

## 全学データサイエンス科目における数学教育と生成 AI 活用

城西大学 数理・データサイエンスセンター 杉谷宜紀

### 1 はじめに

城西大学数理・データサイエンスセンターでは全学向けデータサイエンス教育として独自の教育プログラムを展開しており、ベーシックレベルとして1年生向け講義『データサイエンス入門』を、アドバンスレベルとして3年生向け講義『機械学習とAI』、『データサイエンスと数理統計』、『DS 特別講義 I』、『DS 特別講義 II』をそれぞれ提供している。ベーシックレベルは、各学部ごとに選出された科目と『データサイエンス入門』の単位を合わせて取得する事でデータサイエンスの基礎的素養を認定する教育プログラムであるが、この『データサイエンス入門』は2023年に文部科学省『数理・データサイエンス・AI教育プログラム（リテラシーレベル）』の認定を受けた事で新たに1科目のみのデータサイエンス入門プログラムがベーシックレベルの下位に追加された。アドバンスレベルは、提供する4科目を選択必修に含むより発展的な内容を扱う教育プログラムである。

本稿は上記5科目をこれまでに実施してきた経験を基に本学における数理・データサイエンス・AI教育への取り組みをまとめたものである。第2節は1年生向けの基礎的な数学教育法について、第3節は3年生向けの生成AIを使った新しい教育法について述べる。

### 2 全学1年生向け講義における数学教育

講義『データサイエンス入門』は105分、全13回の表1の様な授業構成で文科省が作成したモデルカリキュラムに沿いながら、同時に数学的な内容をできるだけ削らない様に努めている。しかし、文理選択や高校の過程によっては『数学I』のみで入学する学生も相当数おりベクトルや数列、微分、 $\Sigma$ などを前提とする事が難しい為に、数理・データサイエンス教育として様々な工夫が必要になる。ここではその中で数学が中心的役割を果たす第4週、第7週の2回の講義における工夫をまとめる。

#### 2.1 基本統計量の算出における工夫

第4週『データを読む・扱う』では実データから情報を読み取る為の基本統計量やグラフによる可視化の種類、およびそれを扱う為のExcelの基本的な操作を学ぶ回である。例えば $n$ 個の要素からなるデータ  $X = \{X_1, \dots, X_n\}$  の平均は

$$\frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

となり、これを  $\bar{X}$  と書く。定義式(1)は数学的には厳密であるが $\Sigma$ に馴染みのない学生にとってはかえって混乱の元となるので敢えて説明しない。多くの学生にとってはむしろ

| 週     | 講義タイトル          | 内容                    |
|-------|-----------------|-----------------------|
| 1     | 社会で起こっている変化     | これまでの技術革新と未来社会        |
| 2     | DS・AIで使われるデータとは | デジタルデータの種類、オープンデータ    |
| 3     | データ・AI利活用の現場    | 統計的探求のプロセスとその実践例      |
| 補助    | Excel の基本       | 四則演算、表、関数の基本操作        |
| 4     | データを読む・扱う       | 基本統計量・グラフ、その Excel 操作 |
| 5     | データを説明する        | 地域経済分析システム RESAS の活用  |
| 6     | データを可視化する       | データ視覚化のデザイン、資料作成      |
| 7     | DSに必要な数学        | 最小二乗法、ニュートン法          |
| 8     | DSの分析手法1        | 教師あり学習、深層学習           |
| 9     | DSの分析手法2        | 教師なし学習、生成 AI          |
| 補助    | データ・AI 社会での心構え  | ELSI                  |
| 10～13 | 4 週間の最終課題       | 学んだ事を実践するグループワーク      |

表 1: 講義『データサイエンス入門』の構成

Excel で AVERAGE 関数を使えばデータの平均が計算できるという実用性の方が有用であり、日常会話でも用いられる『平均』を  $\bar{X}$  で表す事にする程度の説明の方が敷居が低くて済む様である。平均・中央値・最頻値などの代表値の説明はまだ簡単だが、データのばらつきを表す分散となると途端に学生にとって困難が生じてくる。例えば、分散  $V(X)$  の定義には3つの表現を用いて説明する。

$$V(X) = \overline{(X - \bar{X})^2} = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (2)$$

