Problem Solutions

Chapter 2

Pierre Paquay

Problem 2.1

Let us begin by extracting the value of N from the $\epsilon(M, N, \delta)$ expression. We have that

$$\sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}} \leq \epsilon \Leftrightarrow N \geq \frac{1}{2\epsilon^2}\ln\frac{2M}{\delta}.$$

(a) So for $M = 1$ and $\delta = 0.03$, to have $\epsilon \leq 0.05$ we need

$$N \geq \frac{1}{2\cdot 0.05^2}\ln\frac{2}{0.03} = 839.9410156.$$

(b) For $M = 100$ and $\delta = 0.03$, to have $\epsilon \leq 0.05$ we need

$$N \geq \frac{1}{2\cdot 0.05^2}\ln\frac{2\cdot 100}{0.03} = 1760.9750528.$$

(c) And for $M = 10000$ and $\delta = 0.03$, to have $\epsilon \leq 0.05$ we need

$$N \geq \frac{1}{2\cdot 0.05^2}\ln\frac{2\cdot 10000}{0.03} = 2682.00909.$$

Problem 2.2

For $N = 4$, if we consider four non aligned points, this $\mathcal{H}$ shatters these points (you only have to efectively enumerate them to see that all dichotomies can be generated), so in this case we have $m_{\mathcal{H}}(4) = 2^4$.

However, for $N = 5$, no matter how you place your five points, some point will be inside a rectangle defined by others. In this case, we are not able to generate all dichotomies and consequently $m_{\mathcal{H}}(5) < 2^5$.

From these two observations, we may conclude that, for poisitive rectangles, we have $d_{VC} = 4$, thus

$$m_{\mathcal{H}}(N) \leq N^4 + 1.$$

Problem 2.3

(a) We already know that the growth function for positive rays is equal to $N + 1$. If we enumerate the dichotomies added by negative rays, we get $N - 1$ new dichotomies (you get the opposite of the ones from positive rays and you have to substract the two dichotomies where all points are $+1$ and where all points are -1). So in total, we get that

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

As the largest value of N for which $m_{\mathcal{H}}(N) = 2^N$ is 2 ($m_{\mathcal{H}}(3) = 6$), we have that $d_{VC} = 2$.

(b) Here, we already know that the growth function for positive intervals is equal to $N^2/2 + N/2 + 1$. If we add the new dichotomies generated by negative intervals, we get $N - 2$ new ones (for example for $N = 3$, we only add the $(+1, -1, +1)$ dichotomy, and for $N = 4$, we add the $(+1, -1, +1, +1)$ and $(+1, +1, -1, +1)$

dichotomies). Of course, this only holds if $N > 1$, in the case where $N = 1$ we already generate the two dichotomies with the positive intervals alone. In conclusion, we may write that

$$m_{\mathcal{H}}(N) = \frac{N^2}{2} + \frac{3N}{2} - 1 \text{ if } N > 1 \text{ and } 2 \text{ if } N = 1.$$

As the largest value of N for which $m_{\mathcal{H}}(N) = 2^N$ is 3 ($m_{\mathcal{H}}(4) = 13$), we have that $d_{VC} = 3$.

(c) To determine the growth function fo concentric circles, we have to map the problem from $\mathbb{R}^d$ to $[0, +\infty[$. To do this, we use the map ϕ defined as

$$\phi : (x_1, \cdots, x_d) \mapsto r = \sqrt{x_1^2 + \cdots + x_d^2}.$$

By doing that, we may see that the problem of concentric circles in $\mathbb{R}^d$ is equivalent to the problem of positive intervals in $\mathbb{R}$ (it is easy to see that ϕ maps points with the same radius to a unique point in $[0, +\infty[$), and consequently we may write that

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

which is independant of d. As the largest value of N for which $m_{\mathcal{H}}(N) = 2^N$ is 2 ($m_{\mathcal{H}}(3) = 7$), we have that $d_{VC} = 2$.

