Homework 2

Algorithm Design 2018-19 - Sapienza

Luigi Russo 1699981

January 5, 2019

Contents

1 Michele's birthday

2 Valerio and Set Cover

Notation: A is the set of required skills; S is the set of all the people available, each one is represented as a set of skills $Sj \subseteq A; n = |A|$. We can express the Set Cover with Redundancies problem using the following ILP formulation:

$$\begin{aligned}
&\min \sum_{S_j \in S} cS_j \cdot x_j \\
&\sum_{S_j | A_i \in S_j} x_j \geq 3, && \forall A_i \in A \\
&x_j \in \{0,1\}, && \forall S_j \in S
\end{aligned}$$

In order to build a randomized approximation consider the associated LP problem where $x_j^* \in [0,1]$. The LP solution is a vector x^* of real values. For each set $S_j \in S$, pick S_j with probability x^*j, the entry corresponding to S_j in x^*. Let C be the collection of sets picked. The expected cost of C is

$$E[c(C)] = \sum_{S_j \in S} Pr[S_j \text{ is picked}] \cdot c(S_j) = \sum_{S_j \in S} x_j^* \cdot c(S_j) = OPT_f.$$

Next, let us compute the probability that a skill $a \in U$ is covered at least 3 times by C. Suppose that a occurs in $k \geq 3$ (otherwise the problem has no solution) sets of S. Let the probabilities associated with these sets be $p_1, ..., pk$. Since a is fractionally covered in the optimal solution, $\sum_{i=1}^{k} p_i \geq 3$. Using elementary calculus, it is easy to show that under this condition, the probability that a is covered by C is minimized when each of the p_i is $\frac{3}{k}$. Thus,

$$Pr[a \text{ is covered}] \geq 1 - \sum_{i=0}^{2} \binom{k}{i} (1-\frac{3}{k})^{k-i} = 1-(1-\frac{3}{k})^k + 3\cdot(1-\frac{3}{k})^{k-1} + \frac{17}{2}\cdot(1-\frac{3}{k})^{k-2}$$

and we can bound this:

$$Pr[a \text{ is covered}] \geq e^{-\frac{5}{6}}$$

To get a complete set cover with the redundancies, independently pick $\frac{6}{5} d \log n$ such subcollections, and compute their union, say C', where d is a constant such that: $(e^{-\frac{5}{6}})^{\frac{6}{5} d \log n} \leq \frac{1}{4n}$. Clearly we have that:

$$Pr[a \text{ is not covered}] \leq \frac{1}{4n}$$

Summing up all a:

$$Pr[C' \text{ is not a valid solution}] \leq n \cdot \frac{1}{4n} = \frac{1}{4}$$

Clearly

$$E[c(C')] \leq \frac{6}{5} \cdot OPT_f \cdot d \log n$$

For Markov we have that:

$$Pr[c(C')] \geq OPT_f \cdot 4 \cdot \frac{6}{5} \log n \leq \frac{1}{4}$$

This implies that:

$$Pr[C' \text{ is valid and has cost } \leq OPT_f \cdot 4 \cdot \frac{6}{5}] \geq \frac{1}{2}$$

3 The "k min-cut" problem

Let F^* be an optimal solution for the problem and let F_i^* be the isolating cut in the optimal solution for s_i. Since F_i is a minimum cut for s_i,

$$\sum_{e \in F_i} c_e \leq \sum_{e \in F_i^*} c_e$$

The cost of our solution is at most

$$\sum_{i=1}^{k} \sum_{e \in F_i} c_e \leq \sum_{i=1}^{k} \sum_{e \in F_i^*} c_e$$

Since each edge in an optimal solution $F*$ can be present in at most 2 different F_i^*, we have that our solution is bounded by:

$$\sum_{i=1}^{k} \sum_{e \in F_i} c_e \leq \sum_{i=1}^{k} \sum_{e \in F_i^*} c_e \leq 2 \cdot \sum_{e \in F*} c_e \leq 2 \cdot OPT$$

and this shows the 2-approximation.

4 Cristina and DNA

Let G be the set of genes and define a factorization f of the string D as an ordered multiset $\{g_1, g_2; ..., g_p\}$ with $g_i \in G$, such that the concatenation of $g_1, ..., g_p$ produces the string D. Let F be the set of all possible factorizations of D and $F_g := \{f \in F : g \in f_g\}$ be the set of all factorizations that contain the gene g. We need $|G|$ boolean variables x_g, set to 1 if gene $g \in G$ is used to produce D, otherwise 0; and $|F|$ boolean variables y_f, indicating whether the factorization $f \in F$ is used or not. The ILP formulation and its dual are:

$$\begin{aligned} \min \quad & \sum_{g \in G} w_g \cdot x_g \\ \text{s. t.} \quad & \sum_{f \in F_g} y_f \leq x_g, \quad \forall g \in G \\ & \sum_{f \in F} y_f \geq 1 \\ & x_g, y_f \in \{0,1\}, \quad \forall g \in G, f \in F \end{aligned}$$

$$(*) \quad \begin{aligned} \max \quad & b \\ \text{s. t.} \quad & \sum_{g \in f} a_g \geq b, \quad \forall f \in F \\ & a_g \leq w_g \quad \forall g \in G \\ & a_g, b \geq 0, \quad \forall g \in G \end{aligned}$$

