The set cover problem

- Decision version: Given sets $S_1, \cdots, S_n$ and an integer $k > 0$, can we pick at most k sets among the given n sets so that their union is $U := \cup_i S_i$?

The set cover problem

- Decision version: Given sets $S_1, \cdots, S_n$ and an integer $k > 0$, can we pick at most k sets among the given n sets so that their union is $U := \cup_i S_i$?
- Optimization version: Given sets $S_1, \cdots, S_n$ with nonnegative *weights* $w_1, \ldots, w_n$, find a set cover that minimizes the total cost.

The set cover problem

- Decision version: Given sets $S_1, \cdots, S_n$ and an integer $k > 0$, can we pick at most k sets among the given n sets so that their union is $U := \cup_i S_i$?
- Optimization version: Given sets $S_1, \cdots, S_n$ with nonnegative *weights* $w_1, \ldots, w_n$, find a set cover that minimizes the total cost.
- Final example of greedy approximation algorithm, with a hint at the *pricing method*.

The set cover problem

- Decision version: Given sets $S_1, \cdots, S_n$ and an integer $k > 0$, can we pick at most k sets among the given n sets so that their union is $U := \cup_i S_i$?
- Optimization version: Given sets $S_1, \cdots, S_n$ with nonnegative *weights* $w_1, \ldots, w_n$, find a set cover that minimizes the total cost.
- Final example of greedy approximation algorithm, with a hint at the *pricing method*.
- A natural greedy approach: for each set S_i, define its *per-item cost* to be $w_i / |S_i|$.

The set cover problem

- Decision version: Given sets $S_1, \cdots, S_n$ and an integer $k > 0$, can we pick at most k sets among the given n sets so that their union is $U := \cup_i S_i$?
- Optimization version: Given sets $S_1, \cdots, S_n$ with nonnegative *weights* $w_1, \ldots, w_n$, find a set cover that minimizes the total cost.
- Final example of greedy approximation algorithm, with a hint at the *pricing method*.
- A natural greedy approach: for each set S_i, define its *per-item cost* to be $w_i/|S_i|$.
- Intuitively picking sets with small per-item cost is a good idea. We just need to update the "effective" per-item cost as we go.

The greedy algorithm

- Initialize $R \leftarrow U$, $C \leftarrow \emptyset$. (R records what has not been covered, and C records the set cover we construct.)

The greedy algorithm

- Initialize $R \leftarrow U$, $C \leftarrow \emptyset$. (R records what has not been covered, and C records the set cover we construct.)
- While $R \neq \emptyset$, do:
  - Pick the S_{i^*} that minimzes $w_i/|S_i \cap R|$.

The greedy algorithm

- Initialize $R \leftarrow U$, $C \leftarrow \emptyset$. (R records what has not been covered, and C records the set cover we construct.)
- While $R \neq \emptyset$, do:
  - Pick the S_{i^*} that minimzes $w_i/|S_i \cap R|$.
  - $C \leftarrow C \cup \{i^*\}, R \leftarrow R - S_{i^*}$.

The greedy algorithm

- Initialize $R \leftarrow U$, $C \leftarrow \emptyset$. (R records what has not been covered, and C records the set cover we construct.)
- While $R \neq \emptyset$, do:
    - Pick the S_{i^*} that minimzes $w_i/|S_i \cap R|$.
    - $C \leftarrow C \cup \{i^*\}, R \leftarrow R - S_{i^*}$.
    - For each element $s \in S_{i^*} \cap R$, record $c_s = w_{i^*}/|S_{i^*} \cap R|$. (For analysis)

The greedy algorithm

- Initialize $R \leftarrow U$, $C \leftarrow \emptyset$. (R records what has not been covered, and C records the set cover we construct.)
- While $R \neq \emptyset$, do:
  - Pick the S_{i^*} that minimzes $w_i/|S_i \cap R|$.
  - $C \leftarrow C \cup \{i^*\}, R \leftarrow R - S_{i^*}$.
  - For each element $s \in S_{i^*} \cap R$, record $c_s = w_{i^*}/|S_{i^*} \cap R|$. (For analysis)
- Return C (which contains indices of the sets we pick).

The greedy algorithm

- Initialize $R \leftarrow U$, $C \leftarrow \emptyset$. (R records what has not been covered, and C records the set cover we construct.)
- While $R \neq \emptyset$, do:
  - Pick the S_{i^*} that minimzes $w_i / |S_i \cap R|$.
  - $C \leftarrow C \cup \{i^*\}, R \leftarrow R - S_{i^*}$.
  - For each element $s \in S_{i^*} \cap R$, record $c_s = w_{i^*} / |S_{i^*} \cap R|$. (For analysis)
- Return C (which contains indices of the sets we pick).