Problem 2.4

We proceed by constructing a specific set of dichotomies for N points so that among the 2^N possible dichotomies on N points, we select those that contain at most $k - 1$ points labelled (-1). More precisely, we consider the following dichotomies.

- The dichotomies that contain no (-1). We have only $1 = \binom{N}{0}$ of those.
- The dichotomies that contain a unique (-1). We have $N = \binom{N}{1}$ of those.
- The dichotomies that contain exactly two (-1)s. We have $\binom{N}{2}$ of those.
- $\cdots$
- The dichotomies that contain exactly $k - 1$ (-1)s. We have $\binom{N}{k-1}$ of those.

In total, we have exactly $\sum_{i=0}^{k-1} \binom{N}{i}$ such dichotomies. Moreover, these dichotomies do not shatter any subset of k variables because to do that, we would need one dichotomy that contains k (-1)s, which is not the case in our set. So, we may conclude that

$$B(N, k) \geq \sum_{i=0}^{k-1} \binom{N}{i}$$

and with Sauer's lemma, we get

$$B(N, k) = \sum_{i=0}^{k-1} \binom{N}{i}.$$

Problem 2.5

To prove the inequality, we begin with the case $D = 0$. Here, it is easy to see that

$$1 = \binom{N}{0} \leq N^0 + 1 = 2.$$

Now, we assume the result is correct for D ($D \geq 1$), and we will prove it for $D + 1$. We may write that

$$\begin{aligned}\sum_{i=0}^{D+1}\binom{N}{i} &= \sum_{i=0}^{D}\binom{N}{i}+\binom{N}{D+1}\\ &\leq N^D+1+\binom{N}{D+1}\\ &\leq N^D+1+\frac{N!}{(D+1)!(N-D-1)!}.\end{aligned}$$

To continue, we have to prove that

$$\frac{N!}{(N-D-1)!}\leq N^{D+1},$$

which is equivalent to

$$\frac{1}{N^{D+1}}\cdot\frac{N!}{(N-D-1)!}\leq 1.$$

To see this, it suffices to note that

$$\frac{1}{N^{D+1}}\cdot\frac{N!}{(N-D-1)!}=\frac{1}{N^{D+1}}\prod_{i=0}^{D}(N-i)=\prod_{i=0}^{D}\frac{N-i}{N^{D+1}}\leq 1.$$

So, we are now able to write that

$$\begin{aligned}\sum_{i=0}^{D+1}\binom{N}{i} &\leq N^D+1+\frac{N!}{(D+1)!(N-D-1)!}\\ &\leq N^D+1+\frac{N^{D+1}}{(D+1)!}.\end{aligned}$$

As $D\geq 1$, we have $(D+1)!\geq 2$, and consequently

$$\frac{1}{(D+1)!}\leq\frac{1}{2},$$

which enables us to write that

$$\begin{aligned}\sum_{i=0}^{D+1}\binom{N}{i} &\leq N^D+1+\frac{N^{D+1}}{(D+1)!}\\ &\leq N^D+1+\frac{N^{D+1}}{2}.\end{aligned}$$

Moreover, as we assumed $N\geq D+1$ (if not, we trivially have the result, as in this case $\binom{N}{D+1}=0$), we get $N\geq 2$ and consequently

$$\frac{1}{N}<\frac{1}{2}\Leftrightarrow\frac{N^D}{N^{D+1}}<\frac{1}{2}\Leftrightarrow N^D<\frac{N^{D+1}}{2},$$

which allows us to get our result as we now have

$$\begin{aligned}\sum_{i=0}^{D+1}\binom{N}{i} &\leq N^D+1+\frac{N^{D+1}}{2}\\ &\leq \frac{N^{D+1}}{2}+1+\frac{N^{D+1}}{2}=N^{D+1}+1.\end{aligned}$$

