Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Topic 4**

VC Dimension and Polynomial Discrimination

Lecturer: Yun Wei *Scribe: Chin-Ling Chang*

Lecture date: October 12, 2026.

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1 Review

Recall the route from empirical risk minimization to suprema of empirical processes, $\|P_n - P\|_{\mathcal{F}}$, and then to Rademacher complexity, $\mathfrak{R}_n(\mathcal{F})$. Polynomial discrimination controls the number of labelings on a sample through a bound of the form

$$\operatorname{card}(\mathcal{F}(x_1^n)) \le (n+1)^d \qquad \text{for every } n \ge n_0,$$

where d is fixed. We now introduce VC classes as a way to obtain this polynomial bound.

2 Shattering and VC Classes

Let $\mathcal{F}$ be a class of binary-valued functions on $\mathcal{X}$, and write $x_1^n = (x_1, \dots, x_n)$.

Definition 2.1 (Shattering and VC dimension)**.** The sample x_1^n is *shattered* by $\mathcal{F}$ if $\operatorname{card}(\mathcal{F}(x_1^n)) = 2^n$. The *VC dimension* $\nu(\mathcal{F})$ is the largest integer n for which some n points can be shattered by $\mathcal{F}$; it is infinite if arbitrarily large finite sets can be shattered.

Definition 2.2 (VC class)**.** A binary-valued function class $\mathcal{F}$ is a *VC class* if $\nu(\mathcal{F}) < \infty$. In particular, for a set class $\mathcal{S}$, we consider

$$\mathcal{F} = \{\mathbf{1}_S : S \in \mathcal{S}\},$$

where $\mathbf{1}_S(x) = 1$ means that $x \in S$, and it is zero otherwise.

The corresponding evaluation vectors are

$$\mathcal{F}(x_1^n) = \{(\mathbf{1}_S(x_1), \dots, \mathbf{1}_S(x_n)) : S \in \mathcal{S}\}. \tag{1}$$

Each distinct intersection of a set S with the sample determines one such labeling. Thus

$$\operatorname{card}(\mathcal{F}(x_1^n)) = \operatorname{card}\big(\{S \cap \{x_1, \dots, x_n\} : S \in \mathcal{S}\}\big). \tag{2}$$

Identifying each intersection with its indicator vector, we sometimes use the notation

$$\mathcal{S}(x_1^n) = \mathcal{F}(x_1^n), \qquad \nu(\mathcal{S}) = \nu(\mathcal{F}). \tag{3}$$

The definition of VC dimension gives the two cases recorded in the notes:

(i) If $n \le \nu(\mathcal{F})$, then

$$\exists\, x_1^n \quad \text{such that} \quad \operatorname{card}(\mathcal{F}(x_1^n)) = 2^n.$$

(ii) If $n > \nu(\mathcal{F})$, then

$$\forall\, x_1^n, \qquad \operatorname{card}(\mathcal{F}(x_1^n)) < 2^n.$$

The bound in (ii) alone would allow $\operatorname{card}(\mathcal{F}(x_1^n)) = 2^n - 1$. The proposition below rules out such exponential growth for a class with fixed finite VC dimension.

Proposition 2.3 (Vapnik–Chervonenkis, Sauer and Shelah)**.** *Let $\mathcal{S}$ be a set class with $\nu(\mathcal{S}) < \infty$. For $n > \nu(\mathcal{S})$ and every collection x_1^n,*

$$\operatorname{card}(\mathcal{S}(x_1^n)) \le \sum_{j=0}^{\nu(\mathcal{S})} \binom{n}{j} \le (n+1)^{\nu(\mathcal{S})}. \tag{4}$$

The proof is referred to the textbook [1, Section 4.3.2]. We use the proposition directly in the examples below.

3 Examples on the Real Line

Example 3.1 (Left half-lines)**.** Consider

$$\mathcal{S}_{\text{left}} = \{(-\infty, a] : a \in \mathbb{R}\}, \qquad \mathcal{F}_{\text{left}} = \{\mathbf{1}_{(-\infty,a]} : a \in \mathbb{R}\}.$$

For $n = 1$, either label 0 or 1 can be obtained, so a one-point set is shattered. For $n = 2$ and $x_1 < x_2$,

$$\mathcal{F}_{\text{left}}(x_1, x_2) = \{(0,0), (1,0), (1,1)\}.$$

There are three labelings, fewer than 2^2; the labeling $(0,1)$ is impossible. Therefore

$$\nu(\mathcal{S}_{\text{left}}) = \nu(\mathcal{F}_{\text{left}}) = 1.$$

Applying Proposition 2.3 directly, for $n > 1$, the cardinality is at most $n + 1$.

