15-496/859X: Computer Science Theory for the Information Age — Spring 2012
Carnegie Mellon University — V. Guruswami & R. Kannan

Problem Set 6

Due in class, Thursday, April 26

Instructions

- You should think about *each* problem by yourself for *at least 30 minutes* before choosing to collaborate with others.
- You are allowed to collaborate with fellow students taking the class in solving the problems (in groups of at most 3 people for each problem). In fact this is encouraged so that you interact with and learn from each other. However, *you must write up your solutions on your own.* If you collaborate in solving problems, you should clearly acknowledge your collaborators for each problem.
- Reference to any external material besides the course text and material covered in lecture is not allowed. In particular, you are not allowed to search for answers or hints on the web. You are encouraged to contact the instructors or the TA for a possible hint if you feel stuck on a problem and require some assistance.
- Solutions typeset in LaTeX are preferred.
- Feel free to email the instructors or the TA if you have any questions or would like any clarifications about the problems.
- You are urged to start work on the problem set early.
- Each problem is worth 10 points unless indicated otherwise.

---

1. We saw that the perceptron algorithm converges in $1/\gamma^2$ steps to a halfspace that correctly classfies a set of examples that is linearly-separable with margin $\gamma > 0$. However, we gave no guarantees on the margin of the resulting halfspace.

   Consider the variant of the perceptron algorithm where an example $\mathbf{a_i}$ (with label $\ell_i \in \{1, -1\}$) is considered "incorrect" w.r.t the current solution $\mathbf{w}$ if $\frac{(\mathbf{w} \cdot \mathbf{a_i})\ell_i}{\|\mathbf{w}\|} < \gamma/2$. In each iteration, if there is an incorrect example $\mathbf{a_i}$ (ties broken arbitrarilty), update $\mathbf{w}$ as in the perceptron algorithm: $\mathbf{w} \leftarrow \mathbf{w} + \ell_i \mathbf{a_i}$.

   Prove that this algorithm terminates in at most $10/\gamma^2$ steps (and therefore the $\mathbf{w}$ at the end has margin at least $\gamma/2$).

2. What is the VC dimension of the family consisting of interiors of triangles in the plane? Prove your answer.

3. Suppose $(U, \mathcal{F})$ is a set system with VC dimension d. Let $\epsilon \in (0, 1)$. Prove that there is a distribution p on U such that one needs at least $\Omega(d/\epsilon)$ i.i.d samples from p before the probability of the sampled set "hitting" every set $S \in \mathcal{F}$ with $p(S) \geq \epsilon$ exceeds $1/2$.

   Hint: First prove a lower bound of $\Omega(d)$ (this will receive partial credit), and then suitably "scale" that construction to get the $\Omega(d/\epsilon)$ lower bound.

4. (a) Compute, with proof, the VC dimension of the class of boolean conjunctions of literals over $\{0, 1\}^n$.

(b) In lecture, we saw a PAC learning algorithm for the concept class of conjunctions of some subset of literals corresponding to n variables. This algorithm used $O\left(\frac{n}{\epsilon}\log\frac{n}{\delta}\right)$ examples to find a conjunction that has error at most ϵ w.r.t the unknown conjunction being learned with confidence $1-\delta$. Prove that in fact $O\left(\frac{1}{\epsilon}\left(n+\log(1/\delta)\right)\right)$ examples are sufficient to guarantee that the hypothesis output by the algorithm has error at most ϵ with confidence at least $1-\delta$.

5. (20 points) In problem set 3, you were asked to give a linear algebraic proof of a combinatorial lower bound on the size of 4-wise independent families of n-bit strings. In this exercise, you are asked to give a linear-algebraic proof of the following combinatorial bound concerning VC dimension (recall that we proved this in class using a different "shifting" method): Suppose $\mathcal{F}$ is a family of subsets of $\{1,2,\ldots,n\}$ that has VC dimension d. Then $|\mathcal{F}| \le \binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{d}$.

   Let $\mathcal{F}$ have VC dimension d. Let $\mathcal{S}_d$ be the set of all subset of $\{1,2,\ldots,n\}$ of size at most d. For each $F \in \mathcal{F}$, define the $|\mathcal{S}_d|$-dimensional vector $v_F : \mathcal{S}_d \to \mathbb{R}$ as $v_F(X) = 1$ if $X \subseteq F$ and $v_F(X) = 0$ otherwise.

   (a) Assuming that the real vectors v_F, $F \in \mathcal{F}$, are linearly independent, prove that

   $$|\mathcal{F}| \le \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} .$$

   So your task now is to prove that the v_F's are linearly independent. To this end, assume there is some linear combination $\sum_{F\in\mathcal{F}} \alpha_F v_F = 0$ with not all coefficients α_F being zero (and the goal would be to arrive at a contradiction).

   For a subset $Z \subseteq \{1,2,\ldots,n\}$, define $\mu(Z) = \sum_{F\in\mathbb{F};Z\subseteq F} \alpha_F$.

   (b) Prove that $\mu(X) = 0$ for each $X \in \mathcal{S}_d$.

   (c) Show that there must exist a subset $T \subseteq \{1,2,\ldots,n\}$ such that $\mu(T) \neq 0$.

   Let Y be a minimum-sized subset of $\{1,2,\ldots,n\}$ such that $\mu(Y) \neq 0$. (Part (b) above implies such Y exists.)

   (d) (If you are stuck on this part, assume this for the next parts) Let $Z \subseteq Y$. Argue that

   $$\sum_{\substack{F\in\mathcal{F}\\ Z=F\cap Y}} \alpha_F = \sum_{W:Z\subseteq W\subseteq Y} (-1)^{|W\setminus Z|}\mu(W) .$$

   <u>Hint</u>: Inclusion/exclusion.

   (e) Use the above to conclude that for every $Z \subseteq Y$, there exists $F \in \mathcal{F}$ such that $F \cap Y = Z$.

   (f) Argue why the above is a contradiction, and conclude that no non-trivial linear combination $\sum_{F\in\mathcal{F}} \alpha_F v_F = 0$ can exist.