Assignment 1

There is a universe set U with $|U| = n$ and a family of subsets $F = \{ S_1, \dots, S_m \}$ where $S_i \subseteq U$ for all i. Each subset S_i has a cost c_i. Given a budget B, a feasible solution is an index set I denoting sets S_i where $i \in I$ selected from F such that the total cost of the selected sets is no more than budget B. The objective is to maximize the number of elements covered which is maximizing $|\cup_{i \in I} S_i|$. Let $M = [m]$ and $f : 2^M \to \mathbb{N}^+$ be the coverage function where $f(I) = |\cup_{i\in I} S_i|$. Define the additive cost function: $c : 2^M \to \mathbb{R}^+$ where $c(I) = \sum_{i \in I} c_i$. The problem is to select $I \subseteq M$ maximizing $f(I)$ such that $c(I) \le B$. In the following, for $i \in M$ and $I \subseteq M$, $I + i$ means $I \cup \{ i \}$ and $I - i$ means $I \backslash \{ i \}$. Consider the following Greedy Algorithm:

Algorithm Greedy Algorithm

Input: M, B, cost function c

1: $I_G \leftarrow \emptyset$
2: **while** $M \ne \emptyset$ **do**
3: $\quad i^* \leftarrow \arg\max_{i \in M} \frac{f(I_G + i) - f(I_G)}{c(i)}$
4: $\quad$ **if** $c(I_G) + c(i^*) \le B$ **then**
5: $\quad\quad I_G \leftarrow I_G + i^*$
6: $\quad$ **end if**
7: $\quad M \leftarrow M - i^*$
8: **end while**
9: **return** I_G

1. Show that Greedy Algorithm's approximation ratio can not be better than $\frac{1}{n}$ (by giving an example).

   Proof. Suppose $U = \{ e_1, \cdots, e_n \}$, $F = \{ S_1 = \{ e_1 \}, S_2 = U \}$; S_1 has cost $\frac{B}{2n}$ and S_2 has cost B. Greedy algorithm will only choose S_1 covering one element but opt chooses S_2 covering n elements. ☐

2. Though Greedy Algorithm's approximation ratio can be very small when the number of elements n is very large, by solving the following questions, we will see it can serve as a building block for a modified algorithm which has constant approximation ratio. Suppose line 4 is always True during the first l iterations of While-loop and let $I_l = \{ i_1, \cdots, i_l \}$ be the index set of sets selected after the l-th time the line 5 is executed. Let I^* be the index set of sets selected by optimal solution.

   (a) Show that $f(I^*) - f(I_l) \le (1 - \frac{c(i_l)}{B})(f(I^*) - f(I_{l-1}))$, which is, Greedy Algorithm is approaching optimal solution step by step.

   (Hint: $\forall I \subseteq M, J \subseteq M$, f satisfies: $f(I) \le f(J) + \sum_{i \in I \backslash J} (f(J + i) - f(J))$)

   Proof.

$$
\begin{aligned}
f(I^*) &\le f(I_{l-1}) + \sum_{i \in I^* \backslash I_{l-1}} (f(I_{l-1} + i) - f(I_{l-1})) && \text{(Hint)} \\
&= f(I_{l-1}) + \sum_{i \in I^* \backslash I_{l-1}} c(i) \frac{f(I_{l-1} + i) - f(I_{l-1})}{c(i)} \\
&\le f(I_{l-1}) + \frac{f(I_l) - f(I_{l-1})}{c(i_l)} \sum_{i \in I^* \backslash I_{l-1}} c(i) && \text{(Greediness: } i_l \text{ maximizes } \tfrac{f(I_{l-1}+i) - f(I_{l-1})}{c(i)}\text{)} \\
&\le f(I_{l-1}) + \frac{B}{c(i_l)} (f(I_l) - f(I_{l-1})) && \textstyle(\sum_{i \in I^* \backslash I_{l-1}} c(i) \le c(I^*) \le B)
\end{aligned}
$$

   Multiply $\frac{c(i_l)}{B}$ both sides, reorder items and get the inequality in (a). ☐

   (b) Show that $f(I_l) \ge (1 - e^{-\frac{c(I_l)}{B}}) f(I^*)$

   (Hint: Solve the inequality in (a) recursively and use inequality $1 - x \le e^{-x}$)

Proof.

$$
\begin{aligned}
f(I^*) - f(I_l) &\leq \prod_{t=1}^{l}(1 - \frac{c(i_t)}{B})f(I^*) && \text{(Solve inequality in (a) recursively)} \\
f(I_l) &\geq (1 - \prod_{t=1}^{l}(1 - \frac{c(i_t)}{B}))f(I^*) \\
&\geq (1 - \prod_{t=1}^{l}(e^{-\frac{c(i_t)}{B}})f(I^*) && (1 - \tfrac{c(i_t)}{B}) \leq e^{-\frac{c(i_t)}{B}}) \\
&= (1 - e^{-\frac{c(I_l)}{B}})f(I^*)
\end{aligned}
$$

□

(c) Show that for the first time the line 4 of Greedy algorithm is False, $f(I_G + i^*) \geq (1 - \frac{1}{e})f(I^*)$

Proof. 2(a) and 2(b) also hold for $I_l = \{i_1, ..., i_l\}$ where i_l is the first element making line 4 evaluated to be False, thus we can substitute I_l in 2(b) with $I_G + i^*$:

$$
\begin{aligned}
f(I_G + i^*) &\geq (1 - e^{-\frac{c(I_G + i^*)}{B}})f(I^*) \\
&\geq (1 - \frac{1}{e})f(I^*) && (c(I_G + i^*) > B \text{ when line 4 is False})
\end{aligned}
$$

□

3. Consider the Modified Greedy Algorithm below, let I be the set returned by it, show that it is a $\frac{1}{2}(1 - \frac{1}{e})$-approximation algorithm.

Proof. Let e^* be the first index making line 4 of Greedy Algorithm evaluated to be False during the running of line 2 of Modified Greedy.

$$
\begin{aligned}
f(I) &\geq \frac{1}{2}(f(I_G) + f(i^*)) && \text{(set } I \text{ is the one with max } f \text{ value)} \\
&\geq \frac{1}{2}(f(I_G) + f(e^*)) && (f(i^*) \geq f(e^*)) \\
&\geq \frac{1}{2}f(I_G + e^*) && \text{(coverage property)} \\
&\geq \frac{1}{2}(1 - \frac{1}{e})f(I^*) && \text{(inequality in (c))}
\end{aligned}
$$

□

Algorithm Modified Greedy Algorithm

Input: M, B, cost function c

1: $i^* \leftarrow \arg\max_{i \in M, c(i) \leq B} f(i)$
2: $I_G \leftarrow$ result of Greedy Algorithm
3: **return** $\arg\max \{ f(I_G), f(i^*) \}$