Problem 2.6

As we have $N \geq d$, we may write that $N/d \geq 1$, and also that $(N/d)^{d-i} \geq 1$ for $i = 0, \cdots, d$. Now, we have that

$$\begin{aligned}
\sum_{i=0}^{d} \binom{N}{i} &= \sum_{i=0}^{d} \binom{N}{i} \cdot 1 \\
&\leq \sum_{i=0}^{d} \binom{N}{i} \left(\frac{N}{d}\right)^{d-i} \\
&\leq \left(\frac{N}{d}\right)^{d} \sum_{i=0}^{d} \binom{N}{i} \left(\frac{d}{N}\right)^{i} \\
&\leq \left(\frac{N}{d}\right)^{d} \sum_{i=0}^{N} \binom{N}{i} \left(\frac{d}{N}\right)^{i}.
\end{aligned}$$

Moreover, we also have that

$$\begin{aligned}
\sum_{i=0}^{N} \binom{N}{i} \left(\frac{d}{N}\right)^{i} &= \sum_{i=0}^{N} \binom{N}{i} 1^{N-i} \cdot \left(\frac{d}{N}\right)^{i} \\
&= \left(1 + \frac{d}{N}\right)^{N} \leq e^{d}.
\end{aligned}$$

In conclusion, we have proven that

$$\sum_{i=0}^{d} \binom{N}{i} \leq \left(\frac{N}{d}\right)^{d} \cdot e^{d} = \left(\frac{eN}{d}\right)^{d}.$$

