

離散数学

授業開始までしばらくお待ちください。

2024

離散数学

Discrete Mathematics



bit.ly/d-math

『集合の数え上げと類似度』

小林裕之

大阪工業大学 RD 学部システムデザイン工学科



OSAKA INSTITUTE OF TECHNOLOGY

2 of 14

a L^AT_EX + Beamer slideshow

授業の受講に関して

- 講義資料（スライド等）は **Google Drive** (<https://bit.ly/d-math>) に置く（紙の配布資料は行わない）。授業前には虫喰い状態のスライドのみを提供するが、授業後に uncovered フォルダに穴埋め版を置くので復習に活用されたい。
- ミニレポートは **Google Forms** (<https://forms.gle/hCyJBbFBMW9AisAt7>) に提出。
- 授業の録画はできるだけスライドと同じフォルダ内のフォルダに置くように努力する（が、必ず置きますとお約束はしません）。
- 授業中に計算間違い等を指摘してくれたらその都度 1 点。（内容に依るけど。）

成績評価について

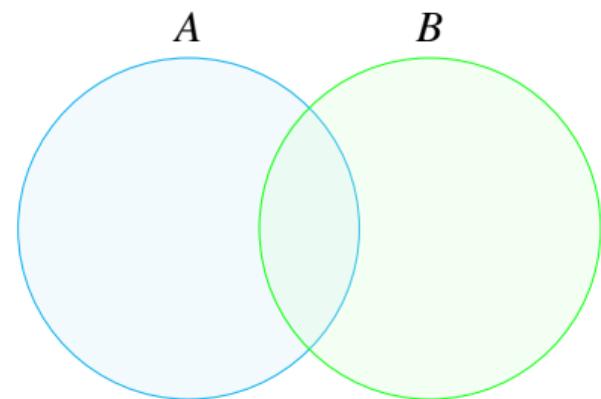
- 出席そのものは評価せず。極論するとテストのみ出席で他は全欠席でも A 評価はあり得る。
- 基本的には**中間演習と期末試験**で評価。
- 毎回ミニレポートを課す。出す者は提出期間を厳守すること。
- 試験の不合格者は**毎回のミニレポートと出席**で少し救済する。
(しっかりした内容のミニレポートを概ね 9 割以上提出し、かつ大学の出欠管理システムで 8 割以上遅刻せず出席していた場合最大 10 点程度の救済。提出数や出席数が少ない場合は救済幅が縮小する。いずれかが 7 割を下回ったら一切救済しない。締め切り後の提出は認めない。)
- 授業中に**スライドの誤りを見つけて指摘してくれた者には、誤り一箇所につき先着一名様限り 100 点満点 1 点相当の加点を行う。(ただしごく軽微なものなど、内容によっては加点しない場合もあり。)

対称差集合 (symmetric difference)

対称差集合

$$A \oplus B = (A \setminus B) \cup (B \setminus A)$$

1. $A \oplus B$ を Venn 図で図示せよ。
2. $A = \{1, 2, 3, 4\}, B = \{1, 3, 5, 7, 9\}$ のとき $A \oplus B$ を求めよ。
3. $|A \oplus B|$ を \oplus を使わない式で示せ。

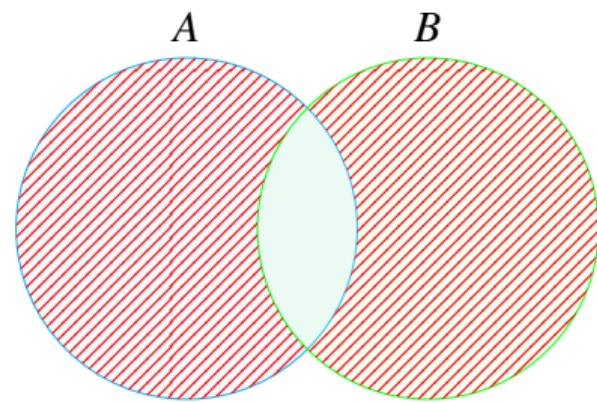


対称差集合 (symmetric difference)

対称差集合

$$A \oplus B = (A \setminus B) \cup (B \setminus A)$$

1. $A \oplus B$ を Venn 図で図示せよ。
2. $A = \{1, 2, 3, 4\}, B = \{1, 3, 5, 7, 9\}$ のとき $A \oplus B$ を求めよ。
3. $|A \oplus B|$ を \oplus を使わない式で示せ。

