Instructor: Adarsh Barik
Scribed by: Dibyajyoti Jena, Vansh Ramani, Serene Rasheed
Lecture 6, COV878
September 2, 2025

LECTURE 6: VAPNIK CHERVONENKIS (VC) DIMENSION

1 Overview of the Lecture

This lecture introduces the Vapnik-Chervonenkis (VC) dimension, a combinatorial measure of the complexity of a hypothesis class of binary classifiers. We will define it, explore its properties, and connect it to the growth function via the Sauer-Shelah Lemma. We then bridge the gap to generalization bounds by going back to Rademacher complexity from prior lectures, a data-dependent measure of complexity, and show how it can be bounded using the VC dimension.

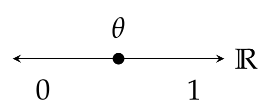
2 VC Dimension

In the previous lecture, we saw that if the Rademacher complexity of a hypothesis class $\mathcal{H}$ is small, Then the true risk of every function in $\mathcal{H}$ can be bounded in terms of its empirical risk and the Rademacher complexity of $\mathcal{H}$ on samples of size n.

In this lecture, we want to characterize hypothesis classes for which (over a given sample space) Rademacher complexity can be bounded. Throughout, we will restrict attention to the binary classification setting and allow the size of $\mathcal{H}$ to be possibly infinite.

A motivating example Consider a dataset $\mathbb{Z} = \{z_1, z_2, ..., z_n\}$. Number of maximum possible labelings in this case is 2^n: every data point z can be classified into two classes and there are n number of points. This holds good when the hypothesis class $\mathcal{H}$ is rich. But what if it is a restrictive class: consider the class of thresholding functions on the real line.

$$\begin{aligned}\mathcal{H} = &\{h : \mathbb{R} \to \{0,1\} \mid h(z) = \mathbb{1}\left[z > \theta\right], \theta \in \mathbb{R}\} \\ &\cup \{h : \mathbb{R} \to \{0,1\} \mid h(z) = \mathbb{1}\left[z \leq \theta\right], \theta \in \mathbb{R}\}\end{aligned}$$



If we sort the points in increasing or decreasing magnitude, the points can be labeled in a maximum number of $2n$ ways. Possible labels:

$$\begin{pmatrix} z_1 & z_{n-2} & z_{n-1}... & z_n \\ 0 & 0 & 0... & 0 \\ 1 & 0 & 0... & 0 \\ 1 & 1 & 0... & 0 \\ ... & ... & ... & ... \\ 1 & 1 & 1... & 0 \end{pmatrix} \text{ or } \begin{pmatrix} z_1... & z_{n-2} & z_{n-1} & z_n \\ 0... & 0 & 0 & 1 \\ 0... & 0 & 1 & 1 \\ 0... & 1 & 1 & 1 \\ ... & ... & ... & ... \\ 1... & 1 & 1 & 1 \end{pmatrix}$$

Any other combination of 1's and 0's for labels can not be achieved by $\mathcal{H}$.

VC dimension provides an easy complexity measure for such hypothesis classes.

Definition 1 (Growth Function). *Given a hypothesis class $\mathcal{H} \subseteq \{h \mid h : \mathbb{Z} \to \{0,1\}\}$ and a dataset $\mathcal{S} = \{z_1, z_2, ..., z_n\}, z \in \mathbb{Z}$, we set $\mathcal{H}(\mathcal{S})$ as $\left\{h(z_1, ...h(z_n) \in \{0,1\}^n \mid h \in \mathcal{H}\right\}$. Then growth function is defined as:*

$$G(\mathcal{H}, n) = \max_{\mathcal{S} \in \mathbb{Z}} |\mathcal{H}(\mathcal{S})|$$

It is upper bounded at 2^n.

Definition 2 (Shattering). *$\mathcal{H}$ shatters a finite set $\mathcal{S} \subset \mathbb{Z}$ if $|\mathcal{H}(\mathcal{S})| = 2^{|\mathcal{S}|}$.*

Definition 3 (VC dimension). *The VC dimension of a hypothesis class $\mathcal{H}$ denoted as VC dim($\mathcal{H}$) is the max size of set $\mathcal{S} \subset \mathbb{Z}$ that can be shattered by $\mathcal{H}$.*

$$VC\ dim(\mathcal{H}) = \max_{n \in \mathbb{N}} \{n \mid G(\mathcal{H}, n) = 2^n\}$$

2.1 Example: VC dimension of threshold functions

Recall the threshold class on $\mathbb{R}$:

$$\mathcal{H} = \left\{h_\theta : \mathbb{R} \to \{0,1\} \;\middle|\; h_\theta(z) = \mathbb{1}_{[z>\theta]} \text{ or } h_\theta(z) = \mathbb{1}_{[z\le\theta]},\ \theta \in \mathbb{R}\right\}.$$

We assume sample points are distinct and write a sorted sample as $z_1 < z_2 < \cdots < z_n$.

