Learning From Data: Homework 03

Topics:

- Lecture 05 (Training Versus Testing)
- Lecture 06 (Theory of Generalization)

For Exercises 1 to 3, the approach is to 'plug and chug' with the Hoeffding Inequality. First, solve the inequality for N:

$$0.03 \geq 2Me^{-2\epsilon^2 N}$$

$$\frac{0.03}{2M} \geq e^{-2\epsilon^2 N}$$

$$\log\frac{0.03}{2M} \geq -2\epsilon^2 N$$

$$\frac{\log\frac{0.03}{2M}}{-2\epsilon^2} \leq N$$

Then make the substitutions as below.

1. B: 1000
   Let $\epsilon = 0.05, M = 1$

$$\frac{\log\frac{0.03}{2(1)}}{-2(0.05)^2} \leq N$$

$$N \geq 840$$

2. C: 1500

Let $\epsilon = 0.05, M = 10$

$$\frac{\log \frac{0.03}{2(10)}}{-2(0.05)^2} \leq N$$

$$N \geq 1300$$

3. D: 2000
   Let $\epsilon = 0.05, M = 100$

$$\frac{\log \frac{0.03}{2(100)}}{-2(0.05)^2} \leq N$$

$$N \geq 1761$$

4. B: 5
   Mentally model the problem using examples from lecture 5 as a guide. A heuristic is evident: surround a $+1$ point with a group of -1 points (or vice versa). For $\mathbb{R}^3$, place a point within a pyramid of triangular base. (Figure **??**)

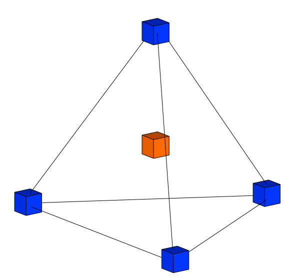

Figure 1: Example layout for $k = 5$ break point for Perceptron Model in $\mathbb{R}^3$

5. B: i, ii, v

   Formulae i and v were seen in lecture as valid growth functions. Formula ii is similar to i, with only a combinatorial addition which is acceptable.

   Formulae iii and iv make use of the floor function, $\lfloor\cdot\rfloor$. Recall that floor$(x) = \lfloor x \rfloor$ is the largest integer not greater than x. For both iii and iv, for low values of N, the results are equal to 1, which is too low to allow a dichotomy on 1 or more points.

6. C: 5

   Solve by drawing several points and looking for a configuration that can't be properly enclosed by either interval. See Figure **??**.

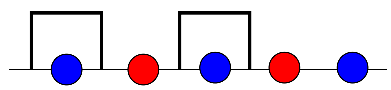

Figure 2: Example layout for $k = 5$ break point for "2-intervals" learning model. Due to the arrangement of the 5 points, the point on the far right cannot be enclosed within an interval, thus breaking the hypothesis set.

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

   Review from lecture 5 how one interval was treated:

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

   This is a combinatoric, stating "from $N+1$ choose 2 items". The above combinatoric was then substituted into $N+1$:

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

   For a "2-intervals" learning model, there are thus 4 sections from which to choose out of $N+1$ total sections of a line containing N points. One might suppose that the solution to this exercise to be $\binom{N+1}{4} + 1$, but this does not account for the combinations that occur when the pair of intervals overlap, effectively acting like a single interval. Thus, we must add the disjoint combinations to arrive at

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

8. D: $2M + 1$
   Establish a pattern based on what we have seen so far:

   - "1-interval" learning model with breakpoint $k = 3$.
   - "2-interval" learning model with breakpoint $k = 5$.
   - ... (determine the breakpoints for several more intervals)
   - "M-interval" learning model with breakpoint $k = 2M + 1$.

9. D: 7
   To approach this problem, start by drawing points, e.g. 7 points, on a circle. Next, paint the points in alternating colors. There will happen to be two adjacent points that will be the same color. Draw a triangle dissecting the circle, enclosing all points of the same color. Repeat this procedure with 8 points to see that it is not possible to shatter the points. Thus, we have determined the breakpoint $k = 8$ and the maximum points that can be shattered is $N = 7$. As seen in lecture, arranging points along a circle is useful for convex shapes. See Figure **??**. If $N \geq 8$, it is not possible to shatter the points for this learning model.

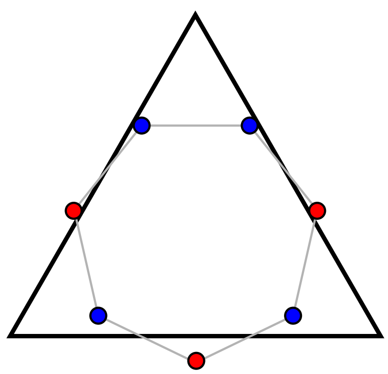

Figure 3: $N = 7$ (depicted as points on a heptagon) is the maximum number of points that can be shattered by the "triangle" learning model.

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

Given the symmetry of the concentric circles, we can simply extend a line from the center outward. Along the radial line are placed points of alternating color. As depicted in Figure **??**, the "concentric circle" learning model is similar to the "1-interval" linear model we saw in lecture 5. The growth functions are identical:

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

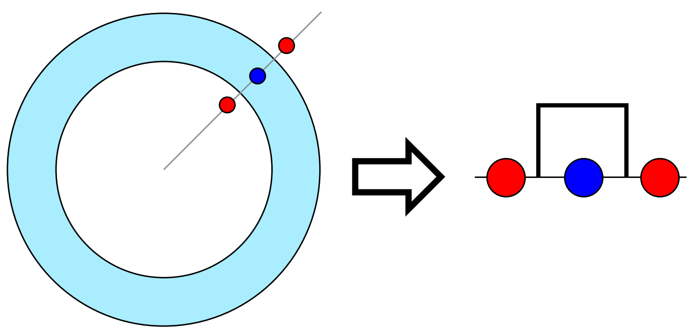

Figure 4: Extending a radial line across the concentric circles shows the similarity of the "concentric circle" learning model to the "1-interval" learning model.