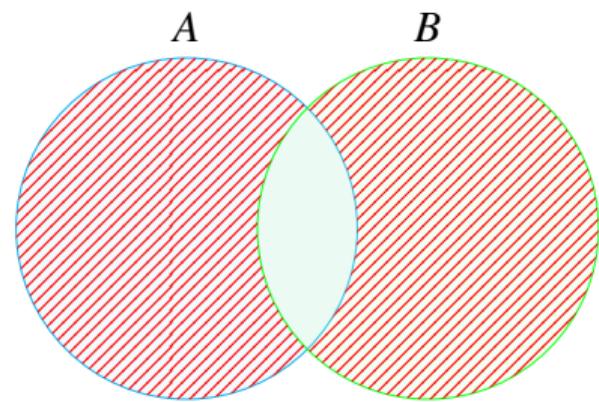


対称差集合 (symmetric difference)

対称差集合

$$A \oplus B = (A \setminus B) \cup (B \setminus A)$$

1. $A \oplus B$ を Venn 図で図示せよ。
2. $A = \{1, 2, 3, 4\}, B = \{1, 3, 5, 7, 9\}$ のとき
 $A \oplus B$ を求めよ。
 $\{2, 4, 5, 7, 9\}$
3. $|A \oplus B|$ を \oplus を使わない式で示せ。



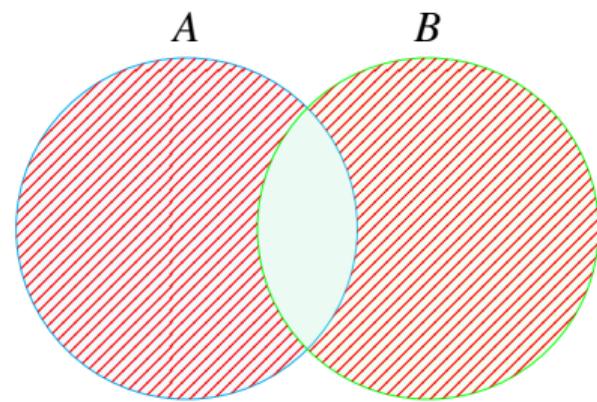
対称差集合 (symmetric difference)

対称差集合

$$A \oplus B = (A \setminus B) \cup (B \setminus A)$$

1. $A \oplus B$ を Venn 図で図示せよ。
2. $A = \{1, 2, 3, 4\}, B = \{1, 3, 5, 7, 9\}$ のとき
 $A \oplus B$ を求めよ。
 $\{2, 4, 5, 7, 9\}$
3. $|A \oplus B|$ を \oplus を使わない式で示せ。

$$|A| + |B| - 2|A \cap B|$$



集合の計算と濃度

- $|\overline{A}| =$
- $|A \cup B| =$
- $|A \cup B \cup C| =$
- $|A \cup B \cup C \cup D| =$

集合の計算と濃度

- $|\overline{A}| = |U| - |A|$
- $|A \cup B| =$
- $|A \cup B \cup C| =$
- $|A \cup B \cup C \cup D| =$

集合の計算と濃度

- $|\bar{A}| = |U| - |A|$
- $|A \cup B| = |A| + |B|$
- $|A \cup B \cup C| =$
- $|A \cup B \cup C \cup D| =$

集合の計算と濃度

- $|\bar{A}| = |U| - |A|$
- $|A \cup B| = |A| + |B| - |A \cap B|$
- $|A \cup B \cup C| =$
- $|A \cup B \cup C \cup D| =$

集合の計算と濃度

- $|\bar{A}| = |U| - |A|$
- $|A \cup B| = |A| + |B| - |A \cap B|$
- $|A \cup B \cup C| = |A| + |B| + |C|$
- $|A \cup B \cup C \cup D| =$

集合の計算と濃度

- $|\overline{A}| = |U| - |A|$
- $|A \cup B| = |A| + |B| - |A \cap B|$
- $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C|$
- $|A \cup B \cup C \cup D| =$

集合の計算と濃度

- $|\overline{A}| = |U| - |A|$
- $|A \cup B| = |A| + |B| - |A \cap B|$
- $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$
- $|A \cup B \cup C \cup D| =$

