LEARNING THEORY STRIKES BACK AGAIN

Matthieu Bloch

February 14, 2019

LOGISTICS

Lecture slides and notes

- Report typos and errors on Piazza (thank you!)

If you are considering dropping the class, please send an email to Dr. Bloch

Problem set #1

- Solutions released once DL student turn their homework

Problem set #2

- Released over the weekend

Registering for office hours

- Solutions released once DL student turn their homework

RECAP: DICHOTOMIES AND GROWTH FUNCTION

Definition (Dichotomy)

For a dataset $\mathcal{D} \triangleq \{\mathbf{x}_i\}_{i=1}^N$ and set of hypotheses $\mathcal{H}$, the set of *dichotomies* generated by $\mathcal{H}$ on $\mathcal{D}$ is

$$\mathcal{H}(\{\mathbf{x}_i\}_{i=1}^N) \triangleq \{\{h(\mathbf{x}_i)\}_{i=1}^N : h \in \mathcal{H}\}$$

By definition $|\mathcal{H}(\{\mathbf{x}_i\}_{i=1}^N)| \leq 2^N$ and in general $|\mathcal{H}(\{\mathbf{x}_i\}_{i=1}^N)| \ll |\mathcal{H}|$

Definition (Growth function)

For a set of hypotheses $\mathcal{H}$, the *growth function* of $\mathcal{H}$ is

$$m_{\mathcal{H}}(N) \triangleq \max_{\{\mathbf{x}_i\}_{i=1}^N} |\mathcal{H}(\{\mathbf{x}_i\}_{i=1}^N)|$$

The growth function does *not* depend on the datapoints $\{\mathbf{x}_i\}_{i=1}^N$ and $m_{\mathcal{H}}(N) \leq 2^N$

RECAP: EXAMPLES OF GROWTH FUNCTIONS

Linear classifiers: $\mathcal{H} \triangleq \{h : \mathbb{R}^2 \to \{\pm 1\} : \mathbf{x} \mapsto \text{sgn}\,(\mathbf{w}^\intercal \mathbf{x} + b) | \mathbf{w} \in \mathbb{R}^2, b \in \mathbb{R}\}$

- $m_{\mathcal{H}}(3) = 8$
- $m_{\mathcal{H}}(4) = 14 < 2^4$

Definition (Shattering)

If $\mathcal{H}$ can generate all dichotomies on $\{\mathbf{x}_i\}_{i=1}^N$, we say that $\mathcal{H}$ *shatters* $\{\mathbf{x}_i\}_{i=1}^N$

Definition (Break point)

If not data set of size k can be shattered by $\mathcal{H}$, then k is a break point for $\mathcal{H}$

BREAKING POINTS

Proposition.

If there exists *any* break point for $\mathcal{H}$, then $m_{\mathcal{H}}(N)$ is polynomial in N

If there is no break point for $\mathcal{H}$, then $m_{\mathcal{H}}(N) = 2^N$

Definition.

Assume $\mathcal{H}$ has break point k. $B(N,k)$ is the maximum number of dichotomies of N points such that *no subset* of size k can be shattered by the dichotomies.

$B(N,k)$ is a purely combinatorial quantity, does *not* depend on $\mathcal{H}$

Example.

Assume $\mathcal{H}$ has break point 2. How many dichotomties can we generate of set of size 3?

By definition, if k is a break point for $\mathcal{H}$, then $m_{\mathcal{H}}(N) \leq B(N,k)$

BREAKING POINTS

Lemma.

$B(N,1) = 1, B(1,k) = 2$ for $k > 1$,

$$\forall k > 1 \quad B(N,k) \leq B(N-1,k) + B(N-1,k-1)$$

Lemma.

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

$B(N,k)$ is polynomial