CS 151 Complexity Theory		**Spring 2004**
	Problem Set 6	
Out: May 13		*Due:* **May 20**

Reminder: you are encouraged to work in groups of two or three; however you must turn in your own write-up and note with whom you worked. You may consult the course materials and text (Papadimitriou). Please attempt all problems.

1. The following problem comes from Learning Theory, where the VC-dimension gives important information about the difficulty of learning a given concept. Given a collection $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ of subsets of a finite set U, the *VC dimension* of $\mathcal{S}$ is the size of the largest set $X \subseteq U$ such that for every $X' \subseteq X$, there is an i for which $S_i \cap X = X'$ (we say that X is *shattered* by $\mathcal{S}$). A Boolean circuit C that computes a function $f : \{0,1\}^m \times \{0,1\}^n \rightarrow \{0,1\}$ succinctly represents a collection $\mathcal{S}$ of 2^m subsets of $U = \{0,1\}^n$ as follows: the set S_i consists of exactly those elements x for which $C(i, x) = 1$. Finally, the language VC-DIMENSION is the set pairs (C, k) for which C represents a collection of subsets $\mathcal{S}$ whose VC dimension is at least k.

   (a) Argue that VC-DIMENSION is in $\mathbf{\Sigma_3^p}$. Hint: what is the size of the largest possible set X shattered by a collection of 2^m subsets?

   (b) Show that VC-DIMENSION is $\mathbf{\Sigma_3^p}$-complete by reducing from QSAT_3. Hint: the universe U should be the set $\{0,1\}^n \times \{1,2,3,\ldots,n\}$. For each n-bit string a, define the subset $U_a = \{a\} \times \{1,2,3,\ldots,n\}$. The sets in your instance of VC-DIMENSION should each be a subset of some U_a; note that the problem definition does not require that sets S_i and S_j to be different for $i \neq j$ — indeed your reduction will probably produce many copies of the same set with different "names."

2. Here is a new class involving alternating quantifiers: $\mathbf{S_2^p}$ (the "S" stands for "symmetric alternation"). A language L is in $\mathbf{S_2^p}$ if and only if there is a language $R \in \mathbf{P}$ for which

$$x \in L \;\Rightarrow\; \exists y\, \forall z\, (x, y, z) \in R$$
$$x \notin L \;\Rightarrow\; \exists z\, \forall y\, (x, y, z) \notin R$$

   where as usual $|y| = \text{poly}(|x|)$ and $|z| = \text{poly}(|x|)$. To make sense of this definition it is useful to think of R as defining for each x a 0/1 matrix M_x whose rows are indexed by y and whose columns are indexed by z. Entry (y, z) of matrix M_x is 1 if $(x, y, z) \in R$ and 0 otherwise. Now, the definition says that $x \in L$ if there is an all-ones row in M_x and $x \notin L$ if there is an all-zeros column in M_x (and it is clear that these configurations are mutually exclusive).

   (a) Argue that $\mathbf{S_2^p} \subseteq (\mathbf{\Sigma_2^p} \cap \mathbf{\Pi_2^p})$.

   (b) Prove that $\mathbf{P^{NP}} \subseteq \mathbf{S_2^p}$. Hint (from Goldreich-Zuckerman): Let M be a deterministic OTM. Call a string T a *valid transcript* of M on input x if it contains a sequence of pairs (q_i, a_i) where q_i is an oracle query and $a_i \in \{\text{yes}, \text{no}\}$, and it correctly describes the step-by-step computation of M on input x in which oracle query q_i is answered by a_i. We say that a valid transcript is *supported* by a sequence S of pairs (q_j, w_j) if for every $a_i = \text{yes}$, there is some j for which $q_i = q_j$ and w_j is an **NP** *witness* for query q_i. We say that a valid transcript is *consistent* with a sequence S of pairs (q_j, w_j) if for every $a_i = \text{no}$, there is no j for which $q_i = q_j$ and w_j is a **NP** witness for query q_i. First argue that for every $x \in L$, there exists a pair (T, S) for which T is a valid transcript of M on input x that ends with M accepting, that is supported by S and consistent with every sequence S'. Similarly, for every $x \notin L$, there exists a pair (T, S) for which T is a valid transcript of M on input x that ends with M rejecting, that is supported by S and consistent with every sequence S'.

(c) Prove a stronger form of the Sipser-Lautemann Theorem: $\mathbf{BPP} \subseteq \mathbf{S_2^P}$.

(d) Prove a stronger form of the Karp-Lipton Theorem: if SAT has polynomial-size circuits then $\mathbf{PH} = \mathbf{S_2^P}$.