Intro to learning theory - ex 1

1. Binary classification - finite realizable case:
   A distribution $\mathcal{D}$ is realizable by $\mathcal{H}$ if there exists some $h^* \in \mathcal{H}$ such that $L_{\mathcal{D}}(h^*) = 0$.

   Assume $\mathcal{Y} = \{\pm 1\}$, $0-1$ loss, and $\mathcal{H}$ is a finite hypothesis class. Prove that $\mathcal{H}$ can PAC learn any *realizable* distribution $\mathcal{D}$ with $\mathfrak{M}(\epsilon, \delta) = \mathcal{O}\left(\frac{\log(|\mathcal{H}|/\delta)}{\epsilon}\right)$. You can use the inequality $1 - x \leq e^{-x}$.

2. Prove a variation of Hoeffdings inequality: If X_i are i.i.d, $X_i \in [0,1]$ and $\mathbb{E}[X_i] = \mu$, then

   $$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2 n}{2(\mu + \epsilon)}\right) \tag{1}$$

   What happens when $\mu \sim 0$?

   Hint: Same as the Hoeffding proof we did, but with an alternative bound for the KL divengence (marked $f(\epsilon)$) in slide 15).

3. Bayes optimal predictor: We define the Bayes-optimal predictor h_b as $h_b(x) = \arg\min_{\bar{y} \in \mathcal{Y}} \mathbb{E}_y[\ell(\bar{y}, y)|x]$.
   a) Show that for classification with $\mathcal{Y} = \{1, ..., k\} = [k]$ and $0 - 1$ loss, $h_b(x) = \arg\max_y P(y|x)$.
   b) Show that for regression, $\mathcal{Y} = \mathbb{R}$ with ℓ_2 loss $\ell(y, \bar{y}) = (y - \bar{y})^2$ that $h_b(x) = \mathbb{E}[y|x]$.
   c) Show that for regression, $\mathcal{Y} = \mathbb{R}$ with ℓ_1 loss $\ell(y, \bar{y}) = |y - \bar{y}|$ that $h_b(x) = median P(y|x)$. You can assume that the distribution $P(y|x)$ is discrete.

4. Show that there exists a hypothesis space $\mathcal{H}$ with $|\mathcal{H}| = 2$ and an unbounded loss ℓ such that $\mathcal{H}$ is not PAC learnable.