Case $n = 1$. For $S = \{z_1\}$ we can realize both labelings:

$$(0): \text{ take } h_\theta(z) = \mathbb{1}_{[z>\theta]} \text{ with } \theta > z_1, \qquad (1): \text{ take } h_\theta(z) = \mathbb{1}_{[z>\theta]} \text{ with } \theta < z_1.$$

Hence $|\mathcal{H}_S| = 2 = 2^1$, so a single point is shattered.

Case $n = 2$. Let $S = \{z_1, z_2\}$ with $z_1 < z_2$. We exhibit hypotheses realizing all four labelings:

- $(0,0)$: choose $h_\theta(z) = \mathbb{1}_{[z>\theta]}$ with $\theta > z_2$.
- $(1,1)$: choose $h_\theta(z) = \mathbb{1}_{[z>\theta]}$ with $\theta < z_1$.
- $(0,1)$: choose $h_\theta(z) = \mathbb{1}_{[z>\theta]}$ with $z_1 < \theta < z_2$.
- $(1,0)$: choose $h_\theta(z) = \mathbb{1}_{[z\le\theta]}$ with $z_1 < \theta < z_2$.

Thus $|\mathcal{H}_S| = 4 = 2^2$, so any two distinct points can be shattered.

No set of size 3 can be shattered. Let $S = \{z_1, z_2, z_3\}$ with $z_1 < z_2 < z_3$. Any $h \in \mathcal{H}$ is either of the form $h(z) = \mathbb{1}_{[z>\theta]}$ (which yields a labeling vector with the pattern $0, \ldots, 0, 1, \ldots, 1$) or of the form $h(z) = \mathbb{1}_{[z\leq\theta]}$ (which yields $1, \ldots, 1, 0, \ldots, 0$). In either case, the labeling on sorted points has at most one sign change. Therefore the patterns $(1, 0, 1)$ and $(0, 1, 0)$ cannot be realized by any $h \in \mathcal{H}$. Hence, no set of size 3 is shattered.

Since some set of size 2 is shattered but no set of size 3 is shattered, we have

$$VCdim(\mathcal{H}) = 2.$$

2.1.1 Example : VC dimension of linear functions in $\mathbb{R}^2$

Let

$$\mathcal{H} = \left\{ h_{w,b} : \mathbb{R}^2 \to \{-1, 1\} \;\middle|\; h_{w,b}(x) = \text{sign}(\langle w, x\rangle + b),\ w \in \mathbb{R}^2,\ b \in \mathbb{R} \right\},$$

with the convention $\text{sign}(t) = 1$ for $t > 0$ and $\text{sign}(t) = -1$ for $t \leq 0$. We show $VCdim(\mathcal{H}) = 3$.

(i) A set of size 3 is shattered. Choose three noncollinear points $S = \{x_1, x_2, x_3\} \subset \mathbb{R}^2$ (for example, the vertices of a triangle). For any desired labeling of these three points, there are only the following types to realize:

- all labels equal: trivial (take any line putting all points on one side);
- exactly one point labeled $+1$ and two labeled -1: separate that single point by a line closely surrounding it on the $+1$ side;
- exactly two points labeled $+1$ and one labeled -1: separate the single -1 point from the other two by a line.

Because the three points are not collinear, for each of the 2^3 labelings, one can place a strict linear separator so that the $+1$-labeled points lie on the positive side and the -1-labeled points lie on the negative side. Hence $|\mathcal{H}_S| = 8 = 2^3$: the set S is shattered. Thus $VCdim(\mathcal{H}) \geq 3$.

