

Assignment 4

Darwin Ding

September 28, 2016

Exercise 2.4

- (a) Back in chapter 1, we defined $h(x) = \text{sign}(w^T x)$. We dealt with 2d space in chapter 1, but the math still applies to $d+1$ dimensional space. Our perceptron is a vector of size $d+1$, which represents a hyperplane in $d+1$ dimensional space.

The question can be reworded to, given any vector y of $d+1$ points:

$$y = [\pm 1, \pm 1, \pm 1, \dots]$$

... and a $d+1$ by $d+1$ matrix X that represents $d+1$ points with $d+1$ dimensions each, can we generate a perceptron w (vector of size $d+1$) such that:

$$y = Xw$$

And in fact, this is quite straightforward to do. It is quite easy to create a $d+1$ by $d+1$ non-singular matrix, even if we require all 0th dimensions of our points to be 1 (see Chapter 1).

Simply see the following matrix, where all the lines are linearly in-

dependent:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & \dots & 1 \end{bmatrix}$$

Due to its non-singularity, X^{-1} exists and we can simply set $w = X^{-1}y$. Because we can do this with at least one example using a non-singular matrix, we have shattered $d+1$ for perceptrons. $\mathbf{d}_{VC} \geq d + 1$

- (b) However, things change when you add the next point. When you have more points than dimensions, linear independence is impossible. Mathematically speaking, this means:

$$\begin{aligned} x_{d+2} &= \sum_{i=1}^{d+1} a_i x_i \\ \implies y_{d+2} &= \text{sign}(w^T x_{d+2}) \\ &= \text{sign}\left(\sum_{i=1}^{d+1} a_i w^T x_i\right) \end{aligned}$$

However the fact that we can now mathematically derive the sign of the $d+2$ nd point based off of the other $d+1$ points shows a glaring hole in the dichotomies that we can implement.

Namely, for each point i from 1 to $d+1$, let's assign it $+1$ if $a_i > 0$, and -1 if $a_i < 0$. It doesn't matter so much if $a_i = 0$, but not all a_i can be 0 due to the linear dependence of point $d+2$.

Simplifying the above, we have $y_i = w^T x_i$, and $a_i w^T x_i = +a_i$ when $a_i > 0$, and $a_i w^T x_i = -a_i$ when $a_i < 0$.

As a result, $\sum_{i=1}^{d+1} a_i w^T x_i$ will always be positive! Therefore, if the $d+2$ nd point is assigned the value -1 , the dichotomy cannot be implemented. Therefore, $d+2$ cannot be shattered, because this process can be done with any set of $d+2$ points.

Problem 2.3

- (a) The positive and negative rays hypothesis set can shatter $N = 2$, unlike just positive rays, which failed on 2. This is because if the left point is positive and the right point is negative, just use a negative ray. On the flipped variant of this, use a positive ray. Otherwise, both points are positive or negative and "raying" these points is trivial.

However, three points cannot be shattered by this hypothesis set. If the points (from left to right) go -1, +1, -1 or +1, -1, +1, then no ray can fulfill this dichotomy. Therefore, the VC dimension of this hypothesis set is **2**.

- (b) Positive intervals shattered up to $N = 2$, but failed on the set +1, -1, +1 for $N = 3$, while fulfilling all others. However, by adding negative intervals to the hypothesis set we can do this dichotomy, and thus with this hypothesis set we shatter $N = 3$.

Unfortunately, we run into issues when doing $N = 4$, because the dichotomy cannot fulfill +1, -1, +1, -1 or -1, +1, -1, +1. Because this hypothesis set cannot shatter $N = 4$, this VC dimension is **3**.

- (c) This can be treated very similarly to the positive interval question. We have points in d -dimensional space but after calculating $a \leq \sqrt{x_1^2 + x_2^2 + \dots + x_d^2} \leq b$ (which is basically asking if the point's distance from the origin in d -dimensions is within the two spheres' radii) this can be converted back into 1d space with points on a number line.

