**Mathematical Tools in Machine Learning** **(11.11.19)**

# Solutions to Exercise Session 3

*Instructor:* Fadoua Balabdaoui *Assistant:* Loris Michel

**Problem (Proof of Sauer's lemma)**

Derived in exercise class. Essentially following closely the proof in the book should be sufficient. Open to suggestions and clarifications.

**Problem 6.1: Monotonicity of VC-dimension**

Let $\mathcal{H}' \subseteq \mathcal{H}$ be two hypothesis classes for binary classification. Since $\mathcal{H}' \subseteq \mathcal{H}$, then for every $C = \{c_1, \dots, c_m\} \subseteq \mathcal{X}$ we have $\mathcal{H}'_C \subseteq \mathcal{H}_C$. In particular, if $C$ is shattered by $\mathcal{H}'$, then $C$ is shattered by $\mathcal{H}$ as well. Thus $\text{VCdim}(\mathcal{H}') \leq \text{VCdim}(\mathcal{H})$.

**Problem 6.4: Sharp bounds**

Let $\mathcal{X} = \mathbb{R}^d$. We will demonstrate all the 4 combinations using the hypothesis classes defined over $\mathcal{X} \times \{0, 1\}$. Remember that the empty set is always considered to be shattered.

$(<, =)$ : Let $d \geq 2$ and consider the class $\mathcal{H} = \{I_{[\|x\|_2 \leq r]} : r \geq 0\}$ of the concentric balls. The VC-dimension of this class is 1. To see this, we first observe that if $x \neq (0, \dots, 0)$, then $\{x\}$ is shattered. Second, if $\|x_1\|_2 = \|x_2\|_2$, then the labeling $y_1 = 0$ and $y_2 = 1$ is not obtained by any hypothesis in $\mathcal{H}$. Let $A = \{e_1, e_2\}$, where $e_1.e_2$ are the first two elements of the standard basis of $\mathbb{R}^d$. Then, $\mathcal{H}_A = \{(0, 0), (1, 1)\}, \{B \subseteq A : \mathcal{H} \text{ shatters } B\} = \{\emptyset, \{e_1\}, \{e_2\}\}$, and $\sum_{i=0}^{d} \binom{|A|}{i} = 3$.

$(=, <)$ : Let $\mathcal{H}$ be the class of axis-aligned rectangles in $\mathbb{R}^2$. We have seen that the VC-dimension of $\mathcal{H}$ is A4. Let $\{x_1, x_2, x_3\}$, where $x_1 = (0, 0)$, $x_2 = (1, 0)$, $x_3 = (2, 0)$. All the labelings expect $(1, 0, 1)$ are obtained. Thus, $|\mathcal{H}_A| = 7$, $|\{B \subseteq A : \mathcal{H} \text{ shatters } B\}| = 7$, and $\sum_{i=0}^{d} \binom{|A|}{i} = 8$.

$(=, =)$ : Let $d \geq 3$, and consider the class $\mathcal{H} = \{\text{sgn}\langle \eta w, x\rangle : w \in \mathbb{R}^d\}$ of homogenous halfspaces. Then $\text{VCdim}(\mathcal{H}) \geq 3$. This fact follows by observing that the set $\{e_1, e_2, e_3\}$ is shattered. Let $A = \{x_1, x_2, x_3\}$, where $x_1 = e_1$, $x_2 = e_2$ and $x_3 = (1, 1, 0, \dots, 0)$. Note that all the labelings expect $(1, 1, -1)$ and $(-1, -1, 1)$ are obtained. It follows that $|\mathcal{H}_A| = 6$, $|\{B \subseteq A : \mathcal{H} \text{ shatters } B\}| = 7$, and $\sum_{i=0}^{d} \binom{|A|}{i} = 8$.

$(=, =)$ : Let $d = 1$, and consider the class $\mathcal{H} = \{I_{[x \geq t]} : t \in \mathbb{R}\}$ of thresholds on the line. We have seen that every singleton is shattered by $\mathcal{H}$, and that every set of size of at least 2 is not shattered by $\mathcal{H}$. Choose any finite set $A \subseteq \mathbb{R}$. Then each of the three terms in "Sauer's inequality" equals $|A| + 1$.