集合の計算と濃度

- $|\bar{A}| = |U| - |A|$
- $|A \cup B| = |A| + |B| - |A \cap B|$
- $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$
- $|A \cup B \cup C \cup D| = |A| + |B| + |C| + |D| - |A \cap B| - |A \cap C| - |A \cap D| - |B \cap C| - |B \cap D| - |C \cap D| + |A \cap B \cap C| + |A \cap B \cap D| + |A \cap C \cap D| + |B \cap C \cap D| - |A \cap B \cap C \cap D|$

$$|A \cup B| = |A| + |B| - |A \cap B|$$

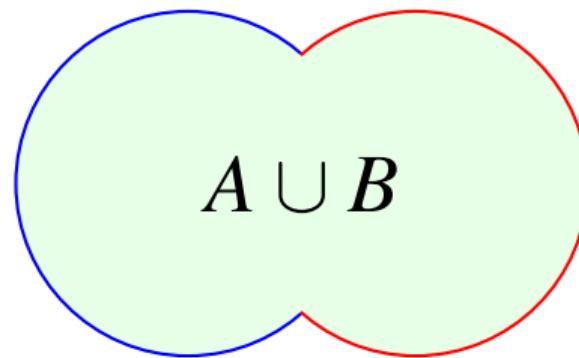
$$A \cup B$$

$$A$$

$$B$$

$$A \cap B$$

$$|A \cup B| = |A| + |B| - |A \cap B|$$

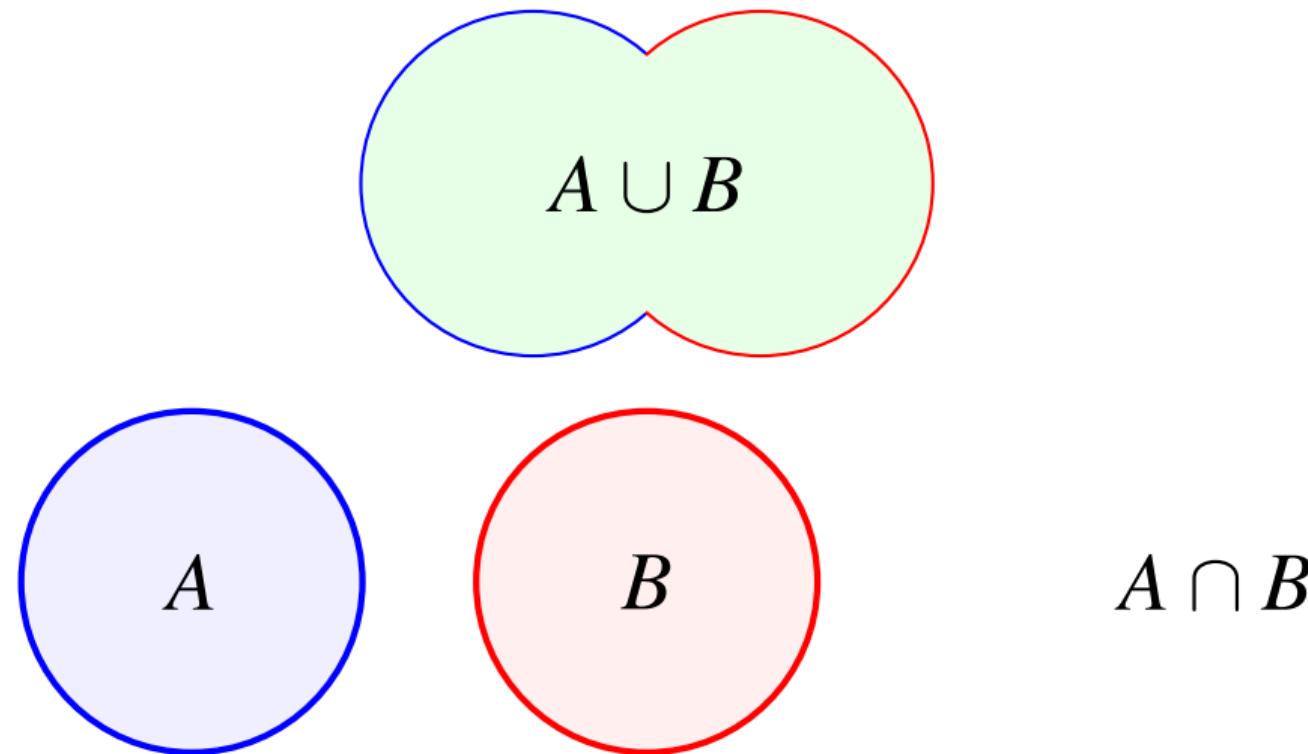


A

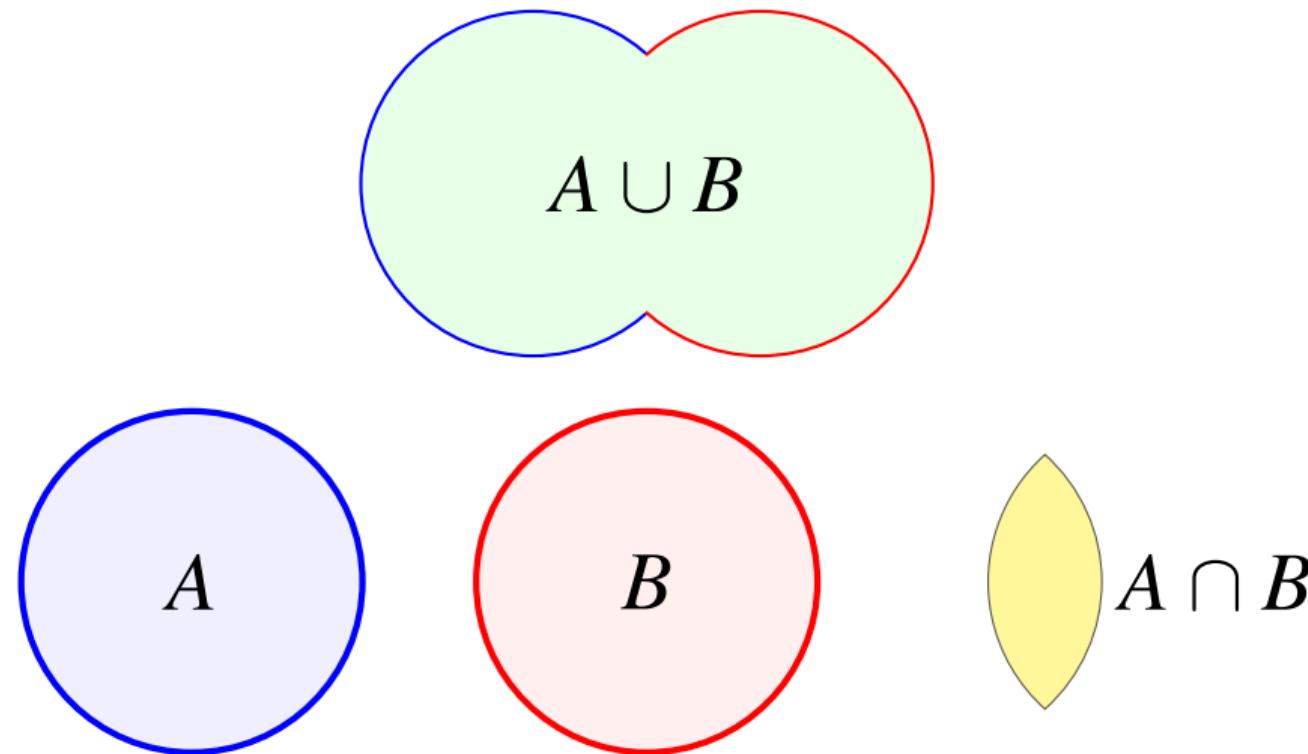
B

$A \cap B$

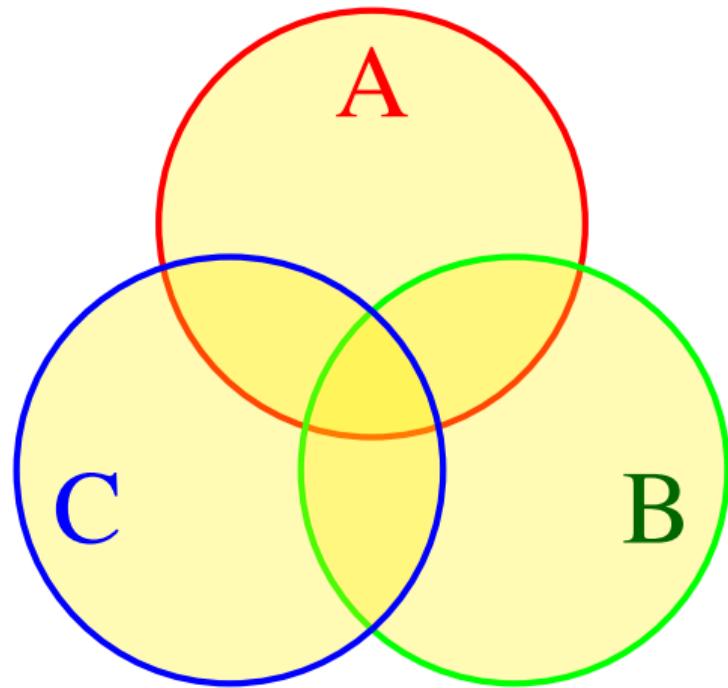
$$|A \cup B| = |A| + |B| - |A \cap B|$$



$$|A \cup B| = |A| + |B| - |A \cap B|$$



$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A| + |A \cup B \cup C|$$



包除原理 (ミニ Q: 各 \sum は何項から成るでしょう?)

$$|\cup_i A_i| =$$

