Generalisation Bounds (2): VC dimension and bounds

Qinfeng (Javen) Shi

The Australian Centre for Visual Technologies,
The University of Adelaide, Australia

27 April 2012

Course Outline

Generalisation Bounds:

1. Basics
2. VC dimensions and bounds (Today)
3. Rademacher complexity and bounds
4. PAC Bayesian Bounds
5. . . .

Recap: finite many hypotheses

For an finite hypothesis set $\mathcal{G} = \{g_1, \cdots, g_N\}$. we have with probability at least $1 - \delta$,

$$\forall g \in \mathcal{G}, R(g) \leq R_n(g) + \sqrt{\frac{\log N + \log(\frac{1}{\delta})}{2n}}$$

$$\forall g \in \mathcal{G}, R(g) \leq R(g^*) + 2\sqrt{\frac{\log N + \log(\frac{1}{\delta})}{2n}}$$

Infinite many hypotheses (1)

Though there are infinite many g in $\mathcal{G}$, there are only two possible outputs for a x (because $g(x) \in \{-1, +1\}$). What matters is the "expressive power" (Blumer *et al.* 1986,1989)(*e.g.* the number of different prediction outputs), not the cardinality of $\mathcal{G}$.

Infinite many hypotheses (2)

For an infinite hypothesis set, for any n training examples, there are at most 2^n different outputs of $g(x)$.
For any finite n, 2^n is finite.
Problem:

$$\sqrt{\frac{\log(2^n) + \log(\frac{1}{\delta})}{2n}} = \sqrt{\frac{1}{2} + \frac{\log(\frac{1}{\delta})}{2n}}$$

is too loose. We need something shrinks to zero as n goes to infinity.

Growth function

Definition (Growth function)

The growth function (a.k.a Shatter coefficient) of $\mathcal{F}$ with n points is

$$S_{\mathcal{F}}(n) = \sup_{(z_1, \cdots, z_n)} |\{(f(z_1), \cdots, f(z_n)) : f \in \mathcal{F}\}|.$$

i.e. maximum number of ways that n points can be classified by the hypothesis set $\mathcal{F}$.

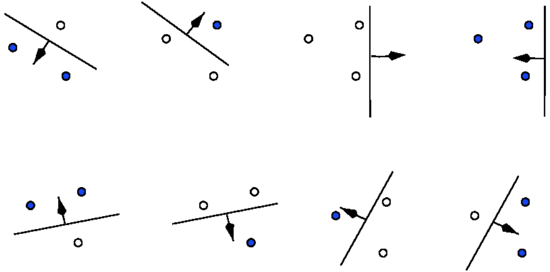
VC dimension (1)

Definition (VC dimension)

The VC dimension of a hypothesis set $\mathcal{G}$, is the largest n such that

$$S_{\mathcal{G}}(n) = 2^n.$$

The growth function $S_{\mathcal{G}}(n) = 8$ for $n = 3$ and $\mathcal{G}$ being the half-space shown in the image below[1].



[1] The image is from http://www.svms.org/vc-dimension/

VC dimension (2)

Lemma

Let $\mathcal{G}$ be a set of functions with finite VC dimension h. Then for all $n \in \mathbb{N}$,

$$S_{\mathcal{G}}(n) \leq \sum_{i=0}^{h} \binom{n}{i},$$

and for all $n \geq h$,

$$S_{\mathcal{G}}(n) \leq \left(\frac{en}{h}\right)^h.$$

VC dimension (3)

Theorem (Growth function bound)

For any $\delta \in (0,1)$, *with probability at least* $1-\delta$, $\forall g \in \mathcal{G}$

$$R(g) \leq R_n(g) + 2\sqrt{2\frac{\log S_{\mathcal{G}}(2n) + \log(\frac{2}{\delta})}{n}}$$

Thus for all $n \geq h$, since $S_{\mathcal{G}}(n) \leq (\frac{en}{h})^h$, we have

Theorem (VC bound)

For any $\delta \in (0,1)$, *with probability at least* $1-\delta$, $\forall g \in \mathcal{G}$

$$R(g) \leq R_n(g) + 2\sqrt{2\frac{h\log\frac{2en}{h} + \log(\frac{2}{\delta})}{n}}.$$

VC dimension (4)

Assume $x \in \mathbb{R}^d, \Phi(x) \in \mathbb{R}^D$ (Note D can be $+\infty$).

