CSE 648: Advanced algorithms

Topic: VC-Dimension & $\epsilon$-nets

April 03, 2003

LECTURER: Piyush Kumar

SCRIBED BY: Gen Ohkawa

Please Help us improve this draft.
If you read these notes, please send feedback.

# 1 Introduction

The lack of a theory of perception has led computer scientists to try to solve more modest problems. Pattern classification is one such subfield of computer science which deals with the assignment of a physical object or event to one or more categories [3]. Let us for now call this assignment (or function) $f$ that takes as input a physical object or event and assigns it to a specified category. To simplify our discussion, let the output be binary and the input be $\mathbb{R}^d$. We take the domain of the function $f$ to be $\mathbb{R}^d$ because pattern classification usually involves a data reduction step that extracts the properties or features of the input data and embeds this data in some space. Statistical learning theory [4] shows that it is important to pick $f$ from a class of functions that has small VC-dimension.

VC-Dimension is also a central concept for the study of Probably(The probability of the bound failing is small) Approximately Correct Model (when the bound holds the classifier $f$ has low error rate), one of the most general models used in computational learning theory. The idea of VC dimension is central in very popular machine learning methods like Support Vector Machines and Boosting [4]. It also helps us answer questions like "How many random samples does a learning procedure need to draw before it has sufficient information to learn an unknown target concept?"

Here is an example to show the power of random sampling and VC dimension. Lets say we want to estimate the number of people who live within 10 miles of a hospital in a given country. We can do this by sampling and scaling up the results. What's amazing is that the same sample size works for both Japan and India! The magic lies in the concept of VC-Dimensions [2].

# 2 VC-Dimension

In order to understand how VC-Dimensions work, we start with a few definitions. Here we will use $|S|$ to denote the cardinality of any set $S$.

**Definition 1** *A* **Hypergraph** *is a pair $H = (V, S)$ where $V$ is a finite set whose elements are called* vertices *and $S$ is a family of subsets of $V$, called* edges.

**Definition 2** *A* **Range Space** *is a pair $\mathcal{R} = (X, \mathcal{R})$ where $X$ is any finite or infinite set whose elements are called* points *and $\mathcal{R}$ is a family of subsets of $X$, called* ranges.

**Definition 3** *A set $A$ is said to be* **Shattered** *by a range space $\mathcal{R} = (X, \mathcal{R})$ if $A \subseteq X$, and all $2^{|A|}$ subsets of $A$ can be obtained by intersecting $A$ with members of $\mathcal{R}$.*

**Definition 4** *The* **Vapnik-Chervonenkis (VC) Dimension** *of a Range Space $(X, R)$ is the cardinality $d$ of the largest set $A \subseteq X$ that can be shattered.*

Now lets take a look at an example of how we might calculate the VC-Dimension.

**Example 5** *VC Dimension of Closed Half-Spaces in $\mathbb{R}^2$*

Let $H = (X, \mathcal{R})$ where $X = \mathbb{R}^2$ and $\mathcal{R} = \{h_{A,B} : A, B \in (-\infty, +\infty)\}$ where $h_{A,B} = \{x : Ax+By+1 \geq 0\}$ (Simply put, $X$ is a plane and $R$ is the set of all half-spaces of $X$.)

What is the largest number of points that we can have in $A$ so that we can shatter it?

- What if we have 3 pts in $A$?
  - As we can see from *Figure 1*, all 8 subsets of the three points are intersected by $X$. Therefore this set of three points is shattered.
- What if we have 4 pts?
  - No, as we can see from *Figure 2*, we can take two points which are at a diagonal and we would not be able to intersect just those two points without intersecting a third point.

In this example of closed half-spaces, the VC-Dimension = 3.
If we were to change $X = \mathbb{R}^2$ to $X = \mathbb{R}^d$, the VC- Dimension = $d + 1$.