包除原理 (ミニ Q: 各 \sum は何項から成るでしょう?)

$$|\cup_i A_i| = + \sum_{1 \leq i \leq n} |A_i|$$

包除原理 (ミニ Q: 各 \sum は何項から成るでしょう?)

$$|\cup_i A_i| = + \sum_{1 \leq i \leq n} |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cap A_j|$$

包除原理 (ミニ Q: 各 \sum は何項から成るでしょう?)

$$\begin{aligned} |\cup_i A_i| &= + \sum_{1 \leq i \leq n} |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cap A_j| \\ &\quad + \sum_{1 \leq i < j < k \leq n} |A_i \cap A_j \cap A_k| - \dots \end{aligned}$$

包除原理 (ミニ Q: 各 \sum は何項から成るでしょう?)

$$\begin{aligned} |\cup_i A_i| &= + \sum_{1 \leq i \leq n} |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cap A_j| \\ &\quad + \sum_{1 \leq i < j < k \leq n} |A_i \cap A_j \cap A_k| - \dots \\ &\quad + (-1)^{n-1} |A_1 \cap A_2 \cap \dots \cap A_n| \end{aligned}$$

問

300名の学生うち、100名がW科、220名が男子学生である。W科には40名の女子学生がいる。この授業の受講生は170名で、そこには150名の男子学生と5名のW科学生がいる。W科ではない男子学生でこの授業を受講していないのは14名である。この授業を受講しているW科女子学生は何名か？