As we already know that

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{VC}} \binom{N}{i},$$

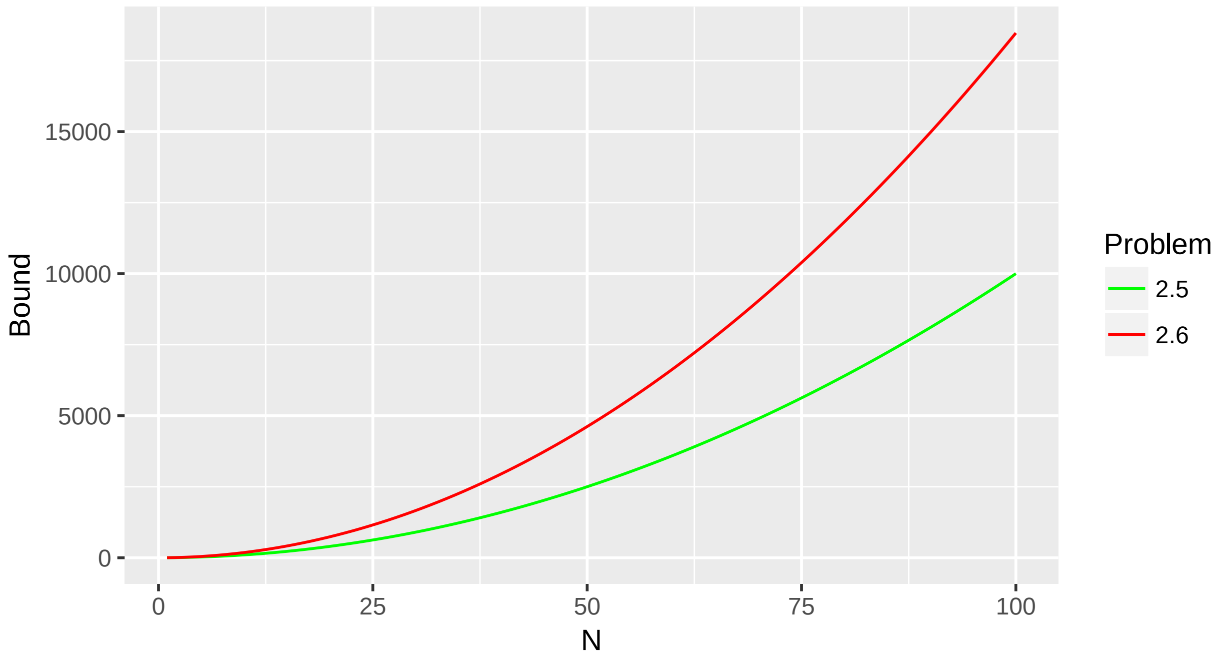
we immediately get that

$$m_{\mathcal{H}}(N) \leq \left(\frac{eN}{d_{VC}}\right)^{d_{VC}}$$

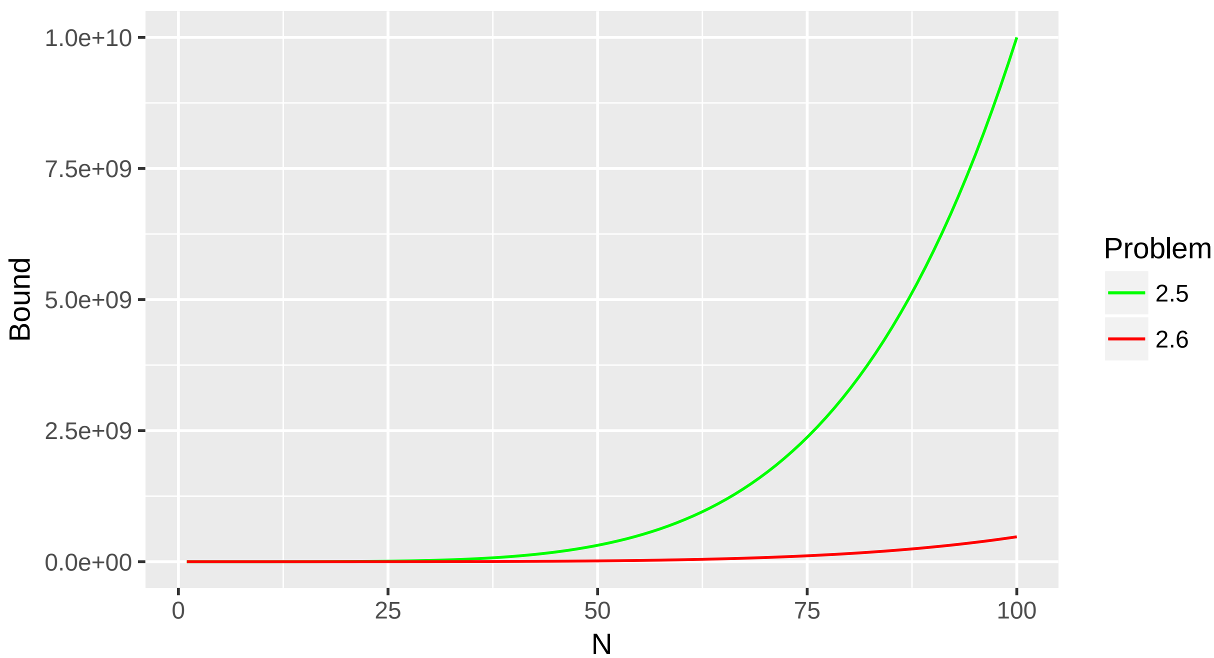
for $N \geq d_{VC}$.

Problem 2.7

We plot below the bounds for $m_{\mathcal{H}}(N)$ given in Problems 2.5 and 2.6 for $d_{VC} = 2$.



Now, we do the same for $d_{VC} = 5$.



For small VC dimensions ($d_{VC} = 2$), it seems that the polynomial bound (Problem 2.5) is better than the exponential one (Problem 2.6); however, for bigger VC dimensions ($d_{VC} = 5$), the exponential bound is way better than the polynomial one.

Problem 2.8

We have only two cases for the growth function : either $d_{VC} = +\infty$ and $m_{\mathcal{H}}(N)$ is equal to 2^N for all N, or d_{VC} is finite and $m_{\mathcal{H}}(N)$ is bounded by $N^{d_{VC}} + 1$.

