

# 推論によるデータ分析における ChatGPT の活用

学習院大学経済学部経営学科 白田由香利

## 要約

数学教育において解法過程における推論力の養成は重要と考える。筆者は、経営数学教育における推論の重要性を感じ、演繹推論を核とする講義を行ってきた。現在、教育現場における ChatGPT の活用に議論が起きている。本稿では、ChatGPT による数学問題解法における、推論での ChatGPT の活用について論じる。ChatGPT の機能の中で、筆者が最も活用すべきと考えた機能はプログラムの提示機能である。プログラミングマニュアルを参照する手間に比較して、ChatGPT への質問のほうが、時間と手間が少なく、質問者の意図に合致する回答が得られることが多いと感じる。この ChatGPT の機能を活用可能な数学教育分野は、データ分析手法の学習と考える。本稿では、データ分析手法の学習における学習者の推論に、ChatGPT をどのように活用できるか提案する。データ分析手法を教える際に、演繹過程を教案として示し、順を追って演繹を行うことで学習を進めてもらう。データ分析においては、ステップごとに導出された結果の可視化を行うことが有益である。従来は、ここでプログラムを書く手間が大きかったため、可視化による結果検証をスキップして先に進めることが多かった。ChatGPT の活用により可視化プログラムが容易に作成可能となれば、分析手法に理解が深まり、分析結果の検証が容易となり、結果として自分のプログラムの間違い及び概念の理解の間違いにも気づくことが容易となる。本稿では、ChatGPT を活用して演繹することでデータ分析手法を学習させる学習方式を提案する。その事例として主成分分析 PCA を学習する演繹過程を示す。

## 1. 始めに

数学教育において解法過程における推論力の養成は重要と考える。筆者は、経営数学教育における推論の重要性を感じ、推論を核とする講義を行ってきた。与えられたデータから未知数に至る演繹推論過程を解法プラングラフとして学生に提示し、解法を演繹推論過程として教えてきた。解法プラングラフを自動生成するシステムも作成し、豊富なドリル問題の生産に貢献した。現在、教育現場における ChatGPT の活用に議論が起きている [1] [2]。本稿では、ChatGPT (以降 GPT と略す) による数学問題解法における推論における GPT の活用について論じる。GPT の機能の中で、筆者が最も活用すべきと考えた機能はプログラムの提示である。プログラミングマニュアルを参照する手間に比較して、GPT への質問のほうが、時間と手間が少なく、質問者の意図に合致する回答が得られることが多いと感じる。この GPT の機能を活用可能な数学教育分野は、データ分析手法の学習と考える。本稿では、データ分析手法の学

習における学習者の推論に、GPT をどのように活用できるか提案する。提案する学習法は、データ分析手法の学習における演繹推論法である。データ分析手法を教える際に、演繹過程を教案として示し、それを順を追って演繹を行うことで、学習を進める。データ分析においては、ステップごとに導出された結果の可視化を行うことが有益である。従来は、ここでプログラムを書く手間が大きかったため、可視化による結果検証をスキップして先に進めることが多かった。GPT の活用により可視化プログラムが容易に作成可能となれば、分析手法に理解が深まり、検証が容易となり、結果として自分のプログラムの間違い及び概念の理解の間違いにも気づくことが容易となる。

次節では、筆者が行ってきた演繹による数学学習法について概説する。第3節は、それを発展させて、GPT を活用してデータ分析手法を学習させるというバージョンの学習方法を提案する。第4節は、その事例として主成分分析 PCA を学習する演繹過程を示す。第5節はそれに関する考察であり、第6節は本稿のまとめである。

## 2. 演繹推論による数学学習手法

本節では、筆者が考案し実践している演繹推論法を簡単に説明する。詳細は [3-13] を参照して頂きたい。

筆者は、経営数学教育における推論の重要性を感じ、講義においても、演繹過程を解法プラングラフとして学生に提示し、推論力養成を目標とする数学教育を行ってきた。そして、解法プラングラフを描くツールとして、解法プラングラフ・ジェネレータというシステムを開発し、解法プラングラフ教材を多数作成し、教育に用いてきた。解法プランジェネレータのシステム構築と評価については [3, 13] を参照して頂きたい。

筆者の将来目標は、学習支援システム (e-Learning システム) において、高度な推論支援機能を実現することである。そのためには、人間の学習者がどのように推論を行って問題を解いているか、考察する必要があるので、簡単に説明する。

推論には、演繹、帰納、アブダクションの3種類がある [14]。演繹では、前提だけから、論理学の教える規則に従って結論を導出する。3段論法は、演繹である。演繹は論理的に仮定しうる理想状態を形式的に図式的に表意し、その理想状態における事象 (論理的対象) の間の必然的関係を言明する記号過程である [15]。数学の例で言うと、連立方程式を解いて解を求める過程は、演繹である [15]。求めたい未知数 (結論) は、与えられた式に既に暗々裏に含意されていて、連立方程式を解く形式的手続きによって、未知数が導かれるからである。アブダクションは個別の事象を最も適切に説明し得る仮説を導出する推論であり、事象間の因果関係の解明に重きを置く [16]。

パース (Charles Sanders Peirce) はアブダクションの推論形式を、以下のように定式化している [15]：驚くべき事実 C が観察される。しかしもし H が真であれば、C は当然の事柄であろう。よって、H が真であると考えらるべき理由がある。C とは、我々の疑念と探究を引き起こすある意外な事実または変則性のことであり、H は C を説明するために考えられた仮説である。但し、アブダクションおよび帰納法においては、前提が真であっても仮説が真であるとは限らない。

経済経営数学の問題の殆どは、証明問題ではなく、未知数の値を発見 (計算) するタイプの問題である。よって、それらの問題においては、問題で与えられたデータ (Given Data) から演繹をすることで、結果として、未知数 (Unknown) の値を求める [23, 24]。求めたい未知数は、Given Data、公式、セオリーを前提として、演繹を行えば、Unknown が結論として導出可能、と言える。Given Data から Unknown に到る演繹過程を図的に表現する手法として、我々は図 1 のような表現を提案した。この演繹過程を我々は「解法プラングラフ」と呼んでいる。

