Problem Solutions

Chapter 5

Pierre Paquay

Problem 5.1

(a) Let f be an arbitrary binary function. If $\mathcal{H}$ shatters $x_1, \cdots, x_N$, there exists $h \in \mathcal{H}$ such that $h(x_n) = f(x_n)$ for all $n = 1, \cdots, N$, and consequently $E_{in}(h) = 0$; so the proposition is not falsifiable in this case.

(b) Since the data is generated from a random (arbitrary) target function, then every dichotomy is equally likely, which means that their probability is $1/2^N$. With that in mind, we have that

$$\begin{aligned}
\mathbb{P}[\text{falsification}] &= 1 - \mathbb{P}[\exists h \in \mathcal{H} : E_{in}(h) = 0] \\
&= 1 - \mathbb{P}[\exists h \in \mathcal{H} : h(x_n) = f(x_n)\ \forall 1 \leq n \leq N] \\
&= 1 - \frac{\text{Number of dichotomies on } x_1, \cdots, x_N \text{ such that } h(x_n) = f(x_n)\ \forall 1 \leq n \leq N}{\text{Number of dichotomies}} \\
&\geq 1 - \frac{m_{\mathcal{H}}(N)}{2^N}.
\end{aligned}$$

(c) If $d_{VC} = 10$ and $N = 100$, we get that

$$\mathbb{P}[\text{falsification}] \geq 1 - \frac{N^{d_{VC}} + 1}{2^N} = 1 - \frac{100^{10} + 1}{2^{100}} \approx 1.$$

Problem 5.2

(a) Since $\mathcal{H}_i \subset \mathcal{H}_{i+1}$, we know that $|\mathcal{H}_i| \leq |\mathcal{H}_{i+1}|$, and

$$E_{in}(g_i) = \min_{h \in \mathcal{H}_i} E_{in}(h) \geq \min_{h \in \mathcal{H}_{i+1}} E_{in}(h) = E_{in}(g_{i+1})$$

for any $i = 1, 2, \cdots$.

(b) Let $p_i = \mathbb{P}[g^* \in \mathcal{H}_i] = \mathbb{P}[g^* = g_i]$, so if p_i is small then $\Omega(\mathcal{H}_i)$ is large, which implies that the model is complex.

(c) It is obvious that

$$g^* \in \mathcal{H}_i \Rightarrow g^* \in \mathcal{H}_{i+1},$$

thus we get that

$$p_i = \mathbb{P}[g^* \in \mathcal{H}_i] \leq \mathbb{P}[g^* \in \mathcal{H}_{i+1}] = p_{i+1}$$

for any $i = 1, 2, \cdots$.

(d) We know from the generalization bound that

$$\begin{aligned}
\mathbb{P}[|E_{in}(g_i) - E_{out}(g_i)| > \epsilon | g^* = g_i] &\leq \frac{\mathbb{P}[|E_{in}(g_i) - E_{out}(g_i)| > \epsilon \cap g^* = g_i]}{\mathbb{P}[g^* = g_i]} \\
&\leq \frac{\mathbb{P}[|E_{in}(g_i) - E_{out}(g_i)| > \epsilon]}{\mathbb{P}[g^* = g_i]} \\
&\leq \frac{4 m_{\mathcal{H}_i}(2N) e^{-\epsilon^2 N/8}}{p_i}.
\end{aligned}$$

Problem 5.3

(a) Here, we consider only one model, so $M = 1$.

(b) For $M = 1$, $N = 10000$, and $\epsilon = 0.02$, the Hoeffding inequality tells us that

$$\mathbb{P}[|E_{in}(g) - E_{out}(g)| > 0.02] \leq 2 \cdot 1 \cdot e^{-2 \cdot 0.02^2 \cdot 10000} = 6.7 \times 10^{-4}.$$

(c) One possible reason is the sampling bias : the data set contains only data about people who did get a credit card, we have no information on people who were rejected. So, when such people are passed to our g function, the results are not as good as predicted.

(d) Yes, we should have used our function g on the entire data set of clients (approved and not approved), in this case only would we have got a meaningful probabilistic guarantee.