Problem 1. Exercise 5 of Chapter 3

Consider some $h \in \mathcal{H}$ s.t. $L_{(\overline{\mathcal{D}_m}, f)}(h) > \epsilon$. Then it follows by definition of $L_{(\overline{\mathcal{D}_m}, f)}$ (the true loss) that

$$\frac{\sum_{i=1}^m \mathbb{P}_{X \sim \mathcal{D}_i}[h(X) = f(X)]}{m} < 1 - \epsilon.$$

Then we bound the probability that $L_S(h) = 0$ as follows:

$$
\begin{aligned}
\mathbb{P}_{S \sim \prod_{i=1}^m \mathcal{D}_i}[L_S(h) = 0] &= \prod_{i=1}^m \mathbb{P}_{X \sim \mathcal{D}_i}[h(X) = f(X)] \\
&= \left(\left(\prod_{i=1}^m \mathbb{P}_{X \sim \mathcal{D}_i}[h(X) = f(X)] \right)^{1/m} \right)^m \\
&\leq \left(\frac{\sum_{i=1}^m \mathbb{P}_{X \sim \mathcal{D}_i}[h(X) = f(X)]}{m} \right)^m \\
&\leq (1 - \epsilon)^m \\
&\leq e^{-\epsilon m}.
\end{aligned}
$$

The first inequality above is the geometric-arithmetic mean inequality. It remains to apply the union bound to conclude that the probability that there exists $h \in \mathcal{H}$ with true loss greater that ϵ and zero loss for the observed samples is upper bounded by $|\mathcal{H}| e^{-\epsilon m}$.

Problem 2. Exercise 2 of Chapter 5

Denote by $\mathcal{H}_d$ the class of axis-aligned rectangles in $\mathbb{R}^d$. Since $\mathcal{H}_2 \subseteq \mathcal{H}_5$, the latter class is more representative and in the presence of sufficiently many samples is likely to have smaller true loss. However, the growth rate (and the VC dimension) of the former is smaller, which gives us better bounds on the approximation error, which makes it more preferable in case of small number of samples.

Problem 3. Exercise 3 of Chapter 5 We modify the proof of NFL theorem in the book. Consider subset $C \subset \mathcal{X}$ of size km. There are $T = 2^{km}$ functions from C to $\{0, 1\}$, $f_1, \ldots, f_T$. Let D_i's be probability distribution defined as in the proof of NFL.

Fix a training set $S_j = \{x_1, \ldots, x_m\}$ and let $v_1, \ldots, v_p$ be the examples in C that do not appear in S_j. Clearly, we have $p \geq (k-1)m$. Therefore, for every $h : C \to \{0, 1\}$ and $i \in [T]$,

1

we have (Eq. 5.5 in the book)

$$L_{D_i}(h) = \frac{1}{km}\sum_{x\in C}\mathbb{1}_{[h(x)\neq f_i(x)]}$$

$$\geq \frac{1}{km}\sum_{r=1}^{p}\mathbb{1}_{[h(v_r)\neq f_i(v_r)]} \tag{1}$$

$$\geq \frac{k-1}{kp}\sum_{r=1}^{p}\mathbb{1}_{[h(v_r)\neq f_i(v_r)]}$$

Then, the right-hand side of Eq. 5.6 becomes

$$\frac{k-1}{k}\min_{r\in[p]}\frac{1}{T}\sum_{i=1}^{T}\mathbb{1}_{[A(S_j^i)(v_r)\neq f_i(v_r)]}$$

From which we can conclude (similar to the proof of NFL) that

$$\mathbb{E}\Big[L_D\big(A(S)\big)\Big] \geq \frac{k-1}{2k}$$

Problem 4. Exercise 2 of Chapter 6

1. We show that $\text{VCdim}(\mathcal{H}_{=k}) = \min(k, |\mathcal{X}| - k)$.

Consider a set of $k+1$ elements. The all-one labeling cannot be obtained, hence $\text{VCdim}(\mathcal{H}) \leq k$. Analogously, for a set of $|\mathcal{X}| - k + 1$ elements, the all-zero labeling cannot be obtained so $\text{VCdim}(\mathcal{H}_{=k}) \leq \min(k, |\mathcal{X}| - k)$.

Take a set C of size $m = \min(k, |\mathcal{X}| - k)$ and a labeling $(y_1, \ldots, y_m)$ with s ones, $0 \leq s \leq m$. We can pick a hypothesis $h \in \mathcal{H}_{=k}$ such that $h(x_i) = y_i$ for all $x_i \in C$ and it has $k - s$ ones on the set $\mathcal{X} \setminus C$ (this is possible because $|\mathcal{X} \setminus C| \geq k \geq k - s$). Therefore, C is shattered and $\text{VCdim}(\mathcal{H}_{=k}) \geq \min(k, |\mathcal{X}| - k)$.

2. We show that $\text{VCdim}(\mathcal{H}_{at-most-k}) = \min(2k + 1, |\mathcal{X}|)$.

Consider a set of $2k + 2$ elements. It is clear that any labeling with $k + 1$ ones and $k + 1$ zeros cannot be obtained, so $\text{VCdim}(\mathcal{H}_{at-most-k}) \leq 2k + 1$. Note that it may happen that $2k + 1 > |\mathcal{X}|$, so the bound should be $\text{VCdim}(\mathcal{H}_{at-most-k}) \leq \min(2k + 1, |\mathcal{X}|)$.

