CS 6150: HW6 – Linear Programs, Review

Submission deadline: Sunday, Dec 11, 2016, 11:59PM

This assignment has 5 questions, for a total of 50 points. Unless otherwise specified, complete and reasoned arguments will be expected for all answers.

Question	Points	Score
Knapsack with small items	6	
Markov's Inequality Revisited	14	
Variant of Max Flow	5	
Linear Programs	15	
Variants of Least Squares	10	
Total:	50	

Question 1: Knapsack with small items .... **[6]**

Let us re-consider the knapsack problem: we have a sack of capacity C, and n objects, with the ith one having size s_i and value v_i (both ≥ 0. The goal is to choose a subset of the objects, such that the total size is at most C, and the total value is maximized.

Consider the greedy heuristic, in which we sort all the elements by "bang for buck", i.e., the v_i/s_i value (in decreasing order), and in every step, add the first remaining element (according to this ordering) that can fit into the sack.

Suppose all the sizes s_i are $\leq C/k$, for some integer k. Then, prove that the total value of the set output by the greedy algorithm is at least $\left(1 - \frac{1}{k}\right)$ times the best-possible total value.

You get partial credit even if you prove it for the case $k = 2$.

[*Hint:* Suppose the objects are ordered such that $\frac{v_1}{s_1} \geq \frac{v_2}{s_2} \geq \dots$. Say we pick the "first r" objects, for some r. Argue that the ratio of the total value to the total size is no smaller than the same ratio for the optimum subset. Next, what can you say about the total size?]

Question 2: Markov's Inequality Revisited .... **[14]**

Markov's inequality says that for a non-negative random variable X whose expectation is μ, we have, for any $t \geq 1$, $\Pr[X \geq t\mu] \leq 1/t$.

(a) **[3]** Prove that the *non-negativity* of X is crucial, by giving an example of a random variable X (which can take negative values), for which: $\mu = 1$, but $\Pr[X > 4] \geq 0.9$.

(b) **[2]** Let X be a random variable which is distributed uniformly in the interval $[0, 1]$. Calculate the expectation, and the probability that $X \geq 3/4$.

(c) **[3]** Now suppose $X_1, X_2, \dots, X_k$ are independent copies of X, and consider $Y = \frac{1}{k}(X_1 + X_2 + \cdots + X_k)$. We will prove that $\Pr[Y \geq 3/4]$ is tiny compared to the probability you computed above, even for moderately big k. This illustrates the effect of 'averaging' independent random variables.
First, for a given i, compute $\mathbf{E}[e^{X_i}]$.

(d) **[2]** Next, compute $\mathbf{E}[e^{X_1+X_2+\cdots+X_k}]$. (*Hint:* For independent random variables, the expectation of the product is the product of expectations.)

(e) **[4]** Observe that $Y \geq 3/4$ iff $e^{X_1+X_2+\ldots X_k} \geq e^{3k/4}$. Now use the calculation from part (d) to prove that $\Pr[Y \geq 3/4] < (0.9)^k$.

Question 3: Variant of Max Flow .... **[5]**

Consider the following variant of max flow: we have a directed graph with capacities, and let s, t be designated source and sink vertices, and let u be a special vertex. Our goal is to find the max flow from s to t, subject to the additional constraint that there is at least a flow of f units flowing 'through' the vertex u.

Design a polynomial time algorithm for this problem.

[*Hint:* Take a look at the next few problems!]

Question 4: Linear Programs .... **[15]**

(a) **[5]** We saw in class how to write various combinatorial problems as 'integer linear programs', i.e., linear optimization problems with variables constrained to be $\in \{0, 1\}$. For the min spanning tree problem (input: weighted undirected graph G), it was tricky to write such a formulation. Eventually, we chose the variables to be x_e, which indicate if edge e is chosen or not, and we had constraints for every subset S of the vertices:

$$\sum_{e \in E(S, \overline{S})} x_e \geq 1.$$

(As usual, $E(S, \overline{S})$ is the set of edges that go between S and its complement.) The objective is then to minimize $\sum_e w_e x_e$, where w_e is the weight of edge e.
Suppose $w_e > 0$ for all edges e. Prove that the optimal solution to the minimization problem (with the constraints $x_e \in \{0, 1\}$) indeed corresponds to a min spanning tree in G.

(b) [**4**] Now suppose we 'relax' the constraint $x_e \in \{0, 1\}$ to $0 \le x_e \le 1$. Then, give an example in which the optimum of the resultant linear program is strictly smaller than the value of the min spanning tree in G.

[*Hint:* Consider the graph to be a triangle.]

(c) [**6**] Consider the following linear program:

$$\begin{aligned} \min x_1 + x_2 \qquad & \text{subject to} \\ x_1 + 2x_2 &\ge 3 \\ x_1 - 2x_2 &\ge -4 \\ x_1 + 7x_2 &\le 6 \\ x_1, x_2 &\ge 0. \end{aligned}$$

Plot the feasible region (draw + scan is OK, if it's almost to scale). Show by manipulating the equations (as we saw in class) that the optimum is ≥ 1. Write down the dual program. Use your figure as a hint to find the optimum of the programs. [Note that it suffices to produce a feasible solution to each, with equal objective function value.]

Question 5: Variants of Least Squares .... [**10**]

Given points (x_i, y_i) in a plane, the least squares problem asks to find the best a, b such that $y_i \approx ax_i + b$ for all i. Formally, the goal is to minimize the squared-error: $\sum_i (y_i - (ax_i - b))^2$. Let us consider some natural variants:

(a) [**4**] Suppose we wish to minimize the "max" (a.k.a. ℓ_∞) error, i.e., find a, b such that $\max_i |y_i - (ax_i - b)|$ is minimized. Show that this can be cast as a linear programming problem.

(b) [**6**] Suppose we wish to minimize the ℓ_1 error, i.e., find a, b such that $\sum_i |y_i - (ax_i - b)|$ is minimized. [*Hint:* How would you capture the absolute values? Perhaps by introducing new variables?]