* この問題はフィクションであり、登場する人物や団体名などはすべて架空のものです。

問

300名の学生うち、100名がW科、220名が男子学生である。W科には40名の女子学生がいる。この授業の受講生は170名で、そこには150名の男子学生と5名のW科学生がいる。W科ではない男子学生でこの授業を受講していないのは14名である。この授業を受講しているW科女子学生は何名か？

* この問題はフィクションであり、登場する人物や団体名などはすべて架空のものです。

ヒント:

$$W = \{x \mid x \text{は W 科学生}\}, F = \{x \mid x \text{は女子学生}\}, C = \{x \mid x \text{はこの授業の受講生}\}$$

問

300名の学生うち、100名がW科、220名が男子学生である。W科には40名の女子学生がいる。この授業の受講生は170名で、そこには150名の男子学生と5名のW科学生がいる。W科ではない男子学生でこの授業を受講していないのは14名である。この授業を受講しているW科女子学生は何名か？

答: 1名

* この問題はフィクションであり、登場する人物や団体名などはすべて架空のものです。

ヒント:

$$W = \{x \mid x \text{は W 科学生}\}, F = \{x \mid x \text{は女子学生}\}, C = \{x \mid x \text{はこの授業の受講生}\}$$

似たような練習

問

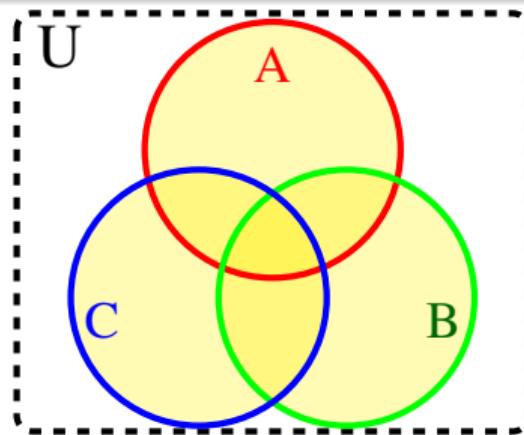
日本には 12623 万人が暮らすが、そのうち東京にいるのが 927 万人、65 歳以上が 3515 万人、外国人が 264 万人である。また東京に住む外国人は 45 万人で、そのうち 2 万人が 65 歳以上である。東京全体では 65 歳以上の人は 296 万人いる。一方、東京以外に住んでいる 65 歳未満の日本人は 8271 万人である^a。日本に 65 歳以上の外国人は何万人いるか。