問題を解こうとする際、我々は始めから、解法プラングラフのような演繹過程が頭に浮かぶわけではない。アブダクションによって、演繹過程の途中途中において仮説を創造し、その仮説を暫定的に採択することで、ステップに分けて推論を行うと考えられる。その際にどのようなステップ分割をするか、定説はないが、我々は、人が経営数学問題を解く推論過程は以下のようなものであると、考えた。経営数学問題を解くため、人は、演繹、帰納、アブダクションのすべてを複合的に駆使して思考を行っているであろう。その詳細は定かではないが、人はその思考の過程において、アブダクションによって、仮説 H を立てて一歩先を予測しながら、Unknown から Given Data 方向にむかって、つまり逆方向に演繹推論を適応して、細かいステップに分けて問題を解いているのではないだろうか。ポイントは、逆向きに推論する、という点と、アブダクションにより部分目標を設定することである。このアブダクションの際、複数の仮説の中から、最ももっともらしい仮説を選ぶことが重要であるが、それは、その時点で与えられたデータと計算済みデータ、および知っている公式およびセオリーを駆使して推測する。人がこのように推論すると仮定した場合、e-Learning システム構築においても、このプロセスをヘルプするような機能を実現すれば、学生の学習に効果があるのではないだろうか、と考えた。

次に具体的事例を示す。

## 文章題

債券 X は、残存年数 2 年、クーポン・レート 2.00%、今日の価格 (額面 100 円当たり) 100.00 円、マコーレー・デュレーションが 1.98 年、修正デュレーションが 1.94、コンベキシティ 5.69、クーポンの支払いが年 1 回の債券である。今日の最終利回り (年 1 回複利) は 2.00% であり、今日は利払い日直後である。スポット・レート・カーブは水平と仮定する。明日、スポット・レートが 1% 低下して、そのまま 1 年間経過したとする。明日の債券価格を、近似式を使って求めなさい。債券 X に今日から 1 年間投資した場合の投資リターンは何%になりますか。

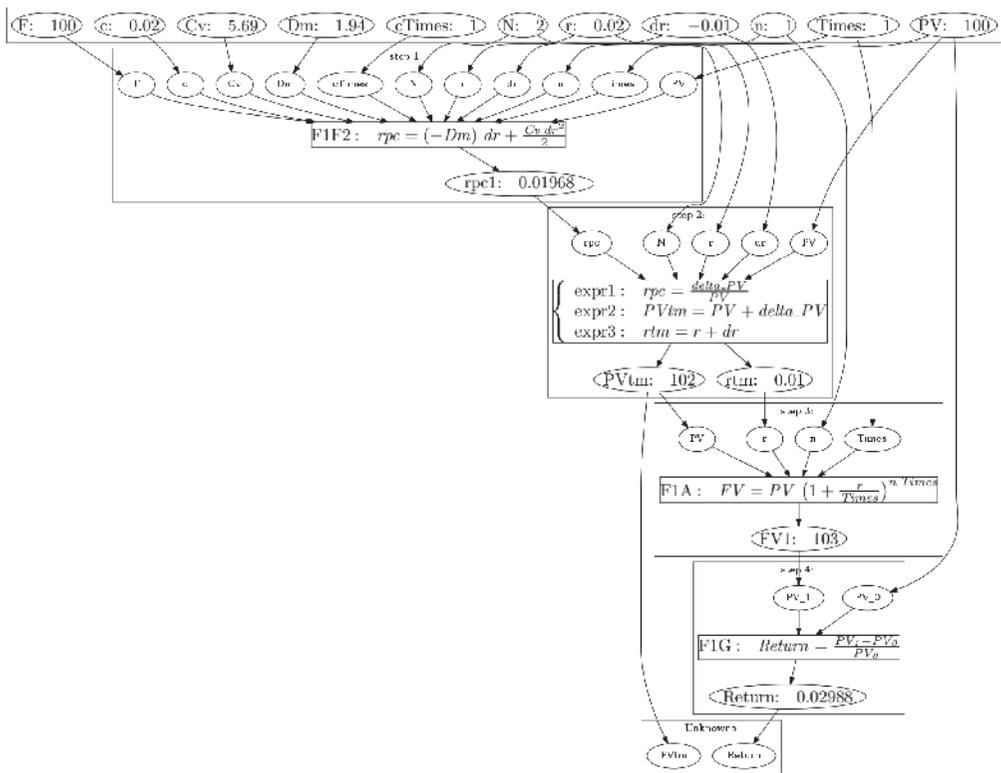


図1：債券数学文章題の解法のための演繹推論過程の例（[13] から引用）

図1に本文章題を解くための解法プランを示した。四角が公式（仮説）を表し、楕円形は変数値を表す。この問題を解くため、Unknownである「リターン値」からの逆向き推論を行うとよいであろう（図1の一番下の四角）。そのためには、今日の債券価格 PV（Present Value）および、1年後の将来価値 FV1（Future Value 1）を求める必要がある（step 4）。金利変動により、今日と明日の金利（スポット・レート）の差は、1%の低下と、文章に記載されている。そこで、「明日の債券価格 PVtm を計算できれば、1年後の債券の将来価値 FV1が計算できることは当然の事柄であろう」という仮説がたてられる（step 4）。「今日と明日の債券価格の差分を計算できるのであれば、明日の債券価格 PVtm が計算できることは当然の事柄であろう」（step 3）。また、「価格変化率 rpc を計算できるのであれば、今日と明日の債券価格の差が計算できることは当然の事柄であろう」（step 3）。つまり、PVtm を求めるためには、価格の差分 dPV が必要であるが、それは価格変化率 rpc の公式を用いることで計算されるであろう（step 3）、と推測する。この rpc を計算するために公式 F1F2のテイラー展開の公式を使えばよいであろう（step 1）。公式 F1F2を用いて rpc を求めるためには、ほしいデータ PV, Dm, Cv がそろっているかが問題となる、GivenDataを確認すると、はたして、PV, Dm, Cv の値は与えられていた。このような逆向きの推論が行われる、と考えることはもっともらしい。この解法における最も重大なひらめきといえるべき仮説は、step 3の「価格変化率 rpc が計算できるならば、債券価

格の差分  $dPV$  が計算できることは当然の事柄であろう」であろう。多くの学生を見ていると、この推論ができない学生が多い。そして、 $rpc$  が **Given Data** を使って求められるかを検証し、その結果、求められることが判明する。 $rpc$  が求められたので、演繹によって、 $dPV$  が求められる、となる。

