

Mathematical Reinforcement Learning

yataka

https://github.com/takaya0/Reinforcement_Learning

目次

第 I 部	数学的知識	5
第 1 章	集合論	7
1.1	様々な集合	7
1.2	集合と写像	8
1.3	集合系と集合族	9
第 2 章	位相空間論	11
2.1	距離空間	11
2.2	位相空間	11
第 3 章	力学系	13
第 4 章	測度論	15
4.1	測度空間	15
4.2	測度空間と位相空間	17
4.3	ルベーグ測度	19
第 5 章	関数解析	21
第 6 章	確率論	23
6.1	確率測度・確率空間	23
第 II 部	最適化とアルゴリズム	25
第 7 章	最適化数学	27
第 8 章	劣モジュラ関数	29
8.1	劣モジュラ関数	29
8.2	単調劣モジュラ関数最大化問題と貪欲法	33
8.3	応用例	36

第 III 部 数理機械学習	37
第 9 章 単回帰分析・重回帰分析	39
第 10 章 再生核ヒルベルト空間とカーネル法	41
第 IV 部 数理強化学習	43
第 11 章 マルコフ決定過程	45
第 12 章 深層強化学習	47
第 V 部 強化学習の応用	49
第 13 章 ゲームへの応用	51
第 14 章 巡回セールスマン問題への応用	53
参考文献	55

第 I 部

数学的知識

第 1 章

集合論

ここでは集合を、ある条件を満たすものを集めたものとして定義する.

1.1 様々な集合

Definition 1.1.1. (全体集合)

その状況における一番大きい集合となる集合を全体集合という.

Definition 1.1.2. (部分集合)

X を全体集合とし, A, B を集合とする. A が B の部分集合であるとは

$$\forall x \in X, x \in A \implies x \in B$$

が成り立つことであり, A が B の部分集合であることを

$$A \subset B \text{ または } B \supset A$$

と表す.

Definition 1.1.3. (空集合)

元を 1 つも含まない集合のことを空集合といい, ϕ と表記する.

Exercise 1.1.4. 空集合は任意の集合の部分集合である.

Proof. A を任意の集合とする. 命題

$$\forall x \in X, x \in \phi \implies x \in A$$

は恒真であるので, 部分集合の定義より $\phi \subset A$ が成立する. □

Definition 1.1.5. (差集合)

X, Y を集合とする. X の元ではあるが Y の元ではないものを集めた集合を差集合といい $X - Y$ と表す. すなわち

$$X - Y = \{x \in X | x \in X \text{ かつ } x \notin Y\}$$

Definition 1.1.6. (補集合)

X を全体集合とする. $A \subset X$ とし差集合 $X - A$ を A の補集合といい A^c で表す. すなわち

$$A^c = \{x \in X \mid x \in X \text{ かつ } x \notin A\}$$

Definition 1.1.7. (和集合)

A, B を集合とする. A の元または B の元を集めた集合を A と B の和集合といい $A \cup B$ と表す. すなわち

$$A \cup B = \{x \in X \mid x \in A \text{ または } x \in B\}$$

Definition 1.1.8. (共通集合)

A, B を集合とする. A の元かつ B の元であるものを集めた集合を A と B の共通集合といい $A \cap B$ と表す. すなわち

$$A \cap B = \{x \in X \mid x \in A \text{ または } x \in B\}$$

Proposition 1.1.9. (ド・モルガンの法則)

A, B を X の部分集合とする. この時, 以下の事が成立する.

1. $(A \cap B)^c = A^c \cup B^c$
2. $(A \cup B)^c = A^c \cap B^c$

Exercise 1.1.10. A, B を X の部分集合とする. この時, 以下の等式を証明せよ.

1. $(A^c)^c = A$
2. $(A - B) \cap B = \phi$

Definition 1.1.11. (直積集合)

X, Y を集合とする. 直積集合 $X \times Y$ を以下のように定義する.

$$X \times Y = \{(x, y) \mid x \in X, y \in Y\}$$

1.2 集合と写像

Definition 1.2.1. (写像)

X, Y を集合とする. f が X の任意の要素を Y の元にただ一つ対応させる操作のことを写像といい, X から Y への写像であるということを

$$f : X \rightarrow Y$$

と表す.

Definition 1.2.2. (単射)

f は X から Y への写像であるとする. f が

$$\forall x_1, x_2 \in X, f(x_1) = f(x_2) \implies x_1 = x_2$$

を満たすとき f は単射であるという.

