15. Growth Function

Friday, February 9, 2024 3:18 PM

Growth Function (§3.2 in Mohri et al., §6 in Shalev-Shwartz + Ben-David)

Def The growth function $\tau_{\mathcal{H}} : \mathbb{N} \to \mathbb{N}$ (of the set of functions $\mathcal{H}$)
← Mohri uses $\Pi_{\mathcal{H}}$

is $\tau_{\mathcal{H}}(m) := \max_{\{x_1,\dots,x_m\} \subseteq X} \left| \{ (h(x_1),\dots,h(x_m)) : h \in \mathcal{H} \} \right|$ ← cardinality, not abs. value

if $\mathcal{H}$ is finite, $\tau_{\mathcal{H}}(m) \le |\mathcal{H}|$

= "maximum number of distinct ways in which m points can be classified via $h \in \mathcal{H}$"

If we define the restriction of $\mathcal{H}$ to a set $C = \{x_1,\dots,x_m\} \subseteq X$ (see Def. 6.2 in [SS])
to be $\mathcal{H}_C = \{ (h(x_1),\dots,h(x_m)) : h \in \mathcal{H} \}$ then, in this notation,

$$\tau_{\mathcal{H}}(m) = \max_{\substack{C \subseteq X \\ |C| = m}} |\mathcal{H}_C|$$

Notes

- Doesn't involve any distribution on data
- Only really useful when Y is finite ($h : X \to Y$)
  - it's a purely combinatorial concept, so less general than Rademacher Complexity
- Ex: $Y = \{-1,1\}$, then $\underbrace{\tau_{\mathcal{H}}(m) \le 2^m}_{\text{trivial bound}}$

Q: for which $\mathcal{H}$ is $\tau_{\mathcal{H}}(m) < 2^m$?

Usefulness

Cor. 3.8 Mohri: $R_m(\mathcal{H}) \le \sqrt{\frac{2 \log(\tau_{\mathcal{H}}(m))}{m}}$ if $Y = \{\pm 1\}$

(proof uses Massart's Lemma) ← Recall: $\hat{R}(A) \le \text{radius}(A) \cdot \frac{1}{m} \cdot \sqrt{2 \log(|A|)}$

i.e. Growth function bounds Rademacher Complexity.

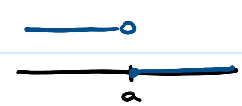
... but $\tau_{\mathcal{H}}$ can still be tricky to calculate.

Example (6.1 in [SS])

$X = \mathbb{R}$, $Y = \{\cancel{\pm}1\}$, $\mathcal{H} = \{ h_a : a \in \mathbb{R} \}$ = threshold functions on $\mathbb{R}$

$Y = \{0,1\}$ is more convenient

i.e. $h_a(x) = \begin{cases} 1 & x < a \\ 0 & x \ge a \end{cases} = \mathbb{1}_{x<a}$



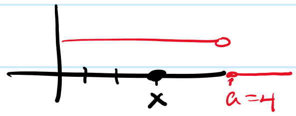
let's compute $\tau_{\mathcal{H}}(m)$ for a few m

15a. Growth Function (p. 2)

Friday, February 9, 2024 4:45 PM

(let $Y = \{0,1\}$)

m = 1. $\tau_{\mathcal{H}}(m) = \max_{|C|=1} |\mathcal{H}_C|$. Let $C = \{x\}$, eg. $x = 3$



Choosing $a = 4$, $h_{a=4}(x=3) = 1$

Choosing $a = 2$, $h_{a=2}(x=3) = 0$

So $\mathcal{H}_C = \{0,1\}$, $|\mathcal{H}_C| = 2 = 2^m$

this was any x,
so also true for $\max_{|C|=1} |\mathcal{H}_C|$.

We say "$\mathcal{H}$ shatters C"
ie. for C, all 2^m possible labels can be explained by $\mathcal{H}$.

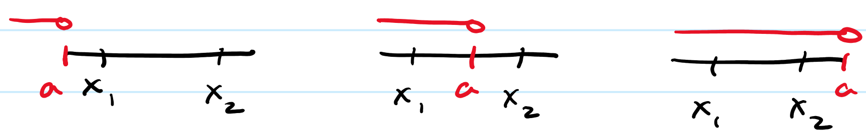
So $\tau_{\mathcal{H}}(1) = 2$

(a bad thing for generalization, ie. "a theory that explains everything explains nothing")

m = 2 $|C| = 2$ so $C = \{x_1, x_2\}$

$x_1 = x_2$ not a wise choice so exclude

wlog let $x_1 < x_2$



labels $\begin{bmatrix}0\\0\end{bmatrix}$ $\begin{bmatrix}1\\0\end{bmatrix}$ $\begin{bmatrix}1\\1\end{bmatrix}$

so $|\mathcal{H}_C| = 3 < 4 = 2^m$

If $x_1 > x_2$, then still only 3 possible labels: $\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}$

So
$\tau_{\mathcal{H}}(2) = \max_{|C|=2} |\mathcal{H}_C| = 3 < 2^m$

(if $x_1 = x_2$, then only 2 possible labels, $\begin{bmatrix}0\\0\end{bmatrix}$ & $\begin{bmatrix}1\\1\end{bmatrix}$)

$\mathcal{H}$ doesn't shatter any sets of size 2

Linking back to our proof of the No Free Lunch thm: we chose $X = 2m$, picked an adversarial hypothesis $h \in \mathcal{H}$ that was still consistent w/ our observed data.

Corollary 6.4 $Y = \{\pm 1\}$, if $\exists$ set $C \subseteq X$ of size $2m$ that is shattered by $\mathcal{H}$, then $\forall$ algo A, $\exists$ a distribution $\mathcal{D}$ and $h \in \mathcal{H}$ st. 1) $L_{\mathcal{D}}(h) = 0$
$|S| = m$
2) wp $\geq 1/7$, $L_{\mathcal{D}}(A(S)) \geq 1/8$