HW 3

CS156, Kai Chang

Question 1

Answer Choice: **B**, 1000
Reasoning: Given the generalization error

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

and our values $\epsilon = 0.05, M = 1$ and we want to achieve a bound of 0.03, we can then solve for what N should be by algebraic manipulation.

So, $0.03 = 2e^{-2*0.05^2*N}$ and by plugging into Mathematica yields $N = 839.941$, and we then need the minimum choice bigger than this value to yield a bound of 0.03 (or at least get close to it). Anything smaller than this N yields to a bound greater than 0.03.

Note what I used for Mathematica was *Solve[0.03 == 2 Exp[-2*0.05^2*N], N]*

Question 2

Answer Choice: **C**, 1500
Reasoning: Plugging into Mathematica for the $M = 10$ case, we get $N = 1300.46$

Question 3

Answer Choice: **D**, 2000
Reasoning: Plugging into Mathematica for the $M = 100$ case, we get $N = 1760.98$.

Question 4

Answer Choice: **B**, 5
Reasoning: A break point is just the minimum number of points needed to be able to fail our classification model. In a Perceptron Model (more specifically a PLA), the decision boundary is linear and the decision are binary. So, in our 2D model, the classification boundary is a line. Thus, with four points, it is possible for us to fail our model (no matter how good it is). For a 3D model, our classification boundary is a plane. Thus, the number of points it takes to fail is 5.

Note that as we increase dimensions, you see a pattern ongoing with classification and points. In a linear case, they are entirely tied to each other (VC + 1).

Look at drawing.

Also, in the case we have 3 points in a 2D case (ie line classification), you may think this is the break point. At first, I was confused and thought this was the break point. However, I realized you can have the line be part of the classification! Thus, with 3 points, you will have at most 3 classification intervals or decision bounds (or shatters).

Question 5

Answer Choice: **B**, i, ii, v
Reasoning: Look at growth formula (it fits a number of conditions, but most importantly $m_H(N-1) \leq m_H(N)$. This means that if there is a break point k, then the growth function is bounded by N^{k-1}. If there is no break point, the growth function is 2^N.

- i) is a growth formula from slides (also $< 2^N$)
- ii) dont be fooled by that binomial, first section is positive interval + second section is positive ray
- iii) not classified in any of the 3 growth functions (ie. not 2^N and not polynomial)
- iv) breaks at k=1. we get $2^0 = 1 \leq 2^1 = 2$, if $N = 2$, then $2^1 = 2 > 2^{1-1} = 1$.
- v) is a growth function!

Question 6

Answer Choice: **C**, 5
Reasoning: This is almost similar to question 4, except now we are mapping or transforming on a new axis scale. Now we have 2 division boundaries but these division boundaries can be classified as subsets of essentially a singular boundary (ie. analogous to a R^3), so it takes 5 points to break this hypothesis.

One may wonder why (unlike question 4), we dont have 3 division boundaries for each one? Well, this is for a number of reasons. 1) the two intervals can share a boundary in between, so it does not double count, and 2) the intervals are inclusive (so unlike a line, which itself can be an interval), we physical have an established interval (with either physical bounds [the line in our 2D case] be either included in the interval or included in the outside).

Question 7

Answer Choice: **C**, $\binom{N+1}{4} + \binom{N+1}{2} + 1$
Reasoning: We have 3 possibilities:

- 2 intervals at different locations and thus 4 decision bounds (boundary judging + or -), $\binom{N+1}{4}$
- 2 intervals overlapping or touching, leaving 2 decision bounds, $\binom{N+1}{2}$
- either the 2 intervals overlapping and spanning the entire space that the points can be in or spanning none of the entire space, 1

Thus, summing those possibilities up yield our growth function: $\binom{N+1}{4} + \binom{N+1}{2} + 1$

Note we place interval ends in n of N+1 spots, giving us a N+1 reasoning on the top of the binomial, and the n on the bottom of the binomial. For more information on these positive intervals and binomials, check out slide 12 of Yaser's CS156 Lecture slides #5.

Question 8

Answer Choice: **D**, $2M + 1$
Reasoning: Now if we consider the M-interval model, we see that we can have at most M decision bounds, meaning we can have up to $2M + 1$ areas if the areas within the intervals were $+1$ and the outside were -1. However, in the case that we have the opposite in our hypothesis dataset, then this would break this learning model. It becomes impossible to classify correctly.

Note if you have $2M$, you may think this is the smallest break point, but actually one can clearly see that a shift in our intevals (remember, it's just a general case) quickly solves our problem, making this possible to shatter.

Question 9

Answer Choice: **D**, 7
Reasoning: Now we deal with convext sets, the third in our possible growth functions. We can see with a triangle, there are essentially 3 boundary lines. However, because we are in a plane, I decided to map the intevals in a radial form (see paper) similar to that on slide 13 of Yaser's lecture slides 5.

We can see that it is possible to get 7 points in a triangle, as you have a very fat or very tall triangle. However, with 8 points, it becomes impossible to do such classifications (ie. our decision bounds have a $(2m + 1) * 2$ max boundary. Thus, we can only shatter up to 7. You will find this relation works for all polygons (try square, pentagon).

Question 10

Answer Choice: **B**, $\binom{N+1}{2} + 1$
Reasoning: If we consider N points on R^2, there are N distances from (0,0) on those points. This question can be seen as selected 2 intervals from $N + 1$ intervals, adding the situation that all distances are -1 not in a, b.