Machine Learning from Data 4

Shane O'Brien

October 2017

Exercise 2.4

a

To show that $d_{vc} \geq d{+}1$, we just need to give a single matrix of size $(d{+}1)x(d{+}1)$ that cannot be shattered. The matrix we will use is the identity matrix.

The identity matrix is used because it is linearly independent by default.

Now, we will look at some math to end with the result that the dimensions of $\mathbf{w}$ is equivalent to the dimensions of $\mathbf{y}$. This means that we can pick any set of $\mathbf{y}$, and get that result from our data $\mathcal{X}$ by picking a set $\mathbf{w}$.

We know by definition:

$\mathbf{w}x = \mathbf{y}$

We multiply each side by the inverse of $\mathcal{X}$:

$\mathbf{w}xx^{-1} = \mathbf{y}x^{-1}$

Because it is the identity matrix, x^{-1} equals one. By definition, xx^{-1} equals one:

$\mathbf{w} = \mathbf{y}$

We have the end result that the dimensions of our weight vector is the same as our "answer" vector.

b

To show that $d_{vc} \leq d + 2$, we need to prove that there is not a single matrix $(d + 2)$x$(d + 1)$ matrix $\mathcal{X}$ that can we shattered by a set of weights $\mathbf{w}$.

By definition, a $(d + 2)$x$(d + 1)$ matrix cannot be linearly independent. This means that our $(d + 2)$ data point will also be dependent on our first $(d + 1)$ data points.

We will define our $(d+2)$ data point as x_i and all of our first $(d+1)$ data points as x_j. So, since x_i is linearly dependent, we know that we can make x_i using some set of constants c_j and the data set x_j.

$x_i = \sum_{i \neq j} c_j x_j$

We get this and multiply each side by w^T:

$x_i w^T = \sum_{i \neq j} c_j w^T x_j$

Now, let's say we want to get a certain result for our point $d+2$, like $y(x_i) = -1$

We know, by definition that $y_j = \text{sign}(w^T x_j)$
and we will define our dichotomy so that $y_j = \text{sign}(c_j)$

If we plug that into $x_i w^T = \sum_{i \neq j} c_j w^T x_j$, we see that our right side is equal to 1. This is because $\text{sign}(w^T x_j)$ and $\text{sign}(c_j)$ must be the same sign, as they both equal y_j. This means that our answer for $y(x_i)$ is 1. This contradicts our earlier statement.

Problem 2.3

a

For positive and negative ray, $m_{\mathcal{H}}(N) = 2N$. I have charted the results (by observation) below:

Analyzing $m_{\mathcal{H}}(N)$					
N	1	2	3	4	5
Pos/Neg Ray	2	4	6	8	10

The result of $2N$ can be figured intuitively. We know that for positive ray, we have $N+1$ dichotomies. For positive and negative ray, we can add $N-1$ more dichotomies by putting a negatively facing ray between each of the points. Adding $N+1$ and $N-1$ gives you $2N$

b

We can start this problem by giving this an upper bound. As solved in the textbook, the $m_{\mathcal{H}}(N)$ of the positive interval is $\frac{1}{2}N^2 + \frac{1}{2}N + 1$. The positive and negative interval could AT MOST be double. This bounds it as a second degree polynomial.

By observation, we can obtain this chart of the positive and negative interval:

Analyzing $m_{\mathcal{H}}(N)$					
N	1	2	3	4	5
Pos/Neg Int	2	4	8	14	22

We can use any three of these points and make a series of equations in the form $aN^2 + bN + c$ to solve for our a, b, and c.

The three equations are:

$a + b + c = 2$
$4a + 2b + c = 4$
$9a + 3b + c = 8$

a, b, and c are $1, -1$, and 2, giving us the answer $N^2 - N + 2$

For this problem, $d_{vc} = 3$

c
We can convert this problem into a positive interval problem like this:

$r = \sqrt{x_1^2 + x_2^2 + ... + x_d^2}$ where $a \leq r \leq b$

This works because our growth function does not depend on d. Our final answer is $\frac{1}{2}N^2 + \frac{1}{2}N + 1$

For this problem, $d_{vc} = 2$

Problem 2.8

Possible growth functions: first, second, third, sixth

Impossible growth functions: fourth, fifth

This is because only two types of growth functions are possible: polynomial and 2^N.

Problem 2.10

To show this, we will look at both types of growth functions:

No Break Point: 2^N

$m_{\mathcal{H}}(N) = 2^N$

$m_{\mathcal{H}}(2N) = 2^{2N} = 4^N$

$m_{\mathcal{H}}(N)^2 = (2^N)^2 = 2^{2N} = 4^N$

$m_{\mathcal{H}}(2N) = m_{\mathcal{H}}(N)^2$

Break Point: $N^{k+1} + 1$

$m_{\mathcal{H}}(N) \leq N^{k+1} + 1$

$m_{\mathcal{H}}(2N) \leq (2N)^{k-1} + 1$ — $m_{\mathcal{H}}(2N) \leq 2^{k-1}N^{k-1} + 1$

$m_{\mathcal{H}}(N)^2 = ((N)^{k-1} + 1)^2$ — $m_{\mathcal{H}}(N)^2 = (N^{k-1})(N^{k-1}) + 2N^{k-1} + 1$

If we compare these two:

$2^{k-1}N^{k-1} + 1 \leq (N^{k-1})(N^{k-1}) + 2N^{k-1} + 1$

$2^{k-1} + 1 \leq N^{k-1} + 2N^{k-1} + 1$

Therefore:

$m_{\mathcal{H}}(2N) \leq m_{\mathcal{H}}(N)^2$ for all types of hypotheses sets.

Using this, we can construct a new generalization bound:

$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} * \ln \frac{4m_{\mathcal{H}}(N)^2}{\delta}}$

Problem 2.12

The specifics given in the problem give us this:

$0.05 = \sqrt{\frac{8}{N} \ln \frac{4(2N)^{10}+1}{0.05}}$

We can rearrange to this:

$N = \frac{8}{0.0025} * \ln \frac{4(2N)^{10}+1}{0.05}$

If we put 5000 in for N, we get $\frac{8}{0.0025} * \ln \frac{4(2*5000)^{10}+1}{0.05} = 308753$

We do this again, putting in our new result... $\frac{8}{0.0025} * \ln \frac{4(2*308753)^{10}+1}{0.05} = 440693$

$\frac{8}{0.0025} * \ln \frac{4(2*440693)^{10}+1}{0.05} = 452079$

$\frac{8}{0.0025} * \ln \frac{4(2*452079)^{10}+1}{0.05} = 452895$

N eventually converges to ≈ 452947