Definition 1.2.3. (全射)

f は X から Y への写像であるとする. f が

$$\forall y \in Y, \exists x \in X \text{ s.t. } y = f(x)$$

を満たすとき f は全射であるという.

Definition 1.2.4. (像)

X, Y を集合とし, f を X から Y への写像であるとする. X の部分集合 S の f による像を

$$f(S) = \{y \in Y \mid \exists s \in S \text{ s.t. } f(s) = y\}$$

と定義する. 定義より $f(S)$ は Y の部分集合である.

Theorem 1.2.5. X, Y を集合とする. 以下の 2 つは同値である.

1. f は全射
2. $f(X) = Y$

Proof. 1.) \implies 2.) $f(X) \subset Y$ は自明なので $Y \subset f(X)$ を示す. 任意に $y \in Y$ を任意にとる. f は全射だから

$$\exists x \in X \text{ s.t. } f(x) = y$$

である. したがって $y \in f(X)$ であるので $Y \subset f(X)$.

2.) \implies 1.) 任意に $y \in Y$ を任意にとる. $f(X) = Y$ だから $y \in f(X)$ である. 定義より

$$\exists x \in X \text{ s.t. } f(x) = y$$

したがって $y \in Y$ は任意なので, f は全射. □

Definition 1.2.6. (全単射)

f は X から Y への写像であるとする. f が単射かつ全射であるとき f は全単射であるという.

1.3 集合系と集合族

ここからは集合を元に持つ集合を考えることにする.

Definition 1.3.1. (集合系)

X を集合とする. X の全ての部分集合の集合を集合系といい $\mathcal{P}(X)$ と表記する. すなわち

$$\mathcal{P}(X) = \{A \mid A \subset X\}$$

Definition 1.3.2. (集合族)

添字づけられた集合の集合のことを集合族といい, $(A_\lambda \mid \lambda \in \Lambda)$ または $(A_\lambda)_{\lambda \in \Lambda}$ と表す. この時の Λ を添字集合という.

Definition 1.3.3. (部分集合族)

X を集合とする. X の部分集合からなる集合族のことを X の部分集合族という.

Definition 1.3.4. (集合族の和集合)

集合族 $(A_\lambda \mid \lambda \in \Lambda)$ の和集合を

$$\bigcup_{\lambda \in \Lambda} A_\lambda = \{x \in X \mid \exists \lambda \in \Lambda \text{ s.t. } x \in A_\lambda\}$$

と定義する.

Definition 1.3.5. (集合族の共通集合)

集合族 $(A_\lambda \mid \lambda \in \Lambda)$ の共通集合を

$$\bigcap_{\lambda \in \Lambda} A_\lambda = \{x \in X \mid \forall \lambda \in \Lambda \text{ s.t. } x \in A_\lambda\}$$

と定義する.

第 2 章

位相空間論

2.1 距離空間

Definition 2.1.1. (距離空間)

X を空でない集合とする. $d: X \times X \rightarrow \mathbb{R}$ が, 任意の $x, y, z \in X$ に対して

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \iff x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$

を満たす時 d を距離関数 (上の 4 つの公理を距離の公理と呼ぶ) といい, 組 (X, d) を距離空間という.

2.2 位相空間

Definition 2.2.1. (位相空間)

X を空でない集合とする. $\mathcal{P}(X)$ の部分集合 \mathcal{O} が

1. $X, \emptyset \in \mathcal{O}$
2. $O_1, O_2, \dots, O_n \in \mathcal{O}$ ならば

$$\bigcap_{i=1}^n O_i \in \mathcal{O}$$

3. $(O_\lambda | \lambda \in \Lambda)$ を \mathcal{O} の元からなる集合族とすれば

$$\bigcup_{\lambda \in \Lambda} O_\lambda \in \mathcal{O}$$

を満たす時, \mathcal{O} を X の位相といい, 組 (X, \mathcal{O}) を位相空間と呼ぶ.

第 3 章

力学系

第 4 章

測度論

まず初めに測度論に関する事項をのべる.

4.1 測度空間

まず, 面積を測れる集合である σ -algebra を定義する.

Definition 4.1.1. (σ -algebra)

X を集合とする. X の部分集合族 \mathcal{F} が

1. $X \in \mathcal{F}$
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
- 3.

$$A_i \in \mathcal{F} (i \in \mathbb{N}) \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

を満たす時 \mathcal{F} を σ -algebra という.