If $m_{\mathcal{H}}(N) = 1 + N$, we have $d_{VC} = 1$ (as $m_{\mathcal{H}}(2) = 3 < 2^2$). So it must be bounded by $N + 1$ for all N, which is obviously the case here. In conclusion, $m_{\mathcal{H}}(N) = N + 1$ is a possible growth function.

If $m_{\mathcal{H}}(N) = 1 + N + N(N-1)/2$, we have $d_{VC} = 2$ (as $m_{\mathcal{H}}(3) = 7 < 2^3$). So it must be bounded by $N^2 + 1$ for all N, which is also the case as $N \geq 1$. In conclusion, $m_{\mathcal{H}}(N) = 1 + N + N(N-1)/2$ is a possible growth function.

Obviously $m_{\mathcal{H}}(N) = 2^N$ is a possible growth function (when $d_{VC} = +\infty$).

If $m_{\mathcal{H}}(N) = 2^{\lfloor \sqrt{N} \rfloor}$, we have $d_{VC} = 1$ (as $m_{\mathcal{H}}(2) = 2 < 2^2$). Consequently, it must be bounded by $N + 1$ for all N, which is not true (for $N = 25$ for example). In conclusion, $m_{\mathcal{H}}(N) = 2^{\lfloor \sqrt{N} \rfloor}$ is not a possible growth function.

If $m_{\mathcal{H}}(N) = 2^{\lfloor N/2 \rfloor}$, we have $d_{VC} = 0$ (as $m_{\mathcal{H}}(1) = 1 < 2^1$). Consequently, it must be bounded by $N^0 + 1 = 2$ for all N, which is not true (for $N = 4$ for example). In conclusion, $m_{\mathcal{H}}(N) = 2^{\lfloor N/2 \rfloor}$ is not a possible growth function.

Problem 2.9

Problem 2.10

Let us begin with an example : let us say we have 3 ways to dichotomize two points x_1, x_2 ($[1,1]$, $[1,-1]$ and $[-1,1]$) and 2 ways to dichotomize another two points x_3, x_4 ($[1,-1]$ and $[-1,-1]$). So, for each of the 3 ways for the first two points there are at most 2 ways to dichotomize the second two points. In this case, we have at most $3 \times 2 = 6$ ways to dichotomize all four points ($[1,1,1,-1]$, $[1,1,-1,-1]$, $[1,-1,1,-1]$,$[1,-1,-1,-1]$, $[-1,1,1,-1]$, $[-1,1,-1,-1]$).

With this reasoning, let us say that $m_{\mathcal{H}}(N) = p$, now if we partition any set of $2N$ points into two sets of N points each, each of these two partitions will produce p dichotomies at most. If we now combine these two sets, then the maximum number of dichotomies will be the product of p by p. We may conclude that

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

If we combine the result above with the VC generalization bound, we get that

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

Problem 2.11

In the case where $N = 100$, the VC generalization bound tells us that

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{100} \ln \frac{4(2 \cdot 100 + 1)}{0.1}} = E_{in}(g) + 0.8481596$$

with probability at least 90%. When $N = 10000$, we get that

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{10000} \ln \frac{4(2 \cdot 10000 + 1)}{0.1}} = E_{in}(g) + 0.1042782$$

with probability at least 90%.

Problem 2.12

We have the following implicit bound for the sample complexity N (with $d_{VC} = 10$, $\epsilon = 0.05$, and $\delta = 0.05$),

$$N \geq \frac{8}{0.05^2} \ln\left(\frac{4[(2N)^{10} + 1]}{0.05}\right).$$

To determine N, we will use an iterative process with an initial guess of $N = 1000$ in the RHS. We get

$$N \geq \frac{8}{0.05^2} \ln\left(\frac{4[(2 \cdot 1000)^{10} + 1]}{0.05}\right) \approx 2.57251 \times 10^5.$$

We then try the new value $N = 2.57251 \times 10^5$ in the RHS and iterate this process, rapidly converging to an estimate of $N \approx 4.52957 \times 10^5$.