P.S. 3

Sunday, October 15, 2023 11:05 PM

1) $P\left[\,|E_{in}(g) - E_{out}(g)| > \varepsilon\,\right] \leq 2Me^{-2\varepsilon^2 N} \leq \ell$ ← error

$$2Me^{-2\varepsilon^2 N} \leq \ell$$

$$\frac{2M}{e^{2\varepsilon^2 N}} \leq \ell$$

$$2M \leq e^{2\varepsilon^2 N}\ell$$

$$N \geq \frac{\ln(2M/\ell)}{2\varepsilon^2}$$

$$N \geq \frac{\ln\left(\frac{2}{0.03} \cdot M\right)}{2(0.05)^2}$$

$$N \geq \frac{\ln\left(\frac{2}{0.03} \cdot 1\right)}{2(0.05)^2}$$

$$N \geq 839.941$$

The least number of samples among the given choices is 1000, so $\boxed{\text{the answer is b.}}$

2)

$$N \geq \frac{\ln\left(\frac{2}{0.03} \cdot M\right)}{2(0.05)^2}$$

$$N \geq \frac{\cdots}{2(0.05)^2}$$

$$N \geq \frac{\ln\left(\frac{2}{0.03} \cdot 10\right)}{2(0.05)^2}$$

$$N \geq 1300.45$$

The least number of samples among the given choices is 1500, so $\boxed{\text{the answer is c.}}$

3)

$$N \geq \frac{\ln\left(\frac{2}{0.03} \cdot M\right)}{2(0.05)^2}$$

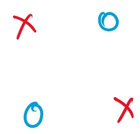
$$N \geq \frac{\ln\left(\frac{2}{0.03} \cdot 100\right)}{2(0.05)^2}$$

$$N \geq 1760.975$$

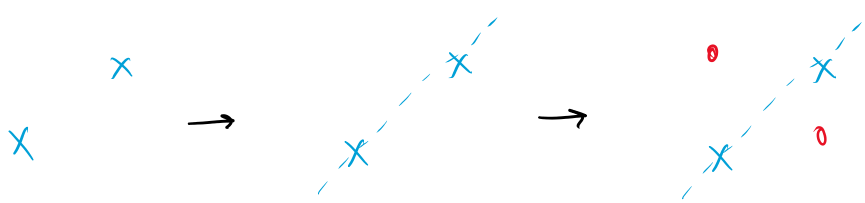
The least number of samples among the given choices is 2000, so $\boxed{\text{the answer is d.}}$

4) In $\mathbb{R}^2$, we see that this configuration is the

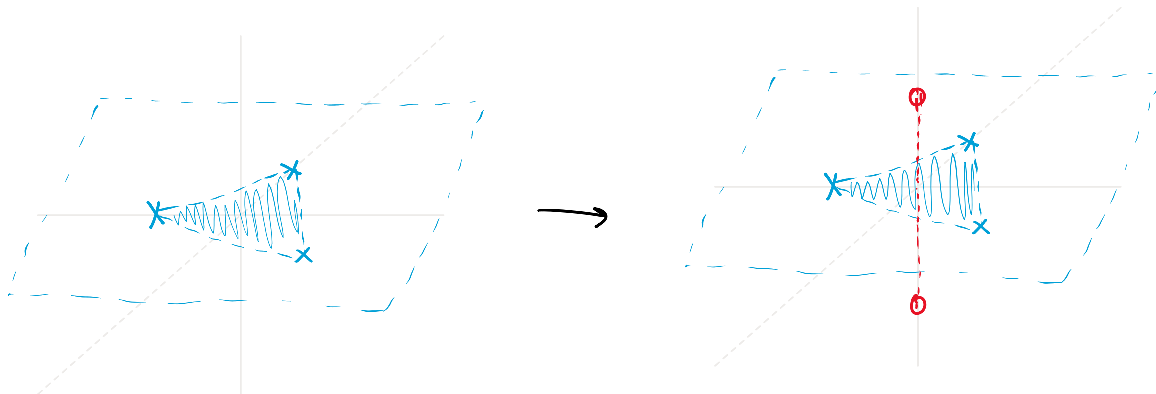
smallest break point.