この step 1 に相当する仮説の設定において、我々は試行錯誤的に仮説を考えている。しかし、その際には、ある理由をもって、複数ある仮説の中から最ももっともらしい仮説を選択している。例えば、文章題として、修正デフレーションおよびコンベキシティの値が与えられているので、 $rpc$  の公式を少なくともどこかで使うのではないかと、ある人は推測するであろう。多くの場合、逆向きに推論が行われるが、逆向き方法からのみ推論が行われるわけではなく、その人の過去の経験などから、部分目標が設定されるのであろう。この step 1 の推論は、機械がやみくもに仮説を検証するのと違い、人間が高いインテリジェンスをもって推論をしている可能性を示す例である

この方式に基づく経済経営数学教科書も筆者は書き [17]、実践している [12]。メインターゲットは経済経営数学の授業であり、そこでは IS-LM 問題、需要供給に係る問題、最適化問題などを扱う。一般的な経済経営数学問題以外にも広く本方式を適応している。

筆者は、演繹推論による数学文章題解法を債券数学教育にも応用した。債券数学に関する公式を整理統合し、債券数学の概念モデルを構築した [9]。本概念モデルは、実体関係 (E-R) ダイアグラムで表わされ、このダイアグラム上で **Unknown** から **GivenData** に至る演繹過程のパスを発見することで、学習者は **Unknown** を求める。この概念モデルだけで、殆どの債券数学の教科書にある文章題が解けることを検証した [9]。

また、シンプル・トピック・モデルのバイズ推論の過程を導出する演繹推論過程も解法プラングラフにより示した [5, 6]。

また、**GivenData** から **Unknown** までのパスが非常に複雑で長い例として、ブラックショールズの公式を求める、という演繹推論過程を示した (図 2 参照)。筆者は、ブラックショールズ方程式を本解法プラングラフを用いて教えている [18]。ブラックショールズ方程式の導出は長いので、学習者は今どの公式をもちいて演繹しているのかが分からなくなりやすいが、解法プラングラフを地図として、現在の推論がどこの部分に位置するのかが分かるという利点がある。

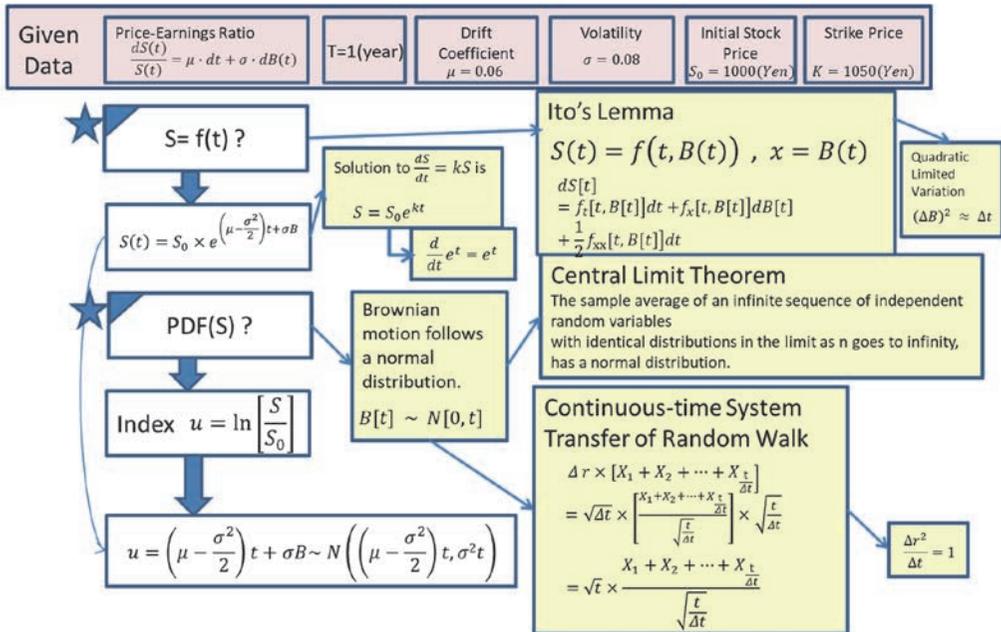


図2：ブラックショールズ方程式を導出するための解法プラングラフ（部分）（[18] から引用）

### 3. 演繹推論によるデータ分析

本節では新たに、演繹推論によるデータ分析学習法を提案する。対象は、データ分析手法の学習である。人間の教師が予め講義を行い、その知識定着のため、学習者（学生）は独習を行ない、独習において学習者はGPTを活用する、と仮定する。以下のプロセスで学習は進む。

- (1) 人間の教師が**教案**を作成し、学習者に提示する。教案は演繹推論による GivenData から Unknown に至るパスとして示されている。
- (2) 学習者は GivenData から Unknown 方向に演繹を行ないながら学習を進める。各仮説は、プログラム実行に対応する。入力データをプログラムに入力し、出力データを得る。そして次の演繹推論に進む。このプログラム作成にGPTを使う。GPTに質問することで、プログラムを返してもらい、そのプログラムを実行する。想定しているプログラミング言語はパイソンである。
- (3) 教案と違う結果が得られた場合、学習者はいずれかを行う。
  - (ア) 同じ内容の質問を言葉を変えて行い、GPTが正しいプログラムを返すまで繰り返す。
  - (イ) 間違った理由を自分でアブダクションする。アブダクションの仮説が思いつかない場合、GPTに質問し、アブダクションを支援してもらう。
- (4) 演繹結果のデータの検証のため、グラフ・プロットする。プロットのためのプログラムはGPTに質問して教えてもらう。

想定しているデータ分析手法は、広く普及している手法、例えば、回帰分析、クラスタリング、次元圧縮、などの枯れている手法を想定している。それは GPT は web 上の集合知を基に回答を作成しているの、普及しているデータ分析手法でないとプログラム例などが web 上に存在しないからである。次節では、具体的な学習の進め方を示す。

#### 4. 演繹推論によるデータ分析事例

本節では、演繹推論によるデータ分析事例として、主成分分析 (PCA) によるデータ分析を GPT との対話を通して学習する事例を紹介する。本事例を使って、GPT 活用の効果を説明する。

