

平成28年度 卒業論文

修学指導支援のための
ロジスティック回帰分析を用いた
要注意学生の推定

指導教員

舟橋 健司 准教授

伊藤 宏隆 助教

名古屋工業大学 工学部第二部 電気情報工学科

平成19年度入学 19213016番

名前 鈴木 博也

目次

第 1 章	はじめに	1
第 2 章	本研究で用いる手法の理論	3
2.1	主成分分析	3
2.2	ワード法	4
2.3	ロジスティック回帰分析	4
2.4	変数選択	5
2.4.1	強制投入法	5
2.4.2	ステップワイズ法	5
2.5	分類問題	5
第 3 章	データの概要と変換	7
3.1	データの概要	7
3.1.1	成績データ	7
3.1.2	打刻データ	7
3.1.3	修学データ	8
3.2	データの変換	8
3.2.1	成績データの変換	8
3.2.2	打刻・出欠データの変換	9
3.2.3	修学データの変換	10
3.3	データ変換の総括	10
第 4 章	検証実験	13
4.1	実験概要	13
4.1.1	実験環境	13
4.1.2	説明変数の選択と離散化	13
4.1.3	検証と評価	13
4.2	実験結果	14
4.2.1	ロジスティック回帰分析を用いた検証 1	15
4.2.2	ロジスティック回帰分析を用いた検証 2	23
4.2.3	ロジスティック回帰分析を用いた検証 3	38
4.3	実験結果と考察	61
第 5 章	むすび	63
	謝辞	64
	参考文献	65
	付録 A VMS の操作方法について	66

第1章 はじめに

近年の情報通信技術の発達やハードウェアの性能向上により、大量のデータが電子化され保持されている。これらのデータの中から有用な知識や傾向を見つける技術はデータマイニングと呼ばれ、商業分野や医療分野で用いられている。例えば、商業分野では顧客の購買履歴からよく売れる商品の傾向を発見することで、商品の開発や発注の指標とすることにより売上向上が図られている。また、医療分野では既存薬の効果から新薬の副作用を予測する研究や、臨床データに基づいた病気の経過予測に用いられている。

教育現場でも、このデータマイニングを活用して学生に関するデータから一人ひとりの修学傾向を読み取り、学習指導を行う指標として用いるという提案がされている。例えば、各学生の成績と授業アンケートを用いて、成績と受講態度の関係、成績と授業の進め方との関係などを調査した研究 [1] や、e-Learning の学習履歴から利用者それぞれに適した教育ナビゲーションの構築を目指す研究 [2] などがある。

名古屋工業大学 (以下、名工大) でも、早期の修学指導を目的とし、コースマネジメントシステム (以下、CMS) と IC カード出欠管理システムを 2007 年より導入している。CMS は情報技術やインターネットを使って e-Learning を支援するシステムである。例えば、課題の提出管理、小テストの実施、教材の作成支援などを Web 上で行う機能を有している。また IC カード出欠管理システムは、IC 化された学生証を講義室に設置されたカードリーダーにかざすこと (以下、打刻) で打刻情報 (ID, 打刻時間, 打刻 IC カードリーダー番号) を管理サーバに保持するシステムである。教員はこのシステムを用いることで、受講生の出席を判定することが可能となる。

ところで、多くの大学では、留年・退学者が存在し大きな問題となっている。大学生が留年・退学する理由は様々であるが、家庭の経済的貧困や学生自身の怪我や病気などのやむを得ない場合、転学や転学科などの積極的理由な場合、就職や大学院入試に失敗した学生の計画的な留年をする場合があげられる。また、大学生活に馴染めない学生や授業についていけない学生が大学に来なくなってしまう場合も存在している。このような学生の助けとなるべく、多くの大学では学生と教職員が直接話し合いを行い、学習面だけではなく生活面も含めアドバイスや相談を行うシステムが導入されている。しかしこの方法では、教職員 1 人あたりの指導量が多く、負荷となってしまう十分な指導が行えない可能性が示唆されている。

そこで、過去の関連研究 [3] では、将来、留年・退学する可能性の高い学生を「要注意学生」と定義し、過去の留年・退学者の履修した講義の成績データや出席時の打刻データを

用いて要注意学生となる学生の傾向の発見・予測を目的としている。

予測により学習指導の対象者を減らすことができ、教職員の負荷軽減と時間的コストの削減が期待できる。さらに、分析により指導の仕方の指標とすることもできる。本研究では、成績の指標となる Grade Point Average(以下, GPA) を用いて要注意学生の定義を行っている。まず、1年前期または1年後期の GPA が 1.0 未満の学生は指導が必要であることは明白であるため、分析対象から除外したうえで指導対象とする。また、転学や計画的留年などの理由で留年・退学した学生は指導をあまり必要としていないと考えられる。さらに、退学理由を見ていくと、「経済的理由」によって退学した学生も修学指導が必要でないと考える。このような学生を予測に含めてしまうと本当に指導が必要な学生を見つける妨げになってしまう可能性があるため要注意学生に含めていない。これらを踏まえた結果、要注意学生の定義を「1年前期または1年後期の GPA が 1.0 以上であり、かつ今後消極的理由により留年・退学してしまう学生」とする。ここで定義した要注意学生を対象学生の中から発見することを目標として様々な研究 [4][5][6] が行われている。

本研究では、さらなる推定精度の向上を目指し、ロジスティック回帰分析を用いている。この分析方法は、ある目的事象が起こる確率をもとめる手法で、ダイレクトマーケティングや医療分野である事象に対するリスクファクターの探索に用いられている。教育分野でも、カリフォルニア大学ロサンゼルス校の高等教育研究所 (HERI) では、ロジスティック回帰分析を用いた、「卒業確率計算ソフト」の開発 [7] が行われている。このソフトは、年齢・性別・SAT または ACT の点数・人種や民族・高校時代の成績の 5 項目を入力することで、過去のデータから作成したモデル式を利用し、卒業確率を計算できるものである。このソフトでは学生自身が気軽に利用できるように 5 項目と少ない入力で予測を行っているが、本研究では、精度の高い予測を期待して、より学業に関係が深いと思われる成績データや出欠席などを用いて分析を行った。本研究の結果、要注意学生を推定するモデル式を生成することができ、要注意学生の傾向を数値で表すことできた。

本論文では、まず第 2 章でデータマイニングによる知識発見法について、第 3 章では分析に用いるデータの概要とその変換について、第 4 章でロジスティック回帰分析を用いた分析・予測を行い、検証を行っている。最後に第 5 章では本研究のまとめと今後の課題や展望を述べる。

ちなみに本研究で用いたデータは、個人を特定できる情報(氏名や学籍番号)は含まれておらず、本研究によって個人情報侵害されることはないことをここに付記する。

第2章 本研究で用いる手法の理論

本研究では、分析及び予測の手法を多く用いている。その多くはデータマイニングの知識発見の手法に含まれている。本章ではそれらについて説明する。

2.1 主成分分析

主成分分析とは、複数の変数の特徴を合成し、新しい変数を生成する手法である。「新しい変数」は主成分と呼ばれ、生成された各主成分に対し解釈を与えることで新しい意味を持つ。ここでは主成分分析の概要を説明するために、変数 x_1, x_2, \dots, x_n があるとして、新しい変数 z_1 を導出することを考える。

$$z_1 = a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (2.1)$$

このとき、各係数のベクトル a_1, a_2, \dots, a_n を z_1 の分散が最大となるように変化させる。ただしベクトル a の大きさを1とする必要がある。

$$\sum_{i=1}^n a_i^2 = 1 \quad (2.2)$$

分散が最大となるとき、この z_1 を第1主成分とする。同様に z_2 を以下の式のように定める。

$$z_2 = b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2.3)$$

このとき、各係数のベクトル b_1, b_2, \dots, b_n を z_2 の分散が z_1 の次に最大となるように変化させる。ただしベクトル b の大きさを1に、かつベクトル a とベクトル b が垂直となる必要がある。

$$\sum_{i=1}^n b_i^2 = 1 \quad (2.4)$$

$$\sum_{i=1}^n a_i b_i = 0 \quad (2.5)$$

この z_2 を第2主成分とする。この計算を繰り返し行い、主成分を作成する。主成分分析は変数の縮小を目的として行っている。そのため主成分の数を利用者で定める必要がある。ここでは一般的な決定指標として、寄与率と累計寄与率を用いる。寄与率とは、データ全体

における, ある主成分がどの程度意味を持つかを表している. ある主成分の固有値を λ_α , 各変数の分散を σ_i としたときの寄与率 C_α は以下の式で求められる.

$$C_\alpha = \frac{\lambda_\alpha}{\sum_{i=1}^n \sigma_i} \quad (2.6)$$

また, 累積寄与率 P は第1主成分からある主成分までの寄与率を合算することで求められる.

$$P = \sum_{i=1}^n C_i \quad (2.7)$$

一般的に, 累積寄与率が 60%~80% になるまで主成分を選択する. 本研究では, 累計寄与率が 70% を超えるまで主成分を選択していくものとする.

2.2 ウォード法

ウォード法とは, あるクラスタ群を結合するとき, 統合されるクラスタ内の平方和が最小となるようにクラスタを形成していく方法である. ここはあるクラスタ C_1 と C_2 を結合するとする. なお $E(C)$ は C の分散を表すとする.

$$E_{C_1, C_2} = E(C_1 \cup C_2) - (E(C_1) + E(C_2)) \quad (2.8)$$

このクラスタ内の平方和 E_{C_1, C_2} が最小になるように結合していく. ウォード法は, 外れ値に強く, 実用性が高い.

2.3 ロジスティック回帰分析

まず, 回帰分析とは, ある変数が他の変数と, どのような相関関係にあるかを推定する手法である. 一般的には, 結果となる変数 y (目的変数) と, それを予測するための変数 x (説明変数) を以下の式で表される. a は, b は最小二乗法を用いて決定する.

$$y = a * x + b \quad (2.9)$$

ロジスティック回帰分析は, 目的変数 y が 0 から 1 の間の値となる場合に用いられる回帰分析である. これは, ある事象が起こる確率を求めることができるということを意味している. 回帰式は, 以下の式で表される.

$$y = \frac{1}{1 + e^l} \quad (2.10)$$

$$l = a + \sum_{i=1}^n (b_i * x_i) \quad (2.11)$$

a を定数項, b を回帰係数と呼び, 最尤法を用いた繰り返しの近似計算によって決定する.

2.4 変数選択

2.4.1 強制投入法

強制投入法とは, 全ての説明変数を一度に投入する方法である. 目的変数の予測における説明変数のそれぞれの寄与がどの程度であるか調べるために利用されることが多い.

2.4.2 ステップワイズ法

ステップワイズ法とは, 統計的データに基づいた最も予測率が高いと考えられる変数から順にモデルに投入する方法である. 適合度が最も高くなる回帰モデルを調べるために利用されることが多い.

2.5 分類問題

本項では2値の分類問題を考える. 2値分類問題は, 入力があるクラスに属している「正例クラス」と, あるクラスに属していない「負例クラス」に分類する問題である. 2×2 の分割表を用いて次のように表せる.

表 2.1: 2×2 分割表

	予測の正例	予測の負例
実際の正例	TP	FN
実際の負例	FP	TN

表の TP, FP, FN, TN の意味は以下の通りである.

TP : True Positive の略. 実際に正例クラスであるデータを正例と予測した件数.

FP : False Positive の略. 実際は負例クラスであるデータを正例と予測した件数.

FN : False Negative の略. 実際に正例クラスであるデータを正例と予測した件数.

TN : True Negative の略. 実際は負例クラスであるデータを正例と予測した件数.

本研究では、これら4つの値を用いて、評価指標として一般的に用いられている、正解率、再現率、適合率、F-measure(以下、F値)を求める。それぞれの計算法は以下の式の通りである。

$$\text{正解率} = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.12)$$

$$\text{再現率} = \frac{TP}{TP + FP} \quad (2.13)$$

$$\text{適合率} = \frac{TP}{TP + FN} \quad (2.14)$$

$$F \text{ 値} = \frac{2 * \text{再現率} * \text{適合率}}{\text{再現率} + \text{適合率}} \quad (2.15)$$

正解率は、分類の予測が的中した割合を表している。再現率は、実際の正例クラスを正例クラスと予測できた割合を表している。適合率は、正例クラスと予想した内の実際の正例クラスの割合を表している。F値は、再現率と適合率の相対的な重要度を表している。F値は再現率と適合率の調和平均値であり、値が大きいほど予測モデルが優れていることを示している。

第3章 データの概要と変換

本章では、本研究に用いるデータの概要とその変換について述べる。データは成績データ、打刻データ、修学データの3種類からなり、これらのデータを分析に利用するため、変換処理を行う。

3.1 データの概要

本研究では、ある年度 A, B に名工大に入学した学生 338 名 (A 年度 171 名, B 年度 167 名) に関するデータを用いる。先述したようにデータは成績データ、打刻データ、修学データの3種類であり、成績データと打刻データは各年度の2年分のデータである。この節では、これら3種類のデータについて述べる。

3.1.1 成績データ

学生の講義別成績データは、暗号化された学籍番号、授業の成績、授業科目名、授業開講期の4つの情報をレコード形式で保存されている。授業の成績は、秀・優・良・可・不可・失格の6つの評価がある。秀が最も優れている評価でそこから順番に成績評価が低くなり、不可・失格は単位が取得できなかったことを表している。不可と失格の違いは、評価可否にあり、試験を欠席した場合や、出席回数が不足していた場合は評価不能として失格判定される。授業科目名はシラバスに載っている授業名ではなく、「専門1」、「演習1」のように改名されており、学生の受講した授業の特定を防ぎ、所属する学科が分からないようにするための処置である。授業開講期は、授業の開かれた年次と時期であり、時期は前期(4月-9月)と後期(10月-12月)を指す。

3.1.2 打刻データ

打刻データは、第1章で述べたICカード出欠管理システムによって保存された情報である。内容は、暗号化された学生番号、打刻した日付(年月日)、打刻時間の3つの情報が記録されている。この打刻データから自動的に生成された「出欠データ」を今回は用いている。これは打刻データに加え、暗号化学生番号を元に、暗号化授業番号、入室打刻、退室打刻、出欠判定の合計6つのデータを持つ。出欠判定は、入室時刻と退室時刻から、出席、遅刻、早退、欠席の4つに分類されている。しかし、授業の開始が遅れたり、授業の終了が早まった