^a少し以前(2019年頃)の政府統計データ(<https://www.e-stat.go.jp>)などを参考に適当に作ったいい加減なデータです。

一見全然違う話に思える練習

問

5名の学生を R, S, W の 3 学科に配属する組み合わせは何通りか。ただし、どの学科にも最低 1 人は配属すること。（ヒント：『配属パターン』一つ一つをそれぞれ元と考える。）



$U :$

$A :$

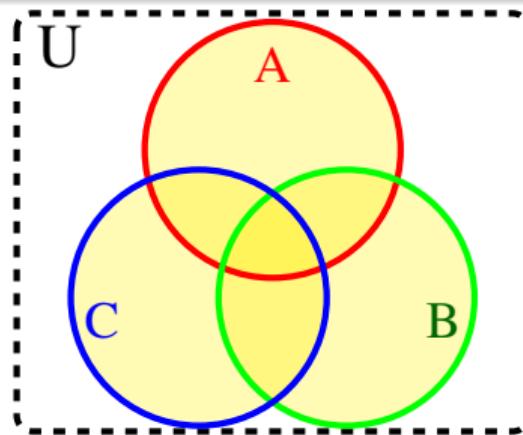
$B :$

$C :$

一見全然違う話に思える練習

問

5名の学生を R, S, W の 3 学科に配属する組み合わせは何通りか。ただし、どの学科にも最低 1 人は配属すること。（ヒント：『配属パターン』一つ一つをそれぞれ元と考える。）



U : あらゆる配属パターン

A :

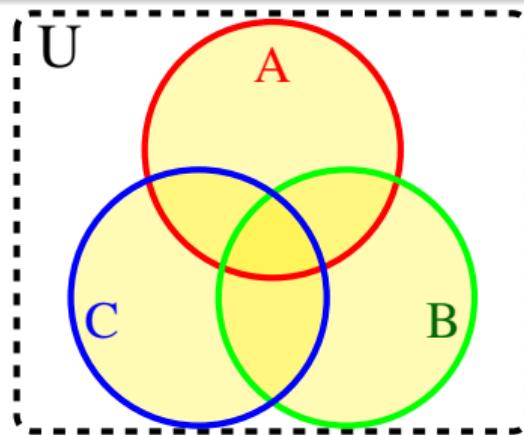
B :

C :

一見全然違う話に思える練習

問

5名の学生を R, S, W の 3 学科に配属する組み合わせは何通りか。ただし、どの学科にも最低 1 人は配属すること。（ヒント：『配属パターン』一つ一つをそれぞれ元と考える。）

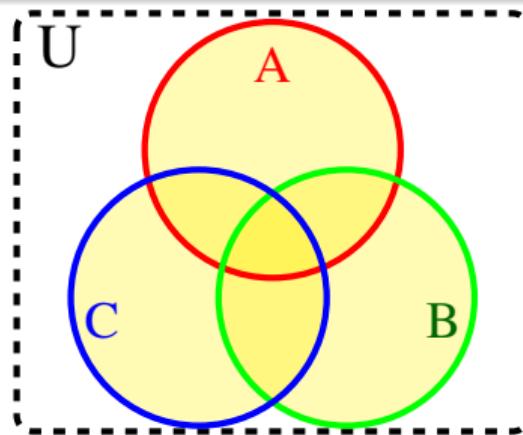


- U* : あらゆる配属パターン
- A* : R 科の配属者 0 のパターン
- B* : S 科の配属者 0 のパターン
- C* : W 科の配属者 0 のパターン

一見全然違う話に思える練習

問

5名の学生を R, S, W の 3 学科に配属する組み合わせは何通りか。ただし、どの学科にも最低 1 人は配属すること。（ヒント：『配属パターン』一つ一つをそれぞれ元と考える。） 答: 150 とおり



