mathcal{A}_d$. If $|\{x_1, \dots, x_n\} \cap A| \neq p$, then $\{x_1, \dots, x_n\} \cap A \neq B$. Else, since $\{x_1, \dots, x_n\} \cap A = \{x_j ; x_j^i \leq a\}$, we have that $x_{i(j)}^i \leq a$ for all $j = 1, \dots, p$, and $x_{i(j)}^i > a$ for all $j = p + 1, \dots, n$. So $\{x_1, \dots, x_n\} \cap A = \{x_{i(1)}; \dots; x_{i(p)}\}$ and $|\{x_{i(1)}; \dots; x_{i(p)}\}| = p$, leading to $\{x_1, \dots, x_n\} \cap A \in \mathcal{B}_p$, and then to $\{x_1, \dots, x_n\} \cap A \neq B$. So, for all $B \in \mathcal{B}_p^c$ and all $A \in \mathcal{A}_d$, $\{x_1, \dots, x_n\} \cap A \neq B$.

So, if $\binom{n}{\lfloor n/2 \rfloor} > d$, (x_1, \dots, x_n) can not be shattered by \mathcal{A}_d . Thus

$$VC(\mathcal{A}_d) \leq \max \left\{ n ; \binom{n}{\lfloor n/2 \rfloor} \leq d \right\}.$$

Let $n \geq 1$ such that $\binom{n}{\lfloor n/2 \rfloor} \leq d$. Let (x_1, \dots, x_n) be n points of \mathbb{R}^d defined as follows: for each coordinate $i = 1, \dots, \binom{n}{\lfloor n/2 \rfloor}$, let $\{i_1; \dots; i_{\lfloor n/2 \rfloor}\}$ be the i^{th} subset of $\lfloor n/2 \rfloor$ indices in $\{1; \dots; n\}$, where the indices are denoted in ascending order, i.e.:

$$1 \leq i_1 < \dots < i_{\lfloor n/2 \rfloor} \leq n.$$

Since $\binom{n}{\lfloor n/2 \rfloor} \leq d$, we obtain $\binom{n}{\lfloor n/2 \rfloor}$ distinct subsets of indices. Hence we take for each such coordinate

$$x_{i_k}^i = k.$$

Then the remaining values of (x_1, \dots, x_n) are taken as follows:

- Since $\binom{n}{\lfloor n/2 \rfloor + 1} \leq d$, for each subset $\{i_1; \dots; i_{\lfloor n/2 \rfloor + 1}\}$ of $\{1; \dots; n\}$ with $\lfloor n/2 \rfloor + 1$ elements, there exists $i' \in \{1; \dots; \binom{n}{\lfloor n/2 \rfloor}\}$ such that $\{i_1; \dots; i_{\lfloor n/2 \rfloor}\} = \{i'_1; \dots; i'_{\lfloor n/2 \rfloor}\}$. Then take $x_{i_{\lfloor n/2 \rfloor + 1}}^{i'} = \lfloor n/2 \rfloor + 1$. Let us note that, if n is odd, there is a bijection between i and i' .
- Let $\{j_1; \dots; j_m\} = \{j \notin \{i_1; \dots; i_{\lfloor n/2 \rfloor + 1}\}\}$, with $j_1 < \dots < j_m$, and let $j_0 = i_{\lfloor n/2 \rfloor + 1}$. Then take $x_{j_k}^{i'} = x_{j_{k-1}}^{i'} + 1$.

If not filled, the last coordinates are set to be equal to n .

Hence, we obtain that, for all $j \notin \{i_1; \dots; i_{\lfloor n/2 \rfloor}\}$, $x_j^i \geq \lfloor n/2 \rfloor + 1$.

Then (x_1, \dots, x_n) is shattered by \mathcal{A}_d : for $p \in \{0; \dots; n\}$, let $B = \{x_{i_1}; \dots; x_{i_p}\} \subset \{x_1, \dots, x_n\}$, with $1 \leq i_1 < i_2 < \dots < i_p \leq n$ as soon as $p \neq 0$.

If $p = 0$, let

$$i_0 = \operatorname{argmin}_{1 \leq i \leq d} \min_j x_j^i,$$

and take $A = \{x^{i_0} \leq \min_j x_j^{i_0} - 1\}$. Then $B = \{x_1, \dots, x_n\} \cap A = \emptyset$.

If $p = n$, let

$$i_n = \operatorname{argmax}_{1 \leq i \leq d} \max_j x_j^i,$$

and take $A = \{x^{i_n} \leq \max_j x_j^{i_n} + 1\}$. Then $B = \{x_1, \dots, x_n\} \cap A = \{x_1, \dots, x_n\}$.

If $0 < p \leq \lfloor n/2 \rfloor$, let $A \in \mathcal{A}_d$ be the subset defined by $A = \{x^i \leq p + 1/2\}$,

with i the coordinate corresponding to a subset of indices $\{i_1; \dots; i_{\lfloor n/2 \rfloor}\}$ containing $\{i_1; \dots; i_p\}$. Then, by definition of (x_1^i, \dots, x_n^i) , $B = \{x_1, \dots, x_n\} \cap A$.

If $\lfloor n/2 \rfloor + 1 \leq p < n$, let i' be the coordinate corresponding to the configuration $\{i_1; \dots; i_{\lfloor n/2 \rfloor + 1}\}$ (as defined by (x_1, \dots, x_n)). Let $A \in \mathcal{A}_d$ be the subset defined by $A = \{x^{i'} \leq p + 1/2\}$. Then, by definition of $(x_1^{i'}, \dots, x_n^{i'})$, $B = \{x_1, \dots, x_n\} \cap A$.

Thus

$$VC(\mathcal{A}_d) \geq \max \left\{ n ; \binom{n}{\lfloor n/2 \rfloor} \leq d \right\}.$$

Hence, since

$$VC(\mathcal{A}_d) = \max \left\{ n ; \binom{n}{\lfloor n/2 \rfloor} \leq d \right\},$$

we have $VC(\mathcal{A}_d) \leq d$.

Then, the lower bound of $VC(\mathcal{A}_d)$ is computed by using the Stirling's formula: for all $n \geq 1$ we have

$$\sqrt{2\pi}e^{-(n+1)}(n+1)^{n+\frac{1}{2}} \leq n! \leq \sqrt{2\pi}e^{-(n+1)}e^{\frac{1}{12(n+1)}}(n+1)^{n+\frac{1}{2}}.$$

If n is even, we obtain that

$$\binom{n}{n/2} \leq e^{\frac{1}{12(n+1)}} \frac{e}{\sqrt{2\pi}} 2^{n+1} \frac{(n+1)^{n+1/2}}{(n+2)^{n+1}} \leq \frac{e^{1+\frac{1}{24}}}{\sqrt{2\pi}} 2^{n+1}.$$

If n is odd, we obtain that

$$\binom{n}{\lfloor n/2 \rfloor} \leq e^{\frac{1}{12(n+1)}} \frac{e}{\sqrt{2\pi}} 2^{n+1} \frac{(n+1)^{(n+1)/2}}{(n+3)^{n/2+1}} \leq \frac{e^{1+\frac{1}{24}}}{\sqrt{2\pi}} 2^{n+1}.$$

So, if

$$\frac{e^{1+\frac{1}{24}}}{\sqrt{2\pi}} 2^{n+1} \leq d,$$

then $\binom{n}{\lfloor n/2 \rfloor} \leq d$. Taking the logarithm leads to the lower bound. □

References