CPSC 531H: Machine Learning Theory
Assignment #2, due Wednesday October 10th, email to Chris

[5] 1. SSBD Exercise 4.1.

[10] 2. **Modifying the No-Free-Lunch Theorem.** Suppose that $m \le |\mathcal{X}|/k$. Fix any learning algorithm. Let h_S be the hypothesis output by the algorithm on input S. Prove that

$$\Pr_S\left[\, L_{\mathcal{D},f}(h_S) \ge \tfrac{1}{2}\big(1-\tfrac{3}{k}\big) \,\right] \;\ge\; \frac{1}{2k}.$$

[10] 3. **VC-dimension practice.**

[4] a. SSBD Exercise 6.2 part 1 only.

[6] b. SSBD Exercise 6.5.

[8] 4. **Improving the Fundamental Theorem.** The proof of the Fundamental Theorem (either SSBD Theorem 6.7, or the in-class version) proves that

$$m_{\mathcal{H}}(\epsilon,\delta) \;\le\; c\frac{d\log(d/\delta\epsilon)}{(\delta\epsilon)^2}$$

for some constant c, where $d = \mathrm{VCdim}(\mathcal{H})$. This sample complexity bound is quite poor, in that it depends polynomially on $1/\delta$. Chapter 28 of the textbook improves this to

$$m_{\mathcal{H}}(\epsilon,\delta) \;\le\; O\Big(\frac{d\log(d/\epsilon)+\log(1/\delta)}{\epsilon^2}\Big),$$

but this analysis uses more sophisticated machinery (Rademacher complexity). In this question, we will use more basic ideas to prove a bound of intermediate quality

$$m_{\mathcal{H}}(\epsilon,\delta) \;\le\; O\Big(\frac{d\log(d/\epsilon)\log(1/\delta)}{\epsilon^2}\Big). \tag{1}$$

[2] a. Define m_{FT} by plugging $\delta = 1/2$ to the Fundamental Theorem to get

$$m_{\mathrm{FT}} \;:=\; 4cd\log(2d/\epsilon)/\epsilon^2.$$

Define $k = \lg(1/\alpha)$. Draw independent training sequences $S_1,\ldots,S_k$, each of which consists of m_{FT} i.i.d. samples. Let $\mathcal{E}$ be the event that at least one of the training sequences is an ϵ-representative sample. Prove that $\Pr[\,\mathcal{E}\,] \ge 1-\alpha$.

[2] b. Let h_i be the output of ERM on training sequence S_i. Draw a validation sequence V which consists of m_v i.i.d. samples. Let $L_V(h_i)$ denote the error of h_i on the validation sequence. Let $\mathcal{F}$ be the event that

$$|L_V(h_i) - L_{\mathcal{D}}(h_i)| \;\le\; \epsilon \qquad \forall i \le k.$$

Choose the number m_v of validation samples to ensure that $\Pr[\,\mathcal{F}\,] \ge 1-\alpha$.

[3] c. Our learning algorithm will pick the hypothesis with minimum validation error, i.e., $\hat{i} = \mathrm{argmin}_{1\le i\le k} L_V(h_i)$. Let $h^* \in \mathrm{argmin}_{h\in\mathcal{H}} L_{\mathcal{D}}(h)$. Prove that

$$\Pr\left[\, L_{\mathcal{D}}(h_{\hat{i}}) \le L_{\mathcal{D}}(h^*) + 4\epsilon \,\right] \;\ge\; 1-2\alpha.$$

[1] d. Conclude that the sample complexity bound (1) holds.

[6] 5. **Small-step Perceptron.** SSBD Exercise 9.5.

[3] 6. **OPTIONAL BONUS! VC-dimension of unions.** SSBD Exercise 6.11 (parts 1 & 2).

Remarks.

- I'm not sure that the constant in part 1 are completely correct. So let's just make the goal of part 1 to prove the bound $O(d \log d + \log r)$.
- After solving part 2, can you see how to improve the bound of part 1 to $O(d + \log r)$?