- U* : あらゆる配属パターン
- A* : R 科の配属者 0 のパターン
- B* : S 科の配属者 0 のパターン
- C* : W 科の配属者 0 のパターン

問

5名の学生を3チームに分ける組み合わせは何通りか。
ただし、どのチームにも最低1人は配属すること。

(ヒント: 前問の結果を使う。)

問

5名の学生を3チームに分ける組み合わせは何通りか。
ただし、どのチームにも最低1人は配属すること。

(ヒント: 前問の結果を使う。)

答: $150 \div 3!$ とおり

問

$A = \{a, b, c\}$ と比べて、
 $B = \{a, b\}$ と $C = \{a, b, c, d, e\}$ は**どちらが似
ているか?**

問

$A = \{a, b, c\}$ と比べて、
 $B = \{a, b\}$ と $C = \{a, b, c, d, e\}$ はどちらが似
ているか?

類似度の定義 が必要

類似度その 1: ジャッカード係数

index: 2 つの集合の和集合に対する積集合の濃度の比率

$$J(A, B) =$$

問

$S = \{\text{シ}, \text{ヨ}, \text{ウ}, \text{ワ}\}$, $T = \{\text{タ}, \text{イ}, \text{シ}, \text{ヨ}, \text{ウ}\}$, $R = \{\text{レ}, \text{イ}, \text{ワ}\}$ とする。 $J(S, T)$ と $J(S, R)$ を求めよ。

類似度その 1: ジャッカード係数

Jaccard index: 2 つの集合の和集合に対する積集合の濃度の比率

$$J(A, B) =$$

問

$S = \{\text{シ, ョ, ウ, ワ}\}, T = \{\text{タ, イ, シ, ョ, ウ}\}, R = \{\text{レ, イ, ワ}\}$ とする。 $J(S, T)$ と $J(S, R)$ を求めよ。

類似度その 1: ジャッカード係数

Jaccard index: 2 つの集合の和集合に対する積集合の濃度の比率

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

問

$S = \{\text{シ, ョ, ウ, ワ}\}, T = \{\text{タ, イ, シ, ョ, ウ}\}, R = \{\text{レ, イ, ワ}\}$ とする。 $J(S, T)$ と $J(S, R)$ を求めよ。

類似度その 1: ジャッカード係数

Jaccard index: 2 つの集合の和集合に対する積集合の濃度の比率

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

問

$S = \{\text{シ, ョ, ウ, ワ}\}, T = \{\text{タ, イ, シ, ョ, ウ}\}, R = \{\text{レ, イ, ワ}\}$ とする。 $J(S, T)$ と $J(S, R)$ を求めよ。

$$J(S, T) = 3/6 = 0.5,$$

類似度その 1: ジャッカード係数

Jaccard index: 2 つの集合の和集合に対する積集合の濃度の比率

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

問

$S = \{\text{シ, ョ, ウ, ワ}\}, T = \{\text{タ, イ, シ, ョ, ウ}\}, R = \{\text{レ, イ, ワ}\}$ とする。 $J(S, T)$ と $J(S, R)$ を求めよ。

$$J(S, T) = 3/6 = 0.5, J(S, R) = 1/6 \simeq 0.17$$

Jaccard 係数だけでいいか?

しっくりこない場合もあります。

問

A 氏は本を 100 冊持っている。図書館 X(蔵書数 400) には A 氏が持つ本と同じ本が 50 冊あり、図書館 Y(蔵書数 1000) には A 氏の持つ 100 冊すべての本がある。A 氏のコレクションと類似度が高いのは X, Y どちらの図書館か?

答: $J(A, X) =$

Jaccard 係数だけでいいか?

しっくりこない場合もあります。

問

A 氏は本を 100 冊持っている。図書館 X(蔵書数 400) には A 氏が持つ本と同じ本が 50 冊あり、図書館 Y(蔵書数 1000) には A 氏の持つ 100 冊すべての本がある。A 氏のコレクションと類似度が高いのは X, Y どちらの図書館か?

答: $J(A, X) = 50/450 = 0.11$

Jaccard 係数だけでいいか?

しっくりこない場合もあります。

問

A 氏は本を 100 冊持っている。図書館 X(蔵書数 400) には A 氏が持つ本と同じ本が 50 冊あり、図書館 Y(蔵書数 1000) には A 氏の持つ 100 冊すべての本がある。A 氏のコレクションと類似度が高いのは X, Y どちらの図書館か?

