# CS 156a Problem Set 3

Amitesh Pandey

October 2024

## Generalization Error

### Problem 1

If we set $M = 1$, $\epsilon = 0.05$, we want an $N$ such that $2Me^{-2\epsilon^2 N} = 2e^{-2\cdot 0.0025N} = 2e^{-0.005N} = 0.03$. We have $-0.005N = \log(0.015) = -4.199 \implies N = 840$. So option **[b]** ($N = 1000$) is correct.

### Problem 2

When $M = 10$, the quantity we have inside the natural log from Problem 1 will simply be 10 times as less, thus we have $-0.005N = \log(0.0015) \implies N = 1300$, so option **[c]** ($N = 1500$) is correct.

### Problem 3

When $M = 100$, obviously, we have $-0.005N = \log(0.00015) \implies N = 1760$, so option **[d]** ($N = 2000$) is correct.

# Break Point

## Problem 4

The correct option is **[b]**, or $k = 5$. Now we will proceed with a proof.

*Proof.* First, we will argue that the breaking point *is not* 4. Recall that $k$ is said to be a *breaking point* for the hypothesis $\mathcal{H}$ if *no* dataset of size $k$ can be *shattered* by $\mathcal{H}$. Observe from this definition that if for a prospective breaking point $k_1$, we can provide a dataset of size $k_1$ that $\mathcal{H}$ can shatter, this is sufficient to rule $k_1$ out as a breaking point. Now also note that the breaking point for the perceptron in $\mathbb{R}^3$ must be at least as large as its breaking point in $\mathbb{R}^2$ since we could just generate some plane from the line creating the dichotomy in $\mathbb{R}^2$ as a valid hypothesis. Now also notice that $k = 4$ is not a breaking point for the perceptron in $\mathbb{R}^3$ because for any failed dichotomy of 4 points in the space $\mathbb{R}^2$, we can simply raise the points causing the problem to a higher level on the $z$-axis and resolve the problem. Thus $k$ must be at least 5. Now we will argue why the pereceptron is always incapable of shattering datasets comprising of 5 arbitrary points. We will show this through a proof by construction. Consider points $p_1, p_2, p_3, p_4, p_5$. First note that it must be that any 4 of these 5 points are mutually non-coplanar, otherwise, it is trivial that they aren't shattered. Now observe that we can always select 4 points and construct a tetrahedron through them such that there is a 5th point that lies opposite to one of the tetrahedron's surfaces (planes). In other words, the vector $\mathbf{r}$ from this point to at least one of the tetrahedron's is guaranteed to have non negative $x$ and $y$ components. We will use the language that this point lies "away" from the tetrahedron. Now simply assign this point and a point of the tetrahedron farthest from the "away" surface described before the same label and the remaining 3 points a different label. Now notice that we can never generate this dichotomy using a perceptron based hypothesis. Thus, $k = 5$. □

# Growth Function

## Problem 5

We know that $m_{\mathcal{H}}(N)$ is either bounded above by a polynomial or it is $2^N$. Immediately, we have (i) and (v) as valid forms of the growth function and (iv) as invalid. Let's now look at

$$m_{\mathcal{H}}(N) = 1 + N + \binom{N}{2}$$

We know by definition that

$$\binom{N}{2} = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2} \leq N^2$$

Then we get

$$m_{\mathcal{H}}(N) = 1 + N + \binom{N}{2} \leq 1 + N + N^2$$

Now we can conclude that (ii) is a valid growth function. Let's now look at (iii), that

$$m_{\mathcal{H}}(N) = \sum_{i=1}^{\lfloor\sqrt{N}\rfloor} \binom{N}{i} = N + \frac{N(N-1)}{2} + \frac{N(N-1)(N-2)}{6} + \cdots + \frac{N(N-1)(N-2)\ldots(N-\lfloor\sqrt{N}\rfloor+1)}{\lfloor\sqrt{N}\rfloor!}$$

The numerator, when expanded, will have an undamped $N^{\lfloor\sqrt{N}\rfloor}$ term which grows exponentially but is not $2^N$, thus (iii) is not a valid growth function. So the correct option is **[b]**.

# Fun with Intervals

## Problem 6

We need to find the smallest size ($k$) of a dataset such that the 2-interval hypothesis does not shatter it. First make the observation that any dichotomy that the hypothesis fails on must have a '+1' classified as '-1,' in other words, we must have run out of intervals to account for all +1 in the dataset. If we have two '+1' points, with however many '-1' points in between them, since we now have two intervals available to us, we do not fail. However, if we have 3 '+1' points, with *at least* one '-1' point in the gap between them, then we'd require three intervals, which we do not have the capacity for. Thus, we would need three '+1' points, and the two gaps filled by at least one '-1' point each. Finally, we can conclude that any dataset of 5 points fails on the dichotomy '+1, -1, +1, -1, +1' when the points are arranged in increasing order. So $k = 5$, the correct option is **[c]**.