PCA は教師なし学習を行い、データの主成分を抽出する [19] [20]。PCA はノイズの多いデータのフィルタリング手段として利用でき、寄与度の大きな主成分のみを選択し、データを再構築すれば、ノイズを除去しデータの本質的性格を抽出することが可能となる [21]。データは、インドネシアの「高校就学率 (2020年度)」及び「携帯電話普及率 (2022年度)」の州別データとする。データ数は31 (州) である。本データはインドネシア統計局のサイトから検索した (<https://www.bps.go.id/>)。

PCA 分析においては、GivenData が上記州別データ、Unknown が2個あり、(1) 各データの主成分値と (2) 主成分の寄与率とする。GivenData から Unknown に至る演繹推論過程 (教案) を図3に示す。図中、四角形は用いる公式あるいは手続きを示し、楕円形はその解 (値あるいは数式) を示す。この推論過程は、学生に予め与えられ、学生はこれを用いた数学の講義も受けている。その理解を深めるため、以下のような学習を独習として行う。

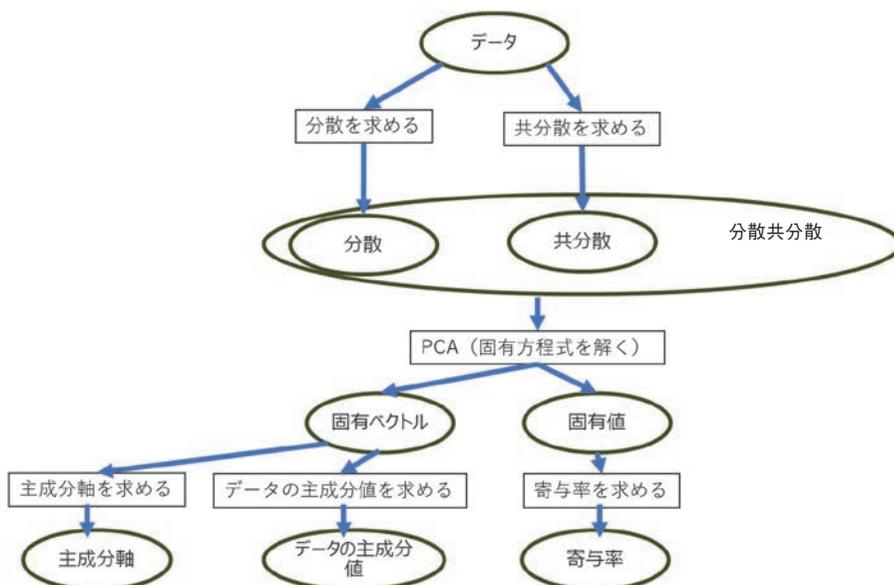


図3 : PCA の演繹推論過程 (教案)

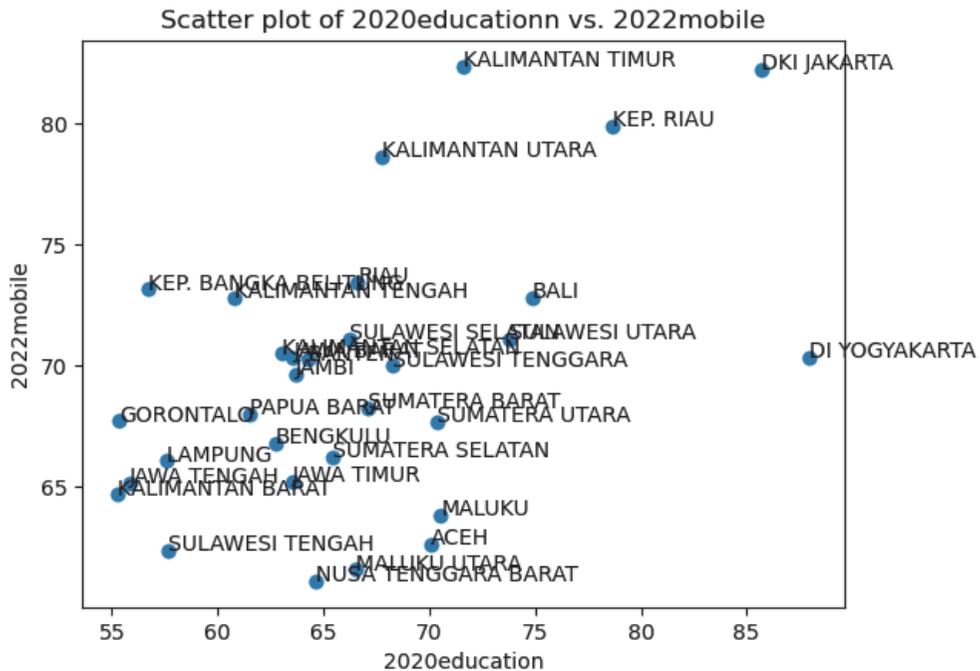


図4：入力データ確認のための散布図

以下では、GPT（バージョン3）とのやり取りをそのまま記す。初めに、学習者の行いたいことを示し、そのためにGPTにどのように質問をし、その結果がどうなったかが書かれている。

① 入力データ確認のため散布図をプロットしたい。

Q：「Province, 2022sanitation, 2020education, 2022mobile, 2022gender という5つの変数がある。このうちの2つの変数を選んで、散布図をプロットするパイソンプログラムを書いてください。各データには、Provinceの値がラベルで示されるようにしたい。」  
 GPTが返してくれたプログラムで容易に散布図をプロットできた（図4参照）。

② 入力データの分散共分散行列を求めたい。

Q：「Province, 2022sanitation, 2020education, 2022mobile, 2022gender という5つの変数がある。2020education, 2022mobileの2つの変数の分散共分散行列を求めるパイソンプログラムを書いてください。」

GPTは入力データとしてarrayを使ったプログラムを返した。著者はデータフレームのプログラムを希望しているので、以下のように質問を変更した。

Q：「Province, 2022sanitation, 2020education, 2022mobile, 2022gender という5つの変数がある。2020education, 2022mobileの2つの変数の分散共分散行列を求めるパイソンプログラムを書いてください。データはread\_csv()で読み込んでほしい。」

今度はデータフレームを使ったプログラムが返された。GPTの回答のプログラムを

見ると、そのメインパートは `cov_mat = np.cov (df [[var1, var2]].T)` であった。T は転置行列の関数であるが、転置行列を取る意味が分からないので、GPT に質問した。