The dual can be obtained with Lagrangian, after relaxing the original problem to LP: first of all the variables x and y are no longer boolean, but have to be ≥ 0. Then we compute

$$L(x,y,a,b,c,d) = \sum_{g \in G} x_g(w_g - a_g - c_g) - \sum_{f \in F} y_f(b + d_f) + \sum_{f \in F_g} y_f \sum_{g \in G} a_g + b$$

and creating the dual (d_1), removing d_f (d_2) and finally removing c_g variables leads to the formulation (*) presented above.

$$(d_1) \quad \begin{aligned} \max \quad & b \\ \text{s. t.} \quad & w_g - a_g - c_g = 0, \quad \forall g \in G \\ & b + d_f = 0, \quad \forall f \in F \\ & \sum_{g \in G} a_g = 0, \quad \forall g \in G \\ & b, a_g, c_g, d_f \geq 0, \quad \forall g \in G, f \in F \end{aligned}$$

$$(d_2) \quad \begin{aligned} \max \quad & b \\ \text{s. t.} \quad & w_g - a_g - c_g = 0, \quad \forall g \in G \\ & b \leq \sum_{g \in f} a_g, \quad \forall f \in F \\ & b, a_g, c_g \geq 0, \quad \forall g \in G, f \in F \end{aligned}$$

5 Comet and Dasher

The problem can be formalized with the following payouts matrix:

	T_C, T_D	T_C, H_D	H_C, T_D	H_C, H_D
Comet	2	-2	-1	4
Dasher	-2	2	1	-4

Let's now define:

- $h_X = Pr(\text{Head})$ for player X
- $t_X = Pr(\text{Tail})$ for player X

We can easily find that:

- $Pr(T_C, T_D) = t_C \cdot t_D$
- $Pr(T_C, H_D) = t_C \cdot h_D$
- $Pr(H_C, T_D) = h_C \cdot t_D$
- $Pr(H_C, H_D) = h_C \cdot h_D$

To guarantee that the game is fair, the expected value of Comet must be equal to the one of Dasher:

$$-2t_c t_D - t_C h_D + 2h_C t_D + 4h_C h_D = 2t_c t_D t_C h_D + -2h_C t_D + -4h_C h_D$$

with $t_C + t_D = 1$ and $h_C + h_D = 1$ since they are probability functions. Resolving the system we obtain

$$9h_C h_D - 3h_C - 4h_D + 2 = 0$$

There are infinite solutions: simple solutions are

- $t_C = 1, h_C = 0, t_D = h_D = 0.5$
- $t_D = 1, h_D = 0, h_C = \frac{2}{3}, t_C = \frac{1}{3}$

6 Drunk Giorgio

Let $P_n = Pr(Home|start = n)$ be the probability Giorgio goes back to home starting from position n and let q = 1 - p the probability to make a step towards home. Let N be the distance from home (Giorgio starts at 0).

$$P_n = \begin{cases} 0, & \text{if } n = -1 \\ p \cdot P_{n-1} + q \cdot P_{n+1}, & \text{if } 0 \leq n < N \\ 1, & \text{if } n = N \end{cases}$$

We can rewrite P_n in this way: $P_n = p \cdot P_n + q \cdot P_n = p \cdot P_{n-1} + q \cdot P_{n+1}$
$\Rightarrow P_{n+1} - P_n = \frac{p}{q} \cdot (P_n - P_{n-1})$.
In particular $P_1 - P_0 = \frac{p}{q} \cdot P_0$; moreover $P_2 - P_1 = (\frac{p}{q})^2 \cdot P_0$. In general we have: $P_{n+1} - P_0 = \sum_{k=0}^{n}(P_{k+1} - P_k) = \sum_{k=0}^{n}((\frac{p}{q})^{k+1} \cdot P_0) = \sum_{k=1}^{n+1}((\frac{p}{q})^k \cdot P_0)$
$\Rightarrow P_{n+1} = P_0 + \sum_{k=1}^{n+1}((\frac{p}{q})^k \cdot P_0) = P_0 \sum_{k=0}^{n+1}(\frac{p}{q})^k$

$$P_{n+1} = \begin{cases} P_0(n+2), & \text{if } p = q = 0.5 \\ P_0(\frac{1-(\frac{p}{q})^{n+2}}{1-\frac{p}{q}}), & \text{if } p \neq q \end{cases}$$

For $n = N - 1$:

$$1 = P_N = \begin{cases} P_0(N+1), & \text{if } p = q = 0.5 \\ P_0(\frac{1-(\frac{p}{q})^{N+1}}{1-\frac{p}{q}}), & \text{if } p \neq q \end{cases}$$

$$P_0 = \begin{cases} \frac{1}{N+1}, & \text{if } p = q = 0.5 \\ \frac{1-\frac{p}{q}}{1-(\frac{p}{q})^{N+1}}, & \text{if } p \neq q \end{cases}$$

The probability to go to hospital starting from 0 is:

$$Pr(Hospital|start = 0) = 1 - \lim_{N \to +\infty} P_0 = \begin{cases} 1, & \text{if } p \geq q \\ \frac{p}{q}, & \text{if } p < q \end{cases}$$

For $p \geq q$, Giorgio always goes (probability = 1) to hospital, instead for $0 \leq p \leq \frac{1}{3}$, Giorgio goes to hospital with probability less than 0.5 (easily obtained by solving the inequality above!).

References