

最適線形判別関数の応用（1）

- 2010年度統計入門中間試験データの分析 -

新 村 秀 一

1. はじめに

統計的判別分析は、線形分離可能（Minimum Number of Misclassifications, MNM=0）なデータを認識できず、変数選択法に問題があることを指摘してきた[7]。線形分離可能なデータとしては、スイス銀行紙幣データ[13]が有名である。このデータは6個の説明変数をもっているが、2個の説明変数（ X_4, X_6 ）でMNM=0になる。MNMの単調減少性から、（ X_4, X_6 ）を含むすべてのモデルでMNM=0であることを新村は初めて指摘した。そして、統計的な判別関数の変数選択法が正しく機能しないことが分かった[6]。

多くの統計家にとって、MNM=0のデータとして判別分析の評価に用いられるFisherのアイリスデータがある[12][17]。セトサ、パーシクル、バージニカという3種類のアイリス各50件の4個の計測値のデータである。セトサが、他の2群から完全に線形分離可能であることが散布図からわかる。そこで筆者を含め統計の研究者は、深く考えることなくセトサを分析から省いて、パーシクルとバージニカの2群判別として扱ってきた。このデータは、4個の説明変数しかなく、スイス銀行紙幣データのように線形分離可能なことに関する「MNMの単調減少性」などの重要な新事実は何も発見できない。

大学の教員にとって身近なMNM=0のデータとして、試験の合否判定（合格群と不合格群）がある。例えば、3科目の合計得点でもって200点以上を合格と判定することは、 $f(x, y, z) = x + y + z - 200$ という自明な判別関数を考えれば、 $f(x, y, z) \geq 0$ であれば合格、 $f(x, y, z) < 0$ であれば不合格になる。すなわち、線形分離可能な代表的なデータであるが、自明なため判別分析の研究対象として意識してこなかった。

一方、個々の試験を考えれば、各設問の得点を説明変数と考えて合否判定を行う場合も、線形分離可能なデータになる。

2010年度から、成蹊大学の経済学部では統計入門を1年生全員に必修とした。筆者も127名の受講生に1変量と2変量（相関、単回帰、分割表）の授業を行うことになった。中間試験と期末試験を10択100問のマークシート試験で行うことにした。授業並びに試験内容の分析は[9][10]を参考にしてほしい。この試験結果の正解/不正解を0/1の100個の説明変数とした判別と、設問群ごとに得点を集計したものを説明変数とした4種類の判別分析を行い検討する。通常の統計分析では、100個の説明変数の判別分析を行うことは稀である。世界最高峰の統計ソフトの

JMP[3][5][15]も、多くの判別手法で分析結果に不満が残った。これはJMPの問題ではなく、判別アルゴリズムの問題でないかと考えている。

比較する手法は、改定IP-OLDFと統計的判別分析（回帰分析によるLDF、2次判別分析、ロジスティック回帰、決定木分析）とハードマージン最大化とソフトマージン最大化SVM（H-SVM, S-SVM）である[18]。統計的判別分析は統計ソフトのJMPを、それ以外は数理計画法ソフトのWhat'sBest!を用いた。LINGOを用いなかったのは、はじめて分析する「合否判定データ」の分析の詳細を検討したいためである[4][16]。

2. 中間試験（100問）の分析

2.1 概略

中間試験の10択100問の正解／不正解を0/1で表わした線形分離可能なデータの分析を試みる。

表2.1は、10%点から90%点まで10%点刻みでまとめた得点の度数表である。46点が9.2%であり、48点以上と46点以下で合否判定を行ったものを「合否10」とする。そして、59点以上と58点以下で合否判定したものを「合否30」、66点以上と65点以下で合否判定したものを「合否50」とする。紙面の都合で、「合否10」と「合否50」の結果のみを示す。

表2.1 得点の度数表

水準	度数	割合	累積割合
31	1	0.008	0.008
46	1	0.008	0.092
48	3	0.025	0.117
54	1	0.008	0.183
55	3	0.025	0.208
58	4	0.033	0.275
59	5	0.042	0.317
62	6	0.050	0.408
65	5	0.042	0.492
66	3	0.025	0.517
70	4	0.033	0.617
75	2	0.017	0.708
79	4	0.033	0.808
82	3	0.025	0.917
91	1	0.008	0.992
93	1	0.008	1
合計	120	1	1

図2.1は、主成分分析の固有値を30個まで出力した。第1主成分の固有値は13.9と大きく、累積寄与率は14.5%である。第2主成分以降はほぼ直線的に減少して、29変数までは固有値が1以上で、累積寄与率は79.2%である。ただし、X9からX12の4変数は、全員が正解であり、一定の値1をとっているの、分析から省いている。