(ii) No set of size 4 is shattered. Let $T = \{y_1, y_2, y_3, y_4\}$ be any set of four distinct points in $\mathbb{R}^2$. There are two mutually exclusive geometric configurations:

Case A: One point lies in the convex hull of the other three.
Let $y_4 \in \text{conv}(\{y_1, y_2, y_3\})$. Consider the labeling that assigns $+1$ to the three outer points y_1, y_2, y_3 and -1 to the interior point y_4. If some linear threshold $h_{w,b}$ realized this labeling, then the positive-labeled points would lie in the open halfspace $\{x : \langle w, x\rangle + b > 0\}$ and the negative-labeled points would lie in the open halfspace $\{x : \langle w, x\rangle + b < 0\}$. But the open halfspace containing y_1, y_2, y_3 is convex, hence it must contain the convex hull of $\{y_1, y_2, y_3\}$, and therefore would contain y_4 as well — contradiction. Thus, this labeling is unrealizable, so T is not shattered.

Case B: All four points are in convex position (vertices of a convex quadrilateral).
Label the points in clockwise order around the quadrilateral as y_1, y_2, y_3, y_4. Consider the labeling that

assigns $+1$ to y_1 and y_3 (two opposite vertices) and -1 to y_2 and y_4.No linear function can output this labeling. The labeling $(+1, -1, +1, -1)$ is unrealizable by any linear function. Hence T is not shattered.

Since every 4-point set falls into Case A or Case B, and in each case we show a labeling that no linear function can realize, no set of size 4 is shattered by $\mathcal{H}$. Thus $VCdim(\mathcal{H}) \leq 3$.

Combining (i) and (ii), we have $VCdim(\mathcal{H}) = 3$.

2.1.2 V C dimension of finite hypothesis classes

Let $\mathcal{H}$ be a finite hypothesis class. Then, for any set S we have $|\mathcal{H}_S| \leq |\mathcal{H}|$ and thus S cannot be shattered if $|\mathcal{H}| < 2^{|s|}$. This implies that $VCdim(\mathcal{H}) \leq \log_2(|\mathcal{H}|)$. Note that there are cases when $VCdim(\mathcal{H}) << \log_2(|\mathcal{H}|)$ (ex: hypothesis class of threshold functions).

3 Growth function upper bound

If the VC dimension of a hypothesis class $\mathcal{H}$ is d, it can shatter at most d data points. This means if there are n data points($n < d$) in a set $\mathcal{S}$, the growth function $G(\mathcal{H}, n)$ increase is bounded at 2^n. It increases exponentially with n, till the limit d is reached.

When n crosses the hypothesis class VC dimension, $\mathcal{H}$ can no longer shatter $\mathcal{S}$. But what is the maximum number of unique labelings $\mathcal{H}$ can achieve on points in $\mathcal{S}$? This is answered in the following lemma.

Lemma 1 (Sauer-Shelah Lemma)**.** *The growth function and VC-dimension of a hypothesis class* $\mathcal{H} \subset \{h \mid h : \mathbb{Z} \to \{0, 1\}\}$ *fulfills*

$$G(H, n) \leq \sum_{i=0}^{d} \binom{n}{i}$$

where VC-dim($\mathcal{H}$*)* $= d < \infty$

Proof. Fix arbitrary $\mathcal{S} = \{z_1, z_2, ..., z_n\}$.

Note: $|\{\mathcal{B} \subseteq \mathcal{S} \mid \mathcal{H} \text{ shatters } \mathcal{B}\}| \leq \binom{n}{0} + \binom{n}{1} ... = \sum_{i=0}^{d} \binom{n}{i}$

We will show:$|\mathcal{H}(\mathcal{S})| \leq |\{\mathcal{B} \subseteq \mathcal{S} \mid \mathcal{H} \text{ shatters } \mathcal{B}\}|$

Proof by induction:

1. Base case: We consider n = 1. Only one data point will yield $|\mathcal{H}(\mathcal{S})|$ as 1 or 2, which satisfies the lemma statement.

