LECTURE 22

Last time

- L_p-testing of monotonicity
- Work investment strategy
- Testing via learning

Today

- PAC learning and VC-dimension

Project Reports are due April 24

PAC Learning [Valiant]

PAC means ``Probably Approximately Correct''

- Let Ω be a (finite or infinite) domain.
- Let $\mathcal{C}$ be a class of Boolean functions on domain Ω, i.e., functions of the form $f: \Omega \to \{0,1\}$
- Let $\mathcal{D}$ be a distribution over Ω.
- The learner $\mathcal{L}$ is given parameters $\varepsilon, \delta \in (0,1)$ and a set S of m *examples* drawn i.i.d. from $\mathcal{D}$ and labeled with a function $f \in \mathcal{C}$:
$$\{(x, f(x)) : x \in S\}.$$
- Goal of $\mathcal{L}$: to find a *hypothesis* $h \in \mathcal{C}$ with error less than ε:
$$err_{\mathcal{D}}(h) := \Pr_{x \sim \mathcal{D}}[f(x) \neq h(x)].$$
- An algorithm $\mathcal{L}$ is a PAC-learner for a class $\mathcal{C}$ if the probability it returns a hypothesis h with $err_{\mathcal{D}}(h) \leq \varepsilon$ is at least $1 - \delta$.

The probability is taken over distribution $\mathcal{D}$ and the coins of $\mathcal{L}$.

ERM

ERM stands for ``empirical risk minimization''

- Empirical error (or empirical risk) of a hypothesis h is
$$err_S(h) := \Pr_{x \sim S}[f(x) \neq h(x)]$$
- An empirical risk minimizer is a hypothesis that has the smallest err_S among all hypothesis in the class, i.e., it mislabels the smallest number of examples in S.
- We will see that returning an ERM hypothesis is a great strategy for the learner.

Big Question

For a given class $\mathcal{C}$,
what sample size m
is sufficient for PAC learning?

Vapnik-Chervonenkis (VC) Dimension

- We will think of functions $f \in \mathcal{C}$ as indicator functions for sets.
- For this part, we will equate them with sets. I.e., now $f \subseteq \Omega$.
- For a finite set $S \subseteq \Omega$, the projection of $\mathcal{C}$ onto S is
$$\Pi_{\mathcal{C}}(S) \coloneqq \{h \cap S : h \in \mathcal{C}\}$$

In the old notation, this is the set of possible labelings of S by hypotheses from $\mathcal{C}$

- A set S is shattered by $\mathcal{C}$ if $|\Pi_{\mathcal{C}}(S)| = 2^{|S|}$, i.e., no labeling is ruled out.
- Note that if S is shattered by $\mathcal{C}$, then so is every subset of S.
- The VC dimension of a class $\mathcal{C}$ is the size of the largest set shattered by $\mathcal{C}$:
$$VC(\mathcal{C}) \coloneqq \max\{|S| : S \text{ shattered by } \mathcal{C}\}$$
- Let $\Pi_{\mathcal{C}}(m) \coloneqq \max_{S \subseteq \Omega, |S|=m}\{|\Pi_{\mathcal{C}}(S)|\}$, i.e., the maximum size of a projection of $\mathcal{C}$ for an m-element set.

Sauer Lemma

Sauer Lemma [Vapnik Chervonenkis]

Let $\mathcal{C}$ be a class of Boolean functions and $d = VC(\mathcal{C})<\infty$.

Then $\Pi_{\mathcal{C}}(m) \le \binom{m}{\le d} \le \left(\frac{em}{d}\right)^d = O\left(m^d\right)$

$\binom{m}{\le d}$ is $\sum_{i=0}^{d}\binom{m}{i}$

If $VC(\mathcal{C})=\infty$, then $\Pi_{\mathcal{C}}(m) = 2^m$ for all m

Proof (by a shifting argument): Fix a set S of size m.