Analysis: Lower bound OPT by the sum of per-item costs it must pay.

The greedy algorithm

- Initialize $R \leftarrow U$, $C \leftarrow \emptyset$. (R records what has not been covered, and C records the set cover we construct.)
- While $R \neq \emptyset$, do:
  - Pick the S_{i^*} that minimzes $w_i / |S_i \cap R|$.
  - $C \leftarrow C \cup \{i^*\}, R \leftarrow R - S_{i^*}$.
  - For each element $s \in S_{i^*} \cap R$, record $c_s = w_{i^*} / |S_{i^*} \cap R|$. (For analysis)
- Return C (which contains indices of the sets we pick).

Analysis: Lower bound OPT by the sum of per-item costs it must pay.
Starting point: say S_1 is the first set picked by Greedy, then we know:

- Intuitively (and not rigorously), Greedy paid the least possible per-item cost for the items in S_1, at least for that step;

The greedy algorithm

- Initialize $R \leftarrow U$, $C \leftarrow \emptyset$. (R records what has not been covered, and C records the set cover we construct.)
- While $R \neq \emptyset$, do:
  - Pick the S_{i^*} that minimzes $w_i / |S_i \cap R|$.
  - $C \leftarrow C \cup \{i^*\}, R \leftarrow R - S_{i^*}$.
  - For each element $s \in S_{i^*} \cap R$, record $c_s = w_{i^*} / |S_{i^*} \cap R|$. (For analysis)
- Return C (which contains indices of the sets we pick).

Analysis: Lower bound OPT by the sum of per-item costs it must pay.
Starting point: say S_1 is the first set picked by Greedy, then we know:

- Intuitively (and not rigorously), Greedy paid the least possible per-item cost for the items in S_1, at least for that step;
- Formally,

$$\frac{w_1}{|S_1|} \leq \frac{\text{OPT}}{|U|}.$$

Proof of first statement

Proposition

Let S_1 *be the first step picked by Greedy, then* $\frac{w_1}{|S_1|} \leq \frac{\text{OPT}}{|U|}$.

Proof of first statement

Proposition

Let S_1 be the first step picked by Greedy, then $\frac{w_1}{|S_1|} \leq \frac{\mathrm{OPT}}{|U|}$.

Intuition: $\frac{\mathrm{OPT}}{|U|}$ is the average per-item cost paid by the optimal cover; it can't be smaller than the least per-item cost we start with, i.e., "average $\geq$ minimum".

Proof of first statement

Proposition

Let S_1 be the first step picked by Greedy, then $\frac{w_1}{|S_1|} \leq \frac{\text{OPT}}{|U|}$.

Intuition: $\frac{\text{OPT}}{|U|}$ is the average per-item cost paid by the optimal cover; it can't be smaller than the least per-item cost we start with, i.e., "average $\geq$ minimum".

Proof.

Let C^* be the optimal set cover, then

$$\frac{\text{OPT}}{|U|} = \frac{\sum_{i \in C^*} w_i}{|U|} \geq \frac{\sum_{i \in C^*} w_i}{\sum_{i \in C^*} |S_i|}$$

□

Proof of first statement

Proposition

Let S_1 be the first step picked by Greedy, then $\frac{w_1}{|S_1|} \leq \frac{\text{OPT}}{|U|}$.

Intuition: $\frac{\text{OPT}}{|U|}$ is the average per-item cost paid by the optimal cover; it can't be smaller than the least per-item cost we start with, i.e., "average $\geq$ minimum".

Proof.

Let C^* be the optimal set cover, then

$$\frac{\text{OPT}}{|U|} = \frac{\sum_{i \in C^*} w_i}{|U|} \geq \frac{\sum_{i \in C^*} w_i}{\sum_{i \in C^*} |S_i|}$$

$$= \sum_{i \in C^*} \frac{w_i}{|S_i|} \cdot \frac{|S_i|}{\sum_{j \in C^*} |S_j|}$$

This is a weighted average of the per-item cost for sets in C^*!

Proof of first statement

Proposition

Let S_1 be the first step picked by Greedy, then $\frac{w_1}{|S_1|} \leq \frac{\text{OPT}}{|U|}$.