2. Induction hypothesis: We assume that for any $\mathbb{T} \subset \mathbb{Z}$, $|\mathbb{T}| < n$,

$$|\mathcal{H}(\mathbb{T})| \leq |\{\mathcal{B} \subseteq \mathbb{T} \mid \mathcal{H} \text{ shatters } \mathcal{B}\}|$$

3. General case: $\mathcal{S}' = \{z_2, z_3, ..., z_n\}$

Define:

$$\begin{aligned}\mathcal{H}(\mathcal{S}') = \mathcal{Y}_0 &= \{(0, y_2, ..y_n) : (y_2, ..y_n) \in \mathcal{H}(\mathcal{S}) \text{ or } (1, y_2, ..y_n) \in \mathcal{H}(\mathcal{S})\} \\ \mathcal{Y}_1 &= \{(0, y_2, ..y_n) : (y_2, ..y_n) \in \mathcal{H}(\mathcal{S}) \text{ and } (1, y_2, ..y_n) \in \mathcal{H}(\mathcal{S})\}\end{aligned}$$

Observe: $\mathcal{H}(\mathcal{S}) = |\mathcal{Y}_0| + |\mathcal{Y}_1|$, and

$$\begin{aligned}|\mathcal{Y}_0| = \left|\mathcal{H}(\mathcal{S}')\right| &\leq \left|\left\{\mathcal{B} \subseteq \mathcal{S}' \mid \mathcal{H} \text{ shatters } \mathcal{B}\right\}\right|, \text{(inductive hypothesis)} \\ &= |\{\mathcal{B} \subseteq \mathcal{S} : z_1 \notin \mathcal{B}, \mathcal{H} \text{ shatters } \mathcal{B}\}|\end{aligned}$$

Define:

$$\mathcal{H}' = \left\{h \in \mathcal{H} : \exists h' \in \mathcal{H}, h(z_1) \neq h'(z_1) \text{ and } h(z_i) = h'(z_i) \forall i = 2, 3.., n\right\}$$

Observe: If $\mathcal{H}'$ shatters $\mathcal{B} \subseteq \mathcal{S}'$, then $\mathcal{H}'$ also shatters $\mathcal{B} \cup \{z_1\} \subseteq \mathcal{S}'$. Therefore $\mathcal{Y}_1 = \mathcal{H}'(\mathcal{S}')$

$$\begin{aligned}|\mathcal{Y}_1| = \left|\mathcal{H}'(\mathcal{S}')\right| &\leq \left|\left\{\mathcal{B} \subseteq \mathcal{S}' \mid \mathcal{H}' \text{ shatters } \mathcal{B}\right\}\right|, \text{ (ind. hypothesis)} \\ &= \left|\left\{\mathcal{B} \subseteq \mathcal{S}' \mid \mathcal{H}' \text{ shatters } \mathcal{B} \cup \{z_1\}\right\}\right| \\ &= \left|\left\{\mathcal{B} \subseteq \mathcal{S} \mid z_1 \in \mathcal{B} \text{ and } \mathcal{H}' \text{ shatters } \mathcal{B}\right\}\right| \\ &\leq |\{\mathcal{B} \subseteq \mathcal{S} \mid z_1 \in \mathcal{B} \text{ and } \mathcal{H} \text{ shatters } \mathcal{B}\}|\end{aligned}$$

Recall:

$$\begin{aligned}\mathcal{H}(\mathcal{S}) &= |\mathcal{Y}_0| + |\mathcal{Y}_1| \\ &\leq |\{\mathcal{B} \subseteq \mathcal{S} : z_1 \notin \mathcal{B} \text{ and } \mathcal{H} \text{ shatters } \mathcal{B}\}| + |\{\mathcal{B} \subseteq \mathcal{S} : z_1 \in \mathcal{B} \text{ and } \mathcal{H} \text{ shatters } \mathcal{B}\}| \\ &= |\{\mathcal{B} \subseteq \mathcal{S} : \mathcal{H} \text{ shatters } \mathcal{B}\}|\end{aligned}$$

(Result)

□

Lemma 2 (Massart's Lemma)**.** *Let $A \subset \mathbb{R}^n$ be a finite set. Let $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ be n independent Rademacher random variables. Then*

$$\mathbb{E}_\sigma\left[\max_{a \in A} \sum_{i=1}^n \sigma_i a_i\right] \leq \sqrt{2 \log |A|} \max_{a \in A} \|a\|_2.$$

Proof. (Exercise from lecture) The proof uses the exponential moment method. Let $Z = \max_{a \in A} \sum_i \sigma_i a_i$. For any $s > 0$, by Jensen's inequality:

$$Z = \frac{1}{s} \log(e^{sZ}) = \frac{1}{s} \log\left(\max_{a \in A} e^{s \sum_i \sigma_i a_i}\right) \leq \frac{1}{s} \log\left(\sum_{a \in A} e^{s \sum_i \sigma_i a_i}\right)$$

