Intro to learning theory - ex 2

1. Find the VC dimension of the following hypothesis spaces (prove your claim):

   (a) Parity functions. $\mathcal{X} = \{0,1\}^m$, for any $S \subset [n]$ define $h_S(x) = \left(\sum_{i\in S} x_i\right) \mod 2$. $\mathcal{H} = \{h_S, \forall S \subset [n]\}$.

   (b) The set of axis aligned rectangles in $\mathbb{R}^d$, i.e. $\mathcal{H} = \{h_{(c,b)} = \mathbb{1}[\forall i |x_i - c_i| \leq b_i] : b, c \in \mathbb{R}^d\}$. We have seen in class the case $d = 2$.

   (c) Let F be a linear space of real valued function with (linear) dimension d, and g be any real valued function. Define $\mathcal{H} = \{sign(f+g) : f \in F\}$.

   (d) * The set of circles in $\mathbb{R}^2$, i.e. $\mathcal{H} = \{h_{(c,r)} = \mathbb{1}[||x - c||_2 \leq r] : c \in \mathbb{R}^2, r < 0\}$

2. For $X = \mathbb{R}$, define $\mathcal{H} = \{h_\theta(x) = \lceil\sin(\theta x)\rceil, \theta \in \mathbb{R}\}$ where we take $\lceil -1 \rceil = 0$. Prove that $VC(\mathcal{H}) = \infty$.
   Hint: prove and use the following lemma - if $x \in (0,1)$ has binary expansion $x = 0.x_1x_2...x_m...$ then for any natural number m, $\lceil\sin(2^m\pi x)\rceil = 1 - x_m$ provided that for some $k > m$ we have $x_k = 1$.

3. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be binary hypothesis spaces over $\mathcal{X}$. define $d_i = VC(\mathcal{H})$, $d = \max(d_1, d_2)$ and assume $d \geq 3$. Prove that $VC(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 2d + 1$.

4. From bounded expected risk to agnostic PAC learning: Let A be an algorithm that guarantees the following: If $m > \mathfrak{M}(\epsilon)$ then for every distribution $\mathcal{D}$ it holds that $\mathbb{E}_S[L_D(A(S))] < \min_{h\in\mathcal{H}} L_D(h) + \epsilon$.

   (a) Show that for every $\delta \in (0,1)$, if $m > \mathfrak{M}(\epsilon \cdot \delta)$ then with probability of at least $1 - \delta$ it holds that $L_{\mathcal{D}}(A(S)) < \min_{h\in\mathcal{H}} L_{\mathcal{D}}(h) + \epsilon$ (hint: Markov's inequality).

   (b) For every $\delta \in (0,1)$ let $k = \lceil \log_2(1/\delta) + 1 \rceil$ and $\bar{\mathfrak{M}}(\epsilon, \delta) = \mathfrak{M}(\epsilon/2)k + \left\lceil 2\frac{\ln(2/\delta)+\ln(k)}{\epsilon^2} \right\rceil$ Suggest a procedure that PAC learns the problem with sample complexity of $\bar{\mathfrak{M}}(\epsilon/2, \delta)$ assuming that the loss function is bounded by 1.
   Hint: Divide the data into $k+1$ chunks, where each of the first k chunks is of size $\mathfrak{M}(\epsilon/2)$ examples.