22C : 153 **Design and Analysis of Algorithms**
Spring 2002
Homework 4
Due on May 9

We are given a set $P = \{p_1, \ldots, p_n\}$ of n points in the d-dimensional space $\mathbb{R}^d$. Let us denote the co-ordinates of a point $p_i \in P$ by $(p_i^1, \ldots, p_i^d)$. For any $1 \leq i \leq n$ and $1 \leq j \leq d$, we have the property that p_i^j is either 0 or 1. (So we can also think of each point as an element of the set $\{0,1\}^d$.) We may assume that $1 \leq d \leq n$, but keep in mind that d is not small and may be quite close to n.

Let us denote by D by the set $\{1, \ldots, d\}$ that indexes the dimensions. We say that a subset $D' \subseteq D$ of dimensions is *good* for a subset $Q \subseteq P$ of points if for any point $p_i \in Q$ and and $j \in D - D'$, we have $p_i^j = 0$. (Notice that we require the j'th co-ordinate for every point in Q to be 0 for every j *not* in D'.) For any $1 \leq k \leq n$, let $\mathrm{Opt}(k)$ denote the size of the smallest possible subset $D' \subseteq D$ that is good for *some* subset Q of P with at least k points. The problem that we want to solve is to compute $\mathrm{Opt}(k)$ given P and k. We also want to report a set of dimensions $D' \subseteq D$ of size $\mathrm{Opt}(k)$ and a set of points $Q \subseteq P$ of size at least k such that D' is good for Q. As an example, consider the following set of 5 points in 3 dimensions.

$$p_1 = (0,\ 1,\ 0)$$
$$p_2 = (1,\ 1,\ 0)$$
$$p_3 = (0,\ 1,\ 1)$$
$$p_4 = (0,\ 0,\ 0)$$
$$p_5 = (1,\ 0,\ 1)$$

Notice that the subset $\{1,3\}$ of dimensions is good for the point set $\{p_4, p_5\}$ whereas the subset $\{1,2\}$ of dimensions is good for the point set $\{p_1, p_2, p_4\}$. We can check that the latter is the smallest subset of dimensions that is good for at least 3 points. Thus $\mathrm{Opt}(3) = 2$.

The Status of this Problem

Given that I don't know much more about the status of this problem than what I have stated above, I would be very happy if you can show any of the following:

- The problem is solvable in polynomial time.
- The problem is NP-hard.
- It is possible in polynomial time to solve the problem approximately in the following sense: for some parameters $c_1, c_2 \geq 1$, there is a subset of dimensions of size at most $c_1 \mathrm{Opt}(k)$ that is good for a subset of points of size at least k/c_2. The parameters c_1, c_2 may depend on d, k, and n, but the result should be non-trivial. For instance, it is easy to achieve the above for $c_1 = d$ and $c_2 = n$.

The minimum expected of you is a working program to solve the above problem. You may also try to improve the speed of the program by using techniques that prune the search space. **I suggest that you develop a working program first and treat the three items listed above as avenues for extra credit.** Please turn in a report describing your program and any additional ideas you have along with a hard-copy of your source code. I will later post instructions for electronically submitting the source code.

You may assume that the input to your progarm is a UNIX file with n, d, and k on the first line (separated by space), followed by n lines. The i'th line gives the co-ordinates of the i'th point (with a space separating two co-ordinates). A sample file is accessible from the web-page.

You may work on this assignment in groups of one or two.