Exercise Sheet 6

Exercise 1

Let $\mathcal{H}$ be a finite hypothesis class of functions $f : \mathcal{X} \to \{\pm 1\}$. Then for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{f \in \mathcal{H}} |R_n(f) - R(f)| > \varepsilon\right) \leq 2\,|\mathcal{H}| \exp\left(-2n\varepsilon^2\right).$$

Show that for any $0 < \delta < 1$ with probability at least $1 - \delta$,

$$R(f_n) - R(f_{\mathcal{H}}) \leq 2\sqrt{\frac{\log|H| + \log 2/\delta}{2n}},$$

where f_n is the model returned by the learner.

Exercise 2

A *break point* of hypothesis class $\mathcal{H}$ is the smallest n such that $m_{\mathcal{H}}(n) < 2^n$

a) Consider the hypothesis classes *decision point*, *interval*, and *convex set* defined in vl05, page 39ff. Find a break point, if any, for each hypothesis class. Plot the growth functions for each hypothesis class (use log scale for the y-axis).

b) By inspection, find a break point for the hypothesis class of axis-aligned rectangles in $\mathbb{R}^2$.

Exercise 3

Consider the following setting: The input space is $\mathcal{X} = [0, 1]$ and the output space is $\mathcal{Y} = \{0, 1\}$. The input data $x \in \mathcal{X}$ is uniformly distributed. The Bayes classifier

$$f^*(x) = \begin{cases} 1 & : \quad x \in [0.2, 0.8] \\ 0 & : \quad \text{otherwise} \end{cases}.$$

has zero Bayes risk $R^* = R(f^*) = 0$. We consider two hypotheses classes. The first hypothesis class

$$\mathcal{H}_T = \left\{f_\theta \,:\, \theta \in [0,1]\right\}$$

consists of threshold functions of the form

$$f_\theta(x) = \begin{cases} 0 & : \quad x < \theta \\ 1 & : \quad x \geq \theta \end{cases}.$$

The second hypothesis class

$$\mathcal{H}_I = \left\{f_{a,b} \,:\, 0 \leq a \leq b \leq 1\right\}$$

consists of interval functions of the form

$$f_{a,b}(x) = \begin{cases} 1 & : \quad a \leq x \leq b \\ 0 & : \quad \text{otherwise} \end{cases}.$$

a) Implement the ERM learners for both hypothesis classes.

b) Consider the following experiment for a given learner A:

1. Sample a training set $\mathcal{S}$ of size n.
2. Fit a model $A(\mathcal{S}) = f_n$
3. Compute the empirical risk $R_n(f_n)$
4. Compute the true risk $R(f_n)$
5. Estimate the probability $\mathbb{P}(|R_n(f_n) - R(f_n)| \geq \varepsilon)$ for $\varepsilon = 0.1$

Repeat the experiment T times for different sizes n and for the ERM learner w.r.t. both hypothesis classes.

Plot the VC inequalities, the probabilities $\mathbb{P}(|R_n(f_n) - R(f_n)| \geq \varepsilon)$, the average empirical risk $R_n(f_n)$ and the true risk $R(f)$ as functions of the sample size n. Use different figure or axes objects to obtain clear and readable plots. Compare the results obtained by ERM w.r.t. both hypothesis classes.