Q : 「分散共分散行列を求めるとき、なぜ以下のように転置行列をとるのですか？」

```
cov_mat = np.cov (df [[var1, var2]] .T) ]
```

A : 分散共分散行列を求める場合、データフレームの行が観測値、列が変数に対応します。しかし、`np.cov` 関数にデータフレームを渡す場合、列が観測値、行が変数に対応するため、転置行列を取って列と行の対応を正しくする必要があります。(以下、略す)

上記の GPT の説明で満足したので、先に進む。

### ③ 固有ベクトルと固有値を計算したい。

Q : 「正方行列を入力して固有ベクトルと固有値を計算するパイソンプログラムを書いてください」

プログラムを動かすと正しそうな結果が返されたので、満足して、次に進む。

### ④ 固有値の寄与率を計算したい。

Q : 「固有値の寄与率を計算するパイソンプログラムを書いてください」

プログラムを動かすと正しそうな結果が返され満足して、次に進む。

### ⑤ 主成分軸を求めて散布図に主成分軸をプロットしたい。

Q : 「2次元データにPCAを行い、各データの第1主成分値と第2主成分値を求めるプログラムをパイソンで書いてください」

質問文中の「2次元データ」の意味を、間違ってデータ数2個と解釈されたため筆者の意図と異なるプログラムが返された。また、`from sklearn.decomposition import PCA` ライブラリを使うプログラムが返され、筆者はこの場合の入力データ形式が不明となった。そこで質問を以下のように変更した。

Q : 「データを `read_csv` で読み込み、データに対してPCAを行い、各データの第1主成分値と第2主成分値を求めるプログラムをパイソンで書いてください」

返されたプログラムは動いた。結果の正当性を確認するために、各データの `PC_1` と `PC_2` で散布図を描いた (図5参照)。散布図のプログラムは、①の問でGPTからプログラムが返されているので、それを利用した。図5から、正しく主成分値が得られたことを確認できた。そこで、主成分軸を書き加えるプログラムを以下のように質問した。

Q : 「データを `read_csv` で読み込み、データに対してPCAを行い、第1主成分軸を赤色で、第2主成分軸を青色で、データの散布図上に描くプログラムをパイソンで書いてください」

GPT のプログラムでは、`plt.scatter (df.iloc[:, 0], df.iloc[:, 1])` というような属性指定をしていたので、データ列を本事例に合うように、自分でプログラムを変更した。プログラムは動いたが、図6のような意図しない図となった。図6から、入力データを加工して平均値からの偏差をプロットすべきこと、つまり、データ平均を0 (原点)

に移動する必要があることに気が付いた。平均を計算する関数を聞くため、以下のよ  
うに質問した。

Q:「パイソンでデータの平均値を求めるにはどうするのか?」

GPT からの回答で, `numpy.mean()` 関数であることを知り, プログラムを修正し, 図  
7のグラフを描けた。次に, 第1主成分軸と第2主成分軸を直交させたいので, デー  
タ範囲を同じにしたいと考えた。そこで, 以下の質問をした。