Example 4.1.2. 集合 X の冪集合 $\mathcal{P}(X)$ は σ -algebra となる.

Proof. 1. $X \in \mathcal{P}(X)$ は $X \subset X$ より言える.

2. $A \in \mathcal{P}(X) \implies A^c \in \mathcal{P}(X)$ を示す.

任意に $A \in \mathcal{P}(X)$ をとる. $X - A \subset X$ より, $A^c \in \mathcal{P}(X)$.

3. $A_i \in \mathcal{P}(X) (i \in \mathbb{N}) \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{P}(X)$ を示す.

任意に $A_i \in \mathcal{P}(X) (i \in \mathbb{N})$ をとる.

$$\bigcup_{i=1}^{\infty} A_i \subset X$$

であるので, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{P}(X)$.

1. 2. 3. より $\mathcal{P}(X)$ は σ -algebra である. □

Theorem 4.1.3. \mathcal{F} を σ -algebra する. この時

1. $\phi \in \mathcal{F}$

2.

$$A_1, A_2, \dots, A_n \in \mathcal{F} \implies \bigcup_{i=1}^n A_i \in \mathcal{F}$$

3.

$$A_1, A_2, \dots, A_n \in \mathcal{F} \implies \bigcap_{i=1}^n A_i \in \mathcal{F}$$

が成立する.

Proof. 1. $X \in \mathcal{F}$ より $\phi = X^c \in \mathcal{F}$.

2. 任意に \mathcal{F} の元 A_1, A_2, \dots, A_n をとる. この時

$$B_i = \begin{cases} A_i & (1 \leq i \leq n) \\ \phi & (i > n) \end{cases}$$

として $B_i \in \mathcal{F}$ ($i \in \mathbb{N}$) を定義すれば, $\phi \in \mathcal{F}$ より,

$$\bigcup_{i=1}^n A_i = \bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$$

3. 任意に \mathcal{F} の元 A_1, A_2, \dots, A_n をとる. この時定義より $A_1^c, A_2^c, \dots, A_n^c \in \mathcal{F}$ であり,

$$\bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c \right)^c \in \mathcal{F}$$

□

Definition 4.1.4. (可測空間)

集合 X と X 上の σ -algebra の組 (X, \mathcal{F}) を可測空間と呼ぶ.

次に面積を測る写像である測度 m を定義する.

Definition 4.1.5. (測度)

\mathcal{F} を σ アルジェブラとする. \mathcal{F} 上の写像 m が,

1. $\forall A \in \mathcal{F}, 0 \leq m(A) \leq \infty$, 特に $m(\phi) = 0$

2. $A_i \in \mathcal{F} (i \in \mathbb{N})$ が互いに排反 ($\forall i, k \in \mathbb{N}, i \neq k \Rightarrow A_i \cap A_k = \phi$) であるならば

$$m \left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} m(A_i)$$

を満たす時 m を \mathcal{F} 上の測度という.

Theorem 4.1.6. m を \mathcal{F} 上の測度とする. この時

1. $A_1, A_2, \dots, A_n \in \mathcal{F}$ が互いに排反である時

$$m\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n m(A_i)$$

2. $A, B \in \mathcal{F}$ で $A \subset B$ の時

$$m(A) \leq m(B)$$

が成立し, 特に $m(A) < \infty$ の時は

$$m(B - A) = m(B) - m(A)$$

が成立する.

Proof. 1. 任意に互いに排反な \mathcal{F} の元 A_1, A_2, \dots, A_n をとる. この時

$$B_i = \begin{cases} A_i & (1 \leq i \leq n) \\ \phi & (i > n) \end{cases}$$

として $B_i \in \mathcal{F}$ ($i \in \mathbb{N}$) を定義すれば, $m(\phi) = 0$ より,

$$m\left(\bigcup_{i=1}^n A_i\right) = m\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} m(B_i) = \sum_{i=1}^n m(A_i)$$

2. 任意に $A, B \in \mathcal{F}$ をとり $A \subset B$ と仮定する. $C = B - A$ とすれば

$$\begin{aligned} m(B) &= m(A \cup C) = m(A) + m(C) \\ &= m(A) + m(B - A) \\ &\leq m(A) \end{aligned}$$

ここで $m(A) < \infty$ とすれば, 上記の式より

$$m(B - A) = m(B) - m(A)$$

□

Definition 4.1.7. (測度空間)

(X, \mathcal{F}) を可測空間とし \mathcal{F} 上の測度を m とする. この時組 (X, \mathcal{F}, m) を測度空間と呼ぶ.

