Exercise 1
Here, we want to consider Non-Metric Facility Location, where facilities $F$ with open cost $f_i$ for $i \in F$, cities $C$ and connection cost $c_{ij}$ are given. Differently from the variant, studied in the lecture, we do not assume anymore, that $c$ is metric.

i) Model the problem as Set Cover problem and obtain a polynomial time $O(\log n)$-approximation $(n := |C|)$ by using the greedy algorithm for Set Cover.

**Hint:** Even if the defined set system has exponentially many sets, under some conditions the greedy algorithm can still be made to run in polynomial time.

ii) A result of Raz and Safra (1997) says the following:

There is a constant $c > 0$ such that, given a Set Cover instance $S_1, \ldots, S_m$ and a parameter $k \in \mathbb{N}$ it is NP-hard to distinguish

- **YES:** $OPT_{SetCover} \leq k$
- **NO:** $OPT_{SetCover} \geq k \cdot c \cdot \log n$

Here $OPT_{SetCover}$ denotes the smallest number of sets that are needed to cover all $n$ elements.

**Remark:** This result means that there is a polynomial time reduction, taking a SAT clause $\mathcal{C}$ as instance and mapping it to a Set Cover instances $\mathcal{F} = \{S_1, \ldots, S_m\}$ such that: If $\mathcal{C}$ is satisfiable, then $OPT_{SetCover}(\mathcal{F}) \leq k$ and $OPT_{SetCover}(\mathcal{F}) \geq k \cdot c \cdot \log n$ otherwise (for more details on gap reductions, I recommend Chapter 29 of Vazirani’s book *Approximation Algorithms*).

Show that it is also NP-hard to approximate Non-Metric Facility Location by a factor better than $c \cdot \log n$.

Solution:

i) For any subset $C' \subseteq C$ of cities and facility $i \in C$, we define a set $S_{C', i} = C'$ of cost $c(S_{C', i}) = f_i + \sum_{j \in C'} c_{ij}$. The Non-Metric Facility Location is equivalent to the arising Set Cover instance. The greedy algorithm now performs as follows:

1. $\mathcal{F}' := \emptyset$

2. WHILE not yet all elements covered DO

3. $\text{price}(S) := \frac{c(S)}{|S \setminus \bigcup_{S' \in \mathcal{F}'} S'|}$
(4) \( \mathcal{S}' := \mathcal{S}' \cup \{ \text{set } S \text{ with minimum price}(S) \} \)

The algorithm gives a \( O(\log n) \)-approximation for \textsc{Set Cover} and hence also for \textsc{Non-Metric Facility Location}, where \( n = |C| \) is the number of elements/cities. The algorithm covers at least one element per iteration, hence the number of iterations is at most \( n \). Just the number of sets is exponentially large. Thus we have to argue, that the set \( S_{C,i} \), minimizing the price can be found efficiently. Consider any iteration and let \( \tilde{C} \subseteq C \) be the not yet covered cities. Note that

\[
\min_{i \in F, C' \subseteq \tilde{C}} \left\{ \frac{c(S_i, C')}{|C'|} \right\} = \min_{i \in F, k \in \{1, \ldots, \tilde{C}\}} \min_{C' \subseteq \tilde{C}} \left\{ \frac{f_i + \sum_{j \in C'} c_{ij}}{k} \right\}
\]

We try out all possibilities for \( i \) and \( k \).

\[
\min_{C' \subseteq \tilde{C}} \left\{ \frac{f_i + \sum_{j \in C'} c_{ij}}{k} \right\} = \frac{f_i}{k} + \frac{1}{k} \min_{C' \subseteq \tilde{C}} \left\{ \sum_{j \in C'} c_{ij} \right\}
\]

But the latter minimum is attained for the \( k \) cities that are closest to \( i \) (which can be easily obtained by sorting the cities according to their distance to \( i \)).

ii) Sei \( S_1, \ldots, S_m \) be the \textsc{Set Cover} instance on elements \( 1, \ldots, n \). Choose the \textsc{Non-Metric Facility Location} instance with \( f_i := 1 \) (one facility per set) and distances

\[
c_{ij} = \begin{cases} 0 & \text{if } j \in S_i \\ m & \text{otherwise} \end{cases}
\]

Then \( OPT_{\text{SetCover}} = OPT_{FL} \). Note that this cost function is in general not metric.

---

**Exercise 2**

We consider the \textsc{Facility Location} problem, with given facilities \( F \), cities \( C \), opening cost \( f_i \) for every facility \( i \). Assume that the cost function \( c_{ij} \) is metric. In this exercise, we want to show that there is no 1.46-approximation algorithm for the (metric) \textsc{Facility Location} problem.