理系学生なら最後の式を難なく理解できるが、総和記号  $\Sigma$  を習っていない学生にとっては他の定義式で理解してもらう事になる。特に、最初の式は集合の2乗や集合と数値の引き算に目を瞑れば式自体が非常にシンプルなので学生にも扱いやすい。この様にデータ  $X$  にバーを付けて平均  $\bar{X}$  を取る演算は、 $\Sigma$  を用いるのが難しい全学向け講義で重要な役割を果たす。このアイデアは [1] に非常によく纏められている。

続いて、偏差値とはデータの変数変換である点を簡潔に述べ、その目的が尺度の異なるデータの比較にある事を強調する。例えば「24.5kg の5歳男子と 68.5kg の17歳男子ではどちらがすごい？」の様な質問を図1の様なヒストグラムと共に問いかける。変数変換によってデータの単位が消失する事で初めて尺度の異なるデータが比較できる事は偏差値の定義式を使って説明する。

$$Z = 50 + 10 \frac{X - \bar{X}}{\sigma}. \quad (3)$$

ここで  $\sigma$  はデータ  $X$  の標準偏差である。右辺第2項の分子と分母が同じ単位を持つ量なのでキャンセルアウトされて単位が消失する。世の中の大抵のデータは単位を持つ事は実験を行う理系学生にはお馴染みだが、公式の暗記に頼りがちな本学数学科の学生の方がむしろ「1kg と 1m は意味が違う」事や「なぜ分散でなく標準偏差を用いるのか」という話は苦手な様である。そう考えると同じ体重データの比較よりも「24.5kg の5歳男子と 175cm の17歳男子」などを例にする方がより偏差値の威力を実感しやすいかもしれない。

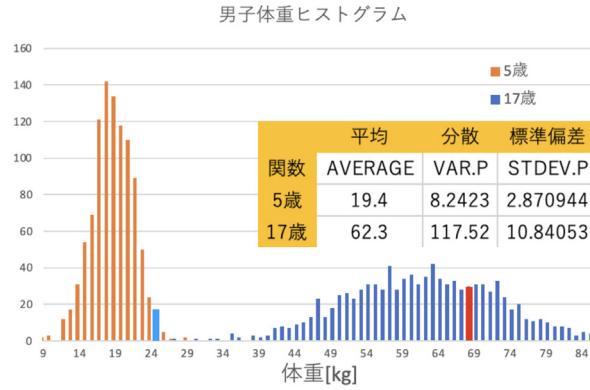


図 1: 日本の男子体重ヒストグラム (e-stat 「R2 学校保険統計」 より作成)

偏差値が本当に平均 50、標準偏差 10（つまり分散 100）を満たすデータの変数変換になっている事はバーを使った平均の演算によって  $\Sigma$  を使わずに簡単に証明できる。実際、

$$\bar{Z} = \overline{\left(50 + 10 \frac{X - \bar{X}}{\sigma}\right)} = 50 + 10 \frac{\bar{X} - \bar{X}}{\sigma} = 50, \quad (4)$$

$$V(Z) = \overline{(Z - \bar{Z})^2} = \overline{\left(10 \frac{X - \bar{X}}{\sigma}\right)^2} = 100 \frac{\overline{(X - \bar{X})^2}}{\sigma^2} = 100. \quad (5)$$

最後に、相関係数  $r$  の定義では形式的に 2 データ  $X$  と  $Y$  の間の比例定数の様なもの  $Y \approx rX$  と説明している。これは後で扱う最小二乗法のアナロジーに向けた説明となっているが、数学的には厳密ではなく  $r$  を比例係数の様なものと言いながら回帰直線の傾きとは一致しない点を補足する必要がある。厳密には、 $r$  は  $X$  と  $Y$  を標準化したデータ間の比例定数（回帰直線の傾き）であり、

$$\frac{Y - \bar{Y}}{\sigma_Y} \approx r \frac{X - \bar{X}}{\sigma_X} \quad (6)$$