場合は正しい出欠判定ができない。この問題の対策は後に説明する「打刻・出欠データの変換」で行っている。ここで、暗号化学生番号と暗号化授業番号は成績データで用いられているものと共通であることを明記しておく。

3.1.3 修学データ

修学データは、本研究の対象である 338 名の学生の、卒業研究に着手した年次、卒業までにかかった年数、退学した学生の退学年次、退学した学生の退学理由が記録されている。名工大では、4 年次に研究室配属が行われる。しかし、3 年次終了までに定められた単位を取得しなければ卒業研究に着手することができない。また、卒業研究に着手できたものであっても、所定の単位を取得しなければ卒業することができない。退学した学生の退学理由は、一身上、就職、転学科、他大学受験、授業料未納、経済的理由、不明の 7 つに分類されている。これら修学データは、要注意学生の定義および判定に利用している。

3.2 データの変換

本研究では、前節で説明した成績データ、打刻・出欠データ、修学データを用いて分析を行う。しかし記録されたままのデータ形式では分析に用いることが困難なため、データの変換を行った。

3.2.1 成績データの変換

成績データは、講義区分別 GPA、獲得成績数、必修科目不合格数、クラス内偏差値別講義数の 4 つに変換した。それぞれの説明は次に示す。

講義区分別 GPA

成績の指標として GPA を用いる。GPA は各科目の成績から以下の式で算出された成績の評価値である。履修登録した講義 n の集合を N とする。 $GP(n)$ は成績ごとの得点であり、秀・優・良・可・不可・失格にそれぞれ 4 点・3 点・2 点・1 点・0 点・0 点と定める。 c_n は講義ごとに定められた単位数を表す。

$$GPA = \frac{\sum_n^N GP(n) * c_n}{\sum_n^N c_n} \quad (3.1)$$

この GPA を、大学が定めている講義区分ごとに計算し、変数を生成する。名工大では、表 3.1 の 6 つの区分に分けられている。

表 3.1: 講義区分

区分名	意味
理系基礎	自然科学基礎に分類される科目区分
外国語	グローバルコミュニケーションに分類される科目区分
人間社会	人間社会に分類される科目区分
体育	健康運動科学に分類される科目区分
専門	専門教育に分類される科目区分
その他	産業・経営リテラシー, または上記区分以外に分類される科目区分

獲得成績数

各成績評価 (秀・優・良・可・不可・失格) ごとの獲得数を表す。例えば, ある2つの講義を受講したある学生の GPA が 2.0 であった場合, 獲得した評価が, 良と良であるのか, 秀と不合格であるかといった判定ができない。そこで評価別の変数を次の獲得成績数として生成することでより細かい傾向を発見できると期待している。

必修科目不合格数

名工大の講義には, 卒業研究着手や卒業するために必ず単位取得しなければならない必修科目と, 合計単位数には反映されるが必ずしも単位取得しなくてもいい選択科目が存在する。選択科目で単位を取得できなかった場合より, 必修科目で単位を取得できなかった場合のほうが留年しやすいと予想を立て, 必修科目を履修し, 不可または失格の判定がされた数を変数として用いる。

クラス内偏差値別講義数

GPA と獲得成績数は絶対的な値であるため, 講義の単位取得難易度が考慮されていない。そこで, 授業クラス内の偏差値を用いることで講義難易度によらない成績を評価する変数が生成できる。本研究では, 偏差値を3段階に分け, それぞれの履修講義数を変数とした。3段階の内訳は, クラス内偏差値が40未満, クラス内偏差値が40以上60未満, クラス内偏差値が60以上となっている。

3.2.2 打刻・出欠データの変換

打刻・出欠データは授業成績と関連付けた変数に変換する。これは, 講義によって出席をとらない教員がいたり, ICカード出欠管理システムのカードリーダーがない講義室を利用している場合があり, 出席数だけでは十分ではないためである。変数は, クラス内における出席数の偏差値 $S_{n,att}$ から正規化する。具体的には, $S_{n,att} < 40$ となる履修講義数を変

数としている。ただし、名工大の1科目あたりの講義数は15回であるため、この変数の標準偏差 $S_{n,att}$ が小さくなる。よって、 $S_{n,att} < 1$ となるデータは除外した。さらに、本来出席回数が既定の値に満たない場合は失格の評価が与えられるが、出席数が少ないにも関わらず、失格以外の評価が与えられている場合がある。この場合、出席数と授業評価の正しい関連付けが行いえないため、このデータは除外した。以上の条件を踏まえた変数を、「出席数の偏差値が40未満、標準偏差が1以上、失格評価を受けていない履修講義数」と定義し、以下に式で表す。なお成績の評価を e_n とし、失格の評価値を X と表す。

$$S_{att} = \{n | (S_{n,att} < 40) \wedge (1 \leq \sigma_n) \wedge (e_n \neq X)\} \quad (3.2)$$

3.2.3 修学データの変換

修学データは要注意学生の定義に用いた。対象学生は338名の内、留年・退学した学生は70名であった。従来研究 [6] で定義した要注意学生は以下である。

要注意学生の定義

1年次前期・1年次後期におけるGPAがともに1.0以上であるにも関わらず、将来的に留年・退学する学生。ただし、3年で卒業研究に着手した学生は対象外、在学年数が3年に満たない退学者と、経済的理由による退学者のデータは除外する。

これを用いることで、調査対象が338名から302名に、要注意学生を25名まで絞り込むことができた。

3.3 データ変換の総括

本節では、これまでの変換したデータを、入力である説明変数と、出力である目的変数に分けて、それぞれ表3.2と表3.3に示す。

表 3.2: 説明変数

番号	変数名	意味
1	1 年次前期理系基礎 GPA	1 年次前期における理系基礎科目の GPA
2	1 年次前期外国語 GPA	1 年次前期における外国語科目の GPA
3	1 年次前期人間社会 GPA	1 年次前期における人間社会科目の GPA
4	1 年次前期体育 GPA	1 年次前期における体育科目の GPA
5	1 年次前期専門 GPA	1 年次前期における専門科目の GPA
6	1 年次前期その他 GPA	1 年次前期におけるその他科目の GPA
7	1 年次前期秀獲得数	1 年次前期における秀評価の数
8	1 年次前期優獲得数	1 年次前期における優評価の数
9	1 年次前期良獲得数	1 年次前期における良評価の数
10	1 年次前期可獲得数	1 年次前期における可評価の数
11	1 年次前期不可獲得数	1 年次前期における不可評価の数
12	1 年次前期失格獲得数	1 年次前期における失格評価の数
13	1 年次前期必修科目不合格	1 年次前期の必修科目で不可・失格評価を受けた数
14	1 年次前期#(S_{low})	1 年次前期においてクラス内偏差値が 40 未満の科目数
15	1 年次前期#(S_{mid})	1 年次前期においてクラス内偏差値が 40 以上 60 未満の科目数
16	1 年次前期#(S_{high})	1 年次前期においてクラス内偏差値が 60 以上の科目数
17	1 年次前期#(S_{att})	1 年次前期に $(S_{n,att} < 40) \wedge (1 \leq \sigma_n) \wedge (e_n \neq X)$ となるの科目数
18	1 年次後期理系基礎 GPA	1 年次後期における理系基礎科目の GPA
19	1 年次後期外国語 GPA	1 年次後期における外国語科目の GPA
20	1 年次後期人間社会 GPA	1 年次後期における人間社会科目の GPA
21	1 年次後期体育 GPA	1 年次後期における体育科目の GPA
22	1 年次後期専門 GPA	1 年次後期における専門科目の GPA
23	1 年次後期その他 GPA	1 年次後期におけるその他科目の GPA
24	1 年次後期秀獲得数	1 年次後期における秀評価の数
25	1 年次後期優獲得数	1 年次後期における優評価の数
26	1 年次後期良獲得数	1 年次後期における良評価の数
27	1 年次後期可獲得数	1 年次後期における可評価の数
28	1 年次後期不可獲得数	1 年次後期における不可評価の数
29	1 年次後期失格獲得数	1 年次後期における失格評価の数
30	1 年次後期必修科目不合格	1 年次後期の必修科目で不可・失格評価を受けた数
31	1 年次後期#(S_{low})	1 年次後期においてクラス内偏差値が 40 未満の科目数
32	1 年次後期#(S_{mid})	1 年次後期においてクラス内偏差値が 40 以上 60 未満の科目数
33	1 年次後期#(S_{high})	1 年次後期においてクラス内偏差値が 60 以上の科目数
34	1 年次後期#(S_{att})	1 年次後期に $(S_{n,att} < 40) \wedge (1 \leq \sigma_n) \wedge (e_n \neq X)$ となるの科目数

表 3.2 の続き

番号	変数名	意味
35	2 年次前期理系基礎 GPA	2 年次前期における理系基礎科目の GPA
36	2 年次前期外国語 GPA	2 年次前期における外国語科目の GPA
37	2 年次前期人間社会 GPA	2 年次前期における人間社会科目の GPA
38	2 年次前期体育 GPA	2 年次前期における体育科目の GPA
39	2 年次前期専門 GPA	2 年次前期における専門科目の GPA
40	2 年次前期その他 GPA	2 年次前期におけるその他科目の GPA
41	2 年次前期秀獲得数	2 年次前期における秀評価の数
42	2 年次前期優獲得数	2 年次前期における優評価の数
43	2 年次前期良獲得数	2 年次前期における良評価の数
44	2 年次前期可獲得数	2 年次前期における可評価の数
45	2 年次前期不可獲得数	2 年次前期における不可評価の数
46	2 年次前期失格獲得数	2 年次前期における失格評価の数
47	2 年次前期必修科目不合格	2 年次前期の必修科目で不可・失格評価を受けた数
48	2 年次前期#(S_{low})	2 年次前期においてクラス内偏差値が 40 未満の科目数
49	2 年次前期#(S_{mid})	2 年次前期においてクラス内偏差値が 40 以上 60 未満の科目数
50	2 年次前期#(S_{high})	2 年次前期においてクラス内偏差値が 60 以上の科目数
51	2 年次前期#(S_{att})	2 年次前期に $(S_{n,att} < 40) \wedge (1 \leq \sigma_n) \wedge (e_n \neq X)$ となるの科目数
52	2 年次後期理系基礎 GPA	2 年次後期における理系基礎科目の GPA
53	2 年次後期外国語 GPA	2 年次後期における外国語科目の GPA
54	2 年次後期人間社会 GPA	2 年次後期における人間社会科目の GPA
55	2 年次後期体育 GPA	2 年次後期における体育科目の GPA
56	2 年次後期専門 GPA	2 年次後期における専門科目の GPA
57	2 年次後期その他 GPA	2 年次後期におけるその他科目の GPA
58	2 年次後期秀獲得数	2 年次後期における秀評価の数
59	2 年次後期優獲得数	2 年次後期における優評価の数
60	2 年次後期良獲得数	2 年次後期における良評価の数
61	2 年次後期可獲得数	2 年次後期における可評価の数
62	2 年次後期不可獲得数	2 年次後期における不可評価の数
63	2 年次後期失格獲得数	2 年次後期における失格評価の数
64	2 年次後期必修科目不合格	2 年次後期の必修科目で不可・失格評価を受けた数
65	2 年次後期#(S_{low})	2 年次後期においてクラス内偏差値が 40 未満の科目数
66	2 年次後期#(S_{mid})	2 年次後期においてクラス内偏差値が 40 以上 60 未満の科目数
67	2 年次後期#(S_{high})	2 年次後期においてクラス内偏差値が 60 以上の科目数
68	2 年次後期#(S_{att})	2 年次後期に $(S_{n,att} < 40) \wedge (1 \leq \sigma_n) \wedge (e_n \neq X)$ となるの科目数

表 3.3: 目的変数

変数名	意味
要注意学生	要注意学生かどうか (True=1, False=0)

第4章 検証実験

本章では、前章で生成した変数を入力とするロジスティック回帰分析を行う。この説明変数はユーザー指定及びステップワイズ法で選択し、ワード法を用いてクラス多数が4になるように離散化する。モデルの評価は、2値分類問題における正例クラスを要注意学生とし、正解率、再現率、適合率、F値を計算することにより行う。

4.1 実験概要

本節では実験の概要について述べる。

4.1.1 実験環境

変数生成はレコード形式のデータを扱いやすい点から Microsoft 社の Excel2010 [8] を、変数の離散化及び検証には NTT データ数理システム社の Visual Mining Studio [9] をそれぞれ利用した。

4.1.2 説明変数の選択と離散化

モデルの精度を決定するのは説明変数であるため、どの説明変数を用いてモデルを作成するかが重要となる。本研究の目的である要注意学生の発見は早期であることが望ましいが、早期の段階では得られるデータが少なく、推定精度は低くなってしまふ。逆に説明変数が多いほど推定精度が高くなるとも限らず、ノイズとなる可能性もある。そこで、本研究では、強制投入法とステップワイズ法を用いて変数選択を行った。また、用いる変数は表 3.2 の通りである。ロジスティック回帰分析では説明変数は離散型である必要があるため、ワード法を用いて各説明変数をクラス数4となるように離散化している。

4.1.3 検証と評価

実験結果の評価に2値分類問題を用いるために、正例クラスと負例クラスの分類をしなければならない。正例クラスは要注意学生であること、負例クラスは要注意学生ではないこと、と定めた。ロジスティック回帰分析では出力が、ある事象の起こる確率であるため、ある一定の閾値を決め、その確率を超えたデータを正例クラスと定めることにする。本実験では、50%、30%、事前確率(対象学生のうち、留年・退学者の割合)の3つの閾値を用意

した。2値分類問題を解くことで、正解率、再現率、適合率、F値を求めることができる。要注意学生をなるべく多く予測できることはもちろんだが、指導対象となる学生を増やしすぎないことが理想である。よって、再現率と適合率の調和平均であるF値が高いモデルを良いモデルと位置付けた。

4.2 実験結果

本節では、3つの方法を用いて実験を行った。1年前期、1年後期、2年前期、2年後期の4つの時期までのデータを用いて、時期ごとの推定精度を調べた。それぞれ、強制投入法とステップワイズ法の2種類の変数選択を行った。いずれも目的変数は要注意学生になると確率である。すべてのパターンにおいて事前確率 pp は、

$$\begin{aligned} pp &= \frac{(\text{実際の要注意学生数})}{(\text{対象となる学生数})} \times 100 \\ &= \frac{25}{302} \times 100 \approx 8.3(\%) \end{aligned}$$

