A. Bounds on VC dimension

[30 points]

1. [15 points] Show that the VC dimension of the set of all closed balls in \( \mathbb{R}^n \), that is sets of the form \( \{ x \in \mathbb{R}^n : \|x - x_0\|^2 \leq r \} \) for some \( x_0 \in \mathbb{R}^n \) and \( r \geq 0 \) is less than or equal to \( n + 2 \).

Let \( B(a, r) \) be the ball of radius \( r \) centered at \( a \in \mathbb{R}^n \). Then \( x \in B(a, r) \) iff

\[
\sum_{i=1}^{n} \|x_i\|^2 - 2 \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} a_i^2 - r \leq 0, \tag{1}
\]

which is equivalent to

\[
\langle W, X \rangle + B \leq 0, \tag{2}
\]

with

\[
W = \begin{bmatrix}
1 \\
-2a_1 \\
\vdots \\
-2a_n
\end{bmatrix},
X = \begin{bmatrix}
\sum_{i=1}^{n} \|x_i\|^2 \\
x_1 \\
\vdots \\
x_n
\end{bmatrix}, \text{ and } B = \sum_{i=1}^{n} a_i^2 - r.
\]

The VC dimension of closed balls in \( \mathbb{R}^n \) is thus at most equal to the VC dimension of hyperplanes in \( \mathbb{R}^{n+1} \), that is \( n + 2 \).

2. [15 points] Determine the VC dimension of the subsets of the real line formed by the union of \( k \) intervals.

It is not hard to see that any \( 2k \) distinct points on the real line can be shattered using \( k \) intervals: it suffices to shatter each of the \( k \) pairs of consecutive points with an interval. Assume now that \( 2k + 1 \) distinct points \( x_1 < \cdots < x_{2k+1} \) are given. For any \( i \in [1, 2k+1] \), label \( x_i \) with \((-1)^{i+1}\), that is alternatively label points with 1 or \(-1\). This leads to \( k + 1 \) points labeled positively and requires \( 2k + 1 \) intervals to shatter the set since no interval can contain two consecutive points. Thus, no set of \( 2k + 1 \) points can be shattered by \( k \) intervals and the VC dimension of the union of \( k \) intervals is \( 2k \).
B. VC dimension of intersection concepts

[30 points]

1. [15 points] Let $C_1$ and $C_2$ be two concept classes. Show that for any concept class $C = \{c_1 \cap c_2 : c_1 \in C_1, c_2 \in C_2\}$, 
\[ \Pi_C(m) \leq \Pi_{C_1}(m) \Pi_{C_2}(m). \] (3)

Fix a set $X$ of $m$ points. Let $\{Y_1, \ldots, Y_k\}$ be the set of intersections of the concepts of $C_1$ with $X$. By definition of $\Pi_{C_1}(X)$, $k \leq \Pi_{C_1}(m)$. By definition of $\Pi_{C_2}(Y_i)$, the intersection of the concepts of $C_2$ with $Y_i$ are at most $\Pi_{C_2}(Y_i) \leq \Pi_{C_2}(m)$. Thus, the number of sets intersections of concepts of $C$ with $X$ is at most 
\[ k \Pi_{C_2}(Y_i) \leq \Pi_{C_1}(m) \Pi_{C_2}(m). \] (4)

2. [15 points] Let $C$ be a concept class with VC dimension $d$ and let $C_s$ be the concept class formed by all intersections of $s$ concepts from $C$, $s \geq 1$. Show that the VC dimension of $C_s$ is bounded by $2ds \log_2(3s)$ (Hint: show that $\log_2(3x) < 9x/(2e)$ for any $x \geq 2$).

In view of the result proved in the previous question, $\Pi_{C_1}(m) \leq (\Pi_{C_1}(m))^s$. By Sauer’s lemma, this implies 
\[ \Pi_{C_s}(m) \leq \left(\frac{em}{d}\right)^{sd}. \] (5)

If $(\frac{em}{d})^{sd} < 2^m$, then the VC dimension of $C_s$ is less than $m$. Thus, it suffices to show this inequality holds with $m = 2ds \log_2(3s)$. Plugging in that value for $m$ and taking the $\log_2$ yield:
\[ ds \log_2 (2es \log_2 (3s)) < 2ds \log_2 (3s) \] (6)
\[ \Leftrightarrow \log_2 (2es \log_2 (3s)) < 2\log_2 (3s) = \log_2 (9s^2) \] (7)
\[ \Leftrightarrow 2es \log_2 (3s) < 9s^2 \] (8)
\[ \Leftrightarrow \log_2 (3s) < \frac{9s}{2e} \] (9)

This last inequality holds for $s = 2$: $\log_2 (6) \approx 2.6 < 9/(2e) \approx 3.3$. Since the functions corresponding to the left-hand-side grows more slowly than the one corresponding to the right-hand-side (compare derivatives for example), this implies that the inequality holds for all $s \geq 2$. 

2
C. Infinite VC dimension

[65 points]

1. [15 points] Show that if a concept class $C$ has infinite VC dimension, then it is not PAC-learnable.

By a theorem of (Ehrenfeucht et al., 1988) presented in class, any algorithm for PAC-learning a concept class $C$ of VC dimension $d$ with accuracy $1 - \epsilon$ and confidence $1 - \delta$ must use a sample size $m = \Omega(\frac{1}{\epsilon} \log \frac{1}{\delta} + \frac{d}{\epsilon})$. For an infinite VC-dimension, this lower bound requires an infinite number of points and thus implies that PAC-learning is not possible.

2. [50 points] In the standard PAC-learning scenario, the learning algorithm receives all examples first and then computes its hypothesis. Within that setting, PAC-learning of concept classes with infinite dimension is not possible as seen in the previous question.

Imagine now a different scenario where the learning algorithm can alternate between drawing more examples and computation. The objective of this problem is to prove that PAC-learning can then be possible for some concept classes with infinite VC dimension.

To do so, consider for example the special case of the concept class $C$ of all subsets of natural numbers.

(a) [0 point] Show that the VC dimension of $C$ is infinite. This is rather straightforward.

(b) [50 points] Professor Vitres has an idea for the first stage of a learning algorithm $L$ PAC-learning $C$. In the first stage, $L$ draws a sufficient number of points $m$ such that the probability of drawing a point beyond the maximum value $M$ observed be small, with high confidence. Can you complete Professor Vitres’ idea by describing the second stage of the algorithm so that it PAC-learns $C$? The description should be augmented with the proof that $L$ can PAC-learn $C$.

Here is a description of the algorithm. Let $M$ be the maximum value observed after drawing $m$ points and let $p$ be the probability that a point greater than $M$ be drawn. The probability that all points drawn be smaller than or equal to $M$ is

$$(1 - p)^m \leq e^{-pm}. \quad (10)$$
Setting $\delta/2$ to match the upper bound, yields $\delta/2 = e^{-pm}$, that is
\[ p = \frac{1}{m} \log \frac{2}{\delta}. \]
(11)

To bound $p$ by $\epsilon/2$, we can impose the following
\[ \frac{1}{m} \log \frac{2}{\delta} \leq \frac{\epsilon}{2}. \]
(12)

Thus, with confidence at least $1 - \delta/2$, the probability that a point greater than $M$ be drawn is at most $\epsilon/2$ if $L$ draws $m \geq \frac{2}{\epsilon} \log \frac{2}{\delta}$ points.

In the second stage, the problem is reduced to a finite VC dimension $M$. Since PAC-learning with $(\epsilon/2, \delta/2)$ is possible for a finite dimension, this guarantees the $(\epsilon, \delta)$-PAC-learning of the full algorithm.