Intro to learning theory - ex 4

1. Prove lemma 2.2 and 2.3 in lecture 7.

2. Define $\ell^{0-1}(y, y') = \frac{1}{2}(sign(y' \cdot y) + 1)$ and $\ell^h(y, y') = \max\{0, 1 - y \cdot y'\}$. Prove that if $\ell(y, y')$ is convex, 1-Lipshitz in y' and $\forall y, y' : \ell^{0-1}(y, y') \leq \ell(y, y')$ then $\forall y, y' : \ell^h(y, y') \leq \ell(y, y')$. THis shows that the hinge loss is the smallest surrogate loss with these properties.

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

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

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

4. Learnability without uniform convergence: Let B_d be the unit ball in $\mathbb{R}^d$. Define $\mathcal{H} = B_d$, $Z = B_d \times \{0,1\}^d$ and The loss finction ℓ is define as $\ell(w, (x, \alpha)) = \sum_{i=1}^{d} \alpha_i (x_i - w_i)^2$. Intuitively, we need to learn the "center of mass" of the distribution only we get another vector α of binary weights which tells us which indices we can ignore for this example.

   (a) Show that this can be learned using regularized risk minimization, with sample complexity independent of d.

   (b) Consider a distribution $\mathcal{D}$ over Z as follows: x is fixed to be some x_0, and each element of α is sampled to be either 1 or 0 with equal probability. Show that the rate of uniform convergence of this problem grows with d.
   Hint: Let m be a training set size. Show that if $d >> 2^m$, then there is a high probability to sample a set of examples such that there exists some $j \in [d]$ for which$\alpha_j = 0$ for all samples in the training set.