Intuition: $\frac{\text{OPT}}{|U|}$ is the average per-item cost paid by the optimal cover; it can't be smaller than the least per-item cost we start with, i.e., "average $\geq$ minimum".

Proof.

Let C^* be the optimal set cover, then

$$\frac{\text{OPT}}{|U|} = \frac{\sum_{i \in C^*} w_i}{|U|} \geq \frac{\sum_{i \in C^*} w_i}{\sum_{i \in C^*} |S_i|}$$

$$= \sum_{i \in C^*} \frac{w_i}{|S_i|} \cdot \frac{|S_i|}{\sum_{j \in C^*} |S_j|} \geq \min_{i \in C^*} \frac{w_i}{|S_i|} \geq \min_i \frac{w_i}{|S_i|} = \frac{w_1}{|S_1|}.$$

□

Generalize the first observation

Let C be the final output of Greedy, note that its weight is just the sum of the "effective" per-item cost $\sum_{s \in U} c_s$.

Generalize the first observation

Let C be the final output of Greedy, note that its weight is just the sum of the "effective" per-item cost $\sum_{s \in U} c_s$.
Suppose the items are covered by Greedy in the order $s_1, \ldots, s_{|U|}$, then we have shown $c_{s_1} \leq \mathrm{OPT}/|U|$ — not a bad start.

Generalize the first observation

Let C be the final output of Greedy, note that its weight is just the sum of the "effective" per-item cost $\sum_{s \in U} c_s$.
Suppose the items are covered by Greedy in the order $s_1, \ldots, s_{|U|}$, then we have shown $c_{s_1} \leq \mathrm{OPT}/|U|$ — not a bad start.

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\mathrm{OPT}}{|U|-j+1}$.

Generalize the first observation

Let C be the final output of Greedy, note that its weight is just the sum of the "effective" per-item cost $\sum_{s \in U} c_s$.
Suppose the items are covered by Greedy in the order $s_1, \ldots, s_{|U|}$, then we have shown $c_{s_1} \leq \mathrm{OPT}/|U|$ — not a bad start.

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\mathrm{OPT}}{|U|-j+1}$.

As a consequence, the cost of C is

$$\sum_{j=1}^{|U|} c_{s_j} \leq \sum_{j=1}^{|U|} \frac{\mathrm{OPT}}{|U|-j+1} = H(|U|) \cdot \mathrm{OPT},$$

where $H(n) := 1 + \frac{1}{2} + \cdots + \frac{1}{n} \approx \ln n$.

Generalize the first observation

Let C be the final output of Greedy, note that its weight is just the sum of the "effective" per-item cost $\sum_{s \in U} c_s$.
Suppose the items are covered by Greedy in the order $s_1, \ldots, s_{|U|}$, then we have shown $c_{s_1} \leq \mathrm{OPT}/|U|$ — not a bad start.

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\mathrm{OPT}}{|U|-j+1}$.

As a consequence, the cost of C is

$$\sum_{j=1}^{|U|} c_{s_j} \leq \sum_{j=1}^{|U|} \frac{\mathrm{OPT}}{|U|-j+1} = H(|U|) \cdot \mathrm{OPT},$$

where $H(n) := 1 + \frac{1}{2} + \cdots + \frac{1}{n} \approx \ln n$.
Anecdote: $\lim_{n\to\infty}(\sum_{k=1}^{n} \frac{1}{k} - \ln n) \approx 0.5772$ is known as Euler's constant.

Proof of proposition

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\text{OPT}}{|U|-j+1}$.

Same intuition as before: when s_j is being covered, there are at least $|U| - j + 1$ elements to cover, and in the optimal solution the "average" per-item cost is at least $\text{OPT}/|U| - j + 1$.

Proof of proposition

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\text{OPT}}{|U|-j+1}$.

Same intuition as before: when s_j is being covered, there are at least $|U| - j + 1$ elements to cover, and in the optimal solution the "average" per-item cost is at least $\text{OPT}/|U| - j + 1$.

Proof.

Apply the previous to the sets in C^* used to cover the remaining items.

Proof of proposition

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\text{OPT}}{|U|-j+1}$.

Same intuition as before: when s_j is being covered, there are at least $|U| - j + 1$ elements to cover, and in the optimal solution the "average" per-item cost is at least $\text{OPT} / |U| - j + 1$.

Proof.

Apply the previous to the sets in C^* used to cover the remaining items. Let R_j be the set of items remaining to be covered (in R) right before Greedy covers s_j, then $|R_j| \geq |U| - j + 1$.