4.2 測度空間と位相空間

ここでは, 位相空間から生成される σ -algebra について述べる.

Lemma 4.2.1. Λ を添字集合とし, \mathcal{F}_λ を σ -algebra の族とする. この時

$$\mathcal{F} = \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$$

とすれば \mathcal{F} も σ -algebra である.

Proof. i) 各 \mathcal{F}_λ は σ -algebra だから

$$\forall \lambda \in \Lambda, X \in \mathcal{F}_\lambda$$

であるので, 定義より $X \in \mathcal{F}$.

ii) 任意に \mathcal{F} の元 A をとる.

$$A \in \mathcal{F}$$

定義より, 任意の $\lambda \in \Lambda$ に対して

$$A \in \mathcal{F}_\lambda$$

が成立する. 各 \mathcal{F}_λ は σ -algebra だから

$$A^c \in \mathcal{F}_\lambda$$

$\lambda \in \Lambda$ は任意なので $A^c \in \mathcal{F}$.

iii) $\{A_n\}_{n \in \mathbb{N}}$ を \mathcal{F} の元からなる可算列とする. 定義より, 任意の $\lambda \in \Lambda$ に対して

$$\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{F}_\lambda$$

が成立する. 各 \mathcal{F}_λ は σ -algebra だから

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_\lambda$$

$\lambda \in \Lambda$ は任意なので

$$\bigcup_{i=1}^{\infty} A_i \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda = \mathcal{F}$$

したがって \mathcal{F} は σ -algebra である. □

Theorem 4.2.2. X の部分集合からなる任意の集合族 \mathcal{U} に対して, \mathcal{U} を含む最小の σ -algebra が存在する. またこの σ -algebra のことを $\sigma(\mathcal{U})$ と表す.

Proof. \mathcal{O} を含む σ -algebra は冪集合が σ -algebra であるので存在する. ここで \mathcal{O} を含む σ -algebra 全ての共通集合をとれば, **Lemma 4.2.1** よりそれは σ -algebra であり, 最小である. □

Definition 4.2.3. (ボレル集合族)

(X, \mathcal{O}) を位相空間とする. この時 \mathcal{O} を含む最小の σ -algebra のことをボレル集合族といい $\mathcal{B}(\mathcal{O})$ で表す. 特に, 位相空間が $(\mathbb{R}, \mathcal{O}_{\mathbb{R}})$ の時は $\mathcal{B}(\mathbb{R})$ と表す.

4.3 ルベーク測度

第 5 章

関数解析

第 6 章

確率論

6.1 確率測度・確率空間

いよいよ今まで書いてきた測度論に基づいて確率空間を定義する.

Definition 6.1.1. (確率測度)

(Ω, \mathcal{F}) を可測空間とする. \mathcal{F} から $[0, 1]$ への写像 P が

1. $P(\Omega) = 1$
2. $A_i \in \mathcal{F} (i \in \mathbb{N})$ が互いに排反であるならば

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

を満たす時, P を確率測度と呼ぶ.

Definition 6.1.2. (確率空間)

(Ω, \mathcal{F}) を可測空間とし P を確率測度とする. この時, 組 (Ω, \mathcal{F}, P) を確率空間と呼ぶ. また, Ω を標本空間といい \mathcal{F} の元を事象と呼ぶ.

測度論の節では σ -algebra を「面積が図れる集合の集まり」, 測度を「集合の面積」を測る写像と言うようなモチベーションで定義したが, 確率空間では σ -algebra を「確率が測れる集合の集まり」確率測度を「確率を測れる」写像としてそれぞれに対する解釈を変える.

Theorem 6.1.3. (Ω, \mathcal{F}, P) を確率空間とする. この時

$$\forall A \in \mathcal{F}, P(A^c) = 1 - P(A)$$

が成立する.

Proof. 任意に $A \in \mathcal{F}$ をとり $X = \Omega - A$ とする. $A \cap X = \phi$, $\Omega = A \cup X$ であるので.

$$\begin{aligned} P(\Omega) &= P(A \cup X) \\ &= P(A) + P(X) \end{aligned}$$

$P(A) < \infty$ より

$$\begin{aligned} P(X) &= P(\Omega) - P(A) \\ &= 1 - P(A) \end{aligned}$$

したがって, $X = \Omega - A = A^c$ で A は任意だったから,

$$\forall A \in \mathcal{F}, P(A^c) = 1 - P(A)$$

が成立する.