と形式的に書ける。さらに両辺に  $\frac{X - \bar{X}}{\sigma_X}$  を掛けてバーをとる操作によって同じく形式的に

$$\overline{\left(\frac{X - \bar{X}}{\sigma_X}\right) \left(\frac{Y - \bar{Y}}{\sigma_Y}\right)} \approx r \overline{\left(\frac{X - \bar{X}}{\sigma_X}\right)^2} \quad (7)$$

$$= r \frac{V(X)}{\sigma_X^2} = r, \quad (8)$$

すなわち

$$r \approx \frac{\overline{(X - \bar{X})(Y - \bar{Y})}}{\sigma_X \sigma_Y} \quad (9)$$

$$= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (10)$$

を得る。 $\text{Cov}(X, Y)$  はデータ  $X$  と  $Y$  の共分散である。実は式 (9) は厳密に等式が成り立つ事が [1] で示されており、相関係数が標準測度の積の平均として説明できるのだが、講義では天下りの式 (10) を相関係数  $r$  の定義として教えている。学生の数学への関心が高ければ非常に良い説明となるだろうが、こちらもやはり Excel で CORREL 関数を使って相関係数を算出する事の方が学生にとっても優先度が高い。

## 2.2 最小二乗法の $\Sigma$ を使わない解法

第 7 週『DS に必要な数学』では後続の教師あり/なし学習や AI に関する回への橋渡しとして最小二乗法とニュートン法を中心に扱う。ニュートン法は微分の定義から説明せざるを得ないが、線形回帰モデルの最小二乗法に関してはバーを使って平均を取る演算がここでも有効である。

まず最初に、この回では AI とは入力と出力の間の対応関係、すなわち写像である事を様々な事例を見ながら学習する。現在の特化型 AI は何かを出力する為に必ず入力項が必要という大前提を学ぶ事は、AI が勝手に何でもやってくれるという誤解を解くきっかけになるだろう。そして、写像をパラメータ付きの関数として数式化する作業を統計モデリングと呼び、そのパラメータをデータから定める事こそが機械学習や AI における「学習」である事を学ぶ。そこから具体的な計算方法である最小二乗法へと展開するにあたって、統計学における重要な考え方としてデータとは確率変数の実現値であり、統計モデル  $y = f(x)$  はあくまで確率変数  $x, y$  の間の対応関係である為に実現値との間には誤差が含まれる事を、実験科学における理論と実験のアナロジーを用いて説明する。すなわち関係式  $y = f(x) + \varepsilon$  の各項の意味を表 2 で対応づける。

|      | $y$ | $f(x)$       | $\varepsilon$ |
|------|-----|--------------|---------------|
| 実験科学 | 観測値 | 自然法則を表す方程式の解 | 測定誤差          |
| 統計学  | 実現値 | 期待値を統計モデルで表す | 各データ点のノイズ     |

表 2: 実験科学と統計学における誤差の考え方のアナロジー

さて、2 つのデータ  $X = (x_1, \dots, x_n), Y = (y_1, \dots, y_n)$  を直線モデル  $y = ax + b$  に当てはめると、各データ点の対  $\{(x_i, y_i)\}_{i=1}^n$  には以下の等式が成り立つ。

$$y_i = ax_i + b + \varepsilon_i, \quad i = 1, \dots, n. \quad (11)$$

ここで  $a, b$  は統計モデルが持つパラメータで  $\varepsilon_i$  は各データ点  $(x_i, y_i)$  におけるモデルとデータの間のずれ、すなわち誤差を表す。最小二乗法とはこの誤差の二乗和が最小となる様にモデルのパラメータを決める計算手法であり、以下の最小化問題として定式化される。

$$\text{Minimize } \sum_{i=1}^n (y_i - (ax_i + b))^2 \quad \text{subject to } a, b. \quad (12)$$

これは図形的には図 2 の様に直線の無限の可能性の中から赤線で表される誤差の総和が最も少ないものを 1 本見つける事を意味する。

最小化問題 (12) を解くには  $\Sigma$  を含んだ量の微分または平方完成が必要となるので、『数学 I』しか知らない学生にとってはやはり敷居が高い。それを回避する為に問題を次の様