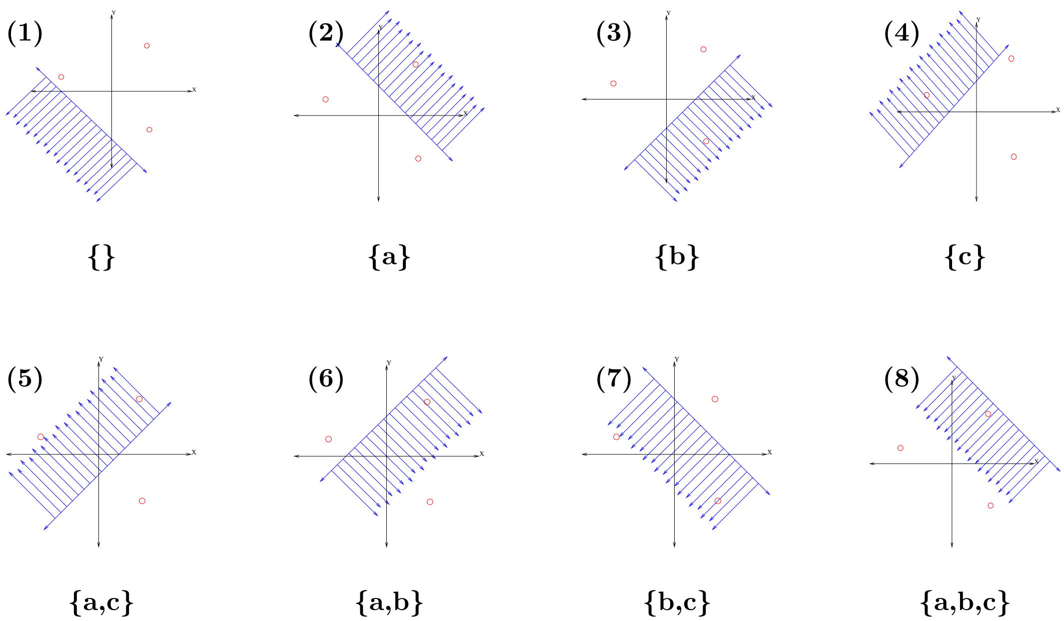


Figure 1: Figure 1: All Subsets are intersected when $|X| = 3$

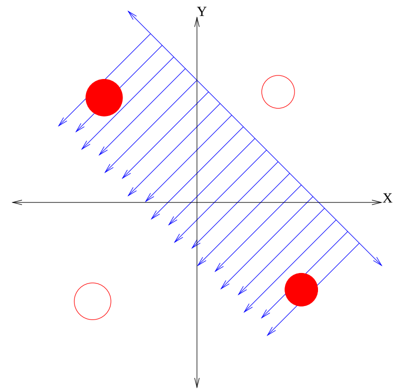


Figure 2: Unable to intersect only the diagonal solid points

**Example 6** *VC Dimension of All Convex Sets of a Plane*

A set $S$ in $n$-dimensional space is called a convex set if the line segment joining any pair of points of $S$ lies entirely in $S$.

Let $H = (X, \mathcal{R})$ where $X = \mathbb{R}^2$ and $\mathcal{R} = \{\text{All Convex Sets of the plane}\}$.
We will choose our $A$ so that our points lie on the unit circle centered at the origin.

*Claim:* The VC-Dimension $= \infty$ for this example.
*Argument (not so mathematical):* Take a subset of 3 points on the circle from $A$.

- If we want to intersect 1 point, we can do so by drawing a point. (Figure 3-1)
- If we want to intersect 2 points, we can draw a line. (Figure 3-2)
- If we want to intersect 3 points, we can draw a triangle. (Figure 3-3)
- What if we want to intersect $n$ points? We can just join them in a convex polygon with $n$ vertices(Figure 4)!

Therefore the VC-Dimension $= \infty$.

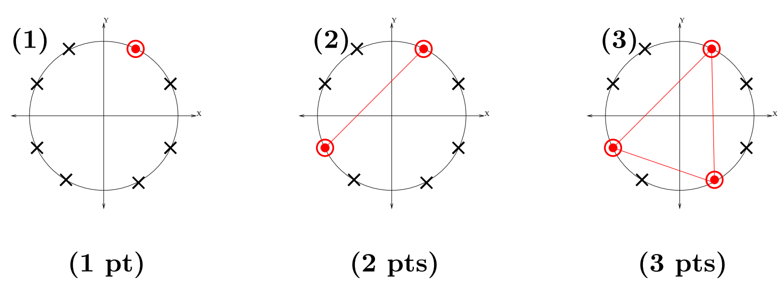


Figure 3: Convex Polygon intersecting 1, 2 and 3 points in unit circle centered at the origin.

Here are some sample exercises about VC-Dimensions. The solutions are placed at the end of the notes.

**Exercise 1** - What is the VC-Dimension of axis aligned rectangles in the plane?
**Exercise 2** - What about convex polygons in the plane with d sides?

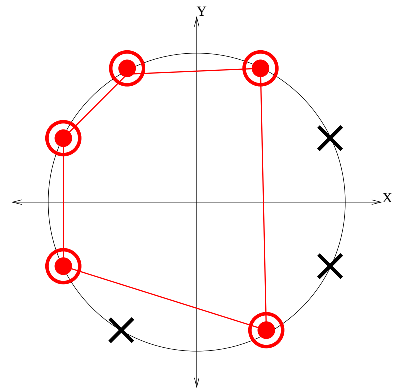


Figure 4: Convex Polygon intersecting $n$ points

### 2.1 $\varepsilon$-Net

**Definition 7** *Let $(X, R)$ be a range space with $X$ finite, and let $\varepsilon \in [0, 1]$. A set $N \subseteq X$ is called an $\varepsilon$-**Net** for the range space $(X, R)$ if $N \cap S \neq \emptyset$ for all $S \in R$ with $|S| \geq \varepsilon|X|$.*

*Comment* : If $\mu$ is a probability measure on $X$, then $N \subseteq X$ is an $\varepsilon$-net with respect to $\mu$ if $N \cap S \neq \emptyset \; \forall \; S \in R$ with $\mu(S) \geq \varepsilon$ $(\mu(X) = 1)$.

