Learning from Data
Homework # 3

Khoi Pham

Generalization Error

The modified Hoeffding Inequality provides a way to characterize the generalization error with a probabilistic bound

$$\mathbb{P}[|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 2Me^{-2\epsilon^2 N}$$

We want the probabilistic bound to be at most 0.03

$$\begin{aligned} 2Me^{-2\epsilon^2 N} &\leq 0.03 \\ \ln M + \ln e^{-2\epsilon^2 N} &\leq \ln 0.015 \\ \frac{1}{2}\ln\frac{M}{0.015} &\leq N \end{aligned}$$

We can now answer questions 1-3.

1. $\epsilon = 0.05$, $M = 1 \implies N \geq 840$

Answer: **[b]**

2. $\epsilon = 0.05$, $M = 10 \implies N \geq 1300$

Answer: **[c]**

3. $\epsilon = 0.05$, $M = 100 \implies N \geq 1761$

Answer: **[d]**

Break Point

3. To solve this problem, we will find the formula for the break point of a perceptron with $d+1$ parameters (including w_0). This perceptron is used to classify data points in the input space $\mathcal{X} = \{1\} \times \mathbb{R}^d$ (these data points contain the constant coordinate $x_0 = 1$). We observe that, given $d+1$, we can construct a nonsingular $(d+1) \times (d+1)$ matrix M, in which each row is used to represent each point in our dataset. Thus, with this nonsingular matrix M and a desired dichotomy, we can construct a $(d+1)$-dimensional weight vector w such that Mw produces the wanted dichotomy. This weight

vector w is our perceptron with $d+1$ parameters.
It is obvious that we cannot construct a set of more than $d+1$ points in the $(d+1)$-dimensional space such that our perceptron can shatter, because at least one of those points must be linearly dependent on the others. Thus, $d+1$ is the maximum of points that a perceptron with $d+1$ parameters can shatter, and $d+2$ is its break point. So the smallest break point for the perceptron model in $\mathbb{R}^3$ is 5.

Answer: [b]

Growth Function

5. (i), (ii), (v) are possible formulas for a growth function $m_{\mathcal{H}}(N)$. This is because a growth function can either be bounded by a polynomial, or be exactly equal to 2^N. It can be seen that (iii) and (iv) are both not 2^N and not bounded by a polynomial.

Answer: [b]

Fun with Intervals

Consider the "2-intervals" learning model, where $h : \mathbb{R} \to \{-1, +1\}$ and $h(x) = +1$ if the point is within either of two arbitrarily chosen intervals and -1 otherwise. To answer question 6 and 7, we will find the growth function for this learning model.
As we have 2 intervals, the dichotomies of N data points can be broken down into 2 cases: the first case is that there exists 2 "separate" contiguous sequences of $+1$ ("separate" means these 2 sequences are separated by one or more -1 in between), and the second case is that there exists only 1 contiguous sequence of $+1$.
For the first case, there are $\binom{n+1}{4}$ ways to choose 2 separate intervals. For the second case, there are $\binom{n+1}{2}$ ways. We also count the case where there exists no $+1$ at all. Thus, the growth function of this "2-intervals" learning model is: $m_{\mathcal{H}}(N) = \binom{n+1}{4} + \binom{n+1}{2} + 1$.

6. The minimum break point is 5 because we are unable to produce the dichotomy $\{+1, -1, +1, -1, +1\}$ (there are 3 intervals of $+1$ but we are only allowed to construct 2 of them).

Answer: [c]

7. Answer: [c]

8. We apply the similar explanation in question 6: for the "M-intervals" learning model, we are only able to construct M separate intervals of $+1$. Thus, we cannot construct the dichotomy with $M+1$ intervals of $+1$. The minimum number of data points such that we can construct a dichotomy with $M+1$ intervals of $+1$ is $2M+1$ (let $+1$ be at the odd positions in our list of data points).

Answer: [d]

Convex Sets: The Triangle

9. Answer: [d]

Non-Convex Sets: Concentric Circles

10. Compute the growth function $m_{\mathcal{H}}(N)$ for the learning model made up of two concentric circles in $\mathbb{R}^2$. Specifically, $\mathcal{H}$ contains the functions which are $+1$ for

$$a^2 \leq x_1^2 + x_2^2 \leq b^2$$

and -1 otherwise. We observe that, by mapping the data points from 2-dimensional to 1-dimensional space $< x_1, x_2 > \Longrightarrow < x_1^2 + x_2^2 >$, the learning model becomes similar to the "1-interval" learning model, in which a data point is assigned with $+1$ when its value is within the specified range and -1 otherwise. There are $\binom{n+1}{2} + 1$ dichotomies in the "1-interval" learning model, and so this is also the answer for this problem.

Answer: [b]