Set 3

Claire Goeckner-Wald

October 17, 2016

Generalization Error

1) B. 1000

$$0.03 \geq 2Me^{-2(\epsilon)^2 N}$$
$$0.03 \geq 2(1)e^{-2(0.05)^2 N}$$
$$0.03 \geq 2e^{(-.005)N}$$
$$\frac{0.03}{2} \geq e^{(-.005)N}$$
$$\ln(\frac{0.03}{2}) \geq (-.005)N$$
$$\frac{\ln(\frac{0.03}{2})}{-.005} \leq N$$
$$\frac{\ln(\frac{0.03}{2})}{-.005} \approx 840$$

2) C. 1500

$$0.03 \geq 2Me^{-2(\epsilon)^2 N}$$
$$0.03 \geq 2(10)e^{-2(0.05)^2 N}$$
$$0.03 \geq 20e^{(-.005)N}$$
$$\frac{0.03}{20} \geq e^{(-.005)N}$$
$$\ln(\frac{0.03}{20}) \geq (-.005)N$$
$$\frac{\ln(\frac{0.03}{20})}{-.005} \leq N$$
$$\frac{\ln(\frac{0.03}{20})}{-.005} \approx 1300$$

3) D. 2000

$$0.03 \geq 2Me^{-2(\epsilon)^2 N}$$
$$0.03 \geq 2(100)e^{-2(0.05)^2 N}$$
$$0.03 \geq 200e^{(-.005)N}$$
$$\frac{0.03}{200} \geq e^{(-.005)N}$$
$$\ln(\frac{0.03}{200}) \geq (-.005)N$$
$$\frac{\ln(\frac{0.03}{200})}{-.005} \leq N$$
$$\frac{\ln(\frac{0.03}{200})}{-.005} \approx 1761$$

Break Point

4) B. 5

Since the break point is four in $\mathbb{R}^2$ (using a line, $\mathbb{R}^1$, to divide), then we only need to add one more point to create the break point in $\mathbb{R}^3$ (using a plane, $\mathbb{R}^2$, to divide points). This is similar to the relationship between $\mathbb{R}^1$ and $\mathbb{R}^2$. For $\mathbb{R}^1$, the break point (using one interval to define +1) was three. Then, when raising to $\mathbb{R}^2$, it was necessary to add only one more point to find the next break point.

In other words, for $\mathbb{R}^3$, the number of dimension is $d = 3$. Then, $d_{vc} = d + 1$. Thus, there exists a plane that separates 4 points into all possible dichotomies. The break point is one more than the d_{vc}. Thus, the break point is 5.

Growth Function

5) B. i, ii, v

The growth function is polynomial in the case that the hypothesis set has a break point. Otherwise, it is 2^N. If the function is nither 2^N nor polynomial, then it is not a possible growth function.

i. $1 + N$

As shown in lecture 5, example 1, the growth function is $1 + N$. Moreover, this function is polynomial, and therefore a potential growth function.

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

This function is a polynomial, and therefore a possible growth function.

iii. $\sum_{i=1}^{\lfloor \sqrt{N} \rfloor} \binom{N}{i}$

On lecture 6, slide 11, it shows that $m_H(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$, and that the maximum power of the sum is N^{k-1}. By this logic, the maximum power of $\sum_{i=1}^{\lfloor\sqrt{N}\rfloor} \binom{N}{i}$ is $N^{\lfloor\sqrt{N}\rfloor}$. Therefore, this is not a polynomial function, nor is it 2^N. Therefore, this is not a possible growth function.

iv. $2^{\lfloor N/2 \rfloor}$

This function is neither polynomial nor 2^N. Therefore, this is not a possible growth function.

v. 2^N

As shown in lecture 5, example 3, the convex set in $\mathbb{R}^2$ shatters N points. Thus, the growth function is 2^N. This is a possible growth function.

Fun with Intervals

6) C. 5

This function has a break point at 5 because there is no way of creating the distribution of points (+1, -1, +1, -1, +1) with two positive intervals. This is intuitive because with only two intervals available to classify points to +1, all one needs is three intervals that require +1 classification to find a break point. The simplest and least-N way to create three +1 intervals is to place three +1 points buffered by -1 points. This is given by (+1, -1, +1, -1, +1). For a more in-depth explanation, see problem eight.

Four points can be shattered by the two-interval model because there is no way to create three +1 intervals. The classification of four points can only create zero, one, or two +1 intervals. This is true no matter how the points are distributed because the distance between the points is arbitrary.