Q:「パイソンで散布図のデータ範囲を -30から30に指定する方法は?」

GPT からの回答で, `plt.xlim(-30,30)` のように指定すればできることを知り, プログ  
ラムを変更し, 図8の図を描けた。以上。

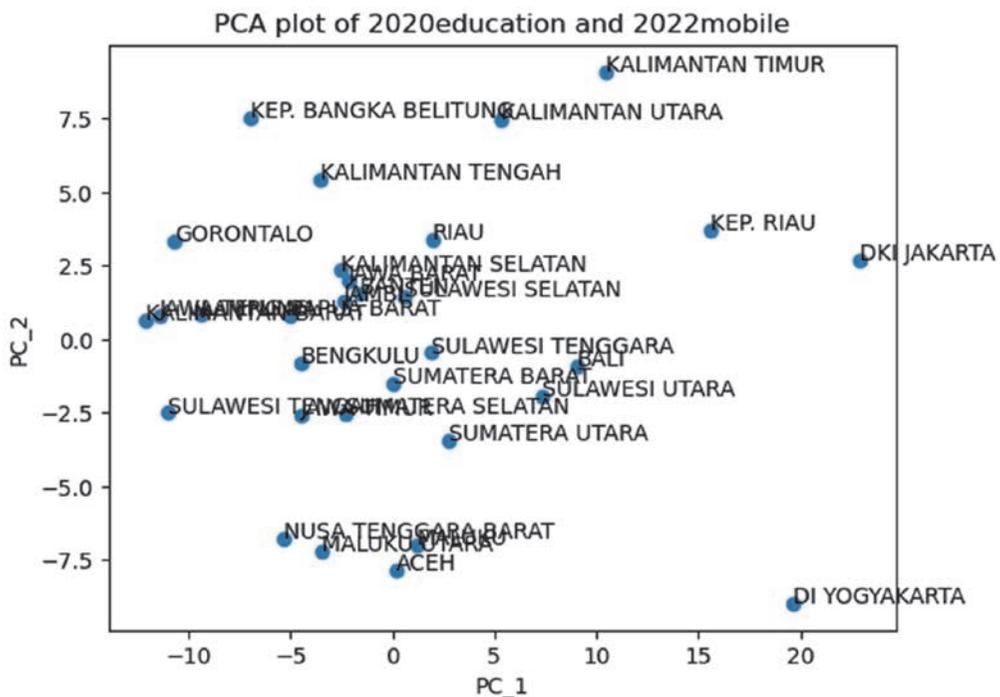


図5：各データの第1主成分値と第2主成分値の散布図

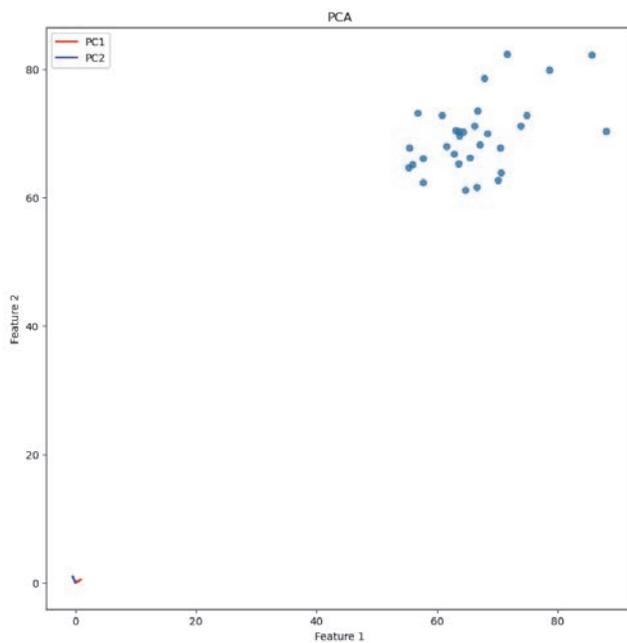


図6：入力データと主成分軸2本（左下に小さく軸が描かれている）

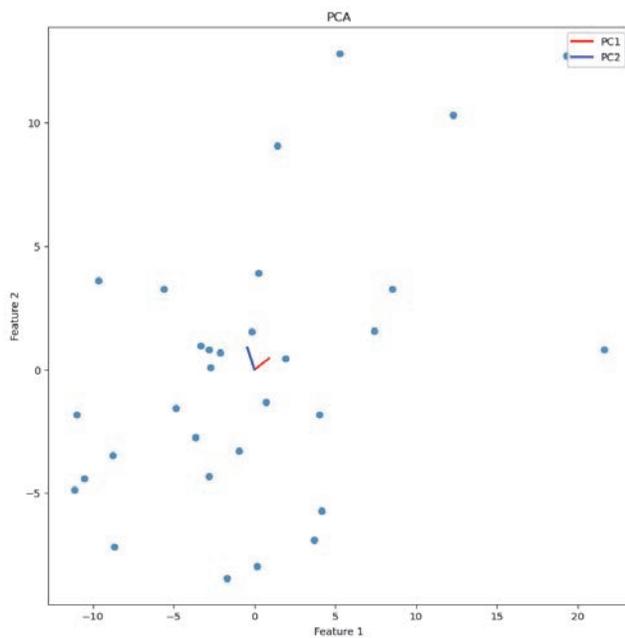


図7：入力データを偏差に直した散布図に主成分軸2本を描いた図

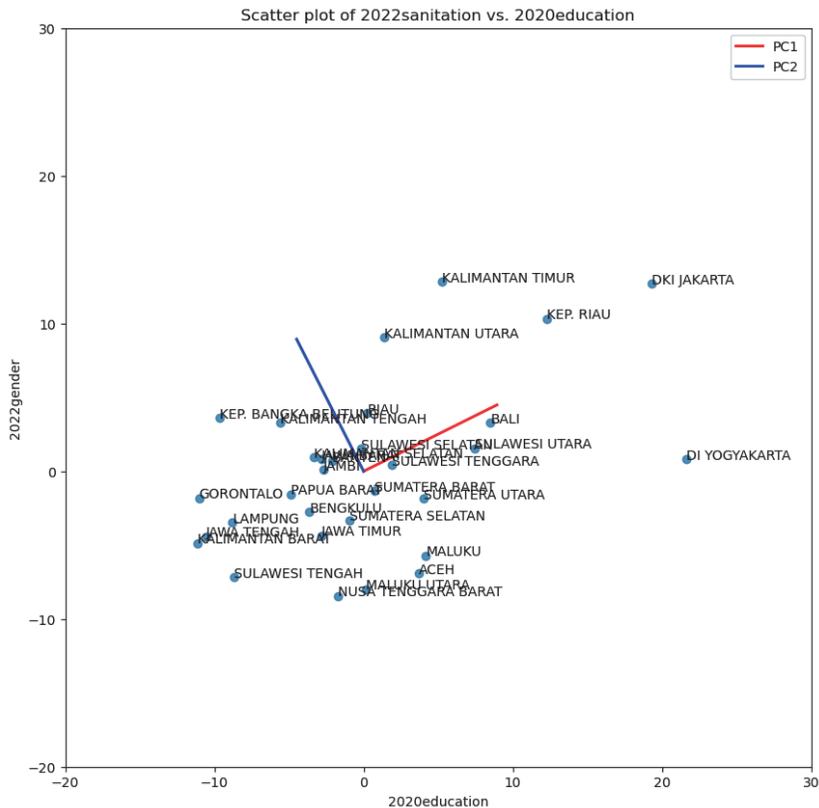


図8：入力データを偏差に直した散布図に、主成分軸を描いた図

## 5. 考察

前節の事例の結果を受けて、GPTを利用した推論によるデータ分析手法学習法について考察する。

GPTを利用することで、自分で一からプログラムを書くのに比較して短い時間でプログラミングを進めることが可能となった。特に、プロットに関する関数のオプション指定に関して、サンプルを変更するだけで目的のプログラムを得ることができる点が優れている。行いたいこと概念が明確になっているときに、単純に関数やオプション、引数など、マニュアルを参照すればわかる問いに対して、GPTにそのプログラムを書いてもらうことは非常に効率を高める。

問題は、自分が知らない複雑な概念に対して、どう質問すべきかである。未知の概念は大別すると、(1) プログラムに関する知識、と (2) データ分析に関する知識となる。まず、パイソンプログラミングについての知識習得に関する考察を行う。前節で、問題となったのは以下の事柄についてであった。

- A) パイソンプログラムにおいて、`array` を使った方式にするか、データフレームを使った方式にするかで、GPT が異なる表現のプログラムを返す。その違いを予め知識として習得していないと、先に進めなくなる。
- B) 「分散共分散行列を求めるとき、なぜ転置行列を取ったのか」についても、ライブラリの仕様がそうになっているから、という理由である。これはマニュアルを参照して得るレベルの知識である。しかし、これを知らないといつまでもプログラムが動かないという困った事態になる。行列計算プログラム作成においてこの問題で動かないことが多い。
- C) 自力でプログラムをカスタマイズする必要がある。GPT が返すプログラムのデータ列および変数名等を、自分の分析ケースの列や変数名に置換できるプログラム能力が必要とされる。これはどのプログラム言語でもよいので、プログラム言語を習得していないと、できない可能性が高い、GPT を活用できるレベルになるためには、最低限のパイソンプログラミング演習が必要である。