を用いた

4.2.1 ロジスティック回帰分析を用いた検証1

説明変数は、その時期までのデータとした。

1年次前期までのデータを用いた推定

表4.1で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.1: 説明変数と回帰係数 (1年次前期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	-24.0	-40.0
1年次前期理系基礎 GPA	-3.72	不採用
1年次前期外国語 GPA	-1.30	不採用
1年次前期人間社会 GPA	-0.530	不採用
1年次前期体育 GPA	-0.580	不採用
1年次前期専門 GPA	-3.24	不採用
1年次前期その他 GPA	0	不採用
1年次前期秀獲得数	2.47	不採用
1年次前期優獲得数	1.70	不採用
1年次前期良獲得数	0.939	不採用
1年次前期可獲得数	0.266	不採用
1年次前期不可獲得数	0.307	0.411
1年次前期失格獲得数	-0.340	不採用
1年次前期必修科目不合格	-0.033	不採用
1年次前期#(S_{low})	2.04	2.84
1年次前期#(S_{mid})	1.83	2.64
1年次前期#(S_{high})	1.47	2.36
1年次前期#(S_{att})	-0.043	不採用

表 4.2: 予測精度 (1年次前期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	273	90.3%	25	1	4%	6	1	16.6%	0.032
	30%	302	270	89.4%	25	3	12%	13	3	23%	0.078
	8.3%	302	237	78.4%	25	7	28%	54	7	12.9%	0.088
ステップワイズ法	50%	302	273	90.3%	25	1	4%	6	1	16.6%	0.032
	30%	302	271	89.7%	25	3	12%	12	3	25%	0.081
	8.3%	302	238	78.8%	25	9	36%	57	9	15.7%	0.109

1 年次後期までのデータを用いた推定

表 4.3 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.3: 説明変数と回帰係数 (1 年次後期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	3.77	2.37
1 年次前期理系基礎 GPA	-52.9	-48.3
1 年次前期外国語 GPA	-20.1	-18.3
1 年次前期人間社会 GPA	-5.94	-4.83
1 年次前期体育 GPA	-12.2	-11.7
1 年次前期専門 GPA	-44.9	-39.8
1 年次前期その他 GPA	0	不採用
1 年次前期秀獲得数	34.4	32.5
1 年次前期優獲得数	23.7	22.1
1 年次前期良獲得数	9.08	8.8
1 年次前期可獲得数	0.423	1.23
1 年次前期不可獲得数	1.79	2.24
1 年次前期失格獲得数	-10.9	-9.63
1 年次前期必修科目不合格	-9.1	-7.39
1 年次前期#(S_{low})	22	21.7
1 年次前期#(S_{mid})	21	21
1 年次前期#(S_{high})	15.5	14.6
1 年次前期#(S_{att})	0.529	-0.227
1 年次後期理系基礎 GPA	17.2	6.17
1 年次後期外国語 GPA	5.77	2.05
1 年次後期人間社会 GPA	1.63	0.358
1 年次後期体育 GPA	2.98	2.05
1 年次後期専門 GPA	7.45	不採用
1 年次後期その他 GPA	6.04	3.99
1 年次後期秀獲得数	-15.9	-22.8
1 年次後期優獲得数	-9.9	-18.2
1 年次後期良獲得数	-7.11	-17.1
1 年次後期可獲得数	-3.32	-15.5
1 年次後期不可獲得数	0.689	-12.1
1 年次後期失格獲得数	0.326	-13.6
1 年次後期必修科目不合格	1.01	不採用
1 年次後期#(S_{low})	-12.6	0.0156
1 年次後期#(S_{mid})	-12.3	不採用
1 年次後期#(S_{high})	-10.3	2.59
1 年次後期#(S_{att})	-0.397	不採用

表 4.4: 予測精度 (1 年次後期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制 投入法	50%	302	255	84.4%	25	7	28%	36	7	19.4%	0.114
	30%	302	253	83.7%	25	7	28%	38	7	18.4%	0.111
	8.3%	302	250	82.7%	25	7	28%	41	7	17%	0.105
ステップ ワイズ法	50%	302	255	84.4%	25	7	28%	36	7	19.4%	0.114
	30%	302	254	84.1%	25	7	28%	37	7	18.9%	0.112
	8.3%	302	249	82.4%	25	7	28%	42	7	16.6%	0.104

2年次前期までのデータを用いた推定

表 4.5 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.5: 説明変数と回帰係数 (2年次前期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	-17.1	-11.1
1年次前期理系基礎 GPA	-58.4	-33.6
1年次前期外国語 GPA	-15.1	-9.04
1年次前期人間社会 GPA	-5.16	-3.11
1年次前期体育 GPA	-12.9	-7.63
1年次前期専門 GPA	-34.5	-20.2
1年次前期その他 GPA	0	不採用
1年次前期秀獲得数	34.8	20.9
1年次前期優獲得数	23.9	14.6
1年次前期良獲得数	16.1	9.79
1年次前期可獲得数	4.5	2.64
1年次前期不可獲得数	10.1	5.55
1年次前期失格獲得数	-19.7	-10.5
1年次前期必修科目不合格	-9.62	-4.91
1年次前期#(S_{low})	25.6	16.1
1年次前期#(S_{mid})	22.1	13.7
1年次前期#(S_{high})	22.1	13.3
1年次前期#(S_{att})	-2.17	-1.36
1年次後期理系基礎 GPA	-25.6	-17.6
1年次後期外国語 GPA	-18.9	-13
1年次後期人間社会 GPA	-8.3	-5.98
1年次後期体育 GPA	-3.43	-2.83
1年次後期専門 GPA	-38.1	-27.1
1年次後期その他 GPA	-6.33	-3.33
1年次後期秀獲得数	10.6	7.57
1年次後期優獲得数	6.22	4.37
1年次後期良獲得数	-0.237	0.314
1年次後期可獲得数	-7.79	-5.17
1年次後期不可獲得数	-5.64	-3.52
1年次後期失格獲得数	-1.67	-1.16
1年次後期必修科目不合格	-7.31	-4.68
1年次後期#(S_{low})	1.64	1.38
1年次後期#(S_{mid})	0.707	0.705
1年次後期#(S_{high})	-0.907	0.316
1年次後期#(S_{att})	-0.727	-0.824

表 4.5 の続き

変数名	回帰係数	
	強制投入法	ステップワイズ法
2年次前期理系基礎 GPA	-12.4	-8.72
2年次前期外国語 GPA	1.13	0.446
2年次前期人間社会 GPA	-3.85	-3.39
2年次前期体育 GPA	0.659	-1.02
2年次前期専門 GPA	-13.3	-10.1
2年次前期その他 GPA	0	不採用
2年次前期秀獲得数	1.27	-1.47
2年次前期優獲得数	-3.27	-5.01
2年次前期良獲得数	-10.1	-9.94
2年次前期可獲得数	-9.83	-9.34
2年次前期不可獲得数	11.1	-1.85
2年次前期失格獲得数	9.08	不採用
2年次前期必修科目不合格	-6.94	-0.0355
2年次前期#(S_{low})	-4.8	7.18
2年次前期#(S_{mid})	5.49	不採用
2年次前期#(S_{high})	-2.48	1.86
2年次前期#(S_{att})	3.33	2.11

表 4.6: 予測精度 (2年次前期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	267	88.4%	25	10	40%	30	10	33.3%	0.181
	30%	302	266	88%	25	10	40%	31	10	32.2%	0.178
	8.3%	302	266	88%	25	10	40%	31	9	29%	0.168
ステップワイズ法	50%	302	264	87.4%	25	9	36%	31	9	29%	0.16
	30%	302	262	86.7%	25	9	36%	34	9	26.4%	0.152
	8.3%	302	261	86.4%	25	9	36%	34	9	26.4%	0.152

2 年次後期までのデータを用いた推定

表 4.7 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.7: 説明変数と回帰係数 (2 年次後期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	1.51	-6.62
1 年次前期理系基礎 GPA	0.824	-56.1
1 年次前期外国語 GPA	3.72	-16.5
1 年次前期人間社会 GPA	5.14	-9.99
1 年次前期体育 GPA	-5.09	-10.4
1 年次前期専門 GPA	14.2	-29
1 年次前期その他 GPA	0	不採用
1 年次前期秀獲得数	7.37	37.3
1 年次前期優獲得数	6.67	28.6
1 年次前期良獲得数	4.86	16.9
1 年次前期可獲得数	1.19	-0.0599
1 年次前期不可獲得数	0.749	20.4
1 年次前期失格獲得数	0.32	-26.2
1 年次前期必修科目不合格	1.07	-5.8
1 年次前期#(S_{low})	12	30.6
1 年次前期#(S_{mid})	6.59	28.6
1 年次前期#(S_{high})	2.56	17.7
1 年次前期#(S_{att})	2	3.93
1 年次後期理系基礎 GPA	4.57	-39
1 年次後期外国語 GPA	-2.06	-21.9
1 年次後期人間社会 GPA	-1.31	-8.95
1 年次後期体育 GPA	8.93	1.19
1 年次後期専門 GPA	-9.89	-43.3
1 年次後期その他 GPA	-1.57	-9.73
1 年次後期秀獲得数	-6.26	13.5
1 年次後期優獲得数	-6.18	3.78
1 年次後期良獲得数	-7.16	-8.35
1 年次後期可獲得数	-5.16	-19.9
1 年次後期不可獲得数	-4.03	-25.3
1 年次後期失格獲得数	2.68	-15.8
1 年次後期必修科目不合格	-1.35	不採用
1 年次後期#(S_{low})	-7.07	9.09
1 年次後期#(S_{mid})	-6.57	1.86
1 年次後期#(S_{high})	-12.5	不採用
1 年次後期#(S_{att})	-0.68	-0.425

表 4.7 の続き

変数名	回帰係数	
	強制投入法	ステップワイズ法
2年次前期理系基礎 GPA	-2.69	0.956
2年次前期外国語 GPA	-3.86	3.59
2年次前期人間社会 GPA	0.12	1.63
2年次前期体育 GPA	0	不採用
2年次前期専門 GPA	-4.23	-4.29
2年次前期その他 GPA	0	不採用
2年次前期秀獲得数	-0.0986	-2.02
2年次前期優獲得数	0.777	-10.3
2年次前期良獲得数	-5.58	-14.3
2年次前期可獲得数	-3.71	-12.3
2年次前期不可獲得数	-2.21	28.3
2年次前期失格獲得数	4.84	21.3
2年次前期必修科目不合格	2.8	-22
2年次前期#(S_{low})	-5.14	-3.64
2年次前期#(S_{mid})	-0.243	4.71
2年次前期#(S_{high})	-0.6	不採用
2年次前期#(S_{att})	0.809	-1.83
2年次後期理系基礎 GPA	1.39	-3.21
2年次後期外国語 GPA	5.99	-4.15
2年次後期人間社会 GPA	-2.14	-0.372
2年次後期体育 GPA	0	不採用
2年次後期専門 GPA	-11.3	-24.2
2年次後期その他 GPA	0	不採用
2年次後期秀獲得数	2.41	2.95
2年次後期優獲得数	1.69	-0.891
2年次後期良獲得数	1.55	1.81
2年次後期可獲得数	2.57	1.43
2年次後期不可獲得数	-1.33	-12.4
2年次後期失格獲得数	0.99	3.84
2年次後期必修科目不合格	-3.41	7.38
2年次後期#(S_{low})	5.73	5.38
2年次後期#(S_{mid})	-1.95	-3.9
2年次後期#(S_{high})	4.1	不採用
2年次後期#(S_{att})	2.89	0.69

表 4.8: 予測精度 (2 年次後期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制 投入法	50%	302	268	88.7%	25	11	44%	31	11	35.4%	0.196
	30%	302	269	89%	25	12	48%	32	12	37.5%	0.21
	8.3%	302	264	87.4%	25	12	48%	37	12	32.4%	0.193
ステップ ワイズ法	50%	302	263	87%	25	12	48%	38	12	31.5%	0.19
	30%	302	259	85.7%	25	12	48%	42	12	28.5%	0.178
	8.3%	302	253	83.7%	25	12	48%	48	12	28.5%	0.178

考察

時期が遅くなり、説明変数が増えるほどの的中数も F 値も向上している。ステップワイズ法による変数選択を見てみると、1 年次前期は 17 変数中 4 変数しか選択されていない。しかし、強制投入法と同程度の結果が出ており、有意な変数選択ができていると考えられる。その他の時期は、選択された変数が多いため結果に大きな違いは見られなかった。また、1 年次前期以外は閾値による的中数の変化がほとんど見られなかった。この結果から推定確率が極端であったことがわかる。

本節では、変換した学生データを変数選択し分析を行ったが、説明変数が多いことが必ずしも良い結果が出るとは限らない。次節では、変換した学生データに対して主成分分析を行い、説明変数を減らして検証を行ってみた。

4.2.2 ロジスティック回帰分析を用いた検証2

説明変数を, その学期までのデータを主成分分析したものとした.

1 年次前期までのデータを用いた推定

表 3.2 の変数 1-17 に対して主成分分析を行い, 表 4.9 の結果が得られた. 累計寄与率が 70% を超えた第 3 主成分までを説明変数として用いた. 各主成分における固有ベクトルを解釈することで, 主成分の意味付けを行った. 固有ベクトルを表 4.10, 意味付けを表 4.11 に示す.

表 4.9: 各主成分の寄与率・累計寄与率 (1 年次前期終了時点)

主成分番号	寄与率	累計寄与率
第 1 主成分	0.355	0.355
第 2 主成分	0.283	0.638
第 3 主成分	0.120	0.758
⋮	⋮	⋮
第 17 主成分	0.000	1.000

表 4.10: 各主成分の固有ベクトル (1 年次前期終了時点)

変数名	主成分番号		
	1	2	3
1 年次前期理系基礎 GPA	-0.175	0.022	0.028
1 年次前期外国語 GPA	-0.073	0.015	-0.037
1 年次前期人間社会 GPA	-0.129	0.023	0.001
1 年次前期体育 GPA	-0.041	0.030	0.029
1 年次前期専門 GPA	-0.147	0.053	-0.025
1 年次前期その他 GPA	0.000	0.000	0.000
1 年次前期秀獲得数	-0.461	0.479	0.157
1 年次前期優獲得数	-0.281	-0.347	-0.635
1 年次前期良獲得数	0.199	-0.214	0.671
1 年次前期可獲得数	0.326	-0.036	-0.046
1 年次前期不可獲得数	0.188	0.098	-0.125
1 年次前期失格獲得数	0.032	0.026	-0.022
1 年次前期必修科目不合格	0.220	0.123	-0.147
1 年次前期#(S_{low})	0.510	0.202	-0.219
1 年次前期#(S_{mid})	-0.245	-0.604	0.165
1 年次前期#(S_{high})	-0.263	0.407	0.055
1 年次前期#(S_{att})	0.111	0.043	-0.012

表 4.11: 各主成分の意味付け (1 年次前期終了時点)

主成分番号	意味付け
第 1 主成分	全体的な成績の悪さ
第 2 主成分	秀評価の多さ
第 3 主成分	落とした必修科目の少なさ

表 4.12 で, それぞれの変数選択法の結果, 得られた説明変数と回帰係数を示す.