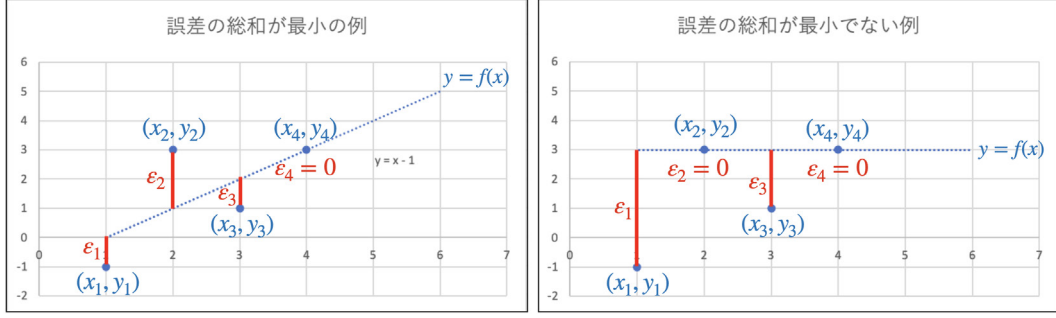


図 2: 最小二乗法の図形的解釈。2 本の直線のうち誤差の二乗和が少ないのは左側の直線。

に再定式化していく。ここで、相関係数のときのアナロジーを考えて 2 データ  $X$  と  $Y$  の間に形式的に直線関係  $Y \approx aX + b$  が成り立つとする。おおよそ等しいの  $\approx$  の部分を厳密に扱う為に両辺の差分

$$E := Y - (aX + b) \quad (13)$$

について考えると、データ  $X = (x_1, \dots, x_n), Y = (y_1, \dots, y_n)$  を集合とみなせば左辺の  $E = \{y_i - ax_i - b\}_{i=1}^n$  も集合となり、特に (11) より  $E$  は誤差の集合となる。この  $E$  の導入によって最小化問題 (12) は  $E$  の集合 (データ) としての最小化

$$\text{Minimize } |\bar{E}| \text{ and } V(E) \text{ subject to } a, b \quad (14)$$

を考えれば良いことになる。<sup>1</sup>

まず  $|\bar{E}|$  の最小化について考えると、バーを使った平均の演算によって

$$\bar{E} = \overline{(Y - (aX + b))} = \bar{Y} - a\bar{X} - b$$

となり  $a$  の値によらず  $b = \bar{Y} - a\bar{X}$  とおけば  $|\bar{E}| = 0$  とできる。次に  $V(E)$  の最小化については  $b = \bar{Y} - a\bar{X}$  のとき  $|\bar{E}| = 0$  を用いると  $E - \bar{E} = (Y - \bar{Y} - a(X - \bar{X}))$  となるから

$$\begin{aligned} V(E) &= \overline{(E - \bar{E})^2} \\ &= \overline{\{Y - \bar{Y} - a(X - \bar{X})\}^2} \\ &= \overline{(Y - \bar{Y})^2 - 2a(Y - \bar{Y})(X - \bar{X}) + a^2(X - \bar{X})^2} \\ &= V(Y) - 2a\text{Cov}(X, Y) + a^2V(X). \end{aligned}$$

ここで (10) などから相関係数  $r$  と標準偏差  $\sigma_X, \sigma_Y$  で  $V(E)$  を表すと

$$\begin{aligned} V(E) &= \sigma_Y^2 - 2\sigma_X\sigma_Yra + \sigma_X^2a^2 \\ &= \sigma_X^2 \left( a - \frac{\sigma_Y}{\sigma_X}r \right)^2 + \sigma_Y^2(1 - r^2). \end{aligned}$$

<sup>1</sup>  $E = \{y_i - ax_i - b\}_{i=1}^n$  より最小化問題 (12) は  $\overline{E^2} = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i - b)^2$  の最小化に等しいが、 $\overline{E^2} = (\bar{E})^2 + V(E)$  が成り立つので (14) と同値になる。