Taking expectations and applying Jensen's inequality again (since log is concave):

$$\mathbb{E}[Z] \leq \frac{1}{s}\mathbb{E}\left[\log\left(\sum_{a\in A} e^{s\sum_i \sigma_i a_i}\right)\right] \leq \frac{1}{s}\log\left(\mathbb{E}\left[\sum_{a\in A} e^{s\sum_i \sigma_i a_i}\right]\right) = \frac{1}{s}\log\left(\sum_{a\in A}\mathbb{E}\left[e^{s\sum_i \sigma_i a_i}\right]\right)$$

By independence of the σ_i, $\mathbb{E}\left[e^{s\sum_i \sigma_i a_i}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{s\sigma_i a_i}\right]$. We use Hoeffding's Lemma, which for a Rademacher variable implies $\mathbb{E}[e^{s\sigma_i a_i}] \leq e^{s^2 a_i^2/2}$. Thus:

$$\prod_{i=1}^n \mathbb{E}\left[e^{s\sigma_i a_i}\right] \leq \prod_{i=1}^n e^{s^2 a_i^2/2} = e^{\frac{s^2}{2}\sum_i a_i^2} = e^{\frac{s^2}{2}\|a\|_2^2}$$

Let $R = \max_{a\in A}\|a\|_2$. We get:

$$\mathbb{E}[Z] \leq \frac{1}{s}\log\left(\sum_{a\in A} e^{\frac{s^2R^2}{2}}\right) = \frac{1}{s}\log\left(|A|e^{\frac{s^2R^2}{2}}\right) = \frac{\log|A|}{s} + \frac{sR^2}{2}$$

This bound holds for any $s > 0$. We minimize it by setting the derivative w.r.t s to zero, which yields $s = \frac{\sqrt{2\log|A|}}{R}$. Plugging this optimal s back in gives the final result:

$$\mathbb{E}[Z] \leq R\sqrt{2\log|A|} = \sqrt{2\log|A|}\cdot\max_{a\in A}\|a\|_2$$

□

Lemma 3. *Let $\mathcal{H} \subseteq \{f \mid f : Z \to \{0,1\}\}$ be a hypothesis class. The Rademacher complexity of $\mathcal{H}$ with n samples is bounded as*

$$\mathfrak{R}_n(\mathcal{H}) \leq \sqrt{\frac{2\log\mathcal{G}(\mathcal{H},n)}{n}}$$

Proof. By definition, the expected Rademacher complexity of $\mathcal{H}$ with n samples is

$$\mathfrak{R}_n(\mathcal{H}) = \mathbf{E}_{S\sim D^n}\left[\hat{\mathfrak{R}}_S(\mathcal{H})\right],$$

where for a fixed sample $S = (z_1,\ldots,z_n)$ the empirical Rademacher complexity is

$$\hat{\mathfrak{R}}_S(\mathcal{H}) \;=\; \mathbf{E}_\sigma\left[\sup_{h\in\mathcal{H}}\frac{1}{n}\sum_{i=1}^n \sigma_i h(z_i)\;\middle|\; S\right].$$

Fix a sample S. Let

$$\mathcal{H}_S := \{(h(z_1),\ldots,h(z_n)) \in \{0,1\}^n : h\in\mathcal{H}\}$$

denote the set of all label vectors realized by hypotheses in $\mathcal{H}$ on S. Then we can rewrite the inner supremum as a maximization over $\mathcal{H}_S$:

$$\sup_{h\in\mathcal{H}}\frac{1}{n}\sum_{i=1}^n \sigma_i h(z_i) \;=\; \frac{1}{n}\sup_{a\in\mathcal{H}_S}\sum_{i=1}^n \sigma_i a_i.$$

Applying Massart's Lemma to the finite set $\mathcal{H}_S \subseteq \mathbb{R}^n$, we obtain

$$\mathbf{E}_\sigma\left[\sup_{h\in\mathcal{H}}\frac{1}{n}\sum_{i=1}^n \sigma_i h(z_i)\;\middle|\; S\right] \;\leq\; \frac{1}{n}\sqrt{2\log|\mathcal{H}_S|}\;\max_{a\in\mathcal{H}_S}\|a\|_2.$$

