Statistical Foundations of Learning - Sample Problems 3

CIT4230004 (Summer Semester 2024)

Sample Problem 3.1: Growth Function and VC Dimension of Unions

Let $H, H' \subseteq \{\pm 1\}^X$ be two hypothesis classes.

(a) Bound the growth function of $H \cup H'$ in terms of the growth functions of H and H'.

The growth function $\tau_H(m)$ of a hypothesis class H is defined as the maximum number of distinct labelings of any m points that can be achieved by hypotheses in H.

For $H \cup H'$, the number of distinct labelings of any m points is at most the sum of the number of distinct labelings by H and H':

$$\tau_{H \cup H'}(m) \leq \tau_H(m) + \tau_{H'}(m)$$

(b) Derive a bound on the VC dimension of $H \cup H'$.

The VC dimension $\mathrm{VC}(H)$ of a hypothesis class H is the maximum number of points that can be shattered by H.

From the growth function bound, we have:

$$\tau_{H \cup H'}(m) \leq \tau_H(m) + \tau_{H'}(m)$$

Using Sauer's lemma:

$$\tau_H(m) \leq \sum_{i=0}^{\mathrm{VC}(H)} \binom{m}{i}$$

$$\tau_{H'}(m) \leq \sum_{i=0}^{\mathrm{VC}(H')} \binom{m}{i}$$

Therefore:

$$\tau_{H \cup H'}(m) \leq \sum_{i=0}^{\mathrm{VC}(H)} \binom{m}{i} + \sum_{i=0}^{\mathrm{VC}(H')} \binom{m}{i}$$

The VC dimension of $H \cup H'$ is at most the maximum of the VC dimensions of H and H':

$$\mathrm{VC}(H \cup H') \leq \mathrm{VC}(H) + \mathrm{VC}(H')$$

Sample Problem 3.2: Growth Function and VC Dimension for Decision Stumps

We generalize the decision stumps of $\mathbb{R}$ to multi-dimensional decision stumps $H_p \subseteq \{\pm 1\}^{\mathbb{R}^p}$. Let $x(i)$ denote the i-th coordinate of $x \in \mathbb{R}^p$.

1. State the corresponding Hypothesis class.

The hypothesis class H_p consists of all decision stumps on $\mathbb{R}^p$, which are functions of the form:

$$h_{i,t,b}(x) = \begin{cases} b & \text{if } x(i) \leq t, \\ -b & \text{if } x(i) > t. \end{cases}$$

where $i \in \{1, \ldots, p\}$, $t \in \mathbb{R}$, and $b \in \{\pm 1\}$.

2. Compute an upper bound on the growth function of the form $\tau_{H_p}(m) \leq$?

For p dimensions, the number of distinct labelings of any m points by H_p is bounded by:

$$\tau_{H_p}(m) \leq 2mp$$

3. Show that the bound is tight for $p = 1$.

For $p = 1$, the hypothesis class H_1 consists of decision stumps on $\mathbb{R}$. Each point can be labeled either $+1$ or -1 depending on its position relative to the threshold t.

The number of distinct labelings of m points is at most $2m$ because each point can be independently labeled as $+1$ or -1, depending on whether it is above or below the threshold.

Thus, the bound $\tau_{H_1}(m) \leq 2m$ is tight.

4. What does this imply on the VC dimension of H_p?

The VC dimension of H_p can be determined by finding the largest m such that $\tau_{H_p}(m) = 2^m$.

For $p = 1$, we have:

$$\tau_{H_1}(m) \leq 2m$$

The largest m such that $2^m \leq 2m$ is $m = 1$. Thus, the VC dimension of H_1 is 1.

For general p, the VC dimension of H_p is at most p, as each dimension can contribute at most one threshold.

Sample Problem 3.3: Graph dimension of multi-class decision stumps

In this problem, we look at 3-class classification over $\mathbb{R}$ using a generalization of decision stumps. Define a function $h_{t_1,t_2,a,b,c} : \mathbb{R} \to \{0, 1, 2\}$ as:

$$h_{t_1,t_2,a,b,c}(x) = a \cdot 1\{x < t_1\} + b \cdot 1\{t_1 \leq x < t_2\} + c \cdot 1\{x \geq t_2\}$$

parametrized by thresholds $t_1, t_2 \in \mathbb{R}$ and integers $a, b, c \in \{0, 1, 2\}$ such that a, b, c are distinct.

Define the hypothesis class $H_{3ds} \subseteq \{0, 1, 2\}^{\mathbb{R}}$ as the set of all above predictors, $H_{3ds} = \{h_{t_1,t_2,a,b,c} : t_1, t_2 \in \mathbb{R}, a, b, c \in \{0, 1, 2\} \text{ distinct}\}$.

1. Let the growth function be defined as in the lecture, $\tau_H(m) = \max_{C:|C|=m} |H|_C|$. Compute the growth function for the above hypothesis class H_{3ds}.

The growth function $\tau_{H_{3ds}}(m)$ counts the maximum number of distinct labelings of any m points by H_{3ds}.

For m points, each point can be in one of three intervals: $(-\infty, t_1)$, $[t_1, t_2)$, or $[t_2, \infty)$.

The number of distinct labelings is at most:

$$\tau_{H_{3ds}}(m) \leq 3^m$$

**2. Graph dimension is a generalization of VC-dimension for multiclass classifiers and is defined as follows. Let $H \subseteq Y^X$ be a hypothesis class. We say that H G-shatters a finite set $C \subseteq X$ if there exists a function $f : C \to Y$ such that for every $S \subseteq C$, there is a $h \in H$ such that $h(x) = f(x)$ for every $x \in S$ and $h(x) \neq f(x)$ for every $x \in C \setminus S$.

Compute the Graph dimension of the hypothesis class H_{3ds}.**

The Graph dimension Graph-dim(H_{3ds}) is the largest m such that H_{3ds} G-shatters any set of m points.

Given that each point can be assigned one of three labels and the class can shatter any configuration of points, the Graph dimension is:

$$\text{Graph-dim}(H_{3ds}) = 3$$