Proof of proposition

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\mathrm{OPT}}{|U|-j+1}$.

Same intuition as before: when s_j is being covered, there are at least $|U| - j + 1$ elements to cover, and in the optimal solution the "average" per-item cost is at least $\mathrm{OPT}\,/|U| - j + 1$.

Proof.

Apply the previous to the sets in C^* used to cover the remaining items.
Let R_j be the set of items remaining to be covered (in R) right before Greedy covers s_j, then $|R_j| \geq |U| - j + 1$.
Let C_j^* be the sets in the optimal solution used to cover R_j, i.e., $\{i \in C^* : S_i \cap R_j \neq \emptyset\}$.

Proof of proposition

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\text{OPT}}{|U|-j+1}$.

Same intuition as before: when s_j is being covered, there are at least $|U| - j + 1$ elements to cover, and in the optimal solution the "average" per-item cost is at least $\text{OPT} / |U| - j + 1$.

Proof.

Apply the previous to the sets in C^* used to cover the remaining items. Let R_j be the set of items remaining to be covered (in R) right before Greedy covers s_j, then $|R_j| \geq |U| - j + 1$.
Let C_j^* be the sets in the optimal solution used to cover R_j, i.e., $\{i \in C^* : S_i \cap R_j \neq \emptyset\}$.

$$\frac{\sum_{i \in C_j^*} w_i}{|R_j|} \geq \frac{\sum_{i \in C_j^*} w_i}{\sum_{i \in C_j^*} |S_i \cap R_j|} = \sum_{i \in C_j^*} \frac{w_i}{|S_i \cap R_j|} \cdot \frac{|S_i \cap R_j|}{\sum_{k \in C_j^*} |S_k \cap R_j|}$$

Proof of proposition

Proposition

For $j = 1, \cdots, |U|$, $c_{s_j} \leq \frac{\text{OPT}}{|U|-j+1}$.

Same intuition as before: when s_j is being covered, there are at least $|U| - j + 1$ elements to cover, and in the optimal solution the "average" per-item cost is at least $\text{OPT}/|U| - j + 1$.

Proof.

Apply the previous to the sets in C^* used to cover the remaining items.
Let R_j be the set of items remaining to be covered (in R) right before Greedy covers s_j, then $|R_j| \geq |U| - j + 1$.
Let C_j^* be the sets in the optimal solution used to cover R_j, i.e., $\{i \in C^* : S_i \cap R_j \neq \emptyset\}$.

$$\frac{\sum_{i \in C_j^*} w_i}{|R_j|} \geq \frac{\sum_{i \in C_j^*} w_i}{\sum_{i \in C_j^*} |S_i \cap R_j|} = \sum_{i \in C_j^*} \frac{w_i}{|S_i \cap R_j|} \cdot \frac{|S_i \cap R_j|}{\sum_{k \in C_j^*} |S_k \cap R_j|}$$

Proof continued

Proof (Cont.)

$$\frac{\sum_{i \in C_j^*} w_i}{|R_j|} \geq \frac{\sum_{i \in C_j^*} w_i}{\sum_{i \in C_j^*} |S_i \cap R_j|} = \sum_{i \in C_j^*} \frac{w_i}{|S_i \cap R_j|} \cdot \frac{|S_i \cap R_j|}{\sum_{k \in C_j^*} |S_k \cap R_j|}$$

$$\geq \min_{i \in C_j^*} \frac{w_i}{|S_i \cap R_j|} \geq \min_{i : S_i \cap R_j \neq \emptyset} \frac{w_i}{|S_i \cap R_j|} = c_{s_j}.$$

Proof continued

Proof (Cont.)

$$\frac{\sum_{i \in C_j^*} w_i}{|R_j|} \geq \frac{\sum_{i \in C_j^*} w_i}{\sum_{i \in C_j^*} |S_i \cap R_j|} = \sum_{i \in C_j^*} \frac{w_i}{|S_i \cap R_j|} \cdot \frac{|S_i \cap R_j|}{\sum_{k \in C_j^*} |S_k \cap R_j|}$$

$$\geq \min_{i \in C_j^*} \frac{w_i}{|S_i \cap R_j|} \geq \min_{i : S_i \cap R_j \neq \emptyset} \frac{w_i}{|S_i \cap R_j|} = c_{s_j}.$$

Finally, note that $\frac{\sum_{i \in C_j^*} w_i}{|R_j|} \leq \frac{\text{OPT}}{|R_j|}$. □