表 4.12: 説明変数と回帰係数 (1 年次前期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	-1.193	-1.444
第 1 主成分	0.219	0.238
第 2 主成分	0.056	不採用
第 3 主成分	-0.160	-0.170

表 4.13: 予測精度 (1 年次前期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	279	92.3%	25	2	8%	2	2	100%	0.148
	30%	302	275	91.1%	25	5	20%	12	5	41.7%	0.270
	8.3%	302	214	70.9%	25	16	64%	95	16	16.9%	0.267
ステップワイズ法	50%	302	278	92.1%	25	1	4%	1	1	100%	0.077
	30%	302	274	90.7%	25	4	16%	11	4	36.4%	0.222
	8.3%	302	212	70.2%	25	18	72%	101	18	17.8%	0.285

1 年次後期までのデータを用いた推定

表 3.2 の変数 1-34 に対して主成分分析を行い, 表 4.14 の結果が得られた. 累計寄与率が 70% を超えた第 5 主成分までを説明変数として用いた. 各主成分における固有ベクトルを解釈することで, 主成分の意味付けを行った. 固有ベクトルを表 4.15, 意味付けを表 4.16 に示す.

表 4.14: 各主成分の寄与率・累計寄与率 (1 年次後期終了時点)

主成分番号	寄与率	累計寄与率
第 1 主成分	0.349	0.349
第 2 主成分	0.170	0.520
第 3 主成分	0.081	0.601
第 4 主成分	0.059	0.660
第 5 主成分	0.052	0.712
⋮	⋮	⋮
第 34 主成分	0.000	1.000

表 4.15: 各主成分の固有ベクトル (1 年次後期終了時点)

変数名	主成分番号				
	1	2	3	4	5
1 年次前期理系基礎 GPA	-0.107	0.024	-0.030	0.062	0.014
1 年次前期外国語 GPA	-0.041	0.025	-0.036	-0.021	-0.036
1 年次前期人間社会 GPA	-0.072	0.037	0.003	0.010	0.003
1 年次前期体育 GPA	-0.025	0.020	0.027	0.009	0.027
1 年次前期専門 GPA	-0.087	0.045	-0.035	0.044	-0.035
1 年次前期その他 GPA	0.000	0.000	0.000	0.000	0.000
1 年次前期秀獲得数	-0.275	0.371	0.128	0.143	0.128
1 年次前期優獲得数	-0.167	-0.221	-0.617	-0.057	-0.617
1 年次前期良獲得数	0.113	-0.176	0.612	0.110	0.612
1 年次前期可獲得数	0.200	-0.046	0.018	-0.106	0.018
1 年次前期不可獲得数	0.111	0.057	-0.120	-0.073	-0.120
1 年次前期失格獲得数	0.020	0.019	-0.020	-0.019	-0.020
1 年次前期必修科目不合格	0.131	0.076	-0.140	-0.092	-0.140
1 年次前期#(S_{low})	0.290	0.116	-0.203	-0.180	-0.203
1 年次前期#(S_{mid})	-0.132	-0.406	0.197	0.029	0.197
1 年次前期#(S_{high})	-0.156	0.294	0.006	0.150	0.006
1 年次前期#(S_{att})	0.071	0.027	-0.050	0.042	-0.050
1 年次後期理系基礎 GPA	-0.145	0.038	-0.017	0.007	-0.017
1 年次後期外国語 GPA	-0.058	0.032	-0.024	-0.082	-0.024
1 年次後期人間社会 GPA	-0.112	0.026	0.078	-0.027	0.078
1 年次後期体育 GPA	-0.048	0.004	0.016	-0.002	0.016
1 年次後期専門 GPA	-0.116	0.052	-0.015	0.001	-0.015
1 年次後期その他 GPA	-0.019	0.023	0.013	-0.015	0.013
1 年次後期秀獲得数	-0.306	0.385	0.074	0.022	0.074
1 年次後期優獲得数	-0.236	-0.114	-0.061	-0.475	-0.061
1 年次後期良獲得数	0.051	-0.257	-0.192	0.737	-0.192
1 年次後期可獲得数	0.188	-0.167	0.211	-0.284	0.211
1 年次後期不可獲得数	0.259	0.110	0.005	-0.012	0.005
1 年次後期失格獲得数	0.039	0.032	-0.027	0.012	-0.027
1 年次後期必修科目不合格	0.298	0.142	-0.022	0.000	-0.022
1 年次後期#(S_{low})	0.388	0.145	-0.037	0.057	-0.037
1 年次後期#(S_{mid})	-0.252	-0.370	0.003	-0.048	0.003
1 年次後期#(S_{high})	-0.141	0.214	0.044	-0.009	0.044
1 年次後期#(S_{att})	0.156	0.034	-0.069	0.112	-0.069

表 4.16: 各主成分の意味付け (1 年次後期終了時点)

主成分番号	意味付け
第1主成分	全体的な成績の悪さ
第2主成分	得手不得手の極端さ
第3主成分	1年次後期の落とした必修の少なさ
第4主成分	良獲得の多さ
第5主成分	クラス内偏差値の高さ

表 4.17 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.17: 説明変数と回帰係数 (1 年次後期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	-1.146	-1.394
第1主成分	0.200	0.215
第2主成分	0.047	不採用
第3主成分	-0.017	不採用
第4主成分	-0.033	不採用
第5主成分	-0.172	-0.184

表 4.18: 予測精度 (1 年次後期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	277	91.7%	25	2	8%	4	2	50%	0.138
	30%	302	272	90%	25	7	28%	19	7	36.8%	0.318
	8.3%	302	233	77.1%	25	17	68%	78	17	21.8%	0.330
ステップワイズ法	50%	302	276	91.3%	25	1	4%	3	1	33.3%	0.071
	30%	302	275	91%	25	8	32%	18	8	44.4%	0.372
	8.3%	302	226	74.8%	25	20	80%	91	20	21.2%	0.345

2 年次前期までのデータを用いた推定

表 3.2 の変数 1-51 に対して主成分分析を行い、表 4.19 の結果が得られた。累計寄与率が 70% を超えた第 7 主成分までを説明変数として用いた。各主成分における固有ベクトルを解釈することで、主成分の意味付けを行った。固有ベクトルを表 4.20、意味付けを表 4.21 に示す。

表 4.19: 各主成分の寄与率・累計寄与率 (2 年次前期終了時点)

主成分番号	寄与率	累計寄与率
第 1 主成分	0.310	0.310
第 2 主成分	0.125	0.434
第 3 主成分	0.095	0.530
第 4 主成分	0.058	0.587
第 5 主成分	0.042	0.629
第 6 主成分	0.040	0.669
第 7 主成分	0.036	0.705
⋮	⋮	⋮
第 51 主成分	0.000	1.000

表 4.20: 各主成分の固有ベクトル (2 年次前期終了時点)

変数名	主成分番号						
	1	2	3	4	5	6	7
1 年次前期理系基礎 GPA	0.086	0.029	-0.065	0.004	-0.061	-0.029	0.020
1 年次前期外国語 GPA	0.034	0.026	0.030	0.025	0.030	-0.020	-0.031
1 年次前期人間社会 GPA	0.055	0.038	-0.001	0.020	-0.001	-0.071	-0.015
1 年次前期体育 GPA	0.019	0.016	-0.011	-0.023	-0.011	-0.024	-0.042
1 年次前期専門 GPA	0.069	0.042	-0.038	-0.024	-0.038	-0.047	-0.030
1 年次前期その他 GPA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1 年次前期秀獲得数	0.225	0.333	-0.095	-0.267	-0.095	-0.074	0.042
1 年次前期優獲得数	0.121	-0.172	-0.002	0.279	-0.002	-0.218	-0.394
1 年次前期良獲得数	-0.087	-0.162	-0.091	0.104	-0.091	0.172	0.459
1 年次前期可獲得数	-0.152	-0.055	0.105	0.011	0.105	0.093	-0.013
1 年次前期不可獲得数	-0.091	0.043	0.071	-0.106	0.071	0.031	-0.080
1 年次前期失格獲得数	-0.017	0.016	0.015	-0.023	0.015	-0.001	-0.014
1 年次前期必修科目不合格	-0.108	0.059	0.086	-0.129	0.086	0.030	-0.094
1 年次前期#(S_{low})	-0.228	0.086	0.132	-0.217	0.132	0.157	-0.130
1 年次前期#(S_{mid})	0.095	-0.339	-0.005	0.470	-0.005	-0.141	0.173
1 年次前期#(S_{high})	0.132	0.257	-0.123	-0.257	-0.123	-0.013	-0.042
1 年次前期#(S_{att})	-0.060	0.023	-0.051	-0.070	-0.051	0.037	0.072

表 4.20 の続き

変数名	主成分番号						
	1	2	3	4	5	6	7
1 年次後期理系基礎 GPA	0.121	0.035	-0.011	-0.023	-0.011	-0.008	0.000
1 年次後期外国語 GPA	0.053	0.027	0.074	0.026	0.074	-0.018	-0.023
1 年次後期人間社会 GPA	0.091	0.019	0.046	0.019	0.046	-0.084	0.076
1 年次後期体育 GPA	0.043	-0.009	0.000	-0.016	0.000	-0.052	0.015
1 年次後期専門 GPA	0.100	0.046	-0.005	-0.013	-0.005	-0.012	0.000
1 年次後期その他 GPA	0.013	0.022	0.023	-0.002	0.023	-0.025	0.025
1 年次後期秀獲得数	0.269	0.352	-0.020	0.081	-0.020	-0.225	0.238
1 年次後期優獲得数	0.189	-0.107	0.387	-0.071	0.387	-0.009	-0.283
1 年次後期良獲得数	-0.040	-0.240	-0.626	-0.053	-0.626	0.340	-0.143
1 年次後期可獲得数	-0.180	-0.152	0.278	-0.143	0.278	-0.056	0.307
1 年次後期不可獲得数	-0.206	0.110	-0.011	0.195	-0.011	-0.068	-0.087
1 年次後期失格獲得数	-0.028	0.027	-0.014	0.004	-0.014	0.008	-0.025
1 年次後期必修科目不合格	-0.235	0.137	-0.025	0.199	-0.025	-0.060	-0.112
1 年次後期#(S_{low})	-0.310	0.145	-0.076	0.254	-0.076	-0.073	-0.101
1 年次後期#(S_{mid})	0.187	-0.348	0.021	-0.349	0.021	0.185	0.001
1 年次後期#(S_{high})	0.127	0.193	0.050	0.108	0.050	-0.121	0.107
1 年次後期#(S_{att})	-0.152	0.067	-0.129	-0.047	-0.129	0.057	0.122
2 年次前期理系基礎 GPA	0.082	0.013	-0.047	0.042	-0.047	0.084	0.021
2 年次前期外国語 GPA	0.052	0.036	-0.016	0.049	-0.016	0.073	-0.035
2 年次前期人間社会 GPA	0.073	0.007	0.062	0.065	0.062	0.024	-0.015
2 年次前期体育 GPA	-0.001	0.000	0.016	0.002	0.016	0.013	0.007
2 年次前期専門 GPA	0.107	0.026	-0.017	0.047	-0.017	0.076	0.012
2 年次前期その他 GPA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2 年次前期秀獲得数	0.220	0.201	-0.134	0.160	-0.134	0.184	0.115
2 年次前期優獲得数	0.174	-0.004	0.331	0.133	0.331	0.417	-0.148
2 年次前期良獲得数	-0.054	-0.171	-0.293	-0.236	-0.293	-0.509	-0.165
2 年次前期可獲得数	-0.190	-0.122	0.156	-0.139	0.156	-0.175	0.319
2 年次前期不可獲得数	-0.129	0.072	0.050	-0.028	0.050	0.009	-0.097
2 年次前期失格獲得数	-0.064	0.065	-0.001	0.023	-0.001	0.028	-0.083
2 年次前期必修科目不合格	-0.191	0.134	0.048	0.003	0.048	0.037	-0.176
2 年次前期#(S_{low})	-0.271	0.135	0.059	-0.031	0.059	0.020	-0.063
2 年次前期#(S_{mid})	0.123	-0.199	0.095	-0.129	0.095	-0.236	-0.024
2 年次前期#(S_{high})	0.105	0.105	-0.045	0.074	-0.045	0.170	0.028
2 年次前期#(S_{att})	-0.192	0.107	0.009	-0.050	0.009	0.042	0.155

表 4.21: 各主成分の意味付け (2年次前期終了時点)

主成分番号	意味付け
第1主成分	全体的な成績の良さ
第2主成分	得手不得手の極端さ
第3主成分	成績の上昇
第4主成分	出席率の高さ
第5主成分	外国語・人間社会の成績の良さ
第6主成分	1年次の成績の悪さ
第7主成分	優獲得の少なさ

表 4.22 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.22: 説明変数と回帰係数 (2年次前期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	-1.137	-1.789
第1主成分	-0.219	-0.250
第2主成分	0.081	不採用
第3主成分	-0.100	-0.103
第4主成分	-0.037	不採用
第5主成分	0.047	不採用
第6主成分	0.067	不採用
第7主成分	-0.175	-0.207

表 4.23: 予測精度 (2年次前期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	279	92.4%	25	7	28%	12	7	58.3%	0.378
	30%	302	280	92.7%	25	12	48%	21	12	57.1%	0.522
	8.3%	302	244	80.8%	25	20	80%	73	20	27.4%	0.408
ステップワイズ法	50%	302	280	92.7%	25	7	28%	11	7	63.6%	0.389
	30%	302	281	93%	25	11	44%	18	11	61.1%	0.512
	8.3%	302	243	80.5%	25	21	84%	76	21	27.6%	0.416

2年次後期までのデータを用いた推定

表 3.2 の変数 1-68 に対して主成分分析を行い, 表 4.24 の結果が得られた. 累計寄与率が 70%を超えた第 10 主成分までを説明変数として用いた. 各主成分における固有ベクトルを解釈することで, 主成分の意味付けを行った. 固有ベクトルを表 4.25, 表 4.26, 意味付けを表 4.27 に示す.

