COL866: Foundations of Data Science

Ragesh Jaiswal, IITD

Machine Learning: Generalization

Theorem

For any hypothesis class $\mathcal{H}$ and distribution D, if a training sample S is drawn from D of size $n \geq \frac{2}{\varepsilon}\left[\log_2\left(2\mathcal{H}[2n]\right) + \log_2\left(1/\delta\right)\right]$. then with probability at least $(1-\delta)$, every $h \in \mathcal{H}$ with error $err_D(h) \geq \varepsilon$ has $err_S(h) > 0$. Equivalently, every $h \in \mathcal{H}$ with $err_S(h) = 0$ has $err_D(h) < \varepsilon$.

Theorem

For any hypothesis class $\mathcal{H}$ and distribution D, if a training sample S is drawn from D of size $n \geq \frac{8}{\varepsilon^2}\left[\log_2\left(2\mathcal{H}[2n]\right) + \log_2\left(2/\delta\right)\right]$. then with probability at least $(1-\delta)$, every $h \in \mathcal{H}$ will have $|err_D(h) - err_S(h)| \leq \varepsilon$.

Theorem (Sauer's Lemma)

If $VCdim(\mathcal{H}) = d$, then $\mathcal{H}[n] \leq \sum_{i=0}^{d} \binom{n}{i} \leq \left(\frac{en}{d}\right)^d$.

Theorem

For any hypothesis class $\mathcal{H}$ and distribution D, a training sample S of size

$$\Omega\left(\frac{1}{\varepsilon}\left[VCdim(\mathcal{H})\log\left(1/\varepsilon\right) + \log 1/\delta\right]\right)$$

is sufficient to ensure that with probability at least $(1-\delta)$, every $h \in \mathcal{H}$ with $err_D(h) \geq \varepsilon$ has $err_S(h) > 0$.

Theorem

For any hypothesis class $\mathcal{H}$ and distribution D, if a training sample S is drawn from D of size

$$n \geq \frac{2}{\varepsilon}\left[\log_2\left(2\mathcal{H}[2n]\right) + \log_2\left(1/\delta\right)\right].$$

then with probability at least $(1-\delta)$, every $h \in \mathcal{H}$ with error $err_D(h) \geq \varepsilon$ has $err_S(h) > 0$. Equivalently, every $h \in \mathcal{H}$ with $err_S(h) = 0$ has $err_D(h) < \varepsilon$.

- We will need the following lemma.

Lemma

Let $\mathcal{H}$ be a concept class over some domain $\mathcal{X}$ and let S and S' be sets of n elements drawn from some distribution D on $\mathcal{X}$, where $n \geq 8/\varepsilon$. Let A be the event that there exists $h \in \mathcal{H}$ with $err_S(h) = 0$ but $err_D(h) \geq \varepsilon$. Let B be the event that there exists $h \in \mathcal{H}$ with $err_S(h) = 0$ but $err_{S'}(h) \geq \varepsilon/2$. Then $\mathbf{Pr}[B] \geq \mathbf{Pr}[A]/2$.

End