答: $J(A, X) = 50/450 = 0.11$, $J(A, Y) = 100/1000 = 0.1$

Jaccard 係数だけでいいか?

しっくりこない場合もあります。

問

A 氏は本を 100 冊持っている。図書館 X(蔵書数 400) には A 氏が持つ本と同じ本が 50 冊あり、図書館 Y(蔵書数 1000) には A 氏の持つ 100 冊すべての本がある。A 氏のコレクションと類似度が高いのは X, Y どちらの図書館か?

答: $J(A, X) = 50/450 = 0.11$, $J(A, Y) = 100/1000 = 0.1$

X の方

Jaccard 係数だけでいいか?

しっくりこない場合もあります。

問

A 氏は本を 100 冊持っている。図書館 X(蔵書数 400) には A 氏が持つ本と同じ本が 50 冊あり、図書館 Y(蔵書数 1000) には A 氏の持つ 100 冊すべての本がある。A 氏のコレクションと類似度が高いのは X, Y どちらの図書館か?

答: $J(A, X) = 50/450 = 0.11$, $J(A, Y) = 100/1000 = 0.1$

X の方 …で腑に落ちたような落ちないような。

類似度その 2: シンプソン係数

小さい方の集合目線で考えてみよう。

index: 小さい方の集合に対する積集合の濃度の比率

$$S(A, B) =$$

問

前のページの図書館の問題をシンプソン係数で再考せよ。

類似度その 2: シンプソン係数

小さい方の集合目線で考えてみよう。

Simpson index: 小さい方の集合に対する積集合の濃度の比率

$$S(A, B) =$$

問

前のページの図書館の問題をシンプソン係数で再考せよ。

類似度その 2: シンプソン係数

小さい方の集合目線で考えてみよう。

Simpson index: 小さい方の集合に対する積集合の濃度の比率

$$S(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}$$

問

前のページの図書館の問題をシンプソン係数で再考せよ。

類似度その 2: シンプソン係数

小さい方の集合目線で考えてみよう。

Simpson index: 小さい方の集合に対する積集合の濃度の比率

$$S(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}$$

問

前のページの図書館の問題をシンプソン係数で再考せよ。

$$S(A, X) = 50/100 = 0.5, S(A, Y) = 100/100 = 1$$

類似度その 3: ダイス係数

中道路線?

index: 2 集合の濃度の平均値に対する積集合の濃度の比率

$$D(A, B) =$$

問

同様に図書館の問題をダイス係数で再々考せよ。

類似度その 3: ダイス係数

中道路線?

Dice index: 2 集合の濃度の平均値に対する積集合の濃度の比率

$$D(A, B) =$$

問

同様に図書館の問題をダイス係数で再々考せよ。

類似度その 3: ダイス係数

中道路線?

Dice index: 2 集合の濃度の平均値に対する積集合の濃度の比率

$$D(A, B) = \frac{|A \cap B|}{(|A| + |B|) / 2}$$

問

同様に図書館の問題をダイス係数で再々考せよ。

類似度その 3: ダイス係数

中道路線?

Dice index: 2 集合の濃度の平均値に対する積集合の濃度の比率

$$D(A, B) = \frac{|A \cap B|}{(|A| + |B|) / 2}$$

問

同様に図書館の問題をダイス係数で再々考せよ。

$$D(A, X) = 50/250 \simeq 0.2, D(A, Y) = 100/550 \simeq 0.18$$

集合の類似度

いろいろわかったところで再チャレンジ！

問

$A = \{a, b, c\}$ と比べて、

$B = \{a, b\}$ と $C = \{a, b, c, d, e\}$ は **どちらが似ているか?** ジャッカード係数、シンプソン係数、ダイス係数で比較せよ。

ミニレポート課題 (提出期間: 本日～次回の授業の前日)

p. 10の問題を解け。

解答を PC 文書や手書きで作成し、PDF にして Google Forms (<https://forms.gle/hCyJBbFBMW9AisAt7>) から提出せよ (要組織アカウントによるログイン)。ただし写真等の画像ファイルの場合は、解像度や露出・照明状態などを十分考慮し、きちんと読解可能なクオリティのものとすること。スマートフォンの場合はスキャナアプリの類の利用を必須とする。