表 4.24: 各主成分の寄与率・累計寄与率 (2年次後期終了時点)

主成分番号	寄与率	累計寄与率
第 1 主成分	0.289	0.289
第 2 主成分	0.108	0.397
第 3 主成分	0.092	0.489
第 4 主成分	0.046	0.535
第 5 主成分	0.041	0.576
第 6 主成分	0.034	0.610
第 7 主成分	0.031	0.641
第 8 主成分	0.028	0.669
第 9 主成分	0.027	0.696
第 10 主成分	0.026	0.722
⋮	⋮	⋮
第 68 主成分	0.000	1.000

表 4.25: 各主成分の固有ベクトル (1-5)(2年次後期終了時点)

変数名	主成分番号				
	1	2	3	4	5
1年次前期理系基礎 GPA	0.071	0.057	-0.042	0.005	-0.039
1年次前期外国語 GPA	0.030	0.030	0.014	0.017	0.014
1年次前期人間社会 GPA	0.043	0.051	-0.030	0.021	-0.030
1年次前期体育 GPA	0.015	0.014	-0.031	-0.018	-0.031
1年次前期専門 GPA	0.059	0.051	-0.068	-0.010	-0.068
1年次前期その他 GPA	0.000	0.000	0.000	0.000	0.000
1年次前期秀獲得数	0.188	0.297	-0.195	-0.262	-0.195
1年次前期優獲得数	0.099	-0.032	0.007	0.306	0.007
1年次前期良獲得数	-0.070	-0.162	0.024	0.099	0.024
1年次前期可獲得数	-0.129	-0.095	0.186	-0.039	0.186
1年次前期不可獲得数	-0.077	-0.010	-0.011	-0.090	-0.011
1年次前期失格獲得数	-0.014	0.004	-0.012	-0.017	-0.012
1年次前期必修科目不合格	-0.091	-0.006	-0.023	-0.107	-0.023
1年次前期#(S_{low})	-0.187	-0.043	0.037	-0.204	0.037
1年次前期#(S_{mid})	0.074	-0.146	0.169	0.440	0.169
1年次前期#(S_{high})	0.112	0.192	-0.207	-0.239	-0.207
1年次前期#(S_{att})	-0.049	-0.001	-0.019	-0.070	-0.019
1年次後期理系基礎 GPA	0.103	0.065	-0.020	-0.021	-0.020
1年次後期外国語 GPA	0.047	0.027	0.043	0.011	0.043
1年次後期人間社会 GPA	0.076	0.039	0.008	0.011	0.008
1年次後期体育 GPA	0.036	0.000	0.030	-0.029	0.030
1年次後期専門 GPA	0.087	0.056	-0.018	-0.011	-0.018
1年次後期その他 GPA	0.009	0.026	0.003	-0.005	0.003
1年次後期秀獲得数	0.234	0.315	0.093	0.009	0.093
1年次後期優獲得数	0.159	-0.003	0.000	-0.049	0.000
1年次後期良獲得数	-0.029	-0.210	-0.242	0.030	-0.242
1年次後期可獲得数	-0.162	-0.107	0.064	-0.127	0.064
1年次後期不可獲得数	-0.175	0.000	0.092	0.145	0.092
1年次後期失格獲得数	-0.023	0.001	0.007	-0.002	0.007
1年次後期必修科目不合格	-0.198	0.001	0.099	0.143	0.099
1年次後期#(S_{low})	-0.259	-0.019	0.094	0.201	0.094
1年次後期#(S_{mid})	0.151	-0.138	-0.149	-0.247	-0.149
1年次後期#(S_{high})	0.111	0.153	0.068	0.052	0.068
1年次後期#(S_{att})	-0.134	0.057	-0.100	-0.019	-0.100

表 4.25 の続き

変数名	主成分番号				
	1	2	3	4	5
2 年次前期理系基礎 GPA	0.076	-0.010	0.056	0.011	0.056
2 年次前期外国語 GPA	0.049	0.017	0.034	0.036	0.034
2 年次前期人間社会 GPA	0.067	-0.017	0.051	0.032	0.051
2 年次前期体育 GPA	-0.002	0.004	0.029	-0.007	0.029
2 年次前期専門 GPA	0.099	0.004	0.051	0.018	0.051
2 年次前期その他 GPA	0.000	0.000	0.000	0.000	0.000
2 年次前期秀獲得数	0.203	0.138	0.000	0.144	0.000
2 年次前期優獲得数	0.161	-0.052	0.251	0.030	0.251
2 年次前期良獲得数	-0.047	-0.188	-0.118	-0.228	-0.118
2 年次前期可獲得数	-0.172	-0.093	0.082	-0.140	0.082
2 年次前期不可獲得数	-0.123	0.110	-0.148	0.049	-0.148
2 年次前期失格獲得数	-0.059	0.078	-0.087	0.065	-0.087
2 年次前期必修科目不合格	-0.180	0.188	-0.227	0.119	-0.227
2 年次前期#(S_{low})	-0.253	0.208	-0.305	0.128	-0.305
2 年次前期#(S_{mid})	0.122	-0.307	0.308	-0.285	0.308
2 年次前期#(S_{high})	0.094	0.092	-0.023	0.078	-0.023
2 年次前期#(S_{att})	-0.185	0.174	-0.077	-0.005	-0.077
2 年次後期理系基礎 GPA	-0.010	-0.098	-0.043	0.122	-0.043
2 年次後期外国語 GPA	0.038	0.028	0.024	0.054	0.024
2 年次後期人間社会 GPA	0.052	-0.032	-0.072	0.017	-0.072
2 年次後期体育 GPA	0.002	-0.007	-0.005	-0.011	-0.005
2 年次後期専門 GPA	0.100	0.007	-0.003	0.038	-0.003
2 年次後期その他 GPA	0.000	0.000	0.000	0.000	0.000
2 年次後期秀獲得数	0.148	0.144	0.043	0.090	0.043
2 年次後期優獲得数	0.125	-0.100	-0.128	0.216	-0.128
2 年次後期良獲得数	-0.004	-0.207	-0.241	-0.073	-0.241
2 年次後期可獲得数	-0.113	-0.131	-0.038	-0.039	-0.038
2 年次後期不可獲得数	-0.145	0.057	0.062	-0.035	0.062
2 年次後期失格獲得数	-0.082	0.087	0.017	0.017	0.017
2 年次後期必修科目不合格	-0.222	0.144	0.088	-0.028	0.088
2 年次後期#(S_{low})	-0.244	0.138	0.121	-0.045	0.121
2 年次後期#(S_{mid})	0.097	-0.369	-0.470	0.182	-0.470
2 年次後期#(S_{high})	0.076	0.083	0.063	0.039	0.063
2 年次後期#(S_{att})	-0.128	0.046	0.021	-0.017	0.021

表 4.26: 各主成分の固有ベクトル (6-10)(2年次後期終了時点)

変数名	主成分番号				
	6	7	8	9	10
1年次前期理系基礎 GPA	-0.012	0.080	-0.001	0.005	-0.067
1年次前期外国語 GPA	-0.019	-0.007	-0.023	-0.026	-0.023
1年次前期人間社会 GPA	-0.126	0.032	-0.079	-0.005	-0.079
1年次前期体育 GPA	-0.073	-0.009	-0.022	0.001	-0.022
1年次前期専門 GPA	-0.047	0.042	-0.026	-0.019	-0.026
1年次前期その他 GPA	0.000	0.000	0.000	0.000	0.000
1年次前期秀獲得数	-0.145	0.106	-0.189	0.110	-0.189
1年次前期優獲得数	-0.122	0.075	0.109	-0.513	0.109
1年次前期良獲得数	0.174	0.068	-0.194	0.500	-0.194
1年次前期可獲得数	0.075	-0.121	0.184	0.010	0.184
1年次前期不可獲得数	0.014	-0.114	0.086	-0.084	0.086
1年次前期失格獲得数	0.003	-0.018	0.002	-0.020	0.002
1年次前期必修科目不合格	0.017	-0.132	0.088	-0.104	0.088
1年次前期#(S_{low})	0.176	-0.253	0.244	-0.162	0.244
1年次前期#(S_{mid})	-0.120	0.179	-0.101	0.168	-0.101
1年次前期#(S_{high})	-0.057	0.070	-0.146	-0.004	-0.146
1年次前期#(S_{att})	0.186	0.090	-0.094	-0.055	-0.094
1年次後期理系基礎 GPA	0.006	0.021	-0.004	-0.027	-0.004
1年次後期外国語 GPA	-0.031	-0.052	0.041	-0.012	0.041
1年次後期人間社会 GPA	-0.119	0.005	0.029	0.088	0.029
1年次後期体育 GPA	-0.057	0.019	0.080	0.020	0.080
1年次後期専門 GPA	0.001	0.013	0.000	-0.024	0.000
1年次後期その他 GPA	-0.039	-0.004	0.008	0.019	0.008
1年次後期秀獲得数	-0.096	0.209	0.211	0.106	0.211
1年次後期優獲得数	-0.172	-0.358	-0.305	-0.127	-0.305
1年次後期良獲得数	0.368	0.258	0.080	-0.228	0.080
1年次後期可獲得数	-0.047	-0.144	0.114	0.317	0.114
1年次後期不可獲得数	-0.084	0.038	-0.070	-0.019	-0.070
1年次後期失格獲得数	0.027	0.011	-0.025	-0.039	-0.025
1年次後期必修科目不合格	-0.057	0.049	-0.095	-0.058	-0.095
1年次後期#(S_{low})	-0.021	0.101	-0.077	-0.052	-0.077
1年次後期#(S_{mid})	0.100	-0.167	0.094	0.013	0.094
1年次後期#(S_{high})	-0.082	0.079	-0.010	0.050	-0.010
1年次後期#(S_{att})	0.269	0.164	-0.165	-0.049	-0.165

表 4.26 の続き

変数名	主成分番号				
	6	7	8	9	10
2 年次前期理系基礎 GPA	0.076	0.028	-0.041	-0.009	-0.041
2 年次前期外国語 GPA	0.044	-0.016	-0.002	-0.049	-0.002
2 年次前期人間社会 GPA	-0.033	-0.046	-0.061	0.024	-0.061
2 年次前期体育 GPA	0.019	-0.009	-0.008	0.000	-0.008
2 年次前期専門 GPA	0.059	0.003	-0.029	-0.012	-0.029
2 年次前期その他 GPA	0.000	0.000	0.000	0.000	0.000
2 年次前期秀獲得数	0.154	0.053	0.084	0.065	0.084
2 年次前期優獲得数	0.131	-0.377	-0.399	-0.101	-0.399
2 年次前期良獲得数	-0.281	0.358	0.122	-0.172	0.122
2 年次前期可獲得数	-0.055	0.036	0.165	0.239	0.165
2 年次前期不可獲得数	-0.096	-0.126	0.029	-0.005	0.029
2 年次前期失格獲得数	0.027	-0.049	-0.013	-0.058	-0.013
2 年次前期必修科目不合格	-0.068	-0.172	0.019	-0.056	0.019
2 年次前期#(S_{low})	-0.077	-0.181	0.075	0.086	0.075
2 年次前期#(S_{mid})	-0.158	0.114	-0.120	-0.151	-0.120
2 年次前期#(S_{high})	0.114	-0.039	0.033	0.035	0.033
2 年次前期#(S_{att})	0.219	0.094	-0.245	-0.035	-0.245
2 年次後期理系基礎 GPA	-0.236	-0.199	0.049	-0.029	0.049
2 年次後期外国語 GPA	0.036	-0.053	0.029	-0.039	0.029
2 年次後期人間社会 GPA	-0.033	-0.029	0.042	0.093	0.042
2 年次後期体育 GPA	0.005	-0.009	0.004	-0.016	0.004
2 年次後期専門 GPA	0.083	-0.012	0.052	0.000	0.052
2 年次後期その他 GPA	0.000	0.000	0.000	0.000	0.000
2 年次後期秀獲得数	-0.028	-0.032	0.132	-0.005	0.132
2 年次後期優獲得数	0.073	-0.116	0.165	0.010	0.165
2 年次後期良獲得数	0.002	0.014	-0.307	0.019	-0.307
2 年次後期可獲得数	-0.223	0.086	0.007	0.078	0.007
2 年次後期不可獲得数	-0.162	0.026	-0.030	0.008	-0.030
2 年次後期失格獲得数	-0.004	-0.016	-0.104	-0.103	-0.104
2 年次後期必修科目不合格	-0.163	0.010	-0.127	-0.090	-0.127
2 年次後期#(S_{low})	-0.201	0.043	-0.106	-0.068	-0.106
2 年次後期#(S_{mid})	-0.210	-0.055	-0.096	0.055	-0.096
2 年次後期#(S_{high})	0.068	-0.027	0.066	0.021	0.066
2 年次後期#(S_{att})	0.135	0.153	-0.186	-0.083	-0.186

表 4.27: 各主成分の意味付け (2 年次後期終了時点)

主成分番号	意味付け
第1主成分	全体的な成績の良さ
第2主成分	成績が徐々に低下
第3主成分	成績が徐々に上昇
第4主成分	文系科目の成績の良さ
第5主成分	1 年前期次の専門 GPA の悪さ
第6主成分	人間社会の悪さ
第7主成分	外国語 GPA の悪さ
第8主成分	出席偏差の悪さ
第9主成分	失格獲得数の少なさ
第10主成分	1 年次前期・2 年次前期の成績の悪さ

表 4.28 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.28: 説明変数と回帰係数 (2 年次後期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	-0.39	0.183
第1主成分	-0.260	-0.238
第2主成分	0.270	0.293
第3主成分	-0.091	不採用
第4主成分	-0.102	不採用
第5主成分	0.162	0.172
第6主成分	0.060	不採用
第7主成分	-0.030	不採用
第8主成分	-0.202	-0.199
第9主成分	-0.455	-0.425
第10主成分	-0.064	不採用

表 4.29: 予測精度 (2 年次後期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	290	96%	25	16	64%	19	16	84.2%	0.727
	30%	302	288	95.3%	25	18	72%	25	18	72%	0.72
	8.3%	302	272	90.1%	25	22	88%	49	22	44.9%	0.595
ステップワイズ法	50%	302	291	96.4%	25	17	68%	20	17	85%	0.756
	30%	302	286	94.7%	25	18	72%	27	18	66.7%	0.692
	8.3%	302	271	89.7%	25	20	80%	46	20	43.5%	0.563

考察

変換された学生データに対して主成分分析を行い、説明変数の数を1/6程度に減らして検証を行った。検証1と比べて、的中数とF値の大幅な向上が見られた。変数選択による大きな差異は見られなかった。しかし、主成分スコアが多くの変数に依存しているため、各主成分の意味付けがわかりづらく傾向の発見が難しかった。

次節では、学生データの変数変換で用いた区分ごとに主成分分析を行うことで傾向の発見を試みた。

4.2.3 ロジスティック回帰分析を用いた検証3

説明変数を, その学期までのデータから第2章のデータ変換を行った小節ごとに主成分分析を行った各主成分とした.