よって  $V(E)$  は  $a = \frac{\sigma_Y}{\sigma_X} r$  のとき最小値  $\sigma_Y^2(1 - r^2)$  となる。以上をまとめると、線形回帰の最小二乗法は  $\Sigma$  を全く使わずに解く事ができて  $a = \frac{\sigma_Y}{\sigma_X} r, b = \bar{Y} - a\bar{X}$  が統計モデルのパラメータと分かる。また、データ  $X, Y$  が標準化されている場合に限り  $\sigma_X = \sigma_Y = 1$  を満たすので  $a = r$  となり、回帰直線の傾きと相関係数が一致する事が分かる。さらに分散の正值性から  $V(E) \geq \sigma_Y^2(1 - r^2) \geq 0$ 、すなわち相関係数が  $-1 \leq r \leq 1$  を満たす事も確認できる。

講義では実データを使って  $a, b$  の値を Excel で計算し、同じく Excel が自動で表示してくれる回帰直線の数式と一致する事を確認する。同時にデータ  $X, Y$  に対する統計モデル  $y = f(x)$  の当てはまりの良さを表す決定係数

$$R^2 := 1 - \frac{V(E)}{V(Y)} = 1 - \frac{\sum_{i=1}^n (y_i - f(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15)$$

が直線モデル  $f(x) = ax + b$  の時は  $V(E) = \sigma_Y^2(1 - r^2)$  より  $R^2 = r^2$  となる事を数値的に確認する。

### 3 全学3年生向け講義における生成 AI を使った教育

2023 年度の講義『データサイエンスと数理統計』では線形計画法をテーマに選んだ。殆ど予備知識が不要でプログラミング実装も容易、何より現実社会の諸問題への応用が広いなどの点で全学向け講義として扱いやすいのが理由である。受講者が数学科と経済学部で 4 名ずつ計 8 人と少人数だったので教科書 [2] の定理の証明を数学科学生に手伝ってもらう半セミナー形式で講義を行った。本学数学科は教員志望の学生が多い一方で、人前で発表する機会は 4 年次科目『数学セミナー』からと圧倒的に経験不足の問題がある。経済学部の学生を生徒に見立て 3 年次からその経験を提供できた点は少人数授業ならではのであった。感想にも「セミナー形式で線形代数を教えるような形で行われていた。その中で、教えることの難しさを知り、どうすれば分かりやすくなるのかを考えることで自分の理解も進んだので良かった。」などの意見が寄せられた。

誰かに教える経験が自分自身の理解にも繋がる事は誰もが知っている事だが、その対象はヒトだけでなく AI にも適用可能だと私は考える。ここでは生成 AI を利用した新しい数理・データサイエンス教育について考えたい。

#### 3.1 AI の間違いをヒトが正解に導く学習法

オムニバス講義『DS 特別講義 II』での担当回『生成 AI 活用』では様々な活用法や諸注意を説明するにあたって、まず第一に会話のラリーが続く対話型 AI としての特徴を強調している。これまでの一問一答型の AI アシスタントと大きく異なり、これからは生成 AI と対話しながら共同作業する事が当たり前となる筈である。その中で次のような質問を ChatGPT に問いかける演習がある。

— 演習問題で ChatGPT に最初に打ち込むプロンプト —

ある映画の試写会を行い、満足度のアンケート調査を行った。試写会に参加したのは 300 人でそのうち女性が 180 人であった。アンケートに満足したと回答したのは男性の 50%、女性の 75%であった。この映画を見て満足しなかったと答えた人が女性である確率はいくらか？

これは簡単な条件付き確率の文章問題だが GPT3.5 が間違いやすい質問となっている。<sup>2</sup>実際に ChatGPT の回答は例えば以下の様になっていた。

