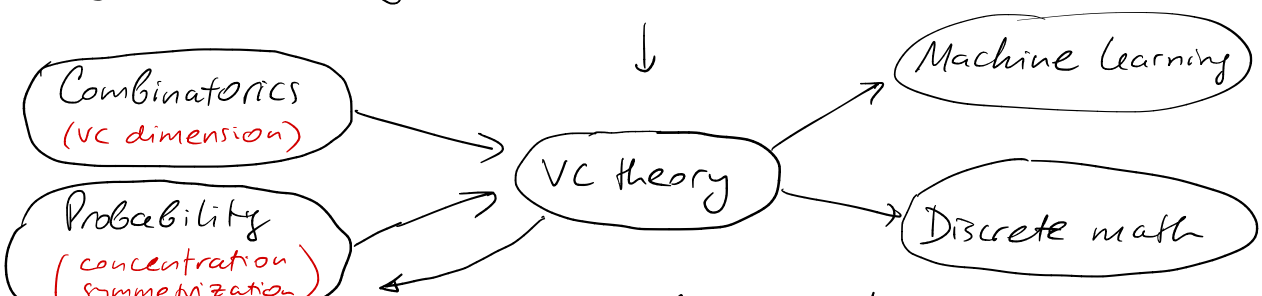
LECTURE 36

• Goal: understand generalization. How much training data is needed?



Combinatorial ingredient:

$\boxed{\text{Lem [Pajor 85]} \ \forall \text{ finite class of Boolean functions } \mathcal{H} \text{ on } \mathcal{X}, \quad |\mathcal{H}| \le \#(\text{subsets of } \mathcal{X} \text{ shattered by } \mathcal{H})}$

Convention: $\emptyset$ is shattered by $\forall$ nonempty $\mathcal{H}$.

Proof WLOG, $\mathcal{X} = \{1, \dots, n\}$. Denote by $sh(\mathcal{H})$ the family of all subsets of $\mathcal{X}$ shattered by $\mathcal{H}$. To prove

$$|\mathcal{H}| \le |sh(\mathcal{H})|,$$

• partition $\mathcal{H}$ according to the value at point n, i.e.

$$\mathcal{H} = \mathcal{H}_0 \sqcup \mathcal{H}_1$$

where $\mathcal{H}_0 = \{h \in \mathcal{H} : h(n) = 0\}$ and $\mathcal{H}_1 = \{h \in \mathcal{H} : h(n) = 1\}$.

• $\forall$ subset $\{i_1, \dots, i_d\} \subset \mathcal{X}$ shattered by $\mathcal{H}_0$ or $\mathcal{H}_1$ is also shattered by $\mathcal{H}$. Thus

$$|sh(\mathcal{H})| \ge |sh(\mathcal{H}_0)| + |sh(\mathcal{H}_1)| \ \ge \qquad (*)$$

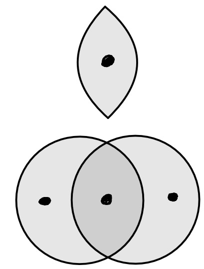
domain $= \{1, \dots, n-1\}$

• Iterate: partition $\mathcal{H}_0$ and $\mathcal{H}_1$ according to the value $h(n-1)$:

$$\ge |sh(\mathcal{H}_{00})| + |sh(\mathcal{H}_{01})| + |sh(\mathcal{H}_{10})| + |sh(\mathcal{H}_{11})| \ \ge$$

... down to single-point classes, each of which shatters one set $\emptyset$ ... $\ge |\mathcal{H}|$

- **MISTAKE**: we double counted in (*) the sets that are shattered by <u>both</u> $\mathcal{H}_0$ and $\mathcal{H}_1$



- **FIX**: suppose $\{i_1, \dots, i_d\}$ is shattered by <u>Both</u> $\mathcal{H}_0$ and $\mathcal{H}_1$ $\Rightarrow$

  $\forall$ label assignment $y_1, \dots, y_d \in \{0,1\}$

  $\exists h \in \mathcal{H}_0$: $\quad h(i_1) = y_{i_1}, \dots, h(i_d) = y_{i_d}, \quad h(n) = 0$

  $\exists g \in \mathcal{H}_1$: $\quad g(i_1) = y_{i_1}, \dots, g(i_d) = y_{i_d}, \quad g(n) = 1$

  $\Rightarrow \underbrace{\{i_1, \dots, i_d, n\}}$ is shattered by $\mathcal{H} = \mathcal{H}_0 \cup \mathcal{H}_1$ (choose either h or g)

  This set is NOT shattered by either $\mathcal{H}_0$ or $\mathcal{H}_1$ $\Rightarrow$ it was <u>NOT</u> counted before

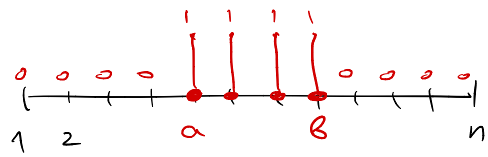
- $\Rightarrow$ $\forall$ set that we <u>double</u> counted, we find a set we <u>never</u> counted

  $\Rightarrow$ (*) is true. Proceed as before. QED

- By def of vc dimension, $\forall$ subset shattered by $\mathcal{H}$ has cardinality $\leq vc(\mathcal{H}) =: d$. So Pajor's lemma yields
$$|\mathcal{H}| \leq \#(\text{subsets of } \{1,\dots,n\} \text{ with card.} \leq d) \leq \sum_{k=0}^{d} \binom{n}{k}.$$
$$\Downarrow$$

Cor (Sauer-Shelah lemma) Let $\mathcal{H}$ be a class of Boolean functions on an n-point domain. Then
$$|\mathcal{H}| \leq \sum_{k=0}^{d} \binom{n}{k} \quad \text{where} \quad d = vc(\mathcal{H})$$

Examples

(a) Integer intervals: $\mathcal{H} = \Big\{$  $: 1 \leq a \leq b \leq n \Big\}$

$vc(\mathcal{H}) = 2$ (as in the previous lecture for real intervals)
$$|\mathcal{H}| = \underbrace{1}_{\text{zero function}} + \underbrace{n}_{a=b} + \underbrace{\binom{n}{2}}_{\#\text{pairs } (a<b)} = \sum_{k=0}^{2} \binom{n}{k} \Rightarrow \text{Pajor lemma is sharp}$$

(b) $\mathcal{H} = \{$all functions on an n-point domain supported by $\leq d$ pts$\}$

$vc(\mathcal{H}) = d$ **(HW?)** and $|\mathcal{H}| = \sum_{k=0}^{d} \binom{n}{k} \Rightarrow$ sharp again!

Remarks ① $\sum_{k=1}^{d} \binom{n}{k} \leq \left(\frac{en}{d}\right)^d$ (HW 3, Problem 3) $\xRightarrow{\text{Pajor}}$
$$d \underset{\text{HW 13}}{\leq} \log|\mathcal{H}| \leq d \log\left(\frac{en}{d}\right), \quad \text{where } d = vc(\mathcal{H}).$$

② Heuristically, $\log|\mathcal{H}| = \#$bits to specify a function in $\mathcal{H}$

$d = vc(\mathcal{F}) \sim \#$parameters that describe functions in $\mathcal{H}$

$\Rightarrow \log|\mathcal{H}| \asymp d$ is expected.