Utility Function

$Food = \{Sushi, Gyu\text{-}don, Pasta, Melon\text{-}pan, Cola\}$

$FoodInOneMeal \subseteq Food$

$f(FoodInOneMeal) :=$ How much happiness you have when you are provided FoodInOneMeal

↑ utility function

$f(\{Sushi, Cola\}) = 100$

$f(\{Melon\text{-}pan, Cola\}) = 10$

Properties of Utility Function

1. Monotonicity $\quad f(S) \leq f(S')$ when $S \subseteq S'$

Ex $\quad f(\{Sushi, Cola\}) \leq f(\{Sushi, Melon\text{-}pan, Cola\})$

2. Submodularity

we are really happy to have more

we are not that happy to have more.

$$f(S \cup \{e\}) - f(S) \geq f(S' \cup \{e\}) - f(S')$$

$f(S \cup \{e\}) - f(S)$: we do not have enough food

$f(S' \cup \{e\}) - f(S')$: we already have enough food

when $S \subseteq S'$

Ex $\quad f(\{Cola\} \cup \{Sushi\}) - f(\{Cola\})$

$\geq f(\{pasta, cola\} \cup \{Sushi\}) - f(\{pasta, cola\})$

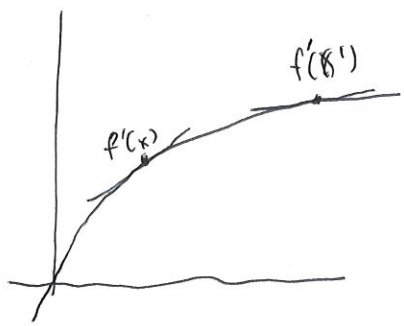
$$\Delta_e f(S) := f(S \cup \{e\}) - f(S) \iff \frac{f(x+\varepsilon) - f(x)}{\varepsilon} = \frac{df(x)}{dx} = f'(x)$$

$$\Delta_e f(S) \geq \Delta_e f(S') \iff f'(x) \geq f'(x')$$

when $S \subseteq S'$ $\quad$ when $x \leq x'$

concave function

Discrete version of concave function $\iff$



By Lovász extension, submodular function is also close to convex function.

More ~~exp~~example : Set of persons $= \{1, 2, 3, \dots, 11\}$

Mailing list $A = \{1, 2, 3\}$

Mailing list $B = \{2, 3, 4, 5\}$

Mailing list $C = \{5, 9, 10, 11\}$

Mailing list $D = \{4, 6, 7, 8\}$

$S :=$ set of mailing lists that we send e-mails to

$f(S) :=$ #users that we can reach by S.

↓

Utility of S

$f(\{A,B,C\}) = \#(\{1,2,3\} \cup \{2,3,4,5\} \cup \{5,9,10,11\}) = |\{1,2,3,4,5,9,10,11\}| = 8$

$f(\{A,B,C,D\}) = 11 \quad f(\{A,C\}) = 7 \quad f(\{A,C,D\}) = 11$

$$f(\{A,B,C,D\}) - f(\{A,B,C\}) \le f(\{A,C,D\}) - f(\{A,C\})$$

$$11 - 8 \le 11 - 8$$

$$3 \le 4$$

$$\Delta_D f(\{A,B,C\}) \le \Delta_D f(\{A,C\})$$

#persons that can reach by Mailing list D but cannot reach by Mailing list A,B,C $\le$ #persons that can reach by Mailing list D but cannot reach by Mailing list A,C

$\therefore f$ is a submodular function.

Also, when $S \subseteq S'$, we can reach more mailing lists with S' than with S (more persons)

$$f(S) \le f(S').$$

$\therefore f$ is a monotone function.

Problem : Maximum Coverage Problem

Input : # persons n (set of persons $= \{1, \dots, n\}$) , positive integer k.

Set of persons that can be reached by each mailing list

$S_1, \dots, S_m \subseteq \{1, \dots, n\}$

→ set of persons that can be reached by mailing list 1.

Output : Set of mailing list to send an e-mail to $S \subseteq \{1, \dots, m\}$

↘ if $2 \in S$, we send an e-mail to Mailing list 2.

Constraint : $|S| \leq k$

[We cannot send e-mails to more than k mailing lists]

Objective Function : Maximize # persons we can reach by S.