We could arrive at that construction as follows. Let's say have two co-linear points with the same label. If we put one point on either side of that line with the same, but opposite labels, we've shown that 4 is the breaking point.



The same can be done with the perceptron in $\mathbb{R}^3$. Take any 3 co-planar points and give them the same label. We only need two other points to reach the breaking point. The 3 co-planar points create a triangular region on the plane, and if the projection of the two additional points lies within the triangular region (one on each side of the plane with the same but opposite label as before), we only need 5 points to reach the breaking point.



$\boxed{\text{The answer is b.}}$

5) We look at 2 facts to analyze these functions. The growth function must either be equal to 2^N or must be a polynomial in N which is bounded by 2^N (a fact that is true for all polynomials).

i) $N+1$ is a polynomial function which is always $\leq 2^N$.

$N+1$ is a growth function.

ii) $1+N+\binom{N}{2} = 1+N+\frac{N(N-1)}{2}$

$= 1+N-\frac{N}{2}+\frac{N^2}{2}$

$= 1+\frac{N}{2}+\frac{N^2}{2} \leq 2^N$

$1 + N + \binom{N}{2}$ is a polynomial which is always $\leq 2^N$, so point making this a growth function.

iii) $\sum_{i=1}^{\lfloor\sqrt{N}\rfloor} \binom{N}{i} = \sum_{i=1}^{\lfloor\sqrt{N}\rfloor} \frac{N!}{i!(N-i)!}$

$$= N + \frac{N(N-1)}{2!} + \frac{N(N-1)(N-2)}{3!} + \cdots + \frac{N(N-1)(N-2)\cdots(N-\lfloor\sqrt{N}\rfloor+1)}{\lfloor\sqrt{N}\rfloor!}$$

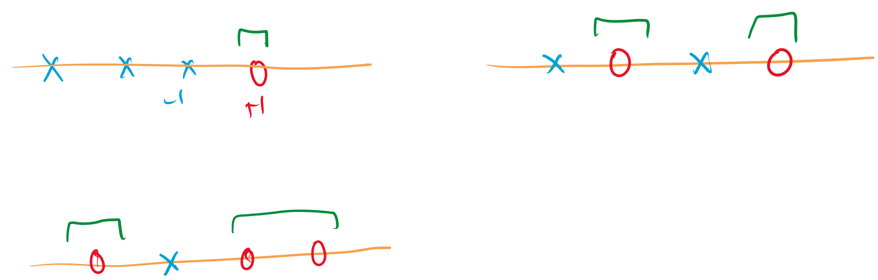
We have shown that the highest degree of the polynomial is $\lfloor\sqrt{N}\rfloor$ which is dependent on N, so the formula is not a polynomial or equal to 2^N, so it's not a growth function

iv) It's easy to see that this formula is not equal to 2^N and it's not a polynomial in N (since it is an exponential), so the formula is not a growth function.

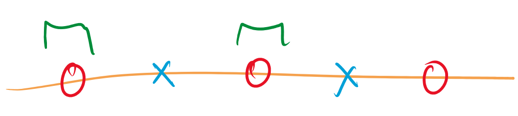
v) This formula is equal to 2^N, so it is a growth function.

Formula i, ii, & v are growth functions, so the answer is b.

6) We'll first show that any set of four points can be leared by h.



This means our breaking point is greater than four. We can see though that if we add a fifth point, we can create a breaking point.

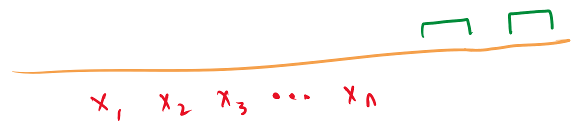


If the points flip one after the other, the two intervals will not be able to learn all the points w/out violating the constraint. As such, the smallest breaking point is 5 and

the answer is c.