次に、データ分析における知識の発見ができるか否かという箇所を上げる。これらの問題解決には、数学知識とプログラミング知識が関連して必要となる。

- a) 2次元データといったときに、「2つの列のデータがある」ことを誤解して、「 $2 \times 2$ 行列」のデータが入力された場合、として GPT がプログラムを返した。著者は、プログラムの返り値がデータ数の32個あるべきところ、2個しかないので、瑕疵に気が付いた。どうしてそのような間違った結果になったのか、アブダクション的推論を行ない理由に至る必要がある。自分の質問の「2次元データ」が誤解されたことに気づかないと、理由が分からないままとなってしまう。
- b) 固有ベクトルをそのまま主成分軸として描いたところ、データ散布図と異なる場所に描かれた (図6参照)。自分の意図と異なるプロットとなった場合、その理由を考える必要がある。この場合は、入力データの平均を求め、平均が原点になるように平行移動する必要があるからである。それに気が付くか否かは、学習者がアブダクション的推論で、その理由に思い至るか否かに依存する。既に「間違いやすい注意すべき点」として、WEB のどこかで先達がコメントしている場合、それを活用して GPT が指摘してくれる可能性がある。こうした、演繹推論過程の中のミッシングリンクとなったアブダクションに対して、GPT に上手に質問をする質問力が向上することが望まれる。
- c) データの散布図上に主成分軸を直交するように描きたい場合、2変数の定義域の幅を同じにしないと直交しない。直交しないという事実から、その理由をアブダクションすることは、簡単にできない可能性が高い。例えば「Q: PCA で主成分1と主成分2の軸を描いたのですが、直交しません。私は、主成分軸は互いに直交すると習いました。私のパイソンプログラムが間違っているのでしょうか?」と質問して、GPT が「軸の定義域の範囲を同じにする必要があります」と答えることが将来可能になるか、筆者は多分可能となると予測する。

これらの事例から、一般的なデータ分析に本手法を適応した場合について考察を進める。

重要なことは、初めに作成する演繹推論過程のダイアグラム、教案の作成である。これは教師が教案を作ることに対応している (図3参照)。学習者はこのダイアグラムに従い、Unknown に到達するまで演繹推論を行う。第2節で示した、文章題を解くための演繹推論過

程と類似しているが、異なる点は以下のようにまとめられる

- (1) **グラフ・プロットの重要性の増加**： 各演繹推論の結果をプロットで可視化することで、その演繹結果を検証できる。また、プロットすることで、その概念に対する理解が深まる。数値だけを見るだけでは、このように理解は深まらない。
- (2) **GPT という生成系 AI の存在の出現**： 上記 (1) を可能とする機能は、GPT がプログラムを返してくれる機能である。可視化のためのプログラミングが容易になったので、可視化を頻繁に行えるようになった。これは従来のデータ分析方法を根本的に変える。  
これは例えると、会社の重役に対してインテリジェント秘書がいて、データ分析をサポートしてくれるような存在である。例えば、重役は「今のデータのうち上位10%のデータに限定して、PCA をしてプロットしてほしい」(対象データの限定)、「次元圧縮アルゴリズムを PCA から UMAP に変えてやってみて、その結果をプロットしてほしい」(アルゴリズムの変更)、「Shapley 値を計算した後、Shapley 値を使って、クラスタリングしてクラスターごとの変数の平均値を求めてほしい」(分析手法の合わせ技)と、データアナリストに依頼するかのようになり、次々と検討したい分析結果をリクエストするのに類似している。従来は、自分でプログラムを組むコストが高いため実現困難であった多面的な分析が、GPT の出現により、少ないコストで実現可能となった。学習システムも、GPT を活用するように変革する必要がある。
- (3) **想定しない結果の場合の理由究明における GPT の活用**： 想定した結果と異なる結果がでた場合、その理由をアブダクションする。そのアブダクションの際に、GPT に質問することで、理由が分かる可能性がある。この仮説の検証は、プロットプログラムの正当性の検証よりも一般に難しい。プロットの場合は、目である程度確認できるが、このような場合の間違った理由は判断が難しいからである。

GPT は同じ目的のプログラムとして、複数の異なるプログラムを返す。例えば、パイソンで array を使うかデータフレームを用いるかで方法は異なるが同じ計算は可能である。また、同じ関数が複数のライブラリで実現できる場合、利用するライブラリに依存してデータ形式は異なるだろう。学習者は複数のプログラムから最適なものを選択する際、単純性と経済性を重視すべきである。これは一般的なアブダクションにおける仮説選択の場合と同様である。単純な仮説ほど、それを実験的にテストするのに費用や時間や思考やエネルギーが節約できるから、単純な仮説を選択すべきである。数学者ジョージ・ポリャ (George Pólya) [22] も、同じ問題を解くのに、知っているからと言って、難しいテクニックを駆使して解くよりも、簡単に解ける方法を選択すべきである、と言っている [23]。その最適選択ができる能力の養成も、学習システムの課題と言える。

データ分析手法学習システムとして、考察をしてきたが、以下ではデータ分析手法を学習しつつ自分のデータでデータ分析を行うことを主眼とした場合、上記の演繹推論過程の位置づけはどうなるのか考察する。アブダクションと帰納の融合による知識発見の自動化が期待されてきている [16]。データ分析の目的を知識発見とするならば、自動化システムに倣ってデータ分析することで人間も効率的に知識発見することが期待できるであろう。Ray によると、知識発展のプロセスは、アブダクション→演繹→実験→インダクション (帰納) →アブダクション

→ (以下これが続く) のスパイラルである [16, 24]。ここで論じている知識とは、科学における知識発見であり、例えば、「高校卒業者のほうが給料が高い」という仮説を立て、演繹により予測や検証を行い、実験を行い、結果を観察して、次に帰納として、複数の事象間に成立する一般化則 (例えば、「高等教育を受けた人のほうが給料が高い」) を発見し、例外として「携帯電話を上手に使える人は他の人に比較して給料が高い」を例えば発見し、それに対して再びアブダクションによって考える、というスパイラルである。

本稿の議論の対象は、上記スパイラルの演繹による仮説検証の部分である。データ分析を教える教育者の立場として、その演繹過程を正しく理解させることが第一義と考えるからである。GPT の活用は望ましい。他方、セキュリティの問題から、仮説立案のアブダクションや帰納に GPT を使用することは今後議論されるべき課題と考えるからである。枯れた技術であるデータ分析手法についての GPT への質問には、新しい知識は含まれていないので、問題ないと考える。

