Solution of Graded Homework 1: 19 March 2019
CS-526 Learning Theory

5.1 We simply apply lemma from the hint to obtain

$$
\begin{aligned}
\mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) \geq 1/8) &= \mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) \geq 1 - 7/8) \\
&\geq \frac{\mathbb{E}[L_{\mathcal{D}}(A(S))] - (1-7/8)}{7/8} \\
&\geq \frac{1/8}{7/8} = 1/7.
\end{aligned}
$$

Alternatively, if you dislike Lemma B.1, you can also prove by contrapositive, i.e., showing that if $\mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) \geq 1/8) < 1/7$ then $\mathbb{E}[L_{\mathcal{D}}(A(S)] < 1/4$. This is easily seen because

$$L_{\mathcal{D}}(A(S)) < 1 \cdot \mathbb{1}_{L_{\mathcal{D}}(A(S)) \geq 1/8} + \frac{1}{8} \cdot \mathbb{1}_{L_{\mathcal{D}}(A(S)) < 1/8}$$

and under the hypothesis

$$\mathbb{E}[L_{\mathcal{D}}(A(S))] < 1 \cdot \frac{1}{7} + \frac{1}{8} \cdot \frac{6}{7} = 1/4.$$

6.2 (a) Consider a set of $k{+}1$ elements. All-one labeling cannot be obtained, so $\text{VCdim}(\mathcal{H}) \leq k$. Analogously, for a set of $|\mathcal{X}| - k + 1$ elements 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 at the set $\mathcal{X} \setminus C$. Therefore, C is shattered and $\text{VCdim}(\mathcal{H}_{=k}) \geq \min(k, |\mathcal{X}| - k)$.

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

6.5 We simply generalize the proof from the two-dimensional case. Let's first formally state the hypothesis class

$$\mathcal{H} = \{h_{(a_i,b_i)} | a_i \leq b_i, h_{(a_i,b_i)}(x_1, \ldots, x_d) = \prod_{i=1}^{d} \mathbb{1}_{a_i \leq x_i \leq b_i}\}$$

Consider set $\{\mathbf{x}_1, \dots, \mathbf{x}_{2d}\}$, where $\mathbf{x}_i = \mathbf{e}_i$ for $1 \le i \le d$ and $\mathbf{x}_i = -\mathbf{e}_{i-d}$ for $d+1 \le i \le 2d$. For any labeling $(y_1, \dots, y_{2d})$, pick $a_i = -2$ if $y_{d+i} = 1$ and $a_i = -0.5$ otherwise. Similarly, pick $b_i = 2$ if $y_i = 1$ and $b_i = 0.5$ otherwise. Then $h_{(a_i,b_i)}(\mathbf{x}_i) = y_i$ and hence $\text{VCdim}(\mathcal{H}) \ge 2d$.

For a set C of size $2d+1$, by the pigeonhole principle there exists an element $\mathbf{x}$ s.t. $\forall j \in [d]$ there exist $\mathbf{x}', \mathbf{x}'' \in C : x'_j \le x_j \le x''_j$. This means that labeling with only $\mathbf{x}$ negative and all other elements positive cannot be obtained and therefore $\text{VCdim}(\mathcal{H}) \le 2d$.

6.8 Let's prove the lemma first.

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

For $x_m = 0$, we know that $\exists k \ge m$ s.t. $x_k = 1$, i.e. the number $0.0x_{m+1} \dots$ is nonzero. This means that $2\pi \cdot (0.0x_{m+1} \dots) \in (0, \pi)$, where $\sin(x)$ is positive, which gives the label 1. For $x_m = 1$, we get $2\pi \cdot (0.1x_{m+1} \dots) \in (\pi, 2\pi)$, where $\sin(x)$ is negative, which gives the label 0. Proof completed.

To prove that $\mathcal{H}$ has infinite VC-dimension, we need to show that for any n there is a set x of n points in $\mathbb{R}$ on which we can obtain all 2^n possible labelings. Consider $x_1, \dots, x_n \in [0,1]$ so that first 2^n bits of their binary expansions give all possible labelings.

<u>Example for $n = 3$:</u>

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	...

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

6.9 $\text{VCdim}(\mathcal{H}) = 3$. In order to prove it, let's recall the unsigned intervals class $\mathcal{H}_+$, which was studied during the class. It can be seen that if labeling $(y_0, y_1, \dots)$ is obtained by $h_{a,b} \in \mathcal{H}_+$, then $h_{a,b,+} \in \mathcal{H}$ gives the same labeling and $h_{a,b,-} \in \mathcal{H}$ gives its inverse $(1-y_0, 1-y_1, \dots)$. Labeling $(0,1,0)$ can be obtained by an interval, so signed intervals can label $(1,0,1)$ and therefore $\text{VCdim}(\mathcal{H}) \ge 3$.

Consider the set of 4 points. Labels $(0,1,0,1)$ and $(1,0,1,0)$ cannot be obtained with any signed interval, so $\text{VCdim}(\mathcal{H}) \le 3$, which concludes the proof.

7.3 (a) For any $h \in \mathcal{H}$ and given $n(h), |\mathcal{H}_{n(h)}|$, we can set $w(h) = \frac{2^{-n(h)}}{|\mathcal{H}_{n(h)}|}$. This gives

$$\sum_{h \in \mathcal{H}} w(h) = \sum_{h \in \mathcal{H}} \frac{2^{-n(h)}}{|\mathcal{H}_{n(h)}|} = \sum_{n \in \mathbb{N}} \frac{2^{-n}}{|\mathcal{H}_n|} \sum_{\substack{h \in \mathcal{H}_n \\ h \notin \mathcal{H}_{n'}, n' < n}} \mathbf{1} \le \sum_{n \in \mathbb{N}} \frac{2^{-n}}{|\mathcal{H}_n|} \sum_{h \in \mathcal{H}_n} \mathbf{1} = \sum_{n \in \mathbb{N}} 2^{-n} = 1.$$

The equality is achieved when all $\mathcal{H}_n$ are disjoint

(b) Since $\mathcal{H}_n$ is countable, we can enumerate all $h \in \mathcal{H}_n$ as $h_{n,1}, h_{n,2}, \ldots$. Consider $w(h_{n,k}) = 2^{-n}2^{-k}$. Similarly to the previous exercise, we get

$$\sum_{h \in \mathcal{H}} w(h) \leq \sum_{n \in \mathbb{N}} 2^{-n} \sum_{k \in \mathbb{N}} 2^{-k} = 1.$$

It should be noted that for some $\mathcal{H}_n$ hypotheses $h_{n,k}$ may not exist for sufficiently big k (e.g. $\mathcal{H}_n$ is finite), but we are only interested in upper bound, so it does not change anything.