CS369N: Problem Set #3

Due to Qiqi Yan by 11:30 AM on Thursday, December 10, 2009

Instructions: same as the first two homeworks.

Problem 11

(12 points) Recall from Lecture #6 that we proved the following (the Leftover Hash Lemma). Suppose X is a random variable with collision probability $cp(X)$ at most $1/K$. Suppose $\mathcal{H}$ is a (2-)universal family of hash functions (from the range of X to the range $\{0, 1, 2, \ldots, M-1\}$), and h is chosen uniformly at random from $\mathcal{H}$. Then the statistical distance between the joint distribution of $(h, h(X))$ and of the uniform distribution (on $\mathcal{H} \times \{0, 1, 2, \ldots, M-1\}$) is at most $\frac{1}{2}\sqrt{M/K}$.

For this problem, assume that you have a sequence $X_1, \ldots, X_T$ of random variables, with the property that for every i and fixed values of $X_1, \ldots, X_{i-1}$, the (conditional) collision probability of X_i is at most $1/K$ (i.e., a "block source"). Prove that the statistical distance between the joint distribution of $(h, h(X_1), \ldots, h(X_T))$ and of the uniform distribution is at most $\frac{T}{2}\sqrt{M/K}$.

[Hint: One high-level approach is to prove, by downward induction on i, a bound of $\frac{(T-i)}{2}\sqrt{M/K}$ on the statistical distance between $(h, h(X_{i+1}), \ldots, h(X_T))$ and the uniform distribution for every fixed value of $X_1, \ldots, X_i$. The increase in statistical distance in the inductive step should come from the Triangle Inequality.]

Problem 12

(20 points) You are given n points $x_1, \ldots, x_n$ in some bounded real interval ($[0, 1]$, if you like) and a parameter k. The goal is to partition the n points into k clusters $C_1, \ldots, C_k$ and designate points $m_1, \ldots, m_k \in \mathcal{R}$ as cluster centers to minimize $\Phi = \sum_{i=1}^{k} \sum_{x_j \in C_i} (x_j - m_i)^2$. One can easily check that, given the C_i's, the optimal thing to do is to set m_i equal to the average value of the points in C_i.

In this problem we will analyze a particular local search heuristic, which works as follows. Iteration 0 begins with an arbitrary clustering $C_1, \ldots, C_k$ with each C_i non-empty. In an odd iteration, we hold the C_i's fixed and re-compute m_i as the average value of the points in C_i. In an even iteration, we independently and simultaneously re-assign each point x_j to the cluster C_i that had mean m_i closest to x_j. You should check that every non-vacuous iteration (i.e., one that makes some change) strictly decreases Φ. Thus, this heuristic is guaranteed to terminate (with a "locally optimal" clustering). Prove that the heuristic has polynomial smoothed complexity, meaning that for every point set $x_1, \ldots, x_n$, if an independent (one-dimensional) Gaussian with standard deviation σ is added to each x_i, then the expected running time (over the perturbation) of the local search heuristic is polynomial in n, k, and $1/\sigma$.

[Hint: You might look to the analysis of the 2-OPT heuristic for TSP for inspiration. Try to identify a sufficient condition on the input that guarantees that every improving local move makes a non-trivial improvement to Φ, and prove probability bounds on the likelihood that the condition is satisfied.]

Problem 13

(15 points) Recall the Balance algorithm for non-clairvoyant online scheduling from Lecture #8. In lecture, we studied the objective of minimizing the average flow (or response) time, $\sum_j (C_j - r_j)$. One concern about such objectives is that minimizing the average might require assigning huge delays to a small number of jobs. This problem proves that this concern is unwarranted for the Balance algorithm.

Precisely, consider the objective of minimizing the maximum idle time of a job, where the idle time is $C_j - r_j - (p_j/s)$, where C_j is the job's completion time, r_j is its release date, p_j is its processing time, and s is the machine speed. Show that the maximum idle time of a job under the Balance algorithm with a machine of speed $1+\epsilon$ is at most $1/\epsilon$ times that of an optimal (clairvoyant and offline) solution with a machine of unit speed.

Problem 14

Recall from Lecture #8 the definition of a selfish routing network, of an equilibrium flow, and of the price of anarchy. For a given network G with continuous and nondecreasing edge cost functions and a traffic rate r between a source s and sink t, let $\pi(G, r)$ denote the ratio of the costs of equilibrium flows at rate r and rate $r/2$. By the resource augmentation result from lecture, the price of anarchy in the network G at rate r is at most $\pi(G, r)$.

(a) (8 points) Prove a "loosely competitive" guarantee using the above resource augmentation bound: for every G and r, and for at least a constant fraction of the traffic rates $\hat{r}$ in $[r/2, r]$, the price of anarchy in G at traffic rate $\hat{r}$ is $O(\log \pi(G, r))$.

(b) (7 points) Prove that for every constant K, there exists a network G with continuous, nondecreasing edge cost functions and a traffic rate r such that the price of anarchy in G is at least K for every traffic rate $\hat{r} \in [r/2, r]$.

Problem 15

(15 points) Recall that in Lecture #9 we studied the problem of selling a good with unlimited supply to n potential buyers to maximize revenue. Now suppose you have only k copies of the good, where $k < n$.

Let's begin with the thought experiment where there is a distribution over inputs, with each valuation v_i drawn IID from a known distribution F. It turns out that the truthful auction that maximizes the expected revenue is the Vickrey auction with a reserve price r (where r depends on F — e.g., it is $\frac{1}{2}$ if F is the uniform distribution on $[0, 1]$).[1] This auction sells to all of the buyers that have a valuation v_i above r and are also amongst the top k valutions overall. All winners pay either r or the $(k+1)$th highest valuation, whichever is larger. As usual, define $\mathcal{C}_D$ as the set of all such auctions (i.e., the Vickrey auction with all possible choices of the reserve r).

Assume that $k \geq 2$ and design a truthful auction that has the same type of guarantee as the RSPE auction from Lecture #9. That is, for every input v, your (randomized) auction should have expected revenue at least a constant fraction of every auction in $\mathcal{C}_D$ that sells to at least 2 buyers. (You don't have to compete with an auction of $\mathcal{C}_D$ that sells to only one bidder on input v, just like in Lecture #9).

[1] Actually, this assertion holds only under a mild technical condition on F, but you don't need to worry about that for this problem.