# Vapnik-Chervonenkis (VC) Dimension

## 1 Definition of VC Dimension

For a given hypothesis set $\mathcal{H}$, the **Vapnik-Chervonenkis (VC) dimension**, denoted as $d_{\text{VC}}(\mathcal{H})$, is defined as the largest value of $N$ (number of data points) for which the growth function $m_{\mathcal{H}}(N)$ satisfies:

$$m_{\mathcal{H}}(N) = 2^N \tag{1}$$

In other words, it is the maximum number of points that can be *shattered* by the hypothesis set.

If $m_{\mathcal{H}}(N) = 2^N$ for all $N$, then the VC dimension is infinite ($d_{\text{VC}} = \infty$).

## 2 The Generalization Bound

If a hypothesis set has a **finite** VC dimension, the gap between the in-sample error ($E_{\text{in}}$) and the out-of-sample error ($E_{\text{out}}$) gets smaller as the number of samples $N$ increases.

The generalization bound is given by:

$$E_{\text{out}} \leq E_{\text{in}} + \sqrt{\frac{1}{2N} \ln\left(\frac{2m_{\mathcal{H}}(N)}{\delta}\right)} \tag{2}$$

where $\delta$ is the probability of the bound failing.

## 3 Growth Function and Polynomial Bound

If $d_{\text{VC}}(\mathcal{H}) < \infty$, the growth function is bounded by a polynomial (Sauer's Lemma):

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{\text{VC}}} \binom{N}{i} \tag{3}$$

This means that for large $N$, the growth function grows polynomially rather than exponentially:

$$m_{\mathcal{H}}(N) = \mathcal{O}\left(N^{d_{\text{VC}}}\right) \tag{4}$$

By taking the natural logarithm, we can approximate the complexity term in the bound:

$$\ln\left(m_{\mathcal{H}}(N)\right) \approx \ln\left(N^{d_{\text{VC}}}\right) = d_{\text{VC}} \ln(N) \tag{5}$$

Substituting this into the penalty term of the generalization bound yields a term proportional to:

$$\sqrt{\frac{d_{\text{VC}} \ln(N)}{2N}} \tag{6}$$

As $N \to \infty$, this penalty term $\epsilon \to 0$, ensuring that $E_{\text{out}} \to E_{\text{in}}$.

## 4 Convergence: Small vs. Large VC Dimension

Consider the penalty term behavior for different finite VC dimensions:

- **Small VC dimension (e.g., $d_{\text{VC}} = 2$):** The penalty term is proportional to $\sqrt{\frac{2 \ln N}{N}}$.
- **Large VC dimension (e.g., $d_{\text{VC}} = 1000$):** The penalty term is proportional to $\sqrt{\frac{1000 \ln N}{N}}$.

While both terms converge to 0 as $N \to \infty$, the model with the smaller VC dimension ($d_{\text{VC}} = 2$) gets smaller **faster**. A smaller VC dimension guarantees faster convergence of $E_{\text{out}}$ to $E_{\text{in}}$.

## 5 The Case of Infinite VC Dimension

What happens if $d_{\text{VC}} = \infty$? In this case, the growth function remains exponential for all $N$:

$$m_{\mathcal{H}}(N) = 2^N \tag{7}$$

If we plug this back into the generalization bound penalty term, we get:

$$\begin{aligned}
\sqrt{\frac{1}{2N} \ln\left(\frac{2(2^N)}{\delta}\right)} &= \sqrt{\frac{1}{2N} \left(\ln(2^{N+1}) - \ln(\delta)\right)} \\
&\approx \sqrt{\frac{N \ln 2}{2N}} \\
&= \sqrt{\frac{\ln 2}{2}} \approx \text{constant}
\end{aligned}$$

Because the penalty term approaches a constant rather than 0, $E_{\text{out}} \approx E_{\text{in}}$ is **not guaranteed**.

**Conclusion**

**Finite VC dimension** is the fundamental property that makes generalization possible within this theoretical framework. Without it, the error bounds do not shrink, and learning from finite data cannot be mathematically guaranteed.