Machine Learning

Symbol	Name
$\mathcal{X}$	Input Set
$\mathcal{D}$	Training Set
$\mathcal{Y}$	Target Set
f	Target Function
$\mathcal{H}$	Hypothesis Set
$\mathcal{A}$	Learning Algorithm
M	Cardinality of Hypothesis Set
g	Final Hypothesis
ϵ	Residual Error
$h(x_k) \exists x_k \in (X)$	Dichotomy
N	Number of Data Points
$m_{\mathcal{H}}(N)$	Growth Function
$m_{\mathcal{H}}(N)$	Growth Function
$B\left(N, k\right)$	Growth Bound with Breakpoint k
d_{vc}	Vapnik-Chervonenkis Dimension

Relations and Sets

$f : \mathcal{X} \rightarrow \mathcal{Y}$

$\{h_1, h_2, \cdots h_M\} \in \mathcal{H}$

Hoeffding Inequality

$$P\left[|\nu - \mu| > \epsilon\right] \leq 2e^{-2\epsilon^2 N} \forall \epsilon > 0 \tag{1}$$

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

For the final hypothesis g selected from $\mathcal{H}$,

$$\begin{aligned}
P\left[|E_{in}(g) - E_{out}(g)| > \epsilon\right] &\leq P\left[|E_{in}(h_1) - E_{out}(h_1)| > \epsilon\right] \\
&\textbf{or} \leq P\left[|E_{in}(h_1) - E_{out}(h_1)| > \epsilon\right] \\
&\cdots \\
&\textbf{or} \leq P\left[|E_{in}(h_M) - E_{out}(h_M)| > \epsilon\right] \\
&\leq \Sigma_{m=1}^{M} P\left[|E_{in}(h_m) - E_{out}(h_m)| > \epsilon\right] \\
&\leq 2Me^{-1\epsilon^2 N}
\end{aligned} \tag{3}$$

In-sample Error

$$E_{in}(h) = \frac{1}{N}\Sigma_{n=1}^{N} P\left[h(x_n) \neq f(x_n)\right] \text{ fraction of D where f \& h disagree} \tag{4}$$

Out-of-sample Error

$$E_{out}(h) = P\left[h(x) \neq f(x)\right] \tag{5}$$

Markov's Inequality

$P\left[x \geq \alpha\right] \leq \frac{E(x)}{\alpha}$ for $\alpha > 0$

Proof:

$$\begin{aligned}
E(x) &= \int xP(x)dx \\
&= \int_0^{\alpha} xP(x)dx + \int_{\alpha}^{\infty} xP(x)dx \\
&\geq \int_{\alpha}^{\infty} xP(x)dx \\
&\geq \int_{\alpha}^{\infty} \alpha P(x)dx \\
&= \alpha \int_{\alpha}^{\infty} P(x)dx \\
&= \alpha P\left[x \geq \alpha\right]
\end{aligned} \tag{6}$$

Generalization Error

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N} ln \frac{2M}{\delta}} \tag{7}$$

Growth Function

$$m_{\mathcal{H}}(N) = \max_{\vec{x} \in \mathcal{X}} |\mathcal{H}(\vec{x})| \tag{8}$$

Bounding the Growth Function

$B\,(N,k)$ is the maximum number of dichotomies on N points such that no subset of size k of the N points can be shattered by these dichotomies.

Vapnik-Chervonenkis Dimension

$$d_{VC}(\mathcal{H}) = \begin{cases} \max N \text{ s.t. } m_{\mathcal{H}}(N) = 2^N & \text{if } m_{\mathcal{H}}(N) < 2^N \exists N \\ \infty & \text{if } m_{\mathcal{H}}(N) = 2^N \forall N \end{cases} \tag{9}$$

$k = d_{VC} + 1$ is a breakpoint for $m_{\mathcal{H}}(N)$

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

$m_{\mathcal{H}}(N) \leq N^{d_{VC}} + 1$

Sauer's Lemma

$$B\,(N,k) \leq \Sigma_{i=0}^{k-1} \binom{N}{i} \tag{10}$$