Two points can be broken, because if you have two points that are not equidistant from the origin, you can place the spheres in such a way that the interval between them either tightly wraps around one of the points, wraps around both or just avoids both.

However, with $N = 3$, you run into issues when the furthest point from the origin is +1, there is a -1 in the middle and then there's a +1 closest to the origin. You cannot split your interval in two ways, so the VC dimension is **2**.

Problem 2.8

All growth functions are bounded by the following theorem: if $m(k) < 2^k$ for any k , then for all N , $m(N) \leq N^{k-1} + 1$.

1. $1 + N$: **Possible**. We showed earlier that positive rays have this growth function.
2. $1 + N + N(N - 1)/2$: **Possible**. This is broken at $N = 3$, since $1 + 2 + 2(1)/2 = 2^2 = 4$, but $1 + 3 + 3(2)/2 = 7 < 2^3$.

$$\begin{aligned} m(N) &= 1 + N + N(N - 1)/2 \\ &= 1 + N + .5N^2 - .5N \\ &= 1 + .5N + .5N^2 \end{aligned}$$

This growth function grows slower than $N^2 + 1$ because of the .5 factor on the N^2 , so since this follows the bounds it is a valid growth function.

3. 2^N : **Possible**. Any hypothesis set that is never broken has this growth function.
4. $2^{\lfloor \sqrt{N} \rfloor}$: **Impossible**. This is broken at $N = 2$, since $2^{\lfloor \sqrt{2} \rfloor} = 2^1 = 2 < 2^2$.

Thus, $m(N)$ should be bounded by $N + 1$. We can verify that this bound clearly does not work by plugging in 121 for N . 2^{11} is clearly way larger than 122.

5. $2^{\lfloor N/2 \rfloor}$: **Impossible**. This is broken at $N = 2$, since $2^{\lfloor 2/2 \rfloor} = 2^1 = 2 < 2^2$.

Again, $m(N)$ should thus be bounded by $N + 1$. We can again verify that this bound clearly does not work by plugging in 100 for N . 2^{50} is clearly way larger than 101.

6. $1 + N + (N(N - 1)(N - 2))/6$: **Impossible**. This is broken at $N = 2$, since $1 + 2 + (2(2 - 1)(2 - 2))/6 = 3 < 2^2$.

$m(N)$ should thus be bounded by $N + 1$. However, while $m(N)$ is not an exponential function, it is a cubic function and thus will not be bounded by this linear function. To verify, we can plug in $N = 100$, where $m(100) = 1718$, which is way larger than 101.

Problem 2.10

By definition, a growth function $m_H(N)$ is the number of dichotomies that a hypothesis set H can implement on N points. Let's say that $m_H(N) = k$, for some k . For any group of $2N$ points, we can split them into N and N points and implement k dichotomies in each. Then for the combined dichotomies, each pairing of dichotomies from each of the k from the two different N -groups of points are possible (but not necessarily all of them are implementable).

As a result, the number of dichotomies that can be implemented is at most $k^2 = m_H^2(N)$. However, if any of these dichotomies are not implementable in the final regrouping of $2N$ points, then they will not be equal. Thus, $m_H(2N) \leq m_H^2(N)$.

We can then use this bound into the VC generalization bound, giving us:

$$\begin{aligned} E_{out}(g) &\leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_H(2N)}{\delta}} \\ &\leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_H^2(N)}{\delta}} \end{aligned}$$

Problem 2.12

$$\begin{aligned} N &\geq \frac{8}{\epsilon^2} \ln \left(\frac{4((2N)^{d_{VC}} + 1)}{\delta} \right) \\ N &\geq \frac{8}{.0025} \ln \left(\frac{4((2N)^{10} + 1)}{.05} \right) \end{aligned}$$

We can, from here, perform the iterative process presented in Example 2.6.

For $N = 1000$, we find $N \geq 257251$

For $N = 257251$, we find $N \geq 434853$

For $N = 434853$, we find $N \geq 451652$

For $N = 451652$, we find $N \geq 452865$

We can see that near the end of these iterations, it seems to start to converge around **453000**.