In fact, for distinct ordered points $x_1 < \cdots < x_n$, writing out the labelings gives

$$\begin{aligned} \operatorname{card}(\mathcal{F}_{\text{left}}(x_1^n)) &= \operatorname{card}\big(\{(0,\ldots,0), (1,0,\ldots,0), \ldots, (1,\ldots,1)\}\big) \\ &= n+1. \end{aligned} \tag{5}$$

This shows the upper bound is sharp.

Example 3.2 (Bounded intervals)**.** Consider

$$\mathcal{S}_{\text{int}} = \{(b, a] : a, b \in \mathbb{R},\ b < a\}, \qquad \mathcal{F}_{\text{int}} = \{\mathbf{1}_{(b,a]} : a, b \in \mathbb{R},\ b < a\}.$$

A one-point set can be shattered. For two distinct points $x_1 < x_2$,

$$\mathcal{F}_{\text{int}}(x_1, x_2) = \{(1,0), (0,1), (1,1), (0,0)\},$$

so all 2^2 labelings are possible. For three points $x_1 < x_2 < x_3$, the labeling $(1,0,1)$ is impossible: an interval containing both x_1 and x_3 must also contain x_2. Thus

$$\nu(\mathcal{S}_{\text{int}}) = \nu(\mathcal{F}_{\text{int}}) = 2.$$

Applying Proposition 2.3 directly, for $n > 2$,

$$\operatorname{card}(\mathcal{F}_{\text{int}}(x_1^n)) \le (n+1)^2. \tag{6}$$

4 Controlling VC Dimension by Vector Space Structure

Let $\mathcal{G}$ be a class of real-valued functions on $\mathcal{X}$. For each $g \in \mathcal{G}$, define

$$S_g = \{x \in \mathcal{X} : g(x) \leq 0\}, \qquad \mathcal{S}(\mathcal{G}) = \{S_g : g \in \mathcal{G}\}. \tag{7}$$

Following the lecture's terminology, $\mathcal{S}(\mathcal{G})$ is the *subgraph class at level zero*. Here each S_g is a zero-sublevel set in $\mathcal{X}$; the threshold is fixed at zero.

Proposition 4.1 (Finite-dimensional vector spaces)**.** *Let $\mathcal{G}$ be a real vector space of functions $g : \mathbb{R}^d \to \mathbb{R}$ with finite dimension m. Then*

$$\nu(\mathcal{S}(\mathcal{G})) \leq m. \tag{8}$$

Proof. Set $n = m + 1$, and fix any n distinct points $x_1, \ldots, x_n$. It suffices to show that these points cannot be shattered by $\mathcal{S}(\mathcal{G})$. Define the linear evaluation map

$$L : \mathcal{G} \longrightarrow \mathbb{R}^n, \qquad L(g) = (g(x_1), \ldots, g(x_n)).$$

Its image is a linear subspace satisfying

$$\dim L(\mathcal{G}) \leq \dim \mathcal{G} = m < n. \tag{9}$$

To see the dimension inequality explicitly, let $g_1, \ldots, g_m$ be a basis of $\mathcal{G}$. If $g = \sum_{j=1}^m a_j g_j$, then $L(g) = \sum_{j=1}^m a_j L(g_j)$. Thus the m vectors $L(g_j)$ span the image. They need not be linearly independent.

Since the image is a proper subspace of $\mathbb{R}^n$, there is a nonzero vector $\gamma = (\gamma_1, \ldots, \gamma_n)$ orthogonal to it:

$$\langle \gamma, L(g) \rangle = \sum_{i=1}^n \gamma_i g(x_i) = 0 \qquad \text{for every } g \in \mathcal{G}. \tag{10}$$

Replacing γ by $-\gamma$ if necessary, we may assume that at least one γ_i is positive.

Rearranging (10) yields

$$\sum_{\{i:\gamma_i \le 0\}} (-\gamma_i) g(x_i) = \sum_{\{i:\gamma_i > 0\}} \gamma_i g(x_i) \qquad \text{for every } g \in \mathcal{G}. \tag{11}$$

Suppose, for a contradiction, that the points are shattered. There must then exist $g \in \mathcal{G}$ whose set S_g contains exactly those sample points with $\gamma_i \le 0$. In other words,

$$g(x_i) \le 0 \quad \text{if } \gamma_i \le 0, \qquad g(x_i) > 0 \quad \text{if } \gamma_i > 0.$$

Every term on the left-hand side of (11) is nonpositive. Every term on the right-hand side is strictly positive, and there is at least one such term. Hence (11) would imply

$$0 \;\ge\; \sum_{\{i:\gamma_i \le 0\}} (-\gamma_i) g(x_i) = \sum_{\{i:\gamma_i > 0\}} \gamma_i g(x_i) \;>\; 0,$$

a contradiction. No $m+1$ points can be shattered, proving (8). □

References

[1] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019. Sections 4.3.1–4.3.3.