□

第II部

最適化とアルゴリズム

第 7 章

最適化数学

第 8 章

劣モジュラ関数

この章は本 PDF の目標と関連性が薄い³が、最適化問題の具体例として紹介する。

8.1 劣モジュラ関数

まず劣モジュラ関数を定義して行く。もしこれらの定義を見てよくわからないと感じた読者は参考文献 [6] の動画が良い参考になる。

Definition 8.1.1. (集合関数)

f を V の冪集合 $\mathcal{P}(V)$ から \mathbb{R} への写像とする。この時 f を集合関数と呼ぶ。

Definition 8.1.2. (離散微分)

f を集合関数とする。 S を V の部分集合とし e を V の元とする。この時 f の S での離散微分 (Discrete derivative) を

$$\Delta_f(e|S) := f(S \cup \{e\}) - f(S)$$

と定義する。

Definition 8.1.3. (劣モジュラ関数)

f を集合関数とする。 f が劣モジュラ関数 (Submodular function) であるとは

$$\forall A, B \subset V, \forall e \in V, (A \subset B \text{ かつ } e \in V - B) \implies \Delta_f(e|A) \geq \Delta_f(e|B)$$

を満たすことをいう。

Example 8.1.4. (カバー関数)

劣モジュラ関数の具体例としてカバー関数 (cover function) というものがある。カバー関数 $f: 2^V \rightarrow \mathbb{R}$ とは

$$f(T) = \left| \bigcup_{i \in T} S_i \right| \quad (T \subset V)$$

で定義される関数である。ただし $S_i (1 \leq i \leq |V|)$ は \mathbb{R}^2 の有限部分集合である。この関数が劣モジユラ関数であることをしめす。

Proof. 任意に V の部分集合 A, B と $e \in V$ を取る。 $A \subset B, e \in V - B$ と仮定する。まず

$$\bigcup_{i \in B \cup \{e\}} S_i - \bigcup_{i \in B} S_i \subset \bigcup_{i \in A \cup \{e\}} S_i - \bigcup_{i \in A} S_i$$

であることを示す。任意に $\bigcup_{i \in B \cup \{e\}} S_i - \bigcup_{i \in B} S_i$ の元 x をとる。

$$x \in \bigcup_{i \in B \cup \{e\}} S_i - \bigcup_{i \in B} S_i$$

$$x \in \bigcup_{i \in B \cup \{e\}} S_i \text{ かつ } x \notin \bigcup_{i \in B} S_i$$

ここで $A \subset B$ であるので

$$x \in e \text{ かつ } \forall i \in x \in A, x \notin S_i$$

$$x \in \bigcup_{i \in A \cup \{e\}} S_i - \bigcup_{i \in A} S_i$$

したがって、

$$\bigcup_{i \in B \cup \{e\}} S_i - \bigcup_{i \in B} S_i \subset \bigcup_{i \in A \cup \{e\}} S_i - \bigcup_{i \in A} S_i$$

である。これより

$$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$$

と言える。したがって f は劣モジユラ関数である。 □

Theorem 8.1.5. f を集合関数とする。この時

1. f は劣モジユラ関数である。
2. $\forall A, B \subset V, f(A \cap B) + f(A \cup B) \leq f(A) + f(B)$

は同値である。

Proof. 1. \implies 2.) 任意に $A, B \subset V$ をとる。 $A - B = \phi$ の時、すなわち $A \subset B$ の時は $A \cap B = A, A \cup B = B$ より成立する。そこで、 $A - B \neq \phi$ の時を考える。 $|A - B| = m$ とし $A - B = \{i_1, i_2, \dots, i_m\}$ とする。 $A_0 = A \cap B, B_0 = B$ として $A_k, B_k (k \in \{1, 2, \dots, m\})$ を以下のように定義する。

$$\begin{aligned} A_1 &= A_0 \cup \{i_1\}, & A_2 &= A_0 \cup \{i_1, i_2\}, & \dots, & & A_m &= A_0 \cup \{i_1, i_2, \dots, i_m\} \\ B_1 &= B_0 \cup \{i_1\}, & B_2 &= B_0 \cup \{i_1, i_2\}, & \dots, & & B_m &= B_0 \cup \{i_1, i_2, \dots, i_m\} \end{aligned}$$

