University of Waterloo
CS 466 — Advanced Algorithm
Spring 2013
Problem Set 7
Siwei Yang - 20258568

1. [13 marks: Weighted Set Cover] In class we looked at the set cover problem where we want a set cover of minimum number of subsets. A variant of the problem, which also has many applications, involves weights on the elements. The input consists of: a set U of n elements, where each u in U has a non-negative weight w(u); a collection $S_1, \ldots, S_t$ where $S_i \subset U$; and a number k. The problem is to pick k of the subsets Si to maximize the sum of the weights of the elements covered.

   The goal of this question is to show that the obvious greedy algorithm has approximation factor $1 - \frac{1}{e}$ (where e is the base of the natural logarithm). The greedy algorithm first picks a set that covers the maximum weight of elements, then deletes those elements and repeats until k sets have been chosen. Let w^* be the weight of the elements covered by the optimum solution. Let w_i, for $i = 1, \ldots, k$, be the weight of the elements covered by the first i sets of the greedy algorithm.

   Thus w_k is the final weight of the greedy solution.

   - (a) [2 marks] Find an example with k = 2 and unit weights where $w_2 = \frac{3}{4} * w^*$.
     Let $U = \{1, 2, 3, 4\}$ and $S_1 = \{1, 2\}$, $S_2 = \{2, 3\}$, $S_3 = \{3, 4\}$
     By choosing S_2 at the first iteration and either S_1, S_3 at the second, we achieve a set cover of three elements. Thus weighting 3.

     $$w_2 = 3 \tag{1}$$

     However the optimal solution is choosing S_1 and S_3 to produce a complete cover with weight of 4.

     $$w^* = 4 \tag{2}$$

- (b) [5 marks] Prove that $w_1 \geq \frac{w^*}{k}$. More generally, prove that $w_i - w_{i-1} \geq \frac{w^* - w_{i-1}}{k}$.

  Assume without loss of generality:

  - $U_0 = U$ and $w_0 = 0$
  - for the optimal algorithm, the subset chosen at iteration i is S_i^*.
  - for the greedy algorithm, that the working set at iteration i is U_i and the subset will chosen at iteration i is S_i(so all subsets will be indexed).

  Then, for the greedy algorithm, at any iteration i, we have the following:

$$weight(S_i \cap Ui) \geq weight(S_j \cap Ui) \text{ where } j \geq i \tag{3}$$

  from the nature of greedy algorithm.

  Then, for the optimal algorithm, we have:

$$weight(\bigcup_1^k S_i^* \cap U_i) + weight(\bigcup_1^k S_i^* \setminus U_i) = weight(\bigcup_1^k S_i^*) = w^* \tag{4}$$

  Consider the optimal k-set cover based off U_i, let its weight be $w^{*'}$ and subset chosen at iteration i be $S_i^{*'}$. Then:

$$weight(\bigcup_1^k S_i^{*'} \cap U_i) = w^{*'} \tag{5}$$

  Since all $weight(S_i \cap Ui) \geq weight(S_i^{*'} \cap Ui)$ by 3

$$k * weight(S_i \cap Ui) \geq w^{*'} \tag{6}$$

  Now, let's observe the relation between w^* and $w^{*'}$. Note that we can carefully choose $S_i^{*'}$s, so that all S_i^*s not chosen by the greedy algorithm is in $S_i^{*'}$s.

$$\bigcup_{j=1}^k S_j^* \subset (\bigcup_{j=1}^i S_j \cup \bigcup_{j=1}^k S_j^{*'}) = (U \setminus U_i) \cup \bigcup_{j=1}^k (S_j^{*'} \cap U_i) \tag{7}$$

Let the weight be $w^{*''}$ for this instance. Then we have:

$$w_{i-1} + w^{*''} \geq w^* \tag{8}$$

Since $w^{*'} \geq w^{*''}$ for optimality, we can combine 6 and 9 to get:

$$k * weight(S_i \cap Ui) \geq w^* - w_{i-1} \tag{9}$$

where $weight(S_i \cap Ui) = w_i - w_{i-1}$. Therefore, **for every iteration i, we have** $w_i - w_{i-1} \geq \frac{w^* - w_{i-1}}{k}$ which resolves to $w_1 \geq \frac{w^*}{k}$ when $i = 1$.

- (c) [6 marks] Prove by induction that $w_i \geq (1 - (1 - \frac{1}{k})^i) * w^*$.
  - **Base case**: $w_1 \geq \frac{w^*}{k} = (1 - (1 - \frac{1}{k})^1) * w^*$
  - **Induction**: Assume we have $w_i \geq (1 - (1 - \frac{1}{k})^i) * w^*$ Combined with $w_{i+1} - w_i \geq \frac{w^* - w_i}{k}$, we have:

$$\begin{aligned} w_{i+1} &\geq \frac{w^*}{k} + (1 - \frac{1}{k}) * w_i \\ &\geq \frac{w^*}{k} + (1 - \frac{1}{k}) * (1 - (1 - \frac{1}{k})^i) * w^* \\ &= \frac{w^*}{k} + (1 - \frac{1}{k}) * (w^* - (1 - \frac{1}{k})^i * w^*) \\ &= w^* - (1 - \frac{1}{k})^{i+1} * w^* \\ &= (1 - (1 - \frac{1}{k})^{i+1}) * w^* \end{aligned}$$

    Therefore, we have $w_{i+1} \geq (1 - (1 - \frac{1}{k})^{i+1}) * w^*$.
  - **Conclusion**: $w_i \geq (1 - (1 - \frac{1}{k})^i) * w^*$ **stands for all** $i > 0$.
- (d) [0 marks] From part (c), the weight of the greedy solution, w_k, is at least $(1 - (1 - \frac{1}{k})^k) * w^*$. Since $\lim_{k \to \infty} 1 - (1 - \frac{1}{k})^k = 1 - \frac{1}{e}$ and $1 - (1 - \frac{1}{k})^k$ is decreasing, thus $1 - (1 - \frac{1}{k})^k \geq 1 - \frac{1}{e}$, so the approximation ratio of the greedy algorithm is $1 - \frac{1}{e}$. In fact $1 - \frac{1}{e}$ is the best approximation factor possible for this problem (unless P=NP).