— GPT3.5 の回答例 —

この問題は条件付き確率を使用します。まず、試写会に参加した男女数を計算します。  
男性の数 = 300 人 - 女性の数 = 300 人 - 180 人 = 120 人  
次に、アンケートに満足した男性と女性の数を計算します。  
男性の満足した人数 = 男性の数 × 満足した割合 = 120 人 × 0.50 = 60 人  
女性の満足した人数 = 女性の数 × 満足した割合 = 180 人 × 0.75 = 135 人  
今、男性が映画を見て満足しなかった確率を  $P(A)$  とすると  
 $P(A) = \text{男性で満足しなかった人数} / \text{全体の人数} = (120 \text{ 人} - 60 \text{ 人}) / 300 \text{ 人} = 1/5$   
次に、女性である条件で映画を見て満足しなかった確率を  $P(B)$  とすると  
 $P(B) = \text{女性で満足しなかった人数} / \text{全女性数} = (180 \text{ 人} - 135 \text{ 人}) / 180 \text{ 人} = 1/4$   
最後に、女性である条件で映画を見て満足しなかった確率  $P(A | B)$  を計算します。  
 $P(A | B) = P(A \cap B) / P(B)$   
 $P(A \cap B)$  は、女性であり映画に満足しなかった人数で、45 人です。したがって、  
 $P(A | B) = 45 \text{ 人} / (1/4) = 45 \text{ 人} \times 4 = 180 \text{ 人}$   
したがって、女性である条件のもとで映画を見て満足しなかった確率は、180 人 / 300 人 = 3/5、または 60% です。

この回答例では条件付き確率  $P(A | B)$  を計算しているが、その場合は事象  $A$  が『女性である』、事象  $B$  が『映画を見て満足しなかった』であるべき所を  $P(A)$ 、 $P(B)$  の両方の計算部分（7、9 行目）で事象を正しく把握できていない為に誤った計算がされていると分かる。LLM によるテキストベースのやり取りだけではこの様な一貫性のないミスがしばしば起こるが、これは確率計算に不慣れた学生が間違える事例とある意味では似ている。

この様な間違った回答が出力された場合に対話を続けて AI を正解に導こうというのが演習の趣旨である。続きのプロンプトをどう与えるかは人によって異なり、例えば「不正解。やり直し」などと伝えると同様の間違いを繰り返したりするが、「答えが違います。男性 or 女性、満足している or 満足していないでマトリクス表を作って考えてみては？」などと伝えると以下の様な表 3 を作成しながら問題を解き直してくれる。

|    | 満足  | 不満足 |
|----|-----|-----|
| 男性 | 60  | 60  |
| 女性 | 135 | 45  |

表 3: 男女別アンケート結果。満足しなかった人が女性である確率は 45/105

<sup>2</sup>この質問は 2023 年度城西大学市民講座の高橋達先生の講演からアイデアを得た。

AI が間違える問題に対して、ヒトが考え方を教える事で AI を正解に導くというアプローチは人間同士のやり取りと同じ効果が期待できると思われる。相手がヒトであれ AI であれ、どの様に説明すればその対象に伝わりやすいかを考える事は非常に重要である。ただし、そもそも AI が間違える質問を見つける事が難しいだけでなく、日進月歩で進化する為に一度見つけた質問もそのうち使えなくなってしまう。特に GPT4 以降ではテキストベースだけでなく背後でプログラム実行を用いて回答が生成される為にもはや簡単には間違えなくなりつつある。それでも新たな数学教育の可能性として、AI が間違えやすく、かつ学生にとっては正解を理解しやすい質問を今後も辛抱強く発見していきたい。

似た様な演習として『データサイエンス入門』の課題の中には Copilot の画像生成を利用した以下の様な問題を出題している。

データサイエンス入門の課題の一例

Microsoft Copilot にヘリコプターの絵を描いてもらいましょう。ただし『ヘリコプター』の単語名を使わずに、自分の言葉だけで説明して見事ヘリコプターの絵を完成させて下さい。入力プロンプトは「実在するある乗り物の絵を描いて下さい。」から始めて、その続きの説明は自分で考える事。

この問題も生成 AI とのコミュニケーション能力を問う問題と言える。この演習問題には続きがあり、作成したヘリコプター画像を今度は物体認識 AI <sup>3</sup>にかけて精度 70%を超える様な画像生成を目標としている。学生ごとにヘリコプターの外観に関するもの、利用用途に関するものなど様々なプロンプトが考えられる。この時の感想には「自分ではヘリコプターの説明をしてつもりなのに飛行機の絵が出てきたり、伝えるのが難しいなと思いました」や「画像を AI に出してもらうには大まかな情報だけでなく一つ一つ細かい説明をしなければいけないことがわかった。これは人とのコミュニケーションでも大切なことだと思った。」などがあつた。

