1. SSBD Exercise 4.1.

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

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

3. **VC-dimension practice.**
   a. SSBD Exercise 6.2 part 1 only.
   b. SSBD Exercise 6.5.

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

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

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

$$m_{\mathcal{H}}(\epsilon, \delta) \leq O\left(\frac{d \log(d/\epsilon) + \log(1/\delta)}{\epsilon^2}\right),$$

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

$$m_{\mathcal{H}}(\epsilon, \delta) \leq O\left(\frac{d \log(d/\epsilon) \log(1/\delta)}{\epsilon^2}\right). \quad (1)$$

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

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

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

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 $F$ be the event that

$$|L_V(h_i) - L_{\mathcal{D}}(h_i)| \leq \epsilon \quad \forall i \leq k.$$

Choose the number $m_v$ of validation samples to ensure that $\Pr[F] \geq 1 - \alpha$.

c. Our learning algorithm will pick the hypothesis with minimum validation error, i.e.,

$$\hat{i} = \text{argmin}_{1 \leq i \leq k} L_V(h_i).$$

Let $h^* \in \text{argmin}_{h \in \mathcal{H}} L_{\mathcal{D}}(h)$. Prove that

$$\Pr \left[ L_{\mathcal{D}}(h_{\hat{i}}) \leq L_{\mathcal{D}}(h^*) + 4\epsilon \right] \geq 1 - 2\alpha.$$
[1] d. Conclude that the sample complexity bound (1) holds.


**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 $O(d + \log r)$?