**Problem 6.8: infinite VC-dimension**

We will begin by proving the lemma provided in the question:

$$\begin{aligned}
& \sin(2^m \pi x) = \sin(2^m \pi (0.x_1 x_2, \dots)) \\
= \;& \sin(2\pi(x_1 x_2, \dots, x_{m-1}.x_m x_{m+1} \dots)) \\
= \;& \sin(2\pi(x_1 x_2, \dots, x_{m-1}.x_m x_{m+1} \dots) - 2\pi(x_1 x_2 \dots x_{m-1}.0)) \\
= \;& \sin(2\pi(0.x_m x_{m+1} \dots)).
\end{aligned}$$

Now, if $x_m = 0$, then $2\pi(0.x_m x_{m+1} \dots) \in (0, \pi)$ (we use the fact that $\exists k \geq m$ s.t. $x_k = 1$), so the expression above is positive, and $\lceil \sin(2^m \pi x) \rceil = 1$. On the other hand, if $x_m = 1$, then $2\pi(0.x_m x_{m+1}) \in [\pi, 2\pi)$, so the expression above is non-positive, and $\lceil \sin(2^m \pi x) \rceil = 0$. In conclusion we have that $\lceil \sin(2^m \pi x) \rceil = 1 - x_m$.

To prove $VC(\mathcal{H}) = \infty$, we need to pick $n$ points which are shattered by $\mathcal{H}$, for any $n$. To do so, we construct $n$ points $x_1, \ldots, x_n \in [0,1]$, such that the set of the $m-th$ bits in the binary expansion, as $m$ ranges from 1 to $2^n$, ranges over all possible labelings of $x_1, \ldots, x_n$:

$$\begin{array}{ccccccccc}
\mathrm{x}_1 & = & 0. & 0 & 0 & 0 & 0 & \ldots & 1 & 1 \\
\mathrm{x}_2 & = & 0. & 0 & 0 & 0 & 0 & \ldots & 1 & 1 \\
 & & & & \vdots & & & & & \\
\mathrm{x}_{n-1} & = & 0. & 0 & 0 & 1 & 1 & \ldots & 1 & 1 \\
\mathrm{x}_n & = & 0. & 0 & 1 & 0 & 1 & \ldots & 0 & 1
\end{array}$$

For example, to give the labeling 1 for all instances, we just pick $h(x) = \lceil \sin(2^1 x) \rceil$, which returns the first bit (column) in the binary expansion. If we wish to give labeling 1 for $x_1, \ldots, x_{n-1}$, and the labeling 0 for $x_n$, we pick $h(x) = \lceil \sin(2^2 x) \rceil$, which returns the second bit in the binary expansion, and so on. We conclude that $x_1, \ldots, x_n$ can be given any labeling by some $h \in \mathcal{H}$, so it is shattered. This can be done for all $n$, so VCdim(H) $= \infty$.

**Problem 6.9 (optional): Signed intervals**

We prove that VCdim$(\mathcal{H}) = 3$. Choose $C = \{1, 2, 3\}$, the following table shows that $C$ is shattered by $\mathcal{H}$.

| 1 | 2 | 3 | a | b | s |
|---|---|---|---|---|---|
| - | - | - | 0.5 | 3.5 | -1 |
| - | - | + | 2.5 | 3.5 | 1 |
| - | + | - | 1.5 | 2.5 | 1 |
| - | + | + | 1.5 | 3.5 | 1 |
| + | - | - | 0.5 | 1.5 | 1 |
| + | - | + | 1.5 | 2.5 | -1 |
| + | + | - | 0.5 | 2.5 | 1 |
| + | + | + | 0.5 | 3.5 | 1 |

We conclude that VCdim$(\mathcal{H}) \geq 3$. Let $C = \{x_1, x_2, x_3, x_4\}$ and assume w.l.o.g that $x_1 < x_2 < x_3 < x_4$. Then the labeling $y_1 = y_3 = -1, y_2 = y_4 = 1$ is not obtained by any hypothesis in $\mathcal{H}$. Thus, VCdim$(\mathcal{H}) \leq 3$.