### 3.2 プログラミングを AI と協創する学習法

講義『DS 特別講義 I』ではプログラミングの基礎を Python を使いながら学んでいく。その中に p5.js でフラクタルを描く回がある。p5.js とは数学関数のグラフを用いたデジタルアートの為の初心者向けプログラミング言語 Processing の java script 版であり、数学教育のみならず Web デザインなどにも応用があり、さらに公式 HP にブラウザ上で動く総合開発環境も提供されていて導入の敷居が非常に低い。フラクタルはその美しい幾何学模様が数学教育的にも学生の関心を集めやすいだけでなく、プログラミング教育的にも再帰関数の使い方を学ぶのに非常に有効である。

講義では p5.js を基礎から学ぶ事はせずいきなり生成 AI を使ってコーディングした。例えば「p5.js でコッホ曲線を描きたい」などとプロンプトを入力すれば以下の様なコードを出力してくれ、それを公式 HP のエディターにコピペして実行すると図 3 の様なコッホ曲線が得られる。

<sup>3</sup>講義では Google Vision API のデモ版を利用している

Listing 1: GPT4 が出力するコッホ曲線のコード例

---

```

1 function setup() {
2   createCanvas(800, 400);
3   background(255);
4   let start = createVector(50, 200);
5   let end = createVector(750, 200);
6   kochLine(start, end, 5);
7 }
8
9 function kochLine(start, end, depth) {
10  if (depth == 0) {
11    line(start.x, start.y, end.x, end.y);
12  } else {
13    let a = start;
14    let b = p5.Vector.lerp(start, end, 1/3);
15    let c = p5.Vector.lerp(start, end, 2/3);
16
17    let dx = c.x - b.x;
18    let dy = c.y - b.y;
19
20    let mid = createVector(
21      b.x + dx / 2 - dy * sqrt(3) / 2,
22      b.y + dy / 2 + dx * sqrt(3) / 2
23    );
24
25    kochLine(a, b, depth - 1);
26    kochLine(b, mid, depth - 1);
27    kochLine(mid, c, depth - 1);
28    kochLine(c, end, depth - 1);
29  }
30 }

```

---

Processing が初心者向けと言うのもあるが、数学的な素養のある学生なら Python で基本的な（オブジェクト指向）プログラミングの仕組みさえ学習していれば java script の経験が無くてもコードが十分理解できる。実際、p5.js の仕様として setup 関数がプログラムが最初に実行されたときに呼び出される関数である事、グラフを描画する為の長方形領域 Canvas は左上の頂点を原点 (0,0) とし、 $x$  軸は右向き、 $y$  軸は下向きに座標が定まる事だけを簡単に説明すれば、あとは線分を 3 分割してそれぞれの左端点の座標を  $a, b, c$  とし、真ん中の部分に正三角形を生やした際にできる新たな頂点を  $mid$  として再起的に kochLine 関数を呼び出している事は簡単に分かる。

このコードは成功例だがそれでも 3、4 行目に start、end という予約語を仕様した為に警告が出ている。失敗例だと例えば、再帰関数を使わずにメソッド関数などで実装してしまう例や、そもそも数値を間違えたりしてコッホ曲線が綺麗に描画されないコードを出力してしまう事も多い。そのような場合に「再帰関数を使ったコードにして」や「パラメータを depth でなく iteration に変更」、「ベクトルを使わないコードは？」など生成 AI と対話しながらコードを修正し、コードを見ながら自分自身でデバッグしていく作業はこれか