この時 $A_m = A$, $B_m = A \cup B$ である. また各 $k \in \{1, 2, \dots, m\}$ に対して

$$A_{k-1} \subset B_{k-1} \text{ かつ } A_k = A_{k-1} \cup \{i_k\} \text{ かつ } B_k = B_{k-1} \cup \{i_k\}$$

が成立するので 1. より

$$\forall k \in \{1, 2, \dots, m\}, f(A_k) - f(A_{k-1}) \geq f(B_k) - f(B_{k-1})$$

が成立する. k についてこの不等式の両辺の和をとると,

$$\sum_{k=1}^m (f(A_k) - f(A_{k-1})) \geq \sum_{k=1}^m (f(B_k) - f(B_{k-1}))$$

両辺を計算すると

$$f(A_m) - f(A_0) \geq f(B_m) - f(B_0)$$

となる. $A_0 = A \cap B$, $A_m = A$, $B_m = A \cup B$, $B_0 = B$ より

$$f(A) - f(A \cap B) \geq f(A \cup B) - f(B)$$

A, B は任意より 2. が成立する.

2. \implies 1.) 任意に $A, B \subset V$ と $i \in V - B$ を $A \subset B$ となるようにとる. $A' = A \cup \{i\}$, $B' = B$ とすると 2. より

$$f(A') + f(B') \geq f(A' \cup B') + f(A' \cap B')$$

が成立する. ここで $A' \cup B' = B \cup \{i\}$, $A' \cap B' = A$ であることから

$$f(A') + f(B') \geq f(B \cup \{i\}) + f(A)$$

A', B' の定義より

$$f(A \cup \{i\}) + f(B) \geq f(B \cup \{i\}) + f(A)$$

式を整理すると

$$f(A \cup \{i\}) - f(A) \geq f(B \cup \{i\}) - f(B)$$

すなわち $\Delta_f(i|A) \geq \Delta_f(i|B)$ である. $A, B \subset V$ と $i \in V - B$ は任意だったので, f は劣モジュラ関数である. \square

Definition 8.1.6. (単調)

f を集合関数とする. f が単調 (Monotone) であるとは

$$\forall S, T \subset V, S \subset T \implies f(S) \leq f(T)$$

を満たすことを言う.

Theorem 8.1.7. f を集合関数とする. このとき

1. f が単調である.
2. $\forall S \subset V, \forall e \in V, \Delta_f(e|S) \geq 0$

は同値である.

Proof. 証明は読者への演習問題とする. □

Theorem 8.1.8. f_1, f_2, f を劣モジユラ関数. α を正の実数とする. この時

1. $f_1 + f_2$ は劣モジユラ関数.
2. αf は劣モジユラ関数.

が成立する.

Proof. 任意に V の部分集合 A, B と V の元 e をとる. $A \subset B, e \in V - B$ と仮定する.

1. f_1, f_2 は劣モジユラ関数 だから

$$\Delta_{f_1}(e|A) \geq \Delta_{f_1}(e|B)$$

$$\Delta_{f_2}(e|A) \geq \Delta_{f_2}(e|B)$$

が成立する. 両辺を足すと

$$\Delta_{f_1}(e|A) + \Delta_{f_2}(e|A) \geq \Delta_{f_1}(e|B) + \Delta_{f_2}(e|B)$$

したがって

$$\Delta_{f_1+f_2}(e|A) \geq \Delta_{f_1+f_2}(e|B)$$

したがって, $f_1 + f_2$ は劣モジユラ関数である.

2. $\alpha > 0$ を任意にとる. f は劣モジユラ関数であるので

$$\Delta_f(e|A) \geq \Delta_f(e|B)$$

両辺を $\alpha > 0$ 倍すると

$$\alpha \Delta_f(e|A) \geq \alpha \Delta_f(e|B)$$

したがって

$$\Delta_{\alpha f}(e|A) \geq \Delta_{\alpha f}(e|B)$$

したがって αf は劣モジユラ関数である. □

Definition 8.1.9. (単調劣モジユラ関数)

集合関数 f が劣モジユラかつ単調であるとき f を単調劣モジユラ関数 (Monotone Submodular function) と言う.

8.2 単調劣モジュラ関数最大化問題と貪欲法

この節では f を非負の単調劣モジュラ関数とする.