- linear $\langle x, w \rangle$, $h = d + 1$
- polynomial $(\langle x, w \rangle + 1)^p$, $h = \binom{d+p-1}{p} + 1$
- Gaussian RBF $\exp\left(-\frac{\|x-x'\|^2}{\sigma^2}\right)$, $h = +\infty$.
- Margin γ, $h \le \min\{D, \lceil \frac{4R^2}{\gamma^2} \rceil\}$, where the radius $R^2 = \max_{i=1}^n \langle \Phi(x_i), \Phi(x_i) \rangle$ (assuming data are already centered)

Proof of growth function/VC bound (1)

One way to prove it is to use Symmetrisation lemma and a variant of Hoeffding inequality.

Lemma (Symmetrisation)

For any $t > 0$, such that $nt^2 \geq 2$,

$$\Pr\left[\sup_{f\in\mathcal{F}}\left(\mathbb{E}_{z\sim P(z)}[f(z)] - \frac{1}{n}\sum_{i=1}^{n} f(z_i)\right) \geq t\right]$$

$$\leq 2\Pr\left[\sup_{f\in\mathcal{F}}\left(\frac{1}{n}\sum_{i=1}^{n} f(z_i') - \frac{1}{n}\sum_{i=1}^{n} f(z_i)\right) \geq \frac{t}{2}\right]$$

Here $\{z_i'\}_{i=1}^{n}$ are called a "ghost sample".

Proof of growth function/VC bound (2)

Theorem (Hoeffding2)

Let $Z_1, \cdots, Z_n, Z_1', \cdots, Z_n'$ be $2n$ i.i.d. random variables with $f(Z) \in [a, b]$. Then for all $\epsilon > 0$, we have

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n} f(Z_i) - \frac{1}{n}\sum_{i=1}^{n} f(Z_i') > \epsilon\right) \leq \exp\left(-\frac{n\epsilon^2}{2(b-a)^2}\right)$$

Proof of growth function/VC bound (3)

$$\begin{aligned}
&\Pr\left[\sup_{g\in\mathcal{G}}\Big(R(g)-R_n(g)\Big)\geq 2\epsilon\right]\\
&\leq 2\Pr\left[\sup_{g\in\mathcal{G}}\Big(R_n'(g)-R_n(g)\Big)\geq \epsilon\right]\\
&= 2\Pr\left[\sup_{g\in\mathcal{G}_{z_1,\cdots,z_n,z_1,\cdots,z_n'}}\Big(R_n'(g)-R_n(g)\Big)\geq \epsilon\right]\\
&\leq 2S_{\mathcal{G}}(2n)\Pr\left[\Big(R_n'(g)-R_n(g)\Big)\geq \epsilon\right]\\
&\leq 2S_{\mathcal{G}}(2n)\exp\Big(\frac{-n\epsilon^2}{2}\Big).
\end{aligned}$$

Let $\delta = 2S_{\mathcal{G}}(2n)\exp(\frac{-n\epsilon^2}{2})$, we have $\epsilon = \sqrt{2\frac{\log S_{\mathcal{G}}(2n)+\log(\frac{2}{\delta})}{n}}$.

$R(g) - R_n(g) \leq 2\epsilon = 2\sqrt{2\frac{\log S_{\mathcal{G}}(2n)+\log(\frac{2}{\delta})}{n}}$ with probability at least $1-\delta$.

Related concepts

- VC Entropy
- Covering Number
- Rademacher complexity (will be covered in the next talk)