A. Structural Risk Minimization

As discussed in class, the Structural Risk Minimization (SRM) technique is based on a hypothesis set $\mathcal{H}$ defined as a countable union of hypothesis sets $\mathcal{H}_n$ with finite VC-dimension or favorable Rademacher complexity. In this problem, we study several questions related to such countable union hypothesis sets.

1. Let $\mathcal{H} = \bigcup_{n=1}^{+\infty} \{h_n\}$ be a countable hypothesis set and assume that the target labeling function is in $\mathcal{H}$. In the standard statistical learning scenario, the learner receives an i.i.d. sample that he uses to train an algorithm and return a predictor. Here, suppose instead that the learner can request more labeled samples drawn i.i.d., as needed. Consider the following algorithm: starting from $t = 1$, at each round $t$, sample $m_t = \frac{1}{\epsilon} \log \frac{1}{\delta_t}$ labeled points; if $h_t$ is consistent with $m_t$, return $h_t$ and stop.

   (a) Prove that the algorithm terminates.

   (b) Fix $\epsilon, \delta > 0$ and choose $\delta_t = \frac{\delta}{2^t}$. Show that with probability $1 - \delta$, the algorithm returns a hypothesis with error at most $\epsilon$. Suppose we use the samples obtained from previous rounds to test consistency, then, what is the maximum number of samples needed by the algorithm?

   (c) Can you generalize these results to the case where $\mathcal{H} = \bigcup_{n=1}^{+\infty} \mathcal{H}_n$ with VCdim($\mathcal{H}_n$) = $d_n < +\infty$?

2. Suppose $S$ is an infinite set that can be fully shattered by $\mathcal{H}$. We wish to show that $\mathcal{H}$ cannot be written as a countable union $\mathcal{H} = \bigcup_{n=1}^{+\infty} \mathcal{H}_n$ with VCdim($\mathcal{H}_n$) = $d_n < +\infty$.

   (a) Show that we can define a family of subsets $(X_n)_{n \geq 1}$ such that $|X_n| = d_n + 1$ and $X_n \subseteq S - \bigcup_{1 \leq k \leq n-1} X_k$. 

1
(b) Show that for any \( n \geq 1 \), there exists a labeling \( X_n^l \) that cannot be obtained using \( \mathcal{H}_n \).

(c) Consider the labeling \( X^l \) of \( X = \bigcup_{n=1}^{+\infty} X_n \) obtained using all the \( X_n^l \)s. Show that no labeling of \( S \) using \( \mathcal{H} \) can be consistent with \( X^l \). Conclude that that \( \mathcal{H} \) cannot be written as a countable union \( \mathcal{H} = \bigcup_{n=1}^{+\infty} \mathcal{H}_n \) with \( \text{VCdim}(\mathcal{H}_n) = d_n < +\infty \).

3. Suppose you only know an upper bound \( \alpha_n \) on \( \text{VCdim}(\mathcal{H}_n) = d_n < +\infty \) with \( \sum_{n=1}^{+\infty} e^{-\alpha_n} < +\infty \). Give a generalization bound for the SRM-type algorithm defined by

\[
\hat{f} = \arg\min_{k \geq 1, h \in \mathcal{H}_k} \hat{R}_S + \sqrt{\frac{32\alpha_k \log(em)}{m}},
\]

for a sample \( S \) of size \( m \).

**B. Learning kernels**

Let \( \mathcal{K} \) be the family of all Gaussian kernels defined over \( \mathbb{R}^N \):

\[
\mathcal{K} = \left\{ K_\gamma : K_\gamma(x,x') = e^{-\gamma \|x-x'\|^2}, \forall x, x' \in \mathbb{R}^N, \gamma > 0 \right\}.
\]

Consider the hypothesis set defined via the reproducing kernel Hilbert space of the kernels in \( \mathcal{K} \):

\[
\mathcal{H} = \left\{ h : h \in \mathbb{H}_K, K \in \mathcal{K}, \|h\|_{\mathbb{H}_K} \leq 1 \right\}.
\]

1. Let \( S = (x_1, \ldots, x_m) \) be a sample of size \( m \). Show that \( \hat{R}_S(\mathcal{H}) = \frac{1}{m} \mathbb{E}_\sigma \left[ \sqrt{\sup_{\gamma > 0} \sigma^T K_\gamma \sigma} \right] \), where \( K_\gamma \) is the Gram matrix of kernel \( K_\gamma \) for the sample \( S \).

2. Suppose \( \|x_i - x_j\| = 1 \) for \( i \neq j \). Compute exactly \( \hat{R}_S(\mathcal{H}) \).