

Vapnik-Chervonenkis Dimension

Yingying Ji

2019/03/17

Introduction

Vapnik-Chervonenkis (VC) dimension is a basic concept in the field of machine learning, which provides a solid theoretical basis for the learnability of many machine learning methods.

We usually evaluate that a good model needs to have small errors. And now we have learned that errors can be taken into two parts: training error (in-sample error) and test error (out-of-sample error). We expect training error to be small, indicating the model fits training data well. In the meanwhile, it can't be too small to generalize the underlying population. So how can we improve E_{in} and E_{out} at the same time? We use the VC dimension to connect them. In statistical learning theory, it can predict a probabilistic upper bound on the test error of a classification model. In this report, I will make a basic introduction for VC dimension.

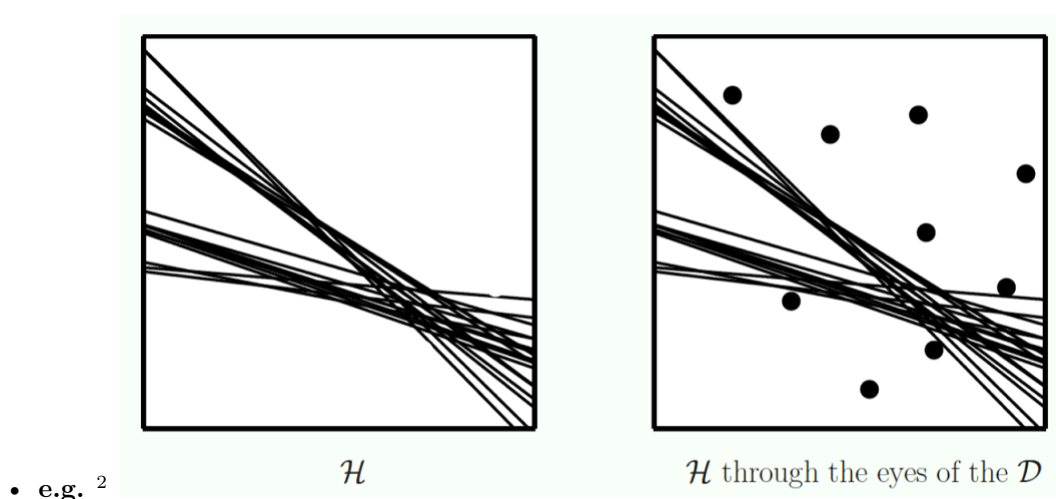
Some Basic Concepts

In order to understand the definition of VC dimension clearly, we need to know some concepts as follows first.

- Dichotomy
- Shatter
- Break Point
- Growth Function

Dichotomy

- Consider binary target functions and hypothesis sets that contain $h : \chi \rightarrow \{-1, +1\}$.¹
- If $h \in H$ is applied to a finite sample $x_1, x_2, \dots, x_N \in \chi$, we get an N -tuple $h(x_1), \dots, h(x_N)$ of ± 1 's.
- Such an N -tuple is called a **dichotomy** since it splits x_1, \dots, x_N into two groups.



- e.g.²

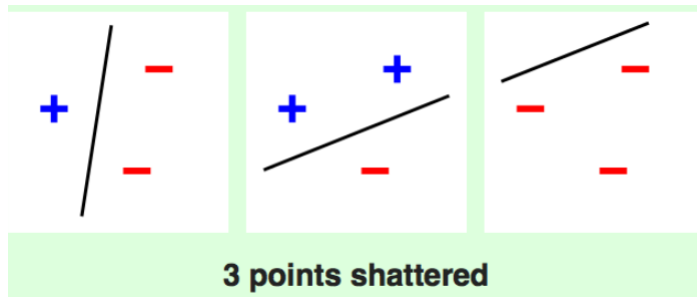
*From the point of view of \mathcal{D} , the entire H is just **one dichotomy**.

¹The following analyses are based on binary target functions.

²The picture of this e.g. is from lecture slides for this course.

Shatter

- If H is capable of generating all possible dichotomies on x_1, x_2, \dots, x_N , and then we say that H **shatters** x_1, x_2, \dots, x_N .

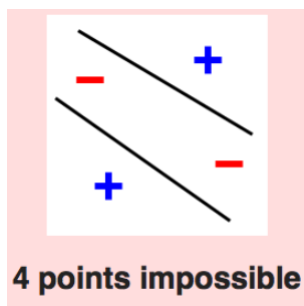


- e.g.³

Let $H = \{\text{linear classifiers over } \mathbb{R}^2\}$. Consider a set S of 3 points in general position. Thus there exists a set of 3 points that can be shattered by H .

Break Point

- The **least** number of inputs that can't be shattered by H for any distribution types, denoted *minimum* k .



- e.g.

*The break point for the last example is equals to 4 since no set of 4 distinct points on the plane can be shattered by H .

Growth Function

- The **growth function** for a hypothesis set H is defined by

$$m_H(N) = \max_{x_1, x_2, \dots, x_N \in \chi} |H(x_1, x_2, \dots, x_N)|$$

- $m_H(N)$ is the **maximum** possible number of dichotomies that H can generate on a data set of N points⁴.
- It measures the complexity of a hypothesis class.
- e.g.** For positive rays, the hypothesis space, applied to one sample point, can generate two dichotomies: $(-1), (+1)$. Applied to 2 sample points, it can generate 3 dichotomies: $(-1, +1), (-1, -1), (+1, +1)$. Applied to 3 sample points, it can generate 4 dichotomies. By analogy, the growth function can be derived as $m_H(N) = N + 1$.

³The pictures of this e.g. are from *wikipedia of VC dimension*.

⁴rather than over the entire input space χ

Definition

The **Vapnic-Chervonenkis (VC) dimension** of H , denoted $d_{vc}(H)$, is the size of the largest data set that H can shatter, as the largest number that is not break point.

- $d_{vc}(H)$ is the largest value of N for which $m_H(N) = 2^N$.
- the **most** inputs that H can shatter.
- If the break point exists, $d_{vc} = \text{minimum } k' - 1$.
- If there is no break point, that is, arbitrarily large finite sets can be shattered by H , then $d_{vc} = \infty$.
- **e.g.** With the above definitions in mind, let's list some examples to put the three concepts: break point, growth function, VC dimension together to understand.

Category	Break Point	Growth Function	VC Dimension
positive rays	2	$m_H(N) = N + 1$	$d_{vc} = 1$
positive intervals	3	$m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$	$d_{vc} = 2$
convex sets	0	$m_H(N) = 2^N$	$d_{vc} = \infty$
1D perception	3	$m_H(N) = 2N$	$d_{vc} = 2$
2D perceptions	4	$m_H(N) < 2^N, \text{ when } N \geq 3$	$d_{vc} = 3$

Summary

After a simple introduction about why we need VC dimension, this report introduce some basic concepts related to the understanding of VC dimension and go over some examples as well. Then explain the definition of VC dimension.

A finite VC dimension can always guarantee that the hypothesis h we find can generalize the underlying data, i.e., make h satisfies $E_{in}(h) \approx E_{out}(h)$, which provides a solid condition for the feasibility of machine learning.