7) C. $\binom{N+1}{4} + \binom{N+1}{2} + 1$

a. $\binom{N+1}{4}$

This function does not equal 2^N for $N < 5$, the break point. Consider N = 1, which should output two. $\binom{1+1}{4} = 0 \neq 2^1$

b. $\binom{N+1}{2} + 1$

This function does not equal 2^N for $N < 5$, the break point, especially for $N = 3$, which should output eight. $\binom{3+1}{2} + 1 = 7 \neq 2^3$

c. $\binom{N+1}{4} + \binom{N+1}{2} + 1$

$\binom{N+1}{4}$ counts all appearances of 2-intervals that are disjoint (so no interval starts and ends in the same space, and no intervals overlap). $\binom{N+1}{2}$ counts all appearances such that intervals overlap or touch, by representing them as continuous intervals (so, much like the 1-interval model). The value 1 represents all intervals such that no points are selected. Thus, this is the correct growth function, because it covers all possibilities of point classification.

d. $\binom{N+1}{4} + \binom{N+1}{3} + \binom{N+1}{2} + \binom{N+1}{1} + 1$

This function does not equal 2^N for $N < 5$, the break point. Consider N = 1, which should output two. $\binom{1+1}{4} + \binom{1+1}{3} + \binom{1+1}{2} + \binom{1+1}{1} + 1 = 4 \neq 2^1$

8) D. $2M + 1$

The distribution of points in 1-d space is generalizable to equidistant points on a line (because the space between them is arbitrary). For M intervals, one cannot shatter a distribution of points that requires $M + 1$ intervals to correctly classify. This can be developed in its least-N form by placing $M + 1$ positive points with negative points in between (such that one interval must be placed to cover each positive point). Then, there must be at least M negative points placed in the spaces between each postive point. (Ex: +1 -1 +1 -1 +1 is an example of negative points serving as "buffers" for the positive points. Positive points will always be on the "ends" of the point distribution, in order to minimize the number of points necessary.) Thus, there are no more points placed than necessary. Summing the positive M+1 and negative M points, we require $(M + 1) + M = 2M + 1$ points to define a break point in the M-interval model.

Moreover, for known M-interval model break points, this functions correctly. For a 0-interval model, the break point is 1. For a 1-interval model, the break point is 3. For a 2-interval model, the break point is 5.

The Triangle

9) D. 7

We arrange the points on a circle to maximize the number of dichotomies available. The circle also gives a rotational symmetry, which can be used to limit the number of patterns of point classification to be considered. Furthermore, since the distance between the points is arbitrary, we can consider the points to be equidistant on the circle.

This problem is similar to a three-interval classification of points on a 2-d line. For any dichotomies where there are only zero, one, two, or three +1 points on the circle, the solution is simple and intuitive. (Place the vertices of the triangle on the +1 points where applicable, and nearby where not.)

We can draw a triangle so long as the +1 points are arranged such that there are no more than three uninterrupted "strips" or intervals of +1 points. (Here, an interval of +1 points is defined as the maximum arc on the circle that contains only +1 points. No interval will border another interval because then we'd just define it as one large interval.) For each of the three intervals of positive points, one can put the vertex of the triangle outside the circle, such that the triangle's sides that lead to the removed vertex pass through (or, just around) the edges of the interval such that the entire arc is captured by the triangle.

Thus, we cannot draw a triangle where there are more than three +1 intervals. To develop the least-N break point, we imagine four +1 intervals, where each interval is a single +1 point. The intervals are buffered by -1 points, so that we cannot combine to create a super-interval. Thus, we have four +1 points and four -1 points. Thus, the break point of this model is eight.

Seven points can still be shattered because one cannot create four +1 intervals. This is unlike 1-d line model simply because the circle wraps around, unlike the line. Thus, the 'edge' points on a line are in fact neighbors on a circle. So, even if there are four +1 points, there are only ever three +1 intervals. This is because we cannot create buffers between the four +1 points with only three -1 points.

Non-Convex Sets: Concentric Circles

10)   B. $\binom{N+1}{2} + 1$

This model is generalizable to the 1-interval classification of points on a line. Each point is dependent on its distance from the origin of the circle, r, and its degree, θ. Because we are classifying using concentric circles centered on the origin in 2-d space, the only characteristic of our points in $\mathbb{R}^2$ that determines their classification is r.

For example, consider two points $p_1 = (r_1, \theta_1)$ and $p_2 = (r_2, \theta_2)$ with $r_1 = r_2$ but $\theta_1 \neq \theta_2$. These points will have the same classification in this model because $r^2 = x_1^2 + x_2^2$. Therefore, since $r_1 = r_2$, then $r_1^2 = r_2^2$. Thus, even if the points are defined differently in rectangular coordinates, only their distance from origin is important to their classification by h.

Thus, one can consider these points to be arranged on a line by r, with 1 positive interval to represent the ring used to classify. Thus, the growth function is equivalent to considering choosing start and end points of intervals, as in the fifth lecture. Thus, we have $\binom{N+1}{2} + 1$. The 1 is to represent intervals that do not classify any points as $+1$.