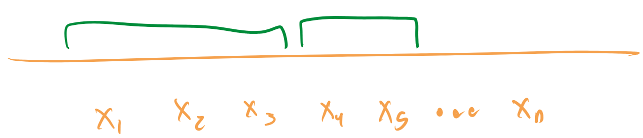
7) There are three possible cases for the intervals. We can look at the possible choices for each case and add them at the end.

i) The intervals are outside the set of finite data points we're looking at:



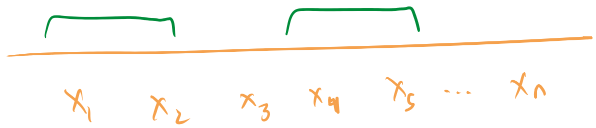
In this case, no matter where the intervals are, only one dichotomy is created.

ii) The intervals are adjacent to one another and appear as one:



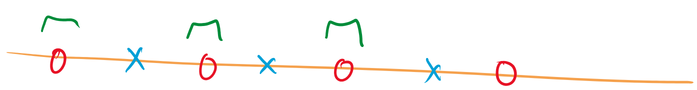
In this case, this is effectively the same as picking one interval w/ 2 endpoints, so there are $\binom{N+1}{2}$ ways to arrange the intervals and create a dichotomy.

iii) The intervals are separated by at least one data point:



In this case, we pick four points to fashion our intervals and there are $\binom{N+1}{4}$ ways to arrange the intervals and create dichotomies.

As such, the growth function is $\binom{N+1}{4} + \binom{N+1}{2} + 1$ which means $\boxed{\text{the answer is c}}$

8) 

1	3
2	5
3	7

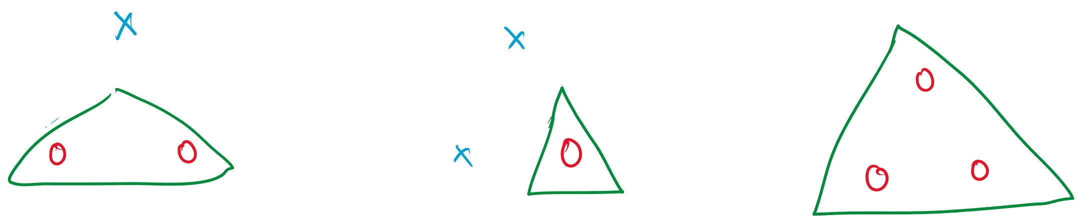
From plotting the above example with three points, we see that we need M pairs of points to create M flips and then add one extra point at the end for the breaking point. As such, the smallest break point of the hypothesis set is $2M+1$ which means $\boxed{\text{the answer is d.}}$