8.2.1 単調劣モジュラ関数最大化問題

前節では劣モジュラ関数の定義および様々な性質を述べてきた. この節での目的は単調劣モジュラ関数がある条件のもと最大化^{*1}する V の部分集合 S を見つけるアルゴリズムを与えることである. そして, この節では特に集合

$$\{f(S) \mid S \subset V \text{ and } |S| \leq k\}$$

が最大値をとるときの V の部分集合 S を見つけることが目標である (そのような S を最適解と呼ぶ). また k は $0 < k \leq |V|$ を満たす定数である. この問題をこの方程では単調劣モジュラ関数最大化問題 (Monotone Submodular function maximization problem) と呼ぶことにする.

8.2.2 近似アルゴリズム

単調劣モジュラ関数最大化問題を解決するための手法について述べる. まず, 最初に全探索アルゴリズムを考える. これは「制約下でのあらゆる可能性を全て探索する^{*2}」アルゴリズムである. このアルゴリズムであれば確実に解が出せるので単調劣モジュラ関数最大化問題を解決することができる. しかしながら, このアルゴリズムでは V の部分集合全てを調べる必要があるため $2^{|V|}$ 回の計算を行う必要がある.(簡単のため $k = |V|$ とした.) $y = 2^x$ という関数は x がある程度大きくなると y は非常に大きな値となる. これではいくらコンピュータを用いたとしても解くのは不可能である. そこで単調劣モジュラ関数最大化問題の真の解を求めることを諦め, 近似解を求めるために「近似アルゴリズム」というものを利用する.

Theorem 8.2.1. 単調劣モジュラ関数最大化問題には最適解が存在する.

Proof. 有限集合だから言える. □

これ以降単調劣モジュラ関数最大化問題の最適解を S_{OPT} と表記する.

Definition 8.2.2. (近似アルゴリズム)

以下の 2 つのことを満たすアルゴリズム \mathcal{A} を近似アルゴリズムと呼ぶ.

^{*1} ここでの最大化とは出来るだけ関数の最大値に近づくような手法のことであり, 実際に関数の最大値が出る保証はない.

^{*2} 実装は bit 全探索というものをよく用いる

1. \mathcal{A} によって近似解が多項式時間^{*3}で得られる.
2. S_{OPT} と \mathcal{A} によって得られる近似解 $S_{\mathcal{A}}$ の f の値の比が一定値より高くなることが理論的に保証される.

Theorem 8.2.3. 近似アルゴリズム \mathcal{A} によって得られる近似解 $S_{\mathcal{A}}$ とする. この時

$$f(S_{\mathcal{A}}) \leq f(S_{OPT})$$

が成立する.

Proof. 最大値の定義より言える. □

近似アルゴリズムが与えられた時, そのアルゴリズムが良い近似アルゴリズムかどうか判定するには色々なものがある. その一つに近似アルゴリズムが最適解にどれだけ近い値を出しているのかというものが考えられる. そこで近似率というものを定義する.

Definition 8.2.4. (近似率)

$0 < \alpha \leq 1$ と定数とし, 近似アルゴリズム \mathcal{A} によって得られる近似解 $S_{\mathcal{A}}$ とする.

$$f(S_{\mathcal{A}}) \leq \alpha f(S_{OPT})$$

が成立する最大の α を \mathcal{A} の近似率と呼ぶ.

8.2.3 貪欲法

今回は単調劣モジュラ関数最大化問題を解くための近似アルゴリズムとして貪欲法 (Greedy algorithm) というアルゴリズムを利用する. 貪欲法とは「その場における最適な解を選択する」というアルゴリズムである. 今回の問題においては

$$S_i = S_{i-1} \cup \arg \max_{e \in V} \Delta_f(e \mid S_{i-1}) \quad (S_0 = \phi)$$

を満たす V の部分集合列 $S_1, S_2, \dots, S_k (0 < k \leq |V|)$ を決定することに他ならない. ただし, $\arg \max_{e \in V} \Delta_f(e \mid S_{i-1})$ の元が2つ以上の場合にはどれか一つを選ぶものとする. また, このアルゴリズムは多項式時間アルゴリズムである. (本方程では多項式時間アルゴリズムの明確な定義はしていないためこの主張は証明をせず認めることにする.)