## Problem 7

Recall that $m_{\mathcal{H}}(N)$ counts not the number of possible hypotheses, but the number of possible dichotomies that could arise on application of the hypotheses picked from the hypothesis set $\mathcal{H}$ on a sample of size $N$. First observe that a hypothesis is entirely identified by $[x_1, x_2], [x_3, x_4]$, the two intervals containing +1 points. Then observe that it is not the exact values of $x_1, x_2, x_3, x_4$ but the *gap* between the points they occupy that dictates the nature of the dichotomy. There are $N+1$ gaps, $-\infty : p_1, p_1 : p_2, \ldots, p_N : \infty$. Without loss of generality, assume $x_1 \leq x_2 \leq x_3 \leq x_4$, i.e. $[x_1, x_2]$ determine the first +1 interval and $[x_3, x_4]$ the second. When $x_2 \neq x_3$, we have 2 non-continuous intervals that can be picked in $\binom{N+1}{4}$ ways, when $x_2 = x_3$, we have $\binom{N+1}{2}$ ways to pick the total interval $[x_1, x_4]$, and finally when $x_1 > p_N$ or $x_4 < p_1$, we have 1 way where there are no '+1's, so $m_{\mathcal{H}}(N) = \binom{N+1}{4} + \binom{N+1}{2} + 1$, the correct option is **[c]**.

## Problem 8

Note that the dataset that fails to be shattered by $M$-intervals will follow the same principle of "running out of +1 intervals for all the +1 points." We construct the smallest possible dichotomy that gives rise to this problem. Trivially if the number of $+1$s in the dichotomy is less than or equal to $M$, then $M$ intervals are more than sufficient. Thus the number of $+1$'s must be at least $M + 1$. Now if these $+1$ points lie with no $-1$ points between each other, then the hypothesis can weaponize creating large intervals that classify multiple $+1$'s at once, so we must the $+1$s with a corresponding $-1$ in the gap. There would be $M$ gaps between $M + 1$ points. So the total minimum size of the dataset would be $M + 1 + M = 2M + 1$ so the correct option is **[d]**.

# Convex Sets: The Triangle

## Problem 9

The correct option is **[d]**, or 7. To prove that 7 is the largest possible set of points that can be shattered by the class of triangles, it is sufficient to prove that the largest possible set is *at least* 7 but no larger than or equal to 8. We will first show that there exists a dataset of 7 points that can be shattered by the class of triangles.

*Proof.* Consider a set of 7 points aligned on a circle. Now note that all possible configurations of these points will include at most three continuous '-1' sequences. Since a triangle has three sides, all three of these sequences are easily separable by individual sides of a triangle. □

This means $VC_{\text{dim}}(\mathcal{H}_{\text{tri}}) \geq 7$. Now we will show why $VC_{\text{dim}}(\mathcal{H}_{\text{tri}}) < 8$.

*Proof.* Notice that when we have 8 points, the convex hull of these points will be a convex irregular octagon otherwise, it's easy to see why the triangle class fails to shatter these 8 points. Let this octagon be $ABCDEFGH$. We will show that the dichotomy $A, C, E, G = +1$, and $B, D, F, H = -1$ is not shattered by the hypotheses class.

**Definition 1** (Minimal Hypotheses)
Let $H \in \mathcal{H}$ be a valid minimal hypothesis if $H$ contains all points $A, C, E, G$ but does **not** contain a hypothesis $H' \in H$ such that $H'$ is also valid. I.e., $H$ is the *smallest* triangle capable of containing $A, C, E, G$.

Let $\Delta PQR$ be the minimal hypothesis that separates $A, C, E, G$ from $B, D, F, H$. Then it's easy to see why (without loss of generality), either $P, Q,$ and/or $R \in \{A, C, E, G\}$ *or* one of $A, C, E, G$ lie on the inner side of segment $PQ$. Then it is easy to see that as a virtue of the convexity of $ABCDEFGH$, at least one point from labelled -1 is bound to be contained in $\Delta PQR$. This is a contradiction. Thus the minimal hypotheses set is empty, which implies that the valid hypotheses set is empty. Points that form a convex octagon are not shattered by the triangle class. □

# Non-Convex Sets: Concentric

## Problem 10

Notice that a point $(x_1, x_2) \in \mathbb{R}^2$ is a distance $\sqrt{x_1 2 + x_2^2}$ away from the origin. This brings us to the observation that the condition $a^2 \leq x_1^2 + x_2^2 \leq b^2$ implies that only points in the region between the two circles are labelled $+1$, and every other point is labelled $-1$. Notably, this is identical to the interval hypotheses, extended to 2D space. So the growth function will trivially depend only on the choices of $a, b$. There can at most be $N + 1$ "gap" regions in circles traced by $N$ points. Therefore we have $N + 1$ *dichotomy-changing* choices for both $a$ and $b$. Thus brings us to $\binom{N+1}{2}$. But, it's also true that one hypothesis we didn't count is when both $a$ and $b$ are less than the smallest circle traced by any of the $N$ points, so we add 1. Finally correct option is **[b]**, $\binom{N+1}{2} + 1$.