Theory of Machine Learning: Homework 6

1. Let $\mathcal{H}_i$ for $i \in [r]$ be hypothesis classes over $\mathcal{X}$ and let $\mathrm{VCdim}(\mathcal{H}_i) \leq d$ for all $i \in [r]$. Assume $d \geq 3$. Let $\mathcal{H} = \bigcup_{i=1}^r \mathcal{H}_i$. We want to show that

$$\mathrm{VCdim}\left(\bigcup_{i=1}^r \mathcal{H}_i\right) < 4d\log_2(2d) + 2\log_2(r)$$

Proof. Let $\mathcal{H}_C = \{(h(x_1), \ldots, h(x_m)) \,:\, h \in \mathcal{H}\}$ be all possible dichotomies of a dataset C under the class $\mathcal{H}$, s.t. $|C| = m$. Then the growth function returns the maximum size of $\mathcal{H}_C$ over all datasets of size m in $\mathcal{X}$: $\tau_{\mathcal{H}}(m) = \max_{C\subseteq\mathcal{X},\, |C|=m} |\mathcal{H}_C|$. So,

$$\tau_{\mathcal{H}}(m) = \max_{C\subseteq\mathcal{X},\, |C|=m} \left|\bigcup_{i=1}^r \mathcal{H}_{i_C}\right|$$

The counting measure $(|\cdot|)$ is subadditive:

$$\left|\bigcup_{i=1}^r \mathcal{H}_{i_C}\right| \leq \sum_{i=1}^r |\mathcal{H}_{i_C}| \implies \tau_{\mathcal{H}}(m) \leq \max_{C\subseteq\mathcal{X},\, |C|=m} \sum_{i=1}^r |\mathcal{H}_{i_C}|$$

Note that

$$\sum_{i=1}^r |\mathcal{H}_{i_C}| \leq \sum_{i=1}^r \max_{C\subseteq\mathcal{X},\, |C|=m} |\mathcal{H}_{i_C}| \implies \max_{C\subseteq\mathcal{X},\, |C|=m} \sum_{i=1}^r |\mathcal{H}_{i_C}| \leq \sum_{i=1}^r \max_{C\subseteq\mathcal{X},\, |C|=m} |\mathcal{H}_{i_C}|$$

So,

$$\tau_{\mathcal{H}}(m) \leq \sum_{i=1}^r \max_{C\subseteq\mathcal{X},\, |C|=m} |\mathcal{H}_{i_C}| = \sum_{i=1}^r \tau_{\mathcal{H}_i}(m)$$

By Sauer's lemma, for $m \geq d$, we have $\tau_{\mathcal{H}_i}(m) \leq (em/d)^d$. Since $d \geq 3 > e$, we find $e/d < 1 \implies \tau_{\mathcal{H}_i}(m) < m^d$. So,

$$\tau_{\mathcal{H}}(m) \leq \sum_{i=1}^r \tau_{\mathcal{H}_i}(m) \implies \tau_{\mathcal{H}}(m) < rm^d$$

The VC-Dimension of $\mathcal{H}$ is the largest m s.t. $\tau_{\mathcal{H}}(m) = 2^m$ (for binary classification, 2^m is the maximum value that the growth function may attain).

As instructed, we assume that Lemma A.2 holds for log base 2 with similar conditions. Given $a \geq 1$, $b > 0$. Then,

$$x \geq 4a\log_2(2a) + 2b \implies x \geq a\log_2(x) + b$$

By contrapositive,

$$x < a\log_2(x) + b \implies x < 4a\log_2(2a) + 2b$$

Let $a = d \geq 1$ and $b = \log_2(r)$. $\log_2(r) > 0$ only when $r > 1$. Assume for now that $r > 1$. It must hold for all $m \leq \mathrm{VCdim}(\mathcal{H})$ that

$$2^m < rm^d \tag{1}$$
$$m < d\log_2(m) + \log_2(r) \tag{2}$$
$$m < 4d\log_2(2d) + 2\log_2(r) \text{ (contrapositive of A.2)} \tag{3}$$

The strict inequality in (2) is only satisfied by finite m, because $\log_2(m) = o(m)$ (little-o). The VC-Dimension must satisfy the inequality, so $\mathrm{VCdim}(\mathcal{H})$ is finite. Since (3) holds for $m \leq \mathrm{VCdim}(\mathcal{H})$,

$$\mathrm{VCdim}(\mathcal{H}) < 4d\log_2(2d) + 2\log_2(r)$$

If $r = 1$, then $\mathcal{H} = \mathcal{H}_1$, so $\mathrm{VCdim}(\mathcal{H}) = \mathrm{VCdim}(\mathcal{H}_1) \leq d$. We need $d < 4d\log_2(2d) + 2\log_2(r) = 4d\log_2(2d)$. This is equivalent to $1 < 4\log_2(2d)$, which would require $2^{1/4} < 2d \implies 2^{-3/4} < d$. This is satisfied by any $d \geq 1$, so the bound holds. QED