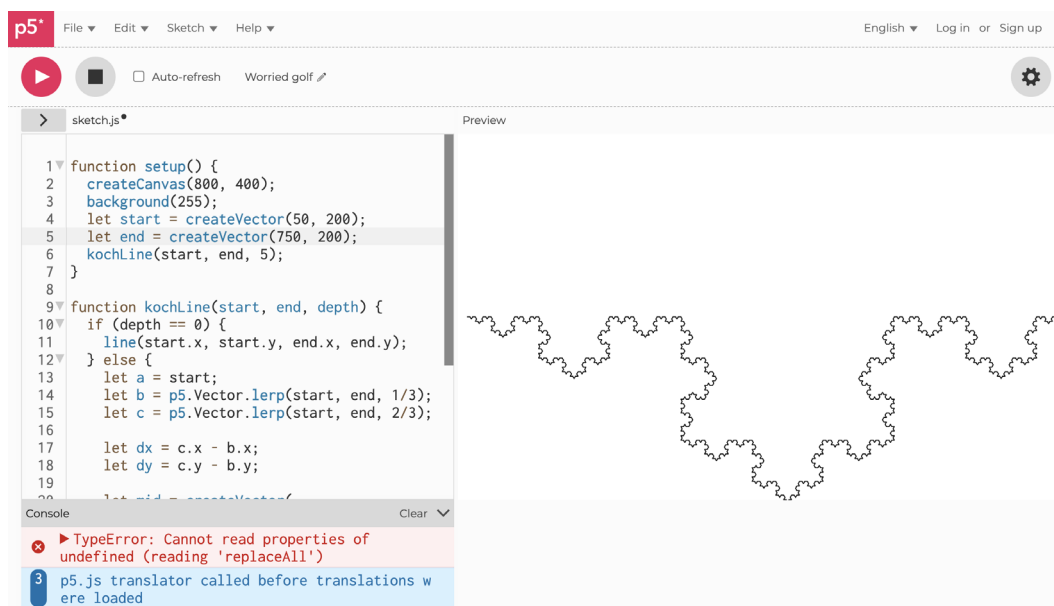


図 3: ChatGPT が出力したコッホ曲線。エラーは 4,5 行目に予約語を変数名に使用した為

らのプログラミング学習、および数学教育に置き換わっていくものと考えられる。講義ではコッホ曲線以外にもドラゴン曲線やシェルピンスキーのギャスケット、マンデルブロ集合、ニュートンフラクタルなど様々なフラクタルを生成 AI と一緒に描く体験をした。

講義『機械学習と AI』も『DS 特別講義 I』と同様に Python プログラミングが主軸であり、どちらもコーディングには環境構築が不要な Google Colaboratory を中心に使っている。しかし、AI アシストが搭載されコードを少し書くだけで残りを自動補完してくれるようになった為に従来の演習問題が機能しなくなりつつある。テキストセルも自動補完の為のプロンプトの範疇なので、AI が問題文を理解して解答プログラムを殆ど全て書いてくれる事態となっている。これからの時代では、各演習問題に対して AI が作成した解答と個々人が作成したものを比較してどちらがより良いプログラムか評価するなど、新しい勉強の仕方を教える事が重要となるだろう。

## 4 おわりに

2021 年後期にスタートした『データサイエンス入門』は 2024 年前期までに各学期 200 名以上、累計 1500 人以上の学生が受講している一方で、2023 年前期にスタートしたアドバンスレベルの講義は各学期 15 名程度に留まっている。2024 年度からは理学部情報数理学科が新設される事も踏まえると、1 年次と 3 年次のデータサイエンス教育を滑らかに接続する様な 2 年次の全学科目を提供できればと考えている。また、変化の激しい分野でもある為これまでの科目についても新しい内容や教授法を引き続き試行錯誤していく所存である。

## 参考文献

- [1] 駒野誠, 必修「数学 I」で統計を学ばせるための教材実践研究, 第 17 回統計教育の方法論ワークショップ・理数系教員授業力向上研修会(東京)および第 9 回科学技術教育フォーラム (2020), [https://estat.sci.kagoshima-u.ac.jp/SESJSS/PDF/JCOTS20\\_komano.pdf](https://estat.sci.kagoshima-u.ac.jp/SESJSS/PDF/JCOTS20_komano.pdf)
- [2] 梅谷俊治, しっかり学ぶ数理最適化, 講談社サイエンティフィク (2020).