### 2.2 Simpler Sampling Lemma

**Lemma 8** *Let $W \subset X$ be chosen randomly such that $|W| \geq \varepsilon|X|$. A set of $O(\frac{1}{\varepsilon} \ln \frac{1}{\delta})$ points sampled independently and uniformly at random from $X$ intersects $W$ with probability at least $1 - \delta$.*

**Proof:** Any particular sample $x$ is in $W$ with probability at least $\varepsilon$, and not in $W$ with probability at most $1 - \varepsilon$. Therefore the probability that all our samples do not intersect with $W$ is at most $\delta$.

$$
\begin{aligned}
Prob(x \,\epsilon\, W) &\geq \varepsilon \\
Prob(x \notin W) &\leq (1 - \varepsilon) \\
Prob\left(x \notin W \text{ for all } \frac{1}{\varepsilon} \ln \frac{1}{\delta} \text{ samples}\right) &\leq (1 - \varepsilon)^{\frac{1}{\varepsilon} \ln \frac{1}{\delta}} \\
&\leq e^{-\ln \frac{1}{\delta}} \\
&= e^{\ln \delta} \\
&= \delta
\end{aligned}
$$

■

The $|W|$ depends on $|X|$, but sampling depends only on $\varepsilon$ !

Problem: This Lemma only applies for a given random sample $W$ of $X$, not for all large random samples...

## 3 $\varepsilon$-Net Theorem

The $\varepsilon$-Net Theorem states the following:

**Theorem 9** *Let VC-Dimension of $(X, R)$ be $d \geq 2$ and $0 \leq \varepsilon \leq \frac{1}{2}$. $\exists$ $\varepsilon$-net for $(X, R)$ of size at most $O(\frac{d}{\varepsilon} \ln \frac{1}{\varepsilon})$.*

A better version of the $\varepsilon$-Net theorem follows:

**Theorem 10** *If we choose $O(\frac{d}{\varepsilon}\log\frac{d}{\varepsilon} + \frac{1}{\varepsilon}\log\frac{1}{\delta})$ points at random from $X$, then the resulting set $N$ is an $\epsilon$-net with probability $\geq \delta$.*

Interested readers should look up the proof in Alon and Spencer [1].

### 3.1 Constructing N in this Case:

Pick $\frac{d}{\varepsilon}\ln\frac{1}{\varepsilon}$ random samples and put them in $N$. The theorem says that $N$ is an $\varepsilon$-net with probability at least $\frac{1}{2}$!!

*Comment :* See references for proofs.

## 4 Applications

As an example of a possible application, let us design a data structure using this algorithm!

Let the set of points $P = \{p_1, p_2, p_3, ...p_n\}$ where $p_i \in \mathbb{R}^d$.
We want to query our structure with any $c \in \mathbb{R}^d$ if $||x - c|| \geq r \;\forall\; p_i \in P$.
If not, the algorithm should produce $p_i \in P$ such that $||p_i - c|| < r$.
Otherwise with high probability $||x - c|| \geq r$ for at least $1 - \varepsilon$ fraction of the points.

We will try to convert the problem into a VC-Dimension problem by converting the spheres of $\mathbb{R}^d$ into half spaces in $\mathbb{R}^{d+1}$. Consider the set system

$$(P, \text{Spheres in } \mathbb{R}^d) = (\text{Lifted}(P), \text{Half Spaces in } \mathbb{R}^{d+1})$$

We know the VC-Dimension of this to be $d + 2$ from our first example!

So by our $\varepsilon$-net theorem we will pick $|N| = O(\frac{d}{\varepsilon}\log\frac{d}{\varepsilon} + \frac{1}{\varepsilon}\log\frac{1}{\delta})$ points. If it fails for some x, we have a witness. Otherwise $F = \{x \in P|\; ||x - c|| < r\}$ i.e. the set where our condition fails to hold. Since $N$ is with high probability a $\varepsilon$-net, we can conclude with high probability that

## 5 Solutions to Practice Excercises

**Exercise 1** - What is the VC-Dimension of axis aligned rectangles in the plane?
**Solution -** 4

**Exercise 2** - What about convex polygons in the plane with $d$ sides?
**Solution -** $2d + 1$

## References


[1] N. Alon and J. Spencer, *The probabilistic method*, ACM Press, 1992.

[2] B. Chazelle, *The discrepancy method: randomness and complexity*, Cambridge University Press, 2000.

[3] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley-Interscience, New York, 1973.

[4] V. N. Vapnik, *The nature of statistical learning theory*, Springer-Verlag New York, Inc., 1995.