## 6. まとめ

本稿では、推論によるデータ分析手法の学習を GPT で加速する方法について論じた。筆者が長年研究し実践している経済経営数学の学習手法演繹推論法との違いは、以下の3である。(1) 演繹推論の各演繹過程におけるプロットの重要性の増加。これはデータ分析の特徴であり、頻繁に可視化したほうがその概念に対する理解が深まる。(2) GPT という AI の存在の出現。GPT が可視化プログラム等のプログラムを返してくれるので、データ可視化を手間をかけずに頻繁に行えるようになった。これは従来のデータ分析方法の学習法を根本的に変えるであろう。(3) 想定しない結果の場合の理由究明における GPT の活用。学習者は自分が間違った理由をアブダクションしようとする際、GPT に質問することで分かる可能性がある。

本稿で述べた推論によるデータ分析手法の学習システムを構築する場合、問題は、学習者自身が知らない複雑な概念に対して、どう質問すべきかである。未知の概念は大別すると、(1) プログラムに関する知識、と (2) データ分析に関する知識となる。想定しない結果をどう解釈するのか、どのように GPT に質問するとよいか、学習者のこれまでの知識とスキルに係ってくる。

パースによると、人間には本来「正しく推測する能力」が備わっている。よって、人間の精神は考えられうるあらゆる仮説の中から、ある有限回の推測によってもっとも正しい仮説を考えあてることができる、とある [15]。データ分析手法学習過程においても、教案と異なる結果がでた場合、その間違った理由を得るため、アブダクションにより仮説を選択する作業において、人間は機械よりも創造的に仮説選択をしたい。GPT の活用はこの有限回の推測の効率化に役立つ。先人達の集合知を基に、学習者がよくやるエラー事例を多数蓄積することで、GPT は多数の仮説を提示することが可能となる。提示された仮説の中から、検討中の問題の現象において、最も理にかなった説明をする仮説を選択するスキルを、学習者はもたなくては GPT を活用することはできない。アブダクションにおいて我々人間は機械よりも創造的である仮説選択を可能とするように、常にスキル向上を図るべきであろう。

## 参考文献

- [1] 週刊東洋経済編集部, “ChatGPT 仕事術革命,” *週刊東洋経済* 2023年4/22号, 2023.
- [2] 浅田義和, “医療教育分野における昨今の e ラーニング動向,” *安全工学*, vol. 62, no. 2, pp. 87-93, 2023.
- [3] P. Stanworth, and Y. Shirota, “Review of the “WebHow2SolveIt” Website,” *Gakushuin Economics Papers*, vol. 50, no. 1, pp. 1-18, 2013
- [4] Y. Shirota, T. Hashimoto, and S. Suzuki, “Knowledge Visualization of Reasoning for Financial Mathematics with Statistical Theorems.” pp. 132-143.
- [5] Y. Shirota, T. Hashimoto, and B. Chakraborty, “Deductive Reasoning for Joint Distribution Probability in Simple Topic Model.” pp. 622-625.
- [6] Y. Shirota, T. Hashimoto, and B. Chakraborty, “Visualization of Deductive Reasoning for Joint Distribution Probability in Simple Topic Model,” *Information Engineering Express*, vol. 3, no. 1, pp. 1-8, 2017.
- [7] Y. Shirota, and T. Hashimoto, “Plausible deductive reasoning plan for business mathematics learners.” pp. 5-8.
- [8] Y. Shirota, and T. Hashimoto, “Learning Material Development Tool to Support Deductive Inference While Solving Business Mathematics Problems,” *Gakushuin Economic Papers*, vol. 48, no. 4, pp. 303-311, 2012.
- [9] Y. Shirota, T. Hashimoto, and T. Kuboyama, “A concept model for solving bond mathematics problems,” *Information Modelling and Knowledge Bases XXIII*, pp. 271-286: IOS Press, 2012.
- [10] 橋本隆子, and 白田由香利, “債券数学における解法プラン導出のための支援システムの提案,” *日本経営数学会誌*, vol. 33, pp. 43-56, 2012.
- [11] 白田由香利, “経営数学問題解法における演繹推論に関する考察,” *学習院大学経済論集*, vol. 49, no. 2, pp. 85-98, 2012.
- [12] 白田由香利, and 橋本隆子, “経営数学問題解決法のための演繹推論支援教材作成ツール,” *学習院大学経済論集*, vol. 48, no. 4, pp. 303-311, 2012.
- [13] Y. Shirota, T. Hashimoto, and P. Stanworth, “Knowledge Visualization of the Deductive Reasoning for Word Problems in Mathematical Economics.” pp. 117-131.
- [14] 米盛裕二, *アブダクション：仮説と発見の論理*：勁草書房, 2007.
- [15] 米盛裕二, *パースの記号学*：勁草書房, 1981.
- [16] 井上克巳, “アブダクションとインダクション (<特集> 論理に基づく推論研究の動向),” *人工知能*, vol. 25, no. 3, pp. 389-399, 2010.
- [17] 白田由香利, *悩める学生のための経済・経営数学入門：3つの解法テクニックで数学アレルギーを克服!*：共立出版, 2009.
- [18] 白田由香利, “演繹推論法によるブラック = ショールズ方程式の導出,” *学習院大学経済論集*, vol. 50, no. 3, pp. 43-55, 2013.
- [19] C. M. Bishop, and N. M. Nasrabadi, *Pattern recognition and machine learning*: Springer, 2006.
- [20] S. Theodoridis, and K. Koutroumbas, *Pattern recognition*: Elsevier, 2006.
- [21] 菊池彰, *Python データサイエンスハンドブック：Jupyter, NumPy, pandas, Matplotlib, scikit-learn を使ったデータ分析*, 機械学習, 2018.

- [22] G. Polya, *Mathematics and plausible reasoning, Volume 1: Induction and analogy in mathematics:* Princeton University Press, 1954.
- [23] G. Polya, *Mathematics and plausible reasoning, Volume 1: Induction and analogy in mathematics:* Princeton University Press, 2020.
- [24] O. Ray, “Hybrid abductive inductive learning,” Department of Computing, Imperial College London, UK, 2005.