Now observe that $|\mathcal{H}_S| \le \mathcal{G}(\mathcal{H}, n)$, since the growth function counts the maximum number of distinct labelings achievable on n points, and each $a \in \{0,1\}^n$ satisfies $\|a\|_2 \le \sqrt{n}$. Therefore,

$$\mathbf{E}_\sigma\left[\sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^n \sigma_i h(z_i) \;\middle|\; S\right] \;\le\; \frac{1}{n}\sqrt{2\log \mathcal{G}(\mathcal{H}, n)} \cdot \sqrt{n}.$$

Taking expectation with respect to $S \sim D^n$ does not change the bound (since the right-hand side no longer depends on S), and we conclude that

$$\mathfrak{R}_n(\mathcal{H}) \;\le\; \sqrt{\frac{2\log \mathcal{G}(\mathcal{H}, n)}{n}}.$$

□

4 Conclusion

Theorem 1 (Generalization Bound via VC-Dimension)**.** *Let $\mathcal{H} \subseteq \{f : Z \to \{0,1\}\}$ be a hypothesis class, and let $d = \mathrm{VCdim}(\mathcal{H})$. Then for any $\delta \in (0,1)$, with probability at least $1-\delta$ over the choice of an i.i.d. sample $S \sim D^n$, every $h \in \mathcal{H}$ satisfies*

$$R(h) \;\le\; \hat{R}_S(h) \;+\; \sqrt{\frac{2d\log\left(\frac{en}{d}\right)}{n}} \;+\; \sqrt{\frac{\log(1/\delta)}{2n}}.$$

Sketch. From the standard Rademacher complexity bound, with probability at least $1-\delta$,

$$R(h) \;\le\; \hat{R}_S(h) \;+\; 2\,\mathfrak{R}_n(\mathcal{H}) \;+\; \sqrt{\tfrac{\log(1/\delta)}{2n}}, \quad \forall h \in \mathcal{H}.$$

By the previous lemma, we showed that

$$\mathfrak{R}_n(\mathcal{H}) \;\le\; \sqrt{\tfrac{2\log \mathcal{G}(\mathcal{H},n)}{n}}.$$

Finally, applying the Sauer–Shelah lemma, we have

$$\mathcal{G}(\mathcal{H}, n) \;\le\; \sum_{i=0}^{d}\binom{n}{i} \;\le\; \left(\tfrac{en}{d}\right)^d.$$

Plugging this into the Rademacher complexity bound yields

$$\mathfrak{R}_n(\mathcal{H}) \;\le\; \sqrt{\frac{2d\log(en/d)}{n}}.$$

Substituting back, we obtain the stated inequality. □

5 Extra Read: Generalizing to Multi-Class Classification

The VC dimension is much more intuitive that Rademacher but it is defined specifically for binary classification ($\{0,1\}$ or $\{-1,1\}$ labels). What if we have more than two classes? The **Natarajan dimension** provides a natural generalization. I read about it from here wiki/Natarajan Dimension and Natarajan, B.K. On learning sets and functions. Mach Learn 4, 67–97 (1989)

Definition 1 (Natarajan Dimension)**.** *Let $\mathcal{H}$ be a class of functions mapping from $\mathcal{Z}$ to $\{1,2,\ldots,k\}$. A set $S \subset \mathcal{Z}$ is **Natarajan-shattered** by $\mathcal{H}$ if there exist two labelings (functions) $y_1 : S \to \{1,\ldots,k\}$ and $y_2 : S \to \{1,\ldots,k\}$ such that $y_1(z) \neq y_2(z)$ for all $z \in S$, and for any subset $B \subseteq S$, there exists a hypothesis $h \in \mathcal{H}$ such that:*

$$h(z) = \begin{cases} y_1(z) & \text{if } z \in B \\ y_2(z) & \text{if } z \in S \setminus B \end{cases}$$

*The **Natarajan dimension** of $\mathcal{H}$ is the size of the largest set S that can be Natarajan-shattered.*

In essence, instead of generating all $2^{|S|}$ binary labelings, the class must be rich enough to generate all $2^{|S|}$ ”hybrid” labelings formed by picking between two pre-defined, distinct labelings for each point. For the binary case ($k = 2$), if we choose y_1 to be all 1s and y_2 to be all 0s, the Natarajan dimension exactly reduces to the VC dimension. Like the VC dimension, the Natarajan dimension can be used to derive generalization bounds for multi-class classification problems, showing that the core idea of shattering is fundamental to learning theory.