CS 156 Problem Set 5

Christopher Zhen

Oct 31, 2016

Problem 1

(C) - We can solve for the smallest possible possible value of N by solving the following equation:

$$\mathbb{E}_{\mathcal{D}}[E_{\text{in}}(\boldsymbol{w}_{\text{lin}})] = \sigma^2(1 - \frac{d+1}{N})$$
$$0.008 = (0.1)^2(1 - \frac{(8)+1}{N})$$
$$N = 45$$

So the lowest answer choice that satisfies the desired expected in-sample error is $N = 100$.

Problem 2

(D) - To achieve the desired boundary, we can use the following formula of a hyperbola:

$$ax_1^2 - bx_2^2 = 1$$
$$ax_1^2 - bx_2^2 - 1 = 0$$

However, because this equation gives the opposite dichotomy as the one desired, we need to multiply everything by -1 to get:

$$-ax_1^2 + bx_2^2 + 1 = 0$$

Which matches with answer choice D.

Problem 3

(C) - The Devroye bound this time is around 2 at $N = 5$ while the Parrond bound is around 1.7.

Problem 4

(E) - The code outputs an $\hat{a} = 1.45$ which doesn't match any of the options.

Problem 5

(B) - The code outputs 0.28 which is closest to 0.3. It makes sense for the bias in this case to be slightly worse than the $\bar{g}(x) = ax + b$ case from class, but better than the $\bar{g}(x) = b$ case.

Problem 6

(A) - The code outputs a value of 0.24 which is closest to 0.2. It makes sense for this method to have smaller variances because there will never be negative values for $\hat{a}$ and the value at $x = 0$ will always be the same.

Problem 7

(B) - In lecture, we already derived the expected value using the bias-variance approach of both (A) and (C) and found that they were 0.75 and 1.90 respectively. The calculated value for (B) is 0.52 which is lower than both. Intuitively, since both (D) and (E) are even functions and we're fitting an odd function, neither should have a lower error than the other 3 functions, leaving (B) as the best choice.

Problem 8

(C) - Since $m_{\mathcal{H}}(1) = 2$ and $m_{\mathcal{H}}(N + 1) = 2m_{\mathcal{H}}(N) - a$ (where a is some value), we'll always have the case where $m_{\mathcal{H}}(N) = 2^N$ for any value where a is 0. The value of a is only non-zero when $q \leq N$, so our first non-zero value will occur when $N = q$ and our VC dimension is q.

Problem 9

(B) - Consider the concrete example of perceptrons. If our hypothesis sets are the different dimensions of perceptrons, the maximum VC dimension of our intersection is simply the maximum VC dimension of the highest dimension perceptron. The minimum is then just the case where the intersection is 0.

Problem 10

(E) - The answer must be D or E because the union of sets contains the hypothesis with the highest dimension, so the dimension of the union must at least be the maximum VC dimension of a single set. Now

consider the case where we have a VC dimension of 0. In this case, the growth function is bounded by:

$$m_{\mathcal{H}} = \sum_{i=0}^{d_{vc}} \binom{N}{i}$$

$$m_{\mathcal{H}} = N^{d_{vc}} + 1$$

$$m_{\mathcal{H}} = 1$$

So the max VC dimension has to include a term to account for the case where we have up to K hypotheses with a VC dimension of 0 by adding on K-1 (since when K = 1, the VC dimension is still 0).