CSE522 - Learning Theory - Homework Exercise 1

Let $\mathcal{X}$ be an instance space, let $\mathcal{Y}$ be a label set, let $\mathcal{D}$ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$, and let ℓ be a loss function that takes values in $[0, C]$.

1. Let $\mathcal{H} = \{h_i\}_{i=1}^{\infty}$ be a countable hypothesis class and let S be a sample of m examples drawn i.i.d. from the distribution $\mathcal{D}$. Prove that for any $\delta > 0$, with probability at least $1 - \delta$, S satisfies

$$\forall i \qquad \ell(h_i; \mathcal{D}) \;\leq\; \ell(h_i; S) \;+\; C\sqrt{\frac{\log(\frac{1}{\delta}) + 2\log(i+1)}{2m}} \quad .$$

2. Let $\mathcal{H} = \{h_i\}_{i=1}^{k}$ be a finite hypothesis space. Let $h^\star$ be the hypothesis with the smallest risk in $\mathcal{H}$, and assume that

$$\forall h \in \mathcal{H} \setminus \{h^\star\} \qquad \ell(h; \mathcal{D}) - \ell(h^\star; \mathcal{D}) > \frac{C}{100} \quad .$$

   What size sample guarantees that $h^\star$ is also the *empirical* risk minimizer, with probability at least 0.99 ?

3. Let h_1 and h_2 be two hypotheses such that

$$|\ell(h_1; \mathcal{D}) - \ell(h_2; \mathcal{D})| \;>\; \frac{C}{100} \quad .$$

   We can sample as many examples as we need, but each example costs \$1. Propose a greedy algorithm that finds the risk minimizer with probability at least 0.99, while spending as little as possible. (Hint: prove a bound that holds uniformly for different sample sizes).

4. Let $\mathcal{H} = \{h_i\}_{i=1}^{k}$ be a finite hypothesis class and let $S_1, \ldots, S_k$ be independent samples from $\mathcal{D}^m$ (namely, each S_i contains m independent examples). Let $\delta > 0$ and $\epsilon > 0$ be such that for any i, with probability $1 - \delta$, we know that

$$|\ell(h_i; \mathcal{D}) - \ell(h_i; S_i)| \;\leq\; \epsilon \quad .$$

   Consider the event

$$\forall\, i \in \{1, \ldots, k\} \quad |\ell(h_i; \mathcal{D}) - \ell(h_i; S_i)| \;\leq\; \epsilon \quad .$$

(a) Lower-bound the probability of this event using a union bound. Note that this bound holds even if all of the empirical risks are computed using one sample, S_1.

(b) Calculate the exact probability of this event (use the fact that the k samples are independent) and upper bound this probability using a second-order Taylor expansion.

(c) Compare the lower bound to the upper-bound and conclude that it doesn't make much difference if we use k independent samples or one sample.