Intro to learning theory - ex 3

1. Let $\mathcal{H}$ be the set of all functions from $[0,1]$ to $[0,1]$ with total variation bounded by V. Prove that $fat_{\mathcal{H}}(\gamma) = 1 + \lfloor \frac{V}{2\gamma} \rfloor$

2. Structural Risk Minimization: If $\mathcal{H}$ has uniform convergence with complexity $\mathfrak{M}(\epsilon,\delta)$ we define the confidence $\epsilon(m,\delta) = \min_{\epsilon>0}\{m > \mathfrak{M}(\epsilon,\delta)\}$, i.e. the best approximation error we can guarantee given m examples with probability δ. Prove the following theorem: Let p_n be a sequence of positive numbers such that $\sum_{i=1}^{\infty} p_n \leq 1$. Let $\mathcal{H} = \cup_{i=1}^{\infty}\mathcal{H}_n$ where $\mathcal{H}_n$ has uniform convergence with complexity $\mathfrak{M}_n(\epsilon,\delta)$ and confidence $\epsilon_n(m,\delta)$. For any distribution $\mathcal{D}$ we have with probability at least $1-\delta$ over $S \sim \mathcal{D}^m$

   $$\forall h \in \mathcal{H}, \quad L_{\mathcal{D}}(h) \leq L_S(h) + \min_{n:h\in\mathcal{H}} \epsilon_n(m, p_n \cdot \delta)$$

   Can you give a specific bound when $p_n = 2^{-n}$ and $VC(\mathcal{H}) = n$?

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. Toy multi-class labelling problem: For every parameter vector $\theta \in \mathbb{R}^k$ 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 $\ell(y,\bar{y}) = |y - \bar{y}|$. For a sample S of m i.i.d examples, compute a (non-trivial) upper bound on the Rademacher complexity of $\mathcal{F} \circ S = \{(\ell(y_1, h_\theta(x_1)), ..., \ell(y_m, h_\theta(x_m))) : \theta \in \mathbb{R}^k\}$

5. The Glivenko-Cantelli theorem (weaker version): Let P be a distribution on X, the cumulative distribution function (CDF) is $F(x) = P(X \leq x)$. Given a sample $S^m = \{x_1, ..., x_m\}$ the empirical CDF is defined as $F_S(x) = \frac{1}{m}\mathbb{1}[x_i \leq x]$. Prove that

   $$P_{S^m}\left(\sup_{x\in\mathbb{R}} |F_{S^m}(x) - F(x)| \geq \epsilon\right) \xrightarrow{m\to 0} 0$$

   (the GC theorem actually claims almost surely convergence) Hint: rewrite the problem as a uniform convergence problem for a set of functions $\mathcal{H}$, then bound using Rademacher complexity and the Massarat lemma.