- Let $\mathcal{F}=\Pi_{\mathcal{C}}(S)$, i.e., it is a family of subsets of $[m]$.
- W.l.o.g. $m > d$

Otherwise, $\binom{m}{\le d}$ is 2^m, and the lemma holds.

- We will transform family $\mathcal{F}$ into a family $\mathcal{G}$ by using shifting

1. Repeat until no further change is possible:
2. **for** $i = 1$ *to* m
3. **for** *all* $F \in \mathcal{F}$ **do**
4. **if** $F - \{i\} \notin \mathcal{F}$ **then** replace F by $F - \{i\}$
5. Set $\mathcal{G} = \mathcal{F}$ and return $\mathcal{G}$

Proof Sauer Lemma

Sauer Lemma [Vapnik Chervonenkis]

Let $\mathcal{C}$ be a class of Boolean functions and $d = VC(\mathcal{C}) < \infty$.

Then $\Pi_{\mathcal{C}}(m) \leq \binom{m}{\leq d} \leq \left(\frac{em}{d}\right)^d = O\left(m^d\right)$

Proof (continued): Properties of the transformation.

1. $|\mathcal{G}| = |\mathcal{F}|$
2. If $A \subset S$ is shattered by $\mathcal{G}$, then A is shattered by $\mathcal{F}$
3. If $A \in \mathcal{G}$, then so is every subset of A.

1. Repeat until no further change is possible:
2. **for** $i = 1$ *to* m
3. **for** *all* $F \in \mathcal{F}$ **do**
4. **if** $F - \{i\} \notin \mathcal{F}$ **then** replace F by $F - \{i\}$
5. Set $\mathcal{G} = \mathcal{F}$ and return $\mathcal{G}$

$\mathcal{G}$ is closed under taking subsets

Instead of upper bounding $|\mathcal{F}|$, we upper bound $|\mathcal{G}|$.

- Every member of $\mathcal{G}$ is shattered by $\mathcal{G}$, so it is also shattered by $\mathcal{F}$.

By (3)

By (2)

- Thus, every member of $\mathcal{G}$ has size at most d, and

$$|\mathcal{G}| \leq \binom{m}{\leq d}$$

Proof Sauer Lemma

Proof (continued): It remains to verify:

2. If $A \subset S$ is shattered by $\mathcal{G}$, then A is shattered by $\mathcal{F}$

Fix some execution of Steps 3-4 for some specific setting of i.

1. Repeat until no further change is possible:
2. **for** $i = 1$ *to* m
3. **for** *all* $F \in \mathcal{F}$ **do**
4. **if** $F - \{i\} \notin \mathcal{F}$ **then** replace F by $F - \{i\}$
5. Set $\mathcal{G} = \mathcal{F}$ and return $\mathcal{G}$

Let $\mathcal{F}$ and $\mathcal{F}'$ be the family before and after the execution, resp.

- Consider A shattered by $\mathcal{F}'$. We need to show: A is shattered by $\mathcal{F}$.
- W.l.o.g. assume $i \in A$.
- Fix $R \subseteq A$. Need to show: $\exists T \in \mathcal{F}$ such that $A \cap T = R$
- Then $\exists F' \in \mathcal{F}'$ such that $A \cap F' = R$
  - If $i \in R$, then $F' \in \mathcal{F}$.
  - Suppose $i \notin R$. Then $\exists T' \in \mathcal{F}'$ such that $A \cap T' = R \cup \{i\}$. Set $T = T' - \{i\}$ $T \in \mathcal{F}$, since T' wasn't replaced in Step 4, and $A \cap T = R$.

Otherwise, intersections of A with members of the family are not affected by the operation (removal of i from the sets)

A shattered by $\mathcal{F}'$

Sauer Lemma

Sauer Lemma [Vapnik Chervonenkis]

Let $\mathcal{C}$ be a class of Boolean functions and $d = VC(\mathcal{C}) < \infty$.

Then $\Pi_{\mathcal{C}}(m) \leq \binom{m}{\leq d} \leq \left(\frac{em}{d}\right)^d = O\left(m^d\right)$