1年次前期までのデータを用いた推定

まず, 表3.2の変数1-17を次の表のように分類する.

表 4.30: 変数の分類 (1年次前期終了時点)

区分名	変数
講義区分別 GPA	変数 1-6
獲得成績数	変数 7-12
必修科目不合格数	変数 13
偏差値別講義数	変数 14-16
出席低偏差講義数	変数 17

この区分ごとに主成分分析を行い, 表4.31の結果が得られた. 累計寄与率が70%を超えた主成分を説明変数として用いた. 各主成分における固有ベクトルを解釈することで, 主成分の意味付けを行った. 固有ベクトルを表4.32, 表4.33, 表4.34, 表4.35, 表4.36, 意味付けを表4.37に示す.

表 4.31: 各主成分の寄与率・累計寄与率(1年次前期終了時点)

主成分番号	寄与率	累計寄与率
講義区分別 GPA		
第1主成分	0.419	0.419
第2主成分	0.215	0.634
第3主成分	0.159	0.793
⋮	⋮	⋮
第6主成分	0.000	1.000
獲得成績数		
第1主成分	0.434	0.434
第2主成分	0.302	0.736
⋮	⋮	⋮
第6主成分	0.000	1.000
必修科目不合格数		
第1主成分	1.000	1.000
偏差値別講義数		
第1主成分	0.638	0.638
第2主成分	0.361	0.999
第3主成分	0.001	1.000
出席低偏差講義数		
第1主成分	1.000	1.000

表 4.32: 各主成分の固有ベクトル(講義区分別 GPA)(1年次前期終了時点)

変数名	主成分番号		
	1	2	3
1年次前期理系基礎 GPA	-0.409	0.564	0.022
1年次前期外国語 GPA	-0.160	0.422	-0.287
1年次前期人間社会 GPA	-0.820	-0.534	-0.202
1年次前期体育 GPA	-0.144	-0.104	0.912
1年次前期専門 GPA	-0.338	0.455	0.209
1年次前期その他 GPA	0.000	0.000	0.000

表 4.33: 各主成分の固有ベクトル (獲得成績数)(1 年次前期終了時点)

変数名	主成分番号	
	1	2
1 年次前期秀獲得数	-0.760	0.441
1 年次前期優獲得数	-0.222	-0.863
1 年次前期良獲得数	0.444	0.129
1 年次前期可獲得数	0.401	0.185
1 年次前期不可獲得数	0.122	0.092
1 年次前期失格獲得数	0.014	0.024

表 4.34: 各主成分の固有ベクトル (必修科目不合格数)(1 年次前期終了時点)

変数名	主成分番号
	1
1 年次前期必修科目不合格	1.000

表 4.35: 各主成分の固有ベクトル (偏差値別講義数)(1 年次前期終了時点)

変数名	主成分番号	
	1	2
1 年次前期#(S_{low})	0.614	0.540
1 年次前期#(S_{mid})	-0.773	0.261
1 年次前期#(S_{high})	0.162	-0.800

表 4.36: 各主成分の固有ベクトル (出席低偏差講義数)(1 年次前期終了時点)

変数名	主成分番号
	1
1 年次前期#(S_{att})	1.000

表 4.37: 各主成分の意味付け (1 年次前期終了時点)

区分名	主成分番号	意味付け
講義区分別 GPA	第 1 主成分	成績の悪さ
	第 2 主成分	人間社会の良さ
	第 3 主成分	体育の良さ
獲得成績数	第 1 主成分	秀・優獲得の少なさ
	第 2 主成分	優獲得の少なさ
必修科目不合格数	第 1 主成分	必修科目不合格の多さ
	第 2 主成分	高偏差の少なさ
偏差値別講義数	第 1 主成分	中偏差の少なさ
	第 2 主成分	高偏差の少なさ
出席低偏差講義数	第 1 主成分	出席率の低さ

表 4.38 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.38: 説明変数と回帰係数 (1 年次前期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	-0.001	-3.93
講義区分別 GPA1	0.373	不採用
講義区分別 GPA2	0.707	不採用
講義区分別 GPA3	-0.236	不採用
獲得成績数 1	-0.354	不採用
獲得成績数 2	0.039	不採用
必修科目不合格数 1	0.197	0.528
偏差値別講義数 1	0.081	不採用
偏差値別講義数 2	0.429	0.311
出席低偏差講義数 1	-0.045	不採用

表 4.39: 予測精度 (1 年次前期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	278	92.1%	25	1	4%	1	1	100%	0.077
	30%	302	274	90.7%	25	5	20%	13	5	38.5%	0.263
	8.3%	302	223	73.8%	25	16	64%	86	16	18.6%	0.288
ステップワイズ法	50%	302	280	92.7%	25	4	16%	5	4	80%	0.267
	30%	302	273	90.4%	25	5	20%	14	5	35.7%	0.256
	8.3%	302	236	78.1%	25	18	72%	77	18	23.4%	0.353

1 年次後期までのデータを用いた推定

まず, 表 3.2 の変数 1-34 を次の表のように分類する.

表 4.40: 変数の分類 (1 年次後期終了時点)

区分名	変数
講義区分別 GPA	変数 1-6, 18-23
獲得成績数	変数 7-12, 24-29
必修科目不合格数	変数 13, 30
偏差値別講義数	変数 14-16, 31-33
出席低偏差講義数	変数 17, 34

この区分ごとに主成分分析を行い, 表 4.41 の結果が得られた. 累計寄与率が 70% を超えた主成分を説明変数として用いた. 各主成分における固有ベクトルを解釈することで, 主成分の意味付けを行った. 固有ベクトルを表 4.42, 表 4.43, 表 4.44, 表 4.45, 表 4.46, 意味付けを表 4.37 に示す.

表 4.41: 各主成分の寄与率・累計寄与率(1年次後期終了時点)

主成分番号	寄与率	累計寄与率
講義区分別 GPA		
第 1 主成分	0.361	0.361
第 2 主成分	0.134	0.496
第 3 主成分	0.122	0.618
第 4 主成分	0.102	0.720
⋮	⋮	⋮
第 12 主成分	0.000	1.000
獲得成績数		
第 1 主成分	0.373	0.373
第 2 主成分	0.168	0.541
第 3 主成分	0.116	0.657
第 4 主成分	0.101	0.759
⋮	⋮	⋮
第 12 主成分	0.000	1.000
必修科目不合格数		
第 1 主成分	0.812	0.812
第 2 主成分	0.188	1.000
偏差値別講義数		
第 1 主成分	0.500	0.500
第 2 主成分	0.243	0.743
⋮	⋮	⋮
第 6 主成分	0.000	1.000
出席低偏差講義数		
第 1 主成分	0.805	0.805
第 2 主成分	0.195	1.000

表 4.42: 各主成分の固有ベクトル (講義区分別 GPA)(1年次後期終了時点)

変数名	主成分番号			
	1	2	3	4
1年次前期理系基礎 GPA	-0.323	0.181	0.196	-0.178
1年次前期外国語 GPA	-0.141	-0.109	0.327	0.498
1年次前期人間社会 GPA	-0.330	0.552	-0.580	0.474
1年次前期体育 GPA	-0.091	0.058	-0.131	-0.081
1年次前期専門 GPA	-0.276	0.123	0.136	-0.163
1年次前期その他 GPA	0.000	0.000	0.000	0.000
1年次後期理系基礎 GPA	-0.444	0.190	0.303	-0.216
1年次後期外国語 GPA	-0.215	-0.220	0.395	0.580
1年次後期人間社会 GPA	-0.522	-0.717	-0.436	-0.082
1年次後期体育 GPA	-0.145	0.023	0.010	-0.175
1年次後期専門 GPA	-0.374	0.176	0.187	-0.208
1年次後期その他 GPA	-0.074	0.048	-0.098	0.005

表 4.43: 各主成分の固有ベクトル (獲得成績数)(1年次後期終了時点)

変数名	主成分番号			
	1	2	3	4
1年次前期優獲得数	-0.154	-0.626	-0.026	0.524
1年次前期良獲得数	0.238	-0.019	0.017	-0.609
1年次前期可獲得数	0.284	0.196	-0.070	0.192
1年次前期不可獲得数	0.105	0.115	-0.024	0.058
1年次前期失格獲得数	0.014	0.028	-0.015	0.011
1年次後期秀獲得数	-0.550	0.297	0.149	0.198
1年次後期優獲得数	-0.265	-0.417	-0.509	-0.408
1年次後期良獲得数	0.201	-0.319	0.752	-0.125
1年次後期可獲得数	0.336	0.111	-0.363	0.213
1年次後期不可獲得数	0.250	0.283	-0.043	0.126
1年次後期失格獲得数	0.026	0.036	0.012	0.000

表 4.44: 各主成分の固有ベクトル (必修科目不合格数)(1年次後期終了時点)

変数名	主成分番号
	1
1年次前期必修科目不合格	0.367
1年次後期必修科目不合格	0.930

表 4.45: 各主成分の固有ベクトル (偏差値別講義数)(1年次後期終了時点)

変数名	主成分番号	
	1	2
1年次前期#(S_{low})	-0.456	-0.009
1年次前期#(S_{mid})	-0.580	-0.340
1年次前期#(S_{high})	0.396	-0.614
1年次後期#(S_{low})	0.541	-0.020
1年次後期#(S_{mid})	0.056	0.623
1年次後期#(S_{high})	0.052	0.345

表 4.46: 各主成分の固有ベクトル (出席低偏差講義数)(1年次後期終了時点)

変数名	主成分番号
	1
1年次前期#(S_{att})	0.493
1年次後期#(S_{att})	0.870

表 4.47: 各主成分の意味付け (1年次後期終了時点)

区分名	主成分番号	意味付け
講義区分別 GPA	第1主成分	成績の悪さ
	第2主成分	外国語の悪さ
	第3主成分	人間社会の悪さ
	第4主成分	専門の良さ
獲得成績数	第1主成分	秀獲得の少なさ
	第2主成分	優獲得の少なさ
	第3主成分	良獲得の多さ
	第4主成分	1年次前期の優の多さ
必修科目不合格数	第1主成分	必修科目不合格の多さ
偏差値別講義数	第1主成分	低偏差の少なさ
	第2主成分	高偏差の多さ
出席低偏差講義数	第1主成分	出席率の低さ

表 4.48 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.48: 説明変数と回帰係数 (1 年次後期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	25.4	11.9
講義区分別 GPA1	4.32	2.26
講義区分別 GPA2	-0.616	不採用
講義区分別 GPA3	-2.44	-1.28
講義区分別 GPA4	0.767	0.398
獲得成績数 1	-1.63	-0.785
獲得成績数 2	-0.164	不採用
獲得成績数 3	0.187	不採用
獲得成績数 4	-0.047	不採用
必修科目不合格数 1	-0.491	不採用
偏差値別講義数 1	-0.065	不採用
偏差値別講義数 2	-0.199	不採用
出席低偏差講義数 1	0.040	不採用

表 4.49: 予測精度 (1 年次後期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	280	92.7%	25	5	20%	7	5	71.4%	0.313
	30%	302	278	92.1%	25	10	40%	19	10	52.6%	0.455
	8.3%	302	244	80.8%	25	18	72%	69	18	26.1%	0.383
ステップワイズ法	50%	302	280	92.7%	25	6	24%	9	6	66.7%	0.353
	30%	302	275	91.1%	25	11	44%	24	11	45.8%	0.449
	8.3%	302	245	81.1%	25	19	76%	70	19	27.1%	0.400

2年次前期までのデータを用いた推定

まず, 表 3.2 の変数 1-51 を次の表のように分類する.

表 4.50: 変数の分類 (2年次前期終了時点)

区分名	変数
講義区分別 GPA	変数 1-6, 18-23, 35-40
獲得成績数	変数 7-12, 24-29, 41-46
必修科目不合格数	変数 13, 30, 47
偏差値別講義数	変数 14-16, 31-33, 48-50
出席低偏差講義数	変数 17, 34, 51

この区分ごとに主成分分析を行い, 表 4.51 の結果が得られた. 累計寄与率が 70%を超えた主成分を説明変数として用いた. 各主成分における固有ベクトルを解釈することで, 主成分の意味付けを行った. 固有ベクトルを表 4.52, 表 4.53, 表 4.54, 表 4.55, 表 4.56, 意味付けを表 4.57 に示す.