9) It can be shown easily that we can shatter the

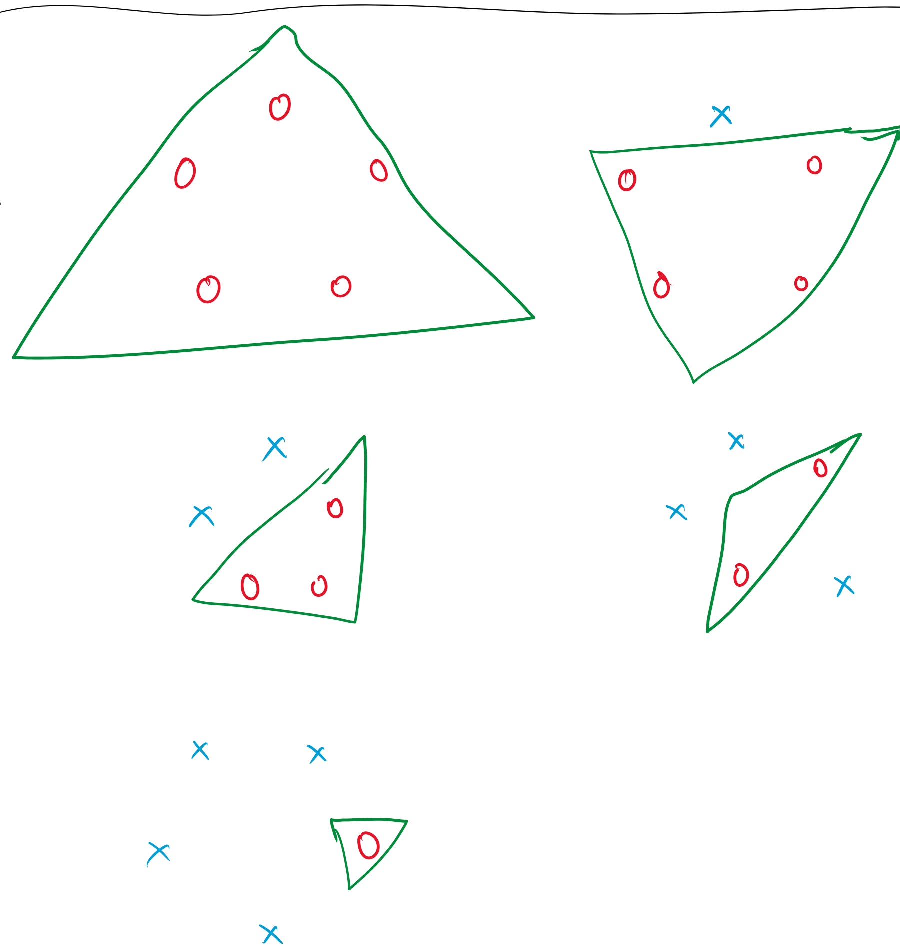
hypothesis set 1 & 3 points, we run through a similar analysis for the rest to show where the points shatter.

N = 1 :

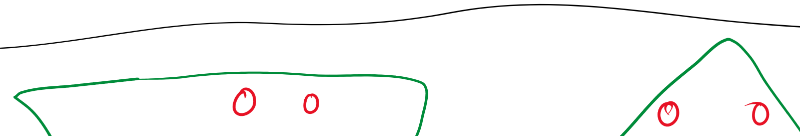
N = 3 :