Take a set of $\min(2k + 1, |\mathcal{X}|)$ elements. Any labeling on this set has either $\leq k$ zeros or $\leq k$ ones, so it is shattered by $\mathcal{H}_{at-most-k}$. Thus, $\text{VCdim}(\mathcal{H}_{at-most-k}) \geq \min(2k + 1, |\mathcal{X}|)$.

Problem 5. Exercise 8 of Chapter 6

Let's first prove the lemma. Let $m \geq 1$ and $0.x_1x_2\ldots$ the binary expansion of $x \in (0, 1)$. Assume that there exists $k \geq m$ such that $x_k = 1$. We have:

$$\sin(2^m\pi x) = \sin(2^m\pi \cdot (0.x_1x_2\ldots)) = \sin(2\pi \cdot (x_1x_2\ldots x_{m-1}.x_mx_{m+1}\ldots))$$
$$= \sin(2\pi \cdot (0.x_mx_{m+1}\ldots)) .$$

If $x_m = 0$ then $\exists k > m$ s.t. $x_k = 1$, i.e., the number $0.0x_{m+1}\ldots$ is nonzero. This means that $2\pi \cdot (0.0x_{m+1}\ldots) \in (0, \pi)$ where $\sin(x)$ is positive, which gives the label 1. If $x_m = 1$ then $2\pi \cdot (0.1x_{m+1}\ldots) \in [\pi, 2\pi)$ where $\sin(x)$ is non positive, which gives the label 0. This ends the proof of the lemma.

To prove that $\mathcal{H}$ has infinite VC-dimension, we need to show that for any n there is a set of n points in $\mathbb{R}$ on which we can obtain all 2^n possible labelings. Let $(x_1, \ldots, x_n) \in (0,1)^n$ such that the first 2^n bits of their binary expansions give all possible labelings and their $(2^n)^{\text{th}}$ bit is always one (i.e., the $(2^n)^{\text{th}}$ bits form the all-one labeling).

Example for $n = 3$:

$$
\begin{array}{llccccccccc}
x_1 & 0. & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
x_2 & 0. & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
x_3 & 0. & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{array}
$$

Using the lemma, invoking the function $\lceil \sin(2^i \pi x) \rceil$ on the set $\{x_1, \ldots, x_n\}$ for $1 \leq i \leq 2^n$ allows to obtain all possible labelings. Hence, $\mathcal{H}$ shatters the set $\{x_1, \ldots, x_n\}$.

Problem 6. VC dimension of circles
We will show that the VC dimension of $\mathcal{H}$ is $d = 3$.

1. Take three points in $\mathbb{R}^2$ located at the corners of an equilateral triangle. It is then clear that a circle can select any single one of these points, but also any pair of points and of course all three points together.

2. We show that $\mathcal{H}$ cannot shatter any set of 4 points. Consider 4 points A, B, C and D.

First, assume that one of the points is in the convex hull of the other 3 points. It is then impossible to label these 3 points with 1 while labeling the point in the convex hull with 0.

If no point is in the convex hull of the three other points then the 4 points form a convex quadrilateral $ABCD$. The line segment $[AC]$ is the first diagonal, and $[BD]$ is the second one. The two diagonals $[AC]$ and $[BD]$ must intersect each other.

We now claim that it is impossible to have circles such that the corresponding functions implement both $(0,1,0,1)$ and $(1,0,1,0)$. This is true since it is impossible to have two circles C_1 and C_2 such that

- C_1 contains only A and C, C_2 contains only B and D, and

- $[AC]$ cuts $[BD]$.

If such C_1 and C_2 existed, it would imply that $(C_1 \cup C_2) \setminus (C_1 \cap C_2)$ has 4 disjoint parts.

Below, we propose another way to show that the 4 points forming the convex quadrilateral $ABCD$ cannot be shattered by $\mathcal{H}$. This is a proof by contradiction. Suppose that $\mathcal{H}$ shatters the four points. The sum of the four interior angles is $360°$. Without loss of generality, we have $\angle ABC + \angle CDA \geq 180°$. Because $\mathcal{H}$ shatters the four points, there is a circle $\mathcal{C}$ that contains A, C but not B, D. Let B', D' be the intersections of the line (BD) with $\mathcal{C}$, and let A', C' be the intersections of the line (AC) with $\mathcal{C}$. Clearly $[B'D'] \subset [BD]$ and $[AC] \subseteq [A'C']$. Note that

$$\angle A'B'C' + \angle C'D'A' > \angle ABC + \angle CDA .$$

Besides, the quadrilateral $A'B'C'D'$ is inscribed in the circle $\mathcal{C}$ so:

$$\angle A'B'C' + \angle C'D'A' = \angle B'C'D' + \angle D'A'B' = 180° .$$

Hence the contradiction: $180° = \angle A'B'C' + \angle C'D'A' > 180°$.