表 4.51: 各主成分の寄与率・累計寄与率(2年次前期終了時点)

主成分番号	寄与率	累計寄与率
講義区分別 GPA		
第 1 主成分	0.319	0.319
第 2 主成分	0.110	0.429
第 3 主成分	0.098	0.528
第 4 主成分	0.081	0.609
第 5 主成分	0.068	0.677
第 6 主成分	0.057	0.734
⋮	⋮	⋮
第 18 主成分	0.000	1.000
獲得成績数		
第 1 主成分	0.335	0.335
第 2 主成分	0.119	0.453
第 3 主成分	0.086	0.539
第 4 主成分	0.084	0.623
第 5 主成分	0.071	0.694
第 6 主成分	0.061	0.755
⋮	⋮	⋮
第 18 主成分	0.000	1.000
必修科目不合格数		
第 1 主成分	0.638	0.638
第 2 主成分	0.253	0.891
第 3 主成分	0.109	1.000
偏差値別講義数		
第 1 主成分	0.399	0.399
第 2 主成分	0.206	0.605
第 3 主成分	0.161	0.766
⋮	⋮	⋮
第 9 主成分	0.000	1.000
出席低偏差講義数		
第 1 主成分	0.705	0.705
第 2 主成分	0.197	0.902
第 3 主成分	0.098	1.000

表 4.53: 各主成分の固有ベクトル (獲得成績数)(2年次前期終了時点)

変数名	主成分番号					
	1	2	3	4	5	6
1年次前期秀獲得数	0.392	0.318	-0.150	-0.101	0.236	-0.429
1年次前期優獲得数	0.123	-0.541	-0.360	0.035	-0.500	0.119
1年次前期良獲得数	-0.177	-0.068	0.349	-0.042	0.392	0.588
1年次前期可獲得数	-0.226	0.157	0.171	0.074	-0.098	-0.204
1年次前期不可獲得数	-0.099	0.113	-0.001	0.028	-0.030	-0.059
1年次前期失格獲得数	-0.014	0.027	-0.009	0.009	0.003	-0.016
1年次後期秀獲得数	0.467	0.322	-0.105	-0.145	-0.157	0.214
1年次後期優獲得数	0.222	-0.391	-0.184	0.426	0.429	0.029
1年次後期良獲得数	-0.147	-0.349	0.302	-0.588	-0.152	-0.247
1年次後期可獲得数	-0.305	0.107	-0.114	0.311	-0.034	-0.069
1年次後期不可獲得数	-0.214	0.270	0.080	0.004	-0.082	0.084
1年次後期失格獲得数	-0.022	0.033	0.019	-0.011	-0.003	0.002
2年次前期秀獲得数	0.354	0.097	-0.103	-0.145	-0.103	0.387
2年次前期優獲得数	0.231	-0.176	-0.038	0.364	-0.038	-0.333
2年次前期良獲得数	-0.144	-0.142	0.494	-0.388	0.494	-0.139
2年次前期可獲得数	-0.306	0.137	-0.135	0.168	-0.135	0.090
2年次前期不可獲得数	-0.127	0.131	-0.097	0.054	-0.097	-0.060
2年次前期失格獲得数	-0.045	0.058	-0.081	0.004	-0.081	0.013

表 4.54: 各主成分の固有ベクトル (必修科目不合格数)(2年次前期終了時点)

変数名	主成分番号	
	1	2
1年次前期必修科目不合格	0.299	0.150
1年次後期必修科目不合格	0.700	0.636
2年次前期必修科目不合格	0.648	-0.757

表 4.55: 各主成分の固有ベクトル (偏差値別講義数)(2年次前期終了時点)

変数名	主成分番号		
	1	2	3
1年次前期#(S_{low})	-0.385	0.236	0.163
1年次前期#(S_{mid})	-0.525	0.083	0.312
1年次前期#(S_{high})	-0.438	-0.483	-0.176
1年次後期#(S_{low})	0.291	-0.489	0.392
1年次後期#(S_{mid})	0.449	-0.272	0.023
1年次後期#(S_{high})	0.270	0.544	0.465
2年次前期#(S_{low})	0.090	0.253	-0.556
2年次前期#(S_{mid})	0.086	0.176	-0.327
2年次前期#(S_{high})	0.096	0.025	-0.242

表 4.56: 各主成分の固有ベクトル (出席低偏差講義数)(2年次前期終了時点)

変数名	主成分番号
	1
1年次前期#(S_{att})	0.299
1年次後期#(S_{att})	0.594
2年次前期#(S_{att})	0.747

表 4.57: 各主成分の意味付け (2年次前期終了時点)

区分名	主成分番号	意味付け
講義区分別 GPA	第1主成分	理系基礎の良さ
	第2主成分	外国語の良さ
	第3主成分	専門の悪さ
	第4主成分	1年次前期の人間社会の良さ
	第5主成分	2年次前期の人間社会の悪さ
	第6主成分	2年次前期の理系基礎の良さ
獲得成績数	第1主成分	秀獲得の多さ
	第2主成分	優獲得の少なさ
	第3主成分	優獲得の多さ, 良獲得の少なさ
	第4主成分	良獲得の少なさ
	第5主成分	1年次前期の優獲得の少なさ
	第6主成分	1年次前期の良獲得の多さ
必修科目不合格数	第1主成分	必修科目不合格の多さ
	第2主成分	2年次前期の必修科目不合格の少なさ
偏差値別講義数	第1主成分	中偏差の多さ
	第2主成分	1年次の中偏差の少なさ
	第3主成分	高偏差の少なさ
出席低偏差講義数	第1主成分	出席率の低さ

表 4.58 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.58: 説明変数と回帰係数 (2 年次前期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	23.0	17.2
講義区分別 GPA1	-3.66	-2.74
講義区分別 GPA2	-0.875	-0.783
講義区分別 GPA3	0.110	不採用
講義区分別 GPA4	0.920	0.725
講義区分別 GPA5	-0.661	不採用
講義区分別 GPA6	0.772	不採用
獲得成績数 1	1.35	0.882
獲得成績数 2	-0.024	不採用
獲得成績数 3	-0.111	不採用
獲得成績数 4	-0.125	不採用
獲得成績数 5	-0.083	不採用
獲得成績数 6	-0.115	不採用
必修科目不合格数 1	-0.011	不採用
必修科目不合格数 2	-0.337	不採用
偏差値別講義数 1	-0.007	不採用
偏差値別講義数 2	0.008	不採用
偏差値別講義数 3	0.250	不採用
出席低偏差講義数 1	0.043	不採用

表 4.59: 予測精度 (2 年次前期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	286	94.7%	25	11	44%	13	11	84.6%	0.579
	30%	302	283	93.7%	25	15	60%	24	15	62.5%	0.612
	8.3%	302	260	86.1%	25	21	84%	59	21	35.6%	0.500
ステップワイズ法	50%	302	287	95%	25	11	44%	12	11	91.7%	0.595
	30%	302	284	94%	25	16	64%	25	16	64%	0.640
	8.3%	302	261	86.4%	25	20	80%	56	20	35.7%	0.494

2年次後期までのデータを用いた推定

まず, 表 3.2 の変数 1-68 を次の表のように分類する.

表 4.60: 変数の分類 (2年次後期終了時点)

区分名	変数
講義区分別 GPA	変数 1-6, 18-23, 35-40, 52-57
獲得成績数	変数 7-12, 24-29, 41-46, 58-63
必修科目不合格数	変数 13, 30, 47, 64
偏差値別講義数	変数 14-16, 31-33, 48-50, 65-67
出席低偏差講義数	変数 17, 34, 51, 68

この区分ごとに主成分分析を行い, 表 4.61 の結果が得られた. 累計寄与率が 70%を超えた主成分を説明変数として用いた. 各主成分における固有ベクトルを解釈することで, 主成分の意味付けを行った. 固有ベクトルを表 4.62, 表 4.63, 表 4.64, 表 4.65, 表 4.66, 意味付けを表 4.67 に示す.

表 4.61: 各主成分の寄与率・累計寄与率(2年次後期終了時点)

主成分番号	寄与率	累計寄与率
講義区分別 GPA		
第 1 主成分	0.244	0.244
第 2 主成分	0.175	0.419
第 3 主成分	0.086	0.506
第 4 主成分	0.077	0.583
第 5 主成分	0.065	0.648
第 6 主成分	0.055	0.703
⋮	⋮	⋮
第 24 主成分	0.000	1.000
獲得成績数		
第 1 主成分	0.311	0.311
第 2 主成分	0.105	0.416
第 3 主成分	0.075	0.491
第 4 主成分	0.066	0.557
第 5 主成分	0.060	0.616
第 6 主成分	0.051	0.668
第 7 主成分	0.045	0.712
⋮	⋮	⋮
第 24 主成分	0.000	1.000
必修科目不合格数		
第 1 主成分	0.631	0.631
第 2 主成分	0.179	0.810
⋮	⋮	⋮
第 4 主成分	0.073	1.000
偏差値別講義数		
第 1 主成分	0.346	0.346
第 2 主成分	0.204	0.550
第 3 主成分	0.131	0.681
第 4 主成分	0.099	0.779
⋮	⋮	⋮
第 12 主成分	0.000	1.000
出席低偏差講義数		
第 1 主成分	0.668	0.668
第 2 主成分	0.157	0.825
⋮	⋮	⋮
第 4 主成分	0.075	1.000

表 4.63: 各主成分の固有ベクトル (獲得成績数)(2年次後期終了時点)

変数名	主成分番号						
	1	2	3	4	5	6	7
1年次前期秀獲得数	0.343	0.302	-0.054	-0.175	0.395	-0.325	0.333
1年次前期優獲得数	0.112	-0.371	-0.524	-0.057	-0.470	-0.056	-0.159
1年次前期良獲得数	-0.152	-0.148	0.330	0.075	0.156	0.545	0.233
1年次前期可獲得数	-0.200	0.103	0.200	0.134	-0.072	-0.135	-0.352
1年次前期不可獲得数	-0.090	0.095	0.046	0.023	-0.011	-0.028	-0.036
1年次前期失格獲得数	-0.013	0.024	0.002	0.004	0.006	-0.002	-0.012
1年次後期秀獲得数	0.424	0.316	0.011	-0.179	-0.085	0.073	-0.128
1年次後期優獲得数	0.196	-0.296	-0.398	0.315	0.365	0.240	0.112
1年次後期良獲得数	-0.124	-0.398	0.284	-0.412	-0.210	-0.361	0.331
1年次後期可獲得数	-0.278	0.111	-0.101	0.245	0.042	-0.028	-0.066
1年次後期不可獲得数	-0.196	0.236	0.165	0.030	-0.103	0.080	-0.256
1年次後期失格獲得数	-0.019	0.025	0.032	-0.003	-0.009	0.004	-0.007
2年次前期秀獲得数	0.334	0.055	-0.153	-0.084	-0.153	0.244	0.024
2年次前期優獲得数	0.216	-0.221	0.068	0.547	0.068	-0.373	-0.174
2年次前期良獲得数	-0.136	-0.144	0.409	-0.511	0.409	0.135	-0.464
2年次前期可獲得数	-0.284	0.149	-0.077	0.078	-0.077	0.059	0.423
2年次前期不可獲得数	-0.124	0.179	-0.083	0.013	-0.083	-0.065	0.074
2年次前期失格獲得数	-0.043	0.077	-0.090	-0.004	-0.090	-0.010	0.030
2年次後期秀獲得数	0.266	0.162	-0.119	-0.001	-0.119	0.063	-0.168
2年次後期優獲得数	0.157	-0.230	-0.160	-0.030	-0.160	0.309	0.006
2年次後期良獲得数	-0.072	-0.228	0.345	-0.043	0.345	-0.225	-0.111
2年次後期可獲得数	-0.198	0.043	0.105	-0.059	0.105	0.002	-0.053
2年次後期不可獲得数	-0.172	0.193	-0.047	0.002	-0.047	-0.011	0.040
2年次後期失格獲得数	-0.068	0.107	-0.087	0.022	-0.087	0.017	0.011

表 4.64: 各主成分の固有ベクトル (必修科目不合格数)(2年次後期終了時点)

変数名	主成分番号	
	1	2
1年次前期必修科目不合格	-0.226	0.219
1年次後期必修科目不合格	-0.523	0.785
2年次前期必修科目不合格	-0.531	-0.501
2年次後期必修科目不合格	-0.636	-0.292

表 4.65: 各主成分の固有ベクトル (偏差値別講義数)(2年次後期終了時点)

変数名	主成分番号			
	1	2	3	4
1年次前期#(S_{low})	-0.313	0.281	0.002	-0.436
1年次前期#(S_{mid})	0.205	-0.333	-0.548	0.373
1年次前期#(S_{high})	0.105	0.052	0.547	0.064
1年次後期#(S_{low})	-0.438	0.305	-0.203	0.134
1年次後期#(S_{mid})	0.345	-0.357	-0.164	-0.418
1年次後期#(S_{high})	0.099	0.037	0.356	0.283
2年次前期#(S_{low})	-0.454	-0.213	-0.060	0.180
2年次前期#(S_{mid})	0.299	0.344	-0.166	-0.436
2年次前期#(S_{high})	0.099	-0.030	0.207	0.158
2年次後期#(S_{low})	-0.403	-0.154	-0.109	-0.166
2年次後期#(S_{mid})	0.232	0.630	-0.294	0.336
2年次後期#(S_{high})	0.080	-0.014	0.183	0.079

表 4.66: 各主成分の固有ベクトル (出席低偏差講義数)(2年次後期終了時点)

変数名	主成分番号	
	1	2
1年次前期#(S_{att})	-0.260	0.528
1年次後期#(S_{att})	-0.515	0.627
2年次前期#(S_{att})	-0.673	-0.521
2年次後期#(S_{att})	-0.462	-0.238

表 4.67: 各主成分の意味付け (2年次後期終了時点)

区分名	主成分番号	意味付け
講義区分別 GPA	第1主成分	専門の悪さ
	第2主成分	2年次後期の理系基礎の悪さ
	第3主成分	人間社会の良さ
	第4主成分	2年次の成績の良さ
	第5主成分	全体的な成績の良さ
	第6主成分	1年次前期の人間社会の良さ
獲得成績数	第1主成分	秀・優獲得の多さ
	第2主成分	不可・失格獲得の多さ
	第3主成分	1年次前期の良・可獲得の多さ
	第4主成分	2年次前期の優獲得の多さ
	第5主成分	成績の緩やかな低下
	第6主成分	1年次前期の良獲得の多さ
	第7主成分	成績の激しい低下
必修科目不合格数	第1主成分	必修科目不合格の少なさ
	第2主成分	1年次の必修科目不合格の多さ
偏差値別講義数	第1主成分	低偏差の少なさ
	第2主成分	2年次の中偏差の多さ
	第3主成分	中偏差の少なさ
	第4主成分	高偏差の少なさ
出席低偏差講義数	第1主成分	出席率の高さ
	第2主成分	1年次の出席率の低さ

表 4.68 で、それぞれの変数選択法の結果、得られた説明変数と回帰係数を示す。

表 4.68: 説明変数と回帰係数 (2 年次後期終了時点)