For the sake of contradiction, suppose that we have a polynomial time algorithm \( \text{algo}(F, C, c_{ij}, f_i) \) that produces a 1.46-approximate solution \( F' \subseteq F \) (note that knowing the set of open facility suffices — the cities are then automatically connected to the nearest such facility).

Let \( S_1, \ldots, S_m \) be a \textsc{Set Cover} instance (with unit cost per set) on elements \( \{1, \ldots, n\} \). We may assume to know the value \( k \) of sets that are contained in an optimum solution. We will now show, how to obtain a \( 0.999 \cdot \ln(n) + O(1) \) approximate \textsc{Set Cover} solution in polynomial time. This would then contradict an inapproximability result of Feige (1998) (given that \( \text{NP} \) is not contained in \( \text{DTIME}(n^{O(\log \log n)}) \)).

We use the following \textsc{Set Cover} algorithm:

1. Let \( C := \{1, \ldots, n\}, F := \{1, \ldots, m\} \) and \( c_{ij} := \begin{cases} 1 & j \in S_i \\ 3 & \text{otherwise} \end{cases} \)

2. \textbf{WHILE } \( C \neq \emptyset \) \textbf{ DO}

3. Let \( f_i := 0.46 \cdot \frac{|C|}{k} \) be the facility cost \( \forall i \in F \)

4. \( F' := \text{algo}(F, C, c_{ij}, f_i) \)
(5) Buy the sets in $F'$.  
(6) $C' :=$ cities covered at cost 1; set $C := C \setminus C'$.  
(7) Return the bought sets

Perform the following analysis:

i) Consider any iteration and let $APX$ be the cost of the FACILITY LOCATION solution $F'$. Show that $APX \leq 1.46^2 \cdot |C|$. 

ii) Suppose the algorithm needs $T$ iterations. For iteration $t \in \{1, \ldots, T\}$, define $\beta_t$ and $\alpha_t$ such that $|F^t| = \beta_t k$ is the number of opened facilities and $|C^t| = \alpha_t |C|$ is the number of elements that are covered in this iteration. Show that $\beta_t \leq 0.999 \cdot \ln(\frac{1}{1-\alpha_t})$ holds for any $t < T$. 

**Hint:** It is OK if your solution contains the phrase “By a Maple/Matlab plot we see that...”.

iii) Why is $\prod_{t=1}^{T-1} (1 - \alpha_t) \geq \frac{1}{n}$?  

iv) Show that the algorithm needs at most $0.999 \cdot \ln(n) \cdot k$ many sets (plus $O(k)$ for the last iteration).

**Solution:**

i) Let $f := f_i$. One could open the $k$ facilities that correspond to sets in the optimum SET COVER solution and connect all clients at cost 1. This would cost in total $|C| + k \cdot f = |C| + 0.46 \cdot k = 1.46 \cdot |C|$. Since we assume to have a 1.46-apx algorithm, we have $APX \leq 1.46^2 \cdot |C|$. 

ii) Suppose that $\beta k$ centers are opened and $c |C| = |C'|$ many clients are connected at cost 1 (the others are connected at cost 3. Then this solution costs 

$$
f \cdot \beta \cdot k + c |C| + 3 (|C| - c |C|) = \beta 0.46 |C| + c |C| + 3 (|C| - c |C|) = |C| \cdot (0.46 \beta + c + 3 - 3c) = |C| \cdot (0.46 \beta - 2c + 3)
$$

On the other hand, we know that the solution costs at most $1.46^2 \cdot |C|$. Hence

$$
|C| \cdot (0.46 \beta - 2c + 3) \leq 1.46^2 \cdot |C| \quad \Rightarrow \quad \beta \leq \frac{4.3479c - 1.8878}{0.999 \ln\left(\frac{1}{1-c}\right)}
$$

iii) After $T - 1$ iterations the number of remaining elements is precisely 

$$
n \cdot \prod_{t=1}^{T-1} (1 - \alpha_t) \geq 1.
$$

Rearranging yields the claim.
iv) The number of chosen sets in iteration $t = 1, \ldots, T - 1$ is

$$
\sum_{t=1}^{T-1} k\beta_t \leq k \sum_{t=1}^{T-1} 0.999 \cdot \ln \left( \frac{1}{1 - \alpha_t} \right) = 0.999k \cdot \ln \left( \prod_{t=1}^{T-1} \frac{1}{1 - \alpha_t} \right) \leq 0.999 \cdot \ln(n) \cdot k
$$

In the last iteration we will not use more than $O(k)$ sets anyway (since $\beta \leq 4.3479c - 1.8878 \leq 3$).