Maximize $f\left(\bigcup_{i \in S} S_i\right)$

Reformulate the problem

Input : # persons n, integer k, $f: 2^{\{1, \dots, m\}} \Rightarrow \mathbb{R}$.

(f is given as a code in Python (oracle))

Output : $S \subseteq \{1, \dots, m\}$

(Ex person reached)

Constraint : $|S| \leq k$

(f is a monotone submodular function)

Objective Function Maximize $f(S)$

Submodular function maximization under size constraint.

Algorithm

1 : $S \leftarrow \emptyset$

2 : for $i = 1$ to k

3 : Find j that maximizes $\Delta_j f(S)$

4 : $S \leftarrow S \cup \{j\}$

Ex $k=3$

Step 1 $S=\phi$

$\Delta_A f(S)=3$ $\Delta_B f(S)=4$ $\Delta_C f(S)=4$ $\Delta_D f(S)=4$

maximum

- Choose B

Step 2 $S=\{B\}$

$\Delta_A f(S)=1$ $\Delta_B f(S)=0$ $\Delta_C f(S)=3$ $\Delta_D f(S)=3$

maximum

- Choose C

Step 3 $S=\{B,C\}$

$\Delta_A f(S)=1$ $\Delta_B f(S)=0$ $\Delta_C f(S)=0$ $\Delta_D f(S)=3$

maximum

- Choose D

Solution from the algorithm $=\{B,C,D\}$ $f(\{B,C,D\})=10$

Best solution $=\{A,C,D\}$ $f(\{A,C,D\})=11$

Theorem If SOL := solution from the algorithm

OPT := best solution, then

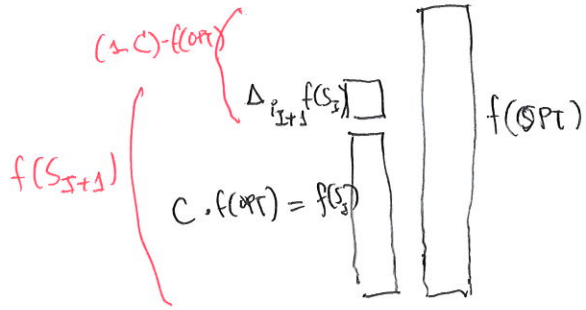
$$f(SOL) \geq 0.63 \cdot f(OPT)$$

Notation i_t := element added at iteration t

S_t := set S at the end of iteration $t = \{i_1, \dots, i_t\}$

$SOL := \{i_1, \dots, i_k\}$, and i_t maximizes $\Delta_{i_t} f(\{i_1, \dots, i_{t-1}\})$

$OPT := \{i_1^*, \dots, i_k^*\}$



$C \geq 1 - (1 - 1/k)^I$

By lemma, $\Delta_{i_{I+1}} f(S_I) \geq 1/k \, (f(OPT) - f(S_I))$

$= 1/k \, (1-C) \cdot f(OPT)$

$f(S_{I+1}) = C \cdot f(OPT) + \Delta_{i_{I+1}} f(S_I)$

$\geq C \cdot f(OPT) + 1/k \, (1-C) \cdot f(OPT)$

$= [C + 1/k (1-C)] \cdot f(OPT)$

$= [(1 - 1/k) C + 1/k] \cdot f(OPT)$

$\geq [(1 - 1/k) \cdot (1 - (1 - 1/k)^I) + 1/k] \cdot f(OPT)$

$= [1 - 1/k - (1 - 1/k)^{I+1} + 1/k] \cdot f(OPT)$

$= [1 - (1 - 1/k)^{I+1}] \cdot f(OPT)$

$f(S_{I+1}) \geq [1 - (1 - 1/k)^{I+1}] \cdot f(OPT)$ ☑

Theorem $f(SOL) = f(S_k) \geq 0.63 \cdot f(OPT)$

Proof $f(S_k) \geq (1 - (1 - 1/k)^k) \cdot f(OPT)$

$(1 - 1/k)^k \leq e^{-1/k}$

$1 + x \leq e^x$

$1 - 1/k \leq e^{-1/k}$

$\geq (1 - (e^{-1/k})^k) \cdot f(OPT)$

$= (1 - e^{-1}) \cdot f(OPT)$

$1 - e^{-1} \geq 0.63$

$\geq 0.63 \cdot f(OPT)$ ☑