Collaborating on homework is encouraged, but you must write your own solutions in your own words and list your collaborators. Copying someone else's solution will be considered plagiarism and may result in failing the whole course.

Please answer clearly and concisely. Explain your answers. Unexplained answers will get lower scores or even no credits.

(1) (15 points) Suppose $\mathcal{C}$ is a concept class of VC dimension d. Consider the concept class $\mathcal{C}_s = \{c_1 \cup \cdots \cup c_s \mid c_i \in \mathcal{C} \text{ for } 1 \leqslant i \leqslant s\}$ of union of s concepts from $\mathcal{C}$. Prove that $\mathrm{VCDim}(\mathcal{C}_s) \leqslant 2ds \log(es)$ for any $s \geqslant 1$.

Hint: Consider growth function. You need results mostly from Notes13, and not much else.

Explain how you get the upper bound, and fill in missing steps (marked as exercise) in lecture notes.

(2) (15 points) In AdaBoost, if base learner A always outputs a hypothesis from hypothesis class $\mathcal{H}$, then the final hypothesis of AdaBoost belongs to the hypothesis class

$$\mathcal{H}_R = \left\{ \mathrm{sign}\left(\sum_{1 \leqslant t \leqslant R} \alpha_t h_t \right) \middle| \alpha_t \in \mathbb{R} \text{ and } h_t \in \mathcal{H} \text{ for } 1 \leqslant t \leqslant R \right\}.$$

Suppose $\mathrm{VCDim}(\mathcal{H}) = d$. Prove that $\mathrm{VCDim}(\mathcal{H}_R) \leqslant O(Rd \log R)$.

(You can model AdaBoost as a neural network and apply results from Notes16. What is the direct acyclic graph G of this neural network? What are the $\mathcal{C}_j$'s?)

(3) (20 points) Suppose concept class $\mathcal{C}$ over $X = \{0,1\}^n$ can be PAC-learned using $\mathrm{poly}(n, \frac{1}{\varepsilon}, \frac{1}{\delta})$ samples. Give a (possibly slow) online algorithm for learning $\mathcal{C}$ making at most $\mathrm{poly}(n)$ mistakes.

Hint: What is the necessary and sufficient condition for $\mathcal{C}$ to be PAC-learnable by a possibly slow algorithm?

This shows that PAC-learnability with polynomial number of samples implies (and is equivalent to, thanks to Online-to-PAC conversion) $\mathrm{poly}(n)$ mistakes bound in the online model, if we allow slow algorithms.

(4) (25 points)
   (a) Consider a variant of the statistical query model in which the learning algorithm, in addition to the oracle $\mathrm{STAT}(c, \mathcal{D})$, is also given access to *unlabeled* random samples from the unknown distribution $\mathcal{D}$.
   Give an efficient algorithm to learn axis-aligned rectangles in $\mathbb{R}^2$ in this variant. Briefly justify why your algorithm works; you may get a low score if we cannot understand your justification.
   (b) Using Theorem 1 in Notes18, briefly argue how to turn your Statistical Query algorithm into an algorithm that efficiently PAC-learn axis-aligned rectangles in $\mathbb{R}^2$ with Random Classification Noise.

(5) (25 points) Give an algorithm to efficiently learn 1-decision lists over n boolean variables from Statistical Queries. Justify why your algorithm works; you may get a low score if we cannot understand your justification.

Hint: Modify the PAC learning algorithm for decision lists and its analysis. We did not cover this algorithm; consult textbook Section 2.4 or online lecture notes from other universities.