Intro to Statistical Learning Theory
Exercise 3

1) Read the lecture notes of lecture 5.

   a) Prove lemma 1.2.

   b) Prove lemma 1.3.

   c) The SVM bound that we got does not depend on the dimension. Why does this not contradict the fundamental theorem for binary classification?

2) Define $\mathcal{H}_1 = \{x \to \langle x, w \rangle : ||w||_1 \leq 1\}$ where $||x||_1 = \sum_{i=1}^{d} |x_i|$. Let $S = \{x_1, ..., x_m\}$ be vectors in $\mathbb{R}^n$. Prove that

   $$R(\mathcal{H}_1 \circ S) = R(\{\langle w, x_1 \rangle, ..., \langle w, x_m \rangle \ ||w||_2 \leq 1\}) \leq \max_i ||x||_\infty \sqrt{\frac{2\log(2n)}{m}}.$$

   Where $||x||_\infty = \max_i |x_i|$.
   Hint: You can use a result from the Holder inequality, $\langle x, y \rangle \leq ||x||_1 \cdot ||y||_\infty$ to reduce the problem to a finite set, and then use the Massart lemma.

3) Let $\mathcal{H}$ and $\mathcal{H}'$ be hypothesis classes. Either prove or give a counter example to $\mathcal{R}_\mathcal{D}(\mathcal{H} \cup \mathcal{H}', m) \leq \mathcal{R}_\mathcal{D}(\mathcal{H}, m) + \mathcal{R}_\mathcal{D}(\mathcal{H}', m)$

4) Multi-class labeling problem: For every parameter vector $\theta \in \mathbb{R}$ define the prediction function $h_\theta(x) = \sum_{i=1}^{k} \mathbb{1}[x \geq \theta_i]$, i.e. k thresholds. The loss function is $\ell(\tilde{y}, y) = |y - \tilde{y}|$. For a sample S of m i.i.d. examples, Compute a (non-trivial) upper bound on the Rademacher complexity of $\mathcal{F} \circ S = \{(\ell(y_1, h_\theta(x_1), ..., \ell(y_m, h_\theta(x_m)) : \theta \in \mathbb{R}^k\}$.