Proposition 8.2.5. S_1, S_2, \dots, S_k を貪欲法によって選ばれた V の部分集合列とする. この時

$$f(S_1) \leq f(S_2) \leq \dots \leq f(S_k)$$

が成立する.

Proof. $S_1 \subset S_2 \subset \dots \subset S_k$ と f の単調性より言える. □

^{*3} この方程では多項式時間アルゴリズムの明確な定義はしない. 簡単に説明するならばある程度実用的な速さで問題の解が求まるアルゴリズムと思ってもらえれば良い.

8.2.4 貪欲法の近似率

いよいよ主定理である貪欲法の近似定理^{*4}について述べて行く.

Lemma 8.2.6. f を劣モジユラ関数とする. この時

$$\forall S, T \subset V, S \subset T \implies f(S) - f(T) \leq \sum_{e \in T-S} \Delta_f(e|S)$$

が成立する.

Lemma 8.2.7. S_1, S_2, \dots, S_k を貪欲法によって選ばれた V の部分集合列とする. この時任意の $l \in \{0, 1, 2, \dots, k-1\}$ に対して

$$\begin{aligned} f(S_{l+1}) - f(S_l) &\geq \frac{1}{|S_{OPT} - S_l|} (f(S_{OPT}) - f(S_l)) \\ &\geq \frac{1}{k} (f(S_{OPT}) - f(S_l)) \end{aligned}$$

が成立する.

この二つの補題の証明については参考文献の [2] 番や [3] 番を見て欲しい.

Lemma 8.2.8.

$$\forall x \in \mathbb{R}, 1 - x \leq e^{-x}$$

が成立する.

Theorem 8.2.9. (貪欲法の近似定理, Nemhauser)

S_1, S_2, \dots, S_k を貪欲法によって選ばれた V の部分集合列とする. この時

$$\forall l \in \{1, 2, \dots, k\}, f(S_l) \geq \left(1 - e^{-\frac{1}{k}}\right) f(S_{OPT})$$

が成立する.

Proof. $l \in \{0, 1, 2, \dots, k-1\}$ を任意にとり, 各 $i \in \{0, 1, 2, \dots, k-1\}$ に対して $\delta_i = f(S_{OPT}) - f(S_i)$ とする. すると **Lemma 3.7.** より

$$\delta_{l+1} \leq \left(1 - \frac{1}{k}\right) \delta_l$$

である. この不等式と f の非負性より $\delta_0 = f(S_{OPT}) - f(\phi) \leq f(S_{OPT})$ であることと, **Lemma 3.8.** を用いれば

$$\delta_l \leq \left(1 - \frac{1}{k}\right)^l \delta_0 \leq e^{-\frac{l}{k}} f(S_{OPT})$$

^{*4} この名前は著者が勝手につけたものである.

が従う。 δ_l の定義より

$$f(S_l) \geq \left(1 - e^{-\frac{l}{k}}\right) f(S_{OPT})$$

が成立する. □

この定理より貪欲法が近似アルゴリズムであることと, 貪欲法の近似率が $1 - \frac{1}{e} \approx 0.62$ ということがわかる.

8.3 応用例

ここでは単調劣モジュラ関数の応用例について紹介をする. 詳細については参考文献を見て欲しい.

- 口コミ効果最大化
- 文章要約問題
- センサー配置問題
- 施設の配置問題
- 機械学習

このように単調劣モジュラ関数には様々な応用がある. 今回紹介した「(個数制約下での) 単調劣モジュラ関数最大化問題」の他にも, ナップザック制約下での単調劣モジュラ関数最大化問題, マトロイド制約下での単調劣モジュラ関数最大化問題などがある. また単調劣モジュラ関数最小化問題というものもある. 興味のある読者は調べてみると面白いかもしれない.

第 III 部

数理機械学習

第 9 章

単回帰分析・重回帰分析

第 10 章

再生核ヒルベルト空間とカーネル法

第 IV 部

数理強化学習

第 11 章

マルコフ決定過程

第 12 章

深層強化学習

第 V 部

強化学習の応用

第 13 章

ゲームへの応用

第 14 章

巡回セールスマン問題への応用

参考文献

- [1] 内田 伏一 (著)「集合と位相」
- [2] 劣モジュラ最適化と機械学習・河原 吉伸, 永野 清仁・2015
- [3] <https://las.inf.ethz.ch/files/krause12survey.pdf>
- [4] 参考文献の名称・著者 2
- [5] 参考文献の名称・著者 N