Problem 1. Sample complexity is defined as the quantity \(N(\epsilon, \delta)\) such that when the learning algorithm is provided with \(n > N(\epsilon, \delta)\) number of training examples, then with probability \(1 - \delta\), the excess risk is no bigger than the quantity \(\epsilon\). Find out the sample complexity for a class of function \(\mathcal{H}\) with finite VC dimension \(d\).

Problem 2. Show that the function class \(\mathcal{H}\) and \(\mathcal{F} = \{f(x, y) : f(x, y) = \ell(y, h(x)), \forall h \in \mathcal{H}\}\) has the same growth function, where \(\ell(y, p) = I(y \neq p)\) is the 0/1 classification error.

Problem 3. Consider the function class \(\mathcal{H} = \{\text{sign} \left( \sum_{i=1}^{K} a_i \phi_i(x) \right) : a_1, \ldots, a_K \in \mathbb{R}\}\) which is the span for \(K\) fixed functions \(\phi_1, \ldots, \phi_K\). What is the VC dimension of \(\mathcal{H}\)? Prove your result.

Problem 4. Let \(\mathcal{H}\) be a class of classifiers \(h: \mathbb{R}^p \mapsto \{0, 1\}\), and assume that \(\hat{h}^*\) is the classifier which minimizes the empirical classification error

\[
\hat{h}^* = \arg\min_{h \in \mathcal{H}} \hat{R}_n(h) = \arg\min_{h \in \mathcal{H}} \sum_{i=1}^{n} I(y_i \neq h(x_i)).
\]

Assume that we have a learning algorithm which outputs a classifier \(\hat{h}_n\) such that

\[
P \left( \hat{R}_n(\hat{h}_n) \leq \inf_{h \in \mathcal{H}} \hat{R}_n(h) + \epsilon_n \right) \geq 1 - \delta_n,
\]

i.e. we allow our output classifier to approximately minimize the empirical error with large probability. \(\{\epsilon_n\}\) and \(\{\delta_n\}\) are sequence of numbers that converge to zero.

1. Show that

\[
P \left( R(\hat{h}_n) - \inf_{h \in \mathcal{H}} R(h) > \epsilon \right) \leq \delta_n + P \left( 2 \sup_{h \in \mathcal{H}} |\hat{R}_n(h) - R(h)| > \epsilon - \epsilon_n \right).
\]

2. Find conditions on \(\{\epsilon_n\}\) and \(\{\delta_n\}\) such that

\[
\mathbb{E}[R(\hat{h}_n)] - \inf_{h \in \mathcal{H}} R(h) = O \left( \sqrt{\frac{\log n}{n}} \right),
\]

i.e. \(\mathbb{E}[R(\hat{h}_n)]\) converges to the optimum at the same order as \(\mathbb{E}[R(\hat{h}^*)]\).