Generalization bounds using covering number.

MATH 637

Feb 20th, 2019

We know that for finite hypothesis space and bounded loss, if we quantify the error in term of number of samples, then

$$|R_n(h) - R(h)| \le \frac{c}{\sqrt{n}}\sqrt{2\log\left(\frac{2}{\delta}\right) + 2\log(|\mathcal{H}|)}, \forall h \in \mathcal{H}$$

with probability at least $1-\delta$.

What about infinite hypothesis class?

Assumption. In this note, we assume that $\mathcal{H}$ is a metric space with distance d defined on it. For $\epsilon > 0$, we denote by $\mathcal{N}(\epsilon, \mathcal{H}, d)$ the *covering number* of $(\mathcal{H}, d)$; that is, $\mathcal{N}(\epsilon, \mathcal{H}, d)$ is the minimal number of balls of radius ϵ needed to cover $\mathcal{H}$. We denote by $\mathcal{H}_\epsilon$ a finite subset of $\mathcal{H}$ such that $\mathcal{H}$ is contained in the union of balls of radius ϵ and $|\mathcal{H}_\epsilon| = \mathcal{N}(\epsilon, \mathcal{H}, d)$.

Note: If $\mathcal{H}$ is a $dk-$dimensional manifold/algebraic surface, then we now that

$$\mathcal{N}(\epsilon, \mathcal{H}, d) = \mathcal{O}\left(\epsilon^{-k}\right)$$

Assume further that the loss function L satisfies:

$$|L(h(x), y) - L(h'(x), y)| \le Cd(h, h') \quad \forall, x \in \mathcal{X}; y \in \mathcal{Y}; h, h' \in \mathcal{H}$$

Generalization bound using covering number.

We first note that if

$$n = \frac{8c^2}{\epsilon^2}\log\left(\frac{2|\mathcal{H}_\epsilon|}{\delta}\right)$$

then the event

$$|R_n(h) - R(h)| \le \epsilon, \forall h \in \mathcal{H}_\epsilon$$

happens with probability at least $1-\delta$.

Under this event, consider any $h \in \mathcal{H}$, then there exists $h_0 \in \mathcal{H}_\epsilon$ such that $d(h, h_0) \le \epsilon$. This means

$$|R_n(h) - R_n(h_0)| \le Cd(h, h_0)$$

and

$$|R(h) - R(h_0)| \le Cd(h, h_0).$$

This implies that

$$|R_n(h) - R(h)| \le (2C+1)\epsilon \quad \forall h \in \mathcal{H}.$$

We conclude that for all $\epsilon > 0$, $\delta > 0$, if

$$n = \frac{8c^2}{\epsilon^2}\log\left(\frac{2\mathcal{N}(\epsilon, \mathcal{H}, d)}{\delta}\right)$$

then

$$|R_n(h) - R(h)| \le (2C+1)\epsilon \quad \forall h \in \mathcal{H}.$$

with probability at least $1-\delta$.