変数名	回帰係数	
	強制投入法	ステップワイズ法
定数項	2.98	2.46
講義区分別 GPA1	3.80	2.87
講義区分別 GPA2	0.441	0.559
講義区分別 GPA3	0.629	0.584
講義区分別 GPA4	0.646	不採用
講義区分別 GPA5	-0.109	不採用
講義区分別 GPA6	-1.38	-1.49
獲得成績数 1	3.01	2.21
獲得成績数 2	-0.331	不採用
獲得成績数 3	-0.761	-0.464
獲得成績数 4	-0.574	-0.391
獲得成績数 5	0.313	不採用
獲得成績数 6	-1.02	-0.805
獲得成績数 7	2.07	1.55
必修科目不合格数 1	-1.46	-0.978
必修科目不合格数 2	1.28	0.932
偏差値別講義数 1	-1.47	-1.08
偏差値別講義数 2	-0.116	不採用
偏差値別講義数 3	-1.99	-1.49
偏差値別講義数 4	-1.34	-1.05
出席低偏差講義数 1	0.598	0.408
出席低偏差講義数 2	0.051	不採用

表 4.69: 予測精度 (2 年次後期終了時点)

変数選択	閾値	正解率			再現率			的中率			F 値
		対象	的中	割合	対象	的中	割合	対象	的中	割合	
強制投入法	50%	302	294	97.4%	25	20	80%	23	20	87%	0.833
	30%	302	294	97.4%	25	24	96%	31	24	77.4%	0.857
	8.3%	302	282	93.4%	25	24	96%	43	24	55.8%	0.706
ステップワイズ法	50%	302	292	96.7%	25	18	72%	21	18	85.7%	0.783
	30%	302	292	96.7%	25	23	92%	31	23	74.2%	0.821
	8.3%	302	281	93%	25	24	96%	44	24	54.5%	0.696

考察

学生データを区分に分け、それぞれに対して主成分分析を行い、説明変数の数を1/3程度に減らして検証を行った。検証2と比べて、さらに的中率とF値の向上が見られてた。変数選択による大きな差異は見られなかった。説明変数として用いた各主成分は、区分ごとに主成分分析を行ったため感覚的にわかりやすくなっている。

4.3 実験結果と考察

実験結果

3つの手法で分析を行った。それぞれの結果と従来研究 [6] で用いられていた、ベイジアンネットワークの出力を Leave-One-Out 法 (以下, 比較手法) で検証して得られた F 値を比較した結果を表 4.70, 表 4.71, 表 4.72, 表 4.73 にそれぞれ示す。比較手法で用いた説明変数は, 本研究と同様に変換された学生データが用いられている。変数選択には, 関連ベース特徴選択を用いている。

表 4.70: 結果の比較 (1 年次前期終了時点)

閾値	検証 1		検証 2		検証 3		比較手法
	強制投入	ステップワイズ	強制投入	ステップワイズ	強制投入	ステップワイズ	
50%	0.032	0.032	0.148	0.077	0.032	0.032	0.351
30%	0.078	0.081	0.270	0.222	0.078	0.081	0.344
8.3%	0.088	0.109	0.267	0.285	0.088	0.109	0.283

表 4.71: 結果の比較 (1 年次後期終了時点)

閾値	検証 1		検証 2		検証 3		比較手法
	強制投入	ステップワイズ	強制投入	ステップワイズ	強制投入	ステップワイズ	
50%	0.114	0.114	0.138	0.071	0.313	0.353	0.406
30%	0.111	0.112	0.318	0.372	0.455	0.449	0.451
8.3%	0.105	0.104	0.330	0.345	0.383	0.400	0.424

表 4.72: 結果の比較 (2 年次前期終了時点)

閾値	検証 1		検証 2		検証 3		比較手法
	強制投入	ステップワイズ	強制投入	ステップワイズ	強制投入	ステップワイズ	
50%	0.181	0.160	0.378	0.389	0.579	0.595	0.484
30%	0.178	0.152	0.522	0.512	0.612	0.610	0.492
8.3%	0.168	0.152	0.408	0.416	0.500	0.494	0.487

表 4.73: 結果の比較 (2 年次後期終了時点)

閾値	検証 1		検証 2		検証 3		比較手法
	強制投入	ステップワイズ	強制投入	ステップワイズ	強制投入	ステップワイズ	
50%	0.196	0.190	0.727	0.756	0.833	0.783	0.656
30%	0.210	0.178	0.720	0.692	0.857	0.821	0.597
8.3%	0.193	0.155	0.595	0.563	0.706	0.696	0.541

考察

変数選択の方法に関しては、大きな差は見られなかった。ただし、一番F値が良かった2年次後期終了時点の検証3においては、強制投入法が優れていた。

検証法1-3に関しては、すべての時期において、

検証3 > 検証2 > 検証1

という結果が得られた。特に検証1と2の間には大きな差が見られた。

従来研究の手法と、本研究で一番結果が優れていた検証3との比較に関しては、1年次前期終了時点では従来研究が優れていたが、1年次後期終了時点でほぼ並び、2年次では3種類すべての閾値において本研究が優れていた。閾値ごとの差に関しては、多くの時期・手法において、

30% > 50% > 事前確率(8.3%)

という結果が得られた。

今回の検証では、F値を再現率と適合率の等分の調和平均としていた。しかし指標となる値は必ずしも等分である必要はない。例えば、再現率に一定の係数を掛けて重み付けをした上で調和平均をとることもできる。この場合は、要注意学生をなるべく多く予測したいときに有効な方法となる。逆に、適合率に重み付けを行った場合は、指導対象を減らしたいときに有効となる。つまり、大学の指導体制に合わせた指標を用いることが重要になるといえるであろう。

回帰係数と主成分の意味付けから、要注意学生となる学生の傾向がわかるはずである。例えば、検証3の「2年次後期までのデータを用いた推定」の強制投入法による結果を例に示す。「講義区分別GPA第1主成分」の回帰係数は3.80であり、意味付けは「専門の悪さ」としている。これは、専門科目のGPAが悪くなると要注意学生となる確率が上がることを意味している。この結果は一般的に予想できる傾向である。そこで予想外であった例を1つ示す。「出席低偏差講義数第1主成分」の回帰係数は0.598であり、意味付けは「出席率の高さ」としている。これは、出席率が高くなると要注意学生となる確率が上がることを意味している。この結果は一般の予想と逆の傾向のように思われるかもしれない。しかし、要注意学生の推定モデルは複数の説明変数が要素となっている。つまり要注意学生の傾向は複数の要素の組み合わせを見る必要があると考えられる。ただし、説明変数が多く正例クラスが少ないため、今回のデータにのみ適したモデルになってしまっている可能性もある。今回の検証では行えなかったが、ロジスティック回帰分析の出力をLeave-One-Out法で検証することで今回の手法の汎用性がわかるのではないかと考えられる。

第5章 むすび

本研究では、過去のある年度に名工大に入学した学生のデータから、将来的に留年・退学してしまう学生の推定・検証をロジスティック回帰分析を用いて行った。第2章では、本研究で用いた分析方法と、その入力となる説明変数の変換法・選択法、検証法についてを説明した。第3章では、用いた学生データの概要と、分析に使いやすくするためのデータ変換について述べた。第4章では、実際に変換したデータを使い、ロジスティック回帰分析を用いた3つの手法で要注意学生の推定・検証を行った。

その結果、従来研究と比較し、1年次後期終了時点以降での推定精度向上が確認できた。特に2年次後期終了時点までのデータを用いた分析・検証では、指導対象候補31人を抽出した。その内には実際の要注意学生24人含まれており要注意学生の96%推定することができた。また、変数選択と閾値についても比較を行い考察した。

今後の課題としては、第4章の考察で述べたモデルの汎用性の検証や、用いるデータの種類や量を増やしての検証が挙げられる。本研究で用いた手法が、実際に現在の本校の学生や他大学の学生の要注意学生の推定に有用であるかは、さらにデータを増やして検証する必要がある。また、本研究や従来研究では用いられていないデータマイニングの手法であったり、データの変換により精度向上が可能であるかもしれない。そして将来的には、より精度の高い予測システムが開発・実用され、要注意学生の早期発見・修学指導が行われることにより、留年・退学してしまう学生が減ることを期待している。

謝辞

本研究を進めるにあたって、日頃から多大な御尽力を頂き、ご指導を賜りました名古屋工業大学 舟橋健司 准教授，伊藤宏隆 助教に心から感謝致します。また，本研究の実験データの提供元である，出欠システム及びコースマネジメントシステムの開発に尽力されました，名古屋工業大学 情報基盤センター長 松尾啓志 教授，内匠逸 教授，情報基盤センター教職員の皆様に深く感謝致します。最後に，本研究に多大な御協力を頂きました舟橋研究室諸氏に心から感謝致します

参考文献

- [1] 原圭司, 高橋健一, 上田祐彰: “ベイジアンネットワークを用いた授業アンケートからの学生行動モデルの構築と考察”, 情報処理学会論文誌, Vol.51, No.4, pp.1215?1226, 2010.
- [2] 寿真田崇志, 松本哲也, 大西昇: “e-Learning におけるベイジアンネットワークを用いた学習者特性の推定”, 電子情報通信学会技術研究報告, Vol.106, No.583, pp.203?208, 2007.
- [3] 伊藤圭佑, 舟橋健司, 伊藤宏隆: “データマイニングによる『要注意学生』の発見に関する研究”, 平成 25 年度名古屋工業大学修士論文, 2013.
- [4] 平田大智, 舟橋健司, 伊藤宏隆: “ベイジアンネットワークによる要注意学生の半期毎の発見精度に関する検証実験”, 平成 26 年度名古屋工業大学卒業研究論文, 2014.
- [5] 稲垣諒, 舟橋健司, 伊藤宏隆: “変数を見直したベイジアンネットワークによる要注意学生の発見手法に関する研究”, 平成 26 年度名古屋工業大学卒業研究論文, 2014.
- [6] 西脇雅弥, 舟橋健司, 伊藤宏隆: “教育支援を目的とした要注意学生の推定精度改善法”, 平成 27 年度名古屋工業大学修士論文, 2015.
- [7] Linda DeAngelo, Ray Franke, Sylvia Hurtado, John H. Pryor, Serge Tran: “Completing College: Assessing Graduation Rates at Four-Year Institutions”, <https://www.heri.ucla.edu/DARCU/CompletingCollege2011.pdf>, 2011
- [8] Microsoft Corporation: “Excel 2010”, <https://products.office.com/ja-jp/excel>, 2010
- [9] NTT データ数理システム: “Visual Mining Studio”, <https://www.msi.co.jp/vmstudio/>, 2017 年 1 月 20 日更新, 2017 年 1 月 20 日参照

付録A VMSの操作方法について

本研究では、ロジスティック回帰分析および主成分分析を用いるためにシェアウェアである Visual Mining Studio(以下, VMS) を採用している。VMS はプロジェクトボードと呼ばれるエリアに、データアイコン・処理アイコン・フロー矢印を設置することでフロー処理を行うことができるソフトである。ただし、VMS の標準機能にはロジスティック回帰分析が含まれていないため、フリーウェアである R と連携することで拡張機能として利用できるようにしている。本研究で用いた VMS は ver8.2, R は ver3.2.3 である。

実際の手順

本研究で一番優れた結果が得られた手法3を例に使い方を紹介していく。

1. あらかじめ Excel で「要注意学生かどうか(以下, SRG)(0/1)」と「変換した学生データ」を対応させたデータセットを作成し、VMS で読み込みデータアイコンを設置する。
2. 「データ操作→サンプリング」を設置し、1 と矢印で繋ぐ。サンプリングのオプションを開き、用いる変数を選択する。
3. 「Visual Mining Studio →多変量解析→主成分分析」を設置し、2 と繋ぐ。主成分分析のオプションを開き、用いる変数を選択する。この時、SRG は選択しない。モデルパラメータは「covariance」、主成分は「全件」を選択する。
4. スクリプトアイコンを設置し、2 と 3 それぞれと矢印で繋ぐ。2 の出力が複数あるので「負荷率」を選択する。この時、2 の他の出力の「重要度」で寄与率・累計寄与率が出力されるので確認して用いる主成分を決定しておく。スクリプトの内容は、2 の出力の変数の値と 3 の出力の各主成分の固有ベクトルの値を乗算して出力するものである。
5. 2 から 4 までを変数を分けた区分別に行う。(5 つの区分に分けたのでスクリプトアイコンは 5 つできる)
6. スクリプトアイコンを設置し、5 で作成した区分別のスクリプトアイコンそれぞれと矢印で繋ぐ。スクリプトの内容は、入力を結合して出力するものである。
7. 「Visual Mining Studio →外部ソフト→R →回帰分析 R →多項ロジスティック回帰 R」を設置し、6 と矢印で繋ぐ。多項ロジスティック回帰 R のオプションを開き、用い

る変数選択法を選択する。本研究で行った強制投入法の場合、変数選択を「ユーザー指定」にして説明変数の選択で SRG 以外の変数にチェックを入れる。ステップワイズ法の場合、変数選択を「ステップワイズ法」にして説明変数の初期選択で SRG 以外の変数にチェックを入れる。どちらも、目的変数は「SRG」、重み付けは「なし」、定数項を含むモデルにチェックを入れる。

8. スクリプトアイコンを設置し、7 と矢印で繋ぐ。7 の出力が複数あるので「result」を選択する。スクリプトの内容は、入力「X1. 確率」列の値をあらかじめ決めておいた閾値と比較してその値より大きければ、SRG. 予測に 1 を、その値以下であれば 0 を代入する。本研究では 3 つの閾値を用いているのでそれぞれの閾値ごとに SRG. 予測を計算し出力する。
9. 「データ操作→集計」を設置し、8 のそれぞれの出力と繋ぐ。集計のオプションを開き、集計方法を「サマリ」、サマリ項目を「件数」、集計キー列名の「SRG と SRG. 予測」にチェックを入れ、集計対象列名の「SRG」にチェックを入れる。マトリックス出力形式のチェックは外す。
10. スクリプトアイコンを設置し、9 のそれぞれと矢印で繋ぐ。スクリプトの内容は、入力から 2 値分類問題を解き、正解率・再現率・適合率・F 値の計算を行うものである。
11. 最後に、「表示→データ&グラフビュー」を設置し、10 と矢印で繋ぐ。グラフビューを実行すると結果が画面に出力される。

以上が本研究で行った分析の一例についての説明である。スクリプトは、NTT 数理システム独自のものであるため VMS 付属のマニュアルを参考されたい。