Lecture 2p1

Sunday, 15 January 2023 4:04 PM

$E_{in}(h) = \frac{1}{N}\sum_{i=1}^{N}[h(x_n) \neq f(x_n)]$

$E_{out}(h) = P[h(x) \neq f(x)]$

Marble Experiment

A bin with red & green marbles

Prob of picking a red marble μ

For green marble it is $1-\mu$

But μ is Unknown!!

Pick N independent marbles and observe ν of red marbles within sample.

Q. How are ν & μ related?

Hoeffding's Inequality:-

$$P[|\nu-\mu| > \epsilon] \leq 2e^{-2\epsilon^2 N}, \epsilon > 0$$

Relate this to Learning

For a fixed hypothesis

$$P[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq 2e^{-2\epsilon^2 N} \quad \epsilon > 0$$

For a fixed set of Hypothesis of Size M

$$P[|E_{in}(h) - E_{out}(g)| > \epsilon] 2Me^{-2\epsilon^2 N} \quad \epsilon > 0$$

With Probability at least $1-\delta$

$$\boxed{E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}}}$$

Generalization Bound

What if $M = \infty$??

Case!! Almost Always the

Some Definitions:-

Def A:- For $x_1, \ldots, x_N \in X$

The dichotomies generated by $\mathcal{H}$ on these points is defined by

$$\mathcal{H}(x_1, \ldots x_N) = \{h(x_1), \ldots h(x_N) \mid h \in \mathcal{H}\}$$

Def B:- The growth function for a hypothesis set $\mathcal{H}$ is

$$m_{\mathcal{H}}(N) = \max_{x_1, \ldots, x_N} |\mathcal{H}(x_1, x_2, \ldots x_N)|$$

$$m_{\mathcal{H}}(N) \leq 2^N$$

If $m_{\mathcal{H}}(N) = 2^N$, $\mathcal{H}$ shatters $x_1, \ldots x_N$

VC-Dimension:- The VC-Dimension of a hypothesis set $\mathcal{H}$, denoted by $d_{vc}(\mathcal{H})$ or d_{vc} is the largest value of N for which $m_{\mathcal{H}}(N) = 2^N$.

If $m_{\mathcal{H}}(N) = 2^N$ for $\forall N$, $d_{vc}(\mathcal{H}) = \infty$

Important Results:-

Thm:- If $m_{\mathcal{H}}(k) < 2^k$ for some value k, then

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1}\binom{N}{i}$$

By definition of VC-Dimension

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{vc}}\binom{N}{i}$$

$d_{vc} \geq N \iff$ there exists data of size N such that $\mathcal{H}$ shatters the data

The VC-Generalization bound

For any tolerance $\delta > 0$

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N}\ln\frac{4m_{\mathcal{H}}(2N)}{\delta}}$$

w.p $1-\delta$