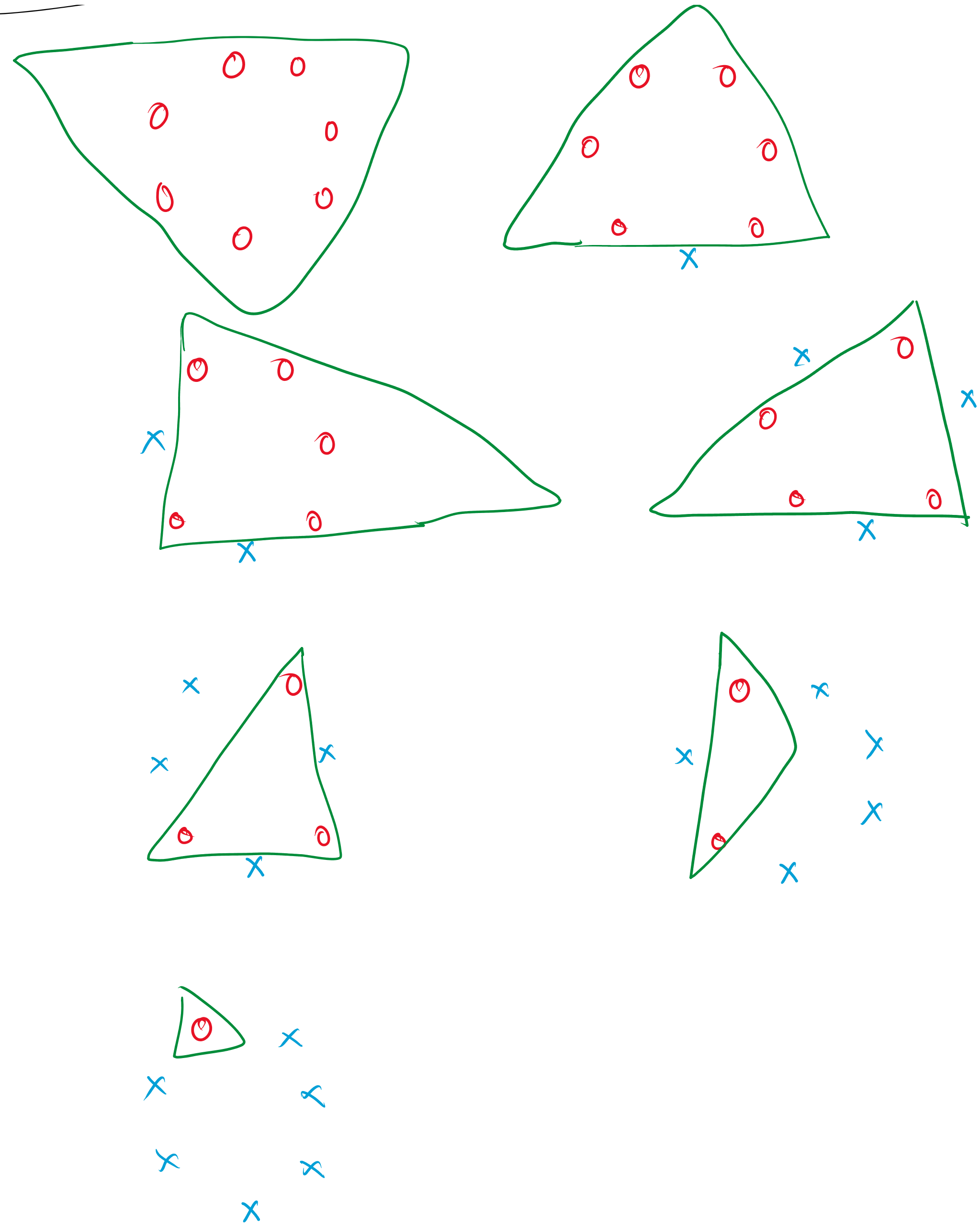
N = 5 :



N = 7 :

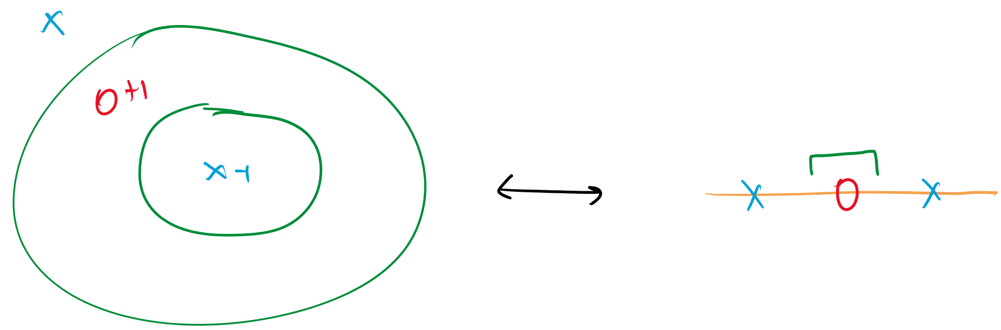


$N = 7$:



By performing a similar analysis for $N=9$, we see that we can't express all possible combinations as each point can be placed inside the triangle, so the answer is d.

10)



This scenario is analagous to the one-interval example in $\mathbb{R}^1$, so the same bound applies. The interval boundary is set by $[a, b]$ where $r = \sqrt{x_1^2 + x_2^2}$ is the input variable. As such, there are $\binom{N+1}{2} + 1$ ways to vary a & b to create dichotomies. As such, $\boxed{\text{the answer is b.}}$