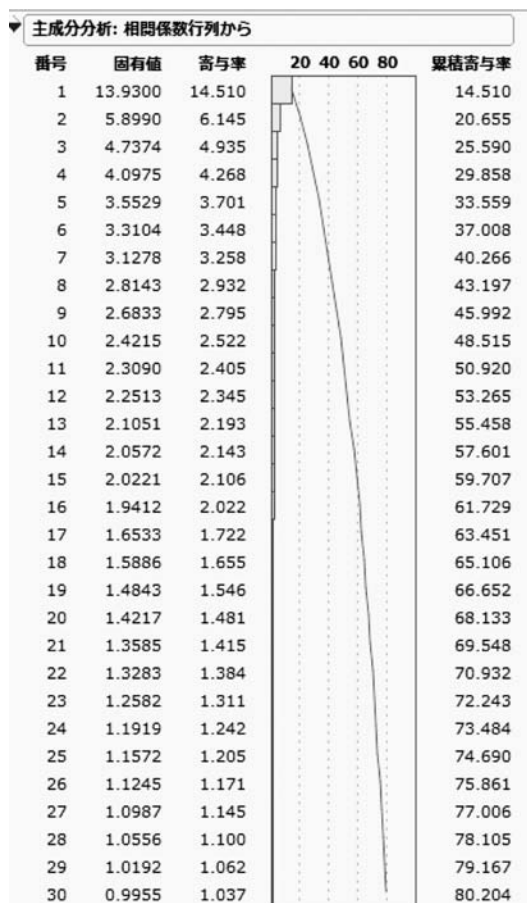


図2.1 固有値

図2.2は因子負荷量プロットで、第1主成分軸をはさんで、4象限から1象限に布置している。すなわち、第1主成分軸は、正は合格者群に負は不合格者群に対応している。

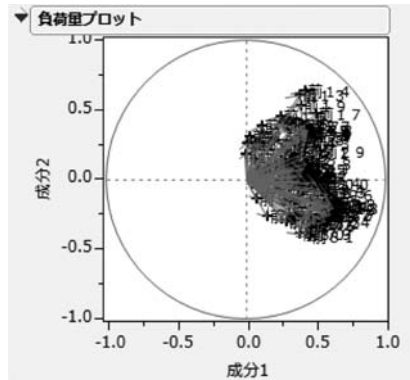


図2.2 因子負荷量プロット

図2.3は、スコアプロットである。0は「合否10」の不合格の学生であり、第2象限から第3象限で第1主成分がほぼ-5以下に布置している。期末試験のスコアプロットと異なり、散布図で線形分離可能なことをうかがわせる。統計的判別関数は、スイス銀行紙幣データで線形分離可能(MNM=0)なデータであることを認識できない。その際の統計家の反応として、線形分離可能なデータは散布図あるいは主成分分析のスコアプロットで把握できるのではないかという意見もみられた。確かに図2.3はそれができる可能性を示す判例であるが、全ての事例で保障されているわけでない。一般的に、3次元以上でMNM=0の場合には、散布図やスコアプロットで線形分離可能な事実は発見できない。

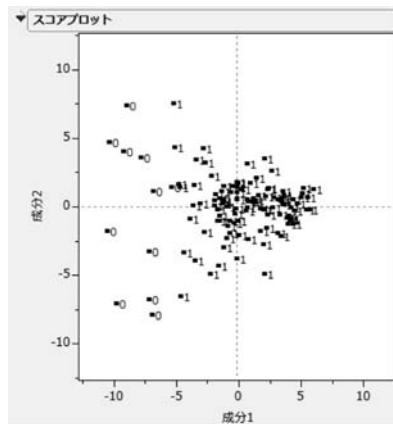


図2.3 スコアプロット

2.2 合否判定を10%点とする場合

「合否10」を目的変数とし、100個の設問を説明変数とした判別を考える。

(1) 逐次変数選択法

表2.2は、合否を1/0の目的変数として、逐次変数増加法と減少法を行った。「Var」と「F/B」列は逐次取り込まれたあるいは掃き出された変数を示す。変数増加法は「X36」から「X41」まで30変数が取り込まれ停止する。その後は、変数減少法でX9からX12までの全員正答の4変数を省いた96変数をフルモデルとしている。「X46」から「X90」まで46個が掃き出され、50変数のモデルが選ばれる。

その後、「p値」、「偏差平方和」、「決定係数」、「Cp値」である。Cp値から4変数のモデルが選ばれる。逐次F検定が30変数を選んだのに対して、26変数少ないモデルを選んだ。これまで研究に用いてきた4種類の実データは、学生データが5変数、アイリスデータが4変数、銀行データが6変数、CPDデータが19変数とそれほど多くなかったので、逐次F検定の結果とAICやCp統計量が選ぶモデルの次数(説明変数の個数)に大きな違いがなかった。説明変数が100個と多くなると、今回のように各変数選択法の選ぶモデルに大きな違いが出てくるか否かは、現時点でははっきりしない。

「LDF」は、JMPの判別分析で求めた誤分類数である。1変数では11個で、8変数で0個になっている。すなわち、LDFでも100個の設問のうち8問で合否判定ができる。これは、説明変数の合計が目的変数の合否と良く対応しているからであろう。逐次変数選択法は、線形分離可能である8変数より22個多い変数を選んだことになる。すなわち、逐次F検定による変数選択法の妥当性が疑われる。

「QD」は2次判別関数であり、7変数目で線形分離可能である。期末試験では、2次判別関数に問題が生じたのと対照的である。

表2.2 逐次変数増加法と減少法

step	Var.	F/B	P値	SS	R2	Cp	LDF	QD	Logi	IP	SVM	決定木
1	X36	追加	0.000	2.643	0.264	91.354	11	11	11	11	11	11
2	X82	追加	0.000	1.771	0.442	43.384	19	19	19	7	7	5
3	X25	追加	0.000	0.861	0.528	21.086	3	19	13	3	3	4
4	X96	追加	0.000	0.727	0.601	2.580	5	5	4	3	3	4
5	X17	追加	0.009	0.236	0.624	-2.071	7	9	3	2	2	
6	X61	追加	0.007	0.235	0.648	-6.708	2	1	0	0	0	
7	X38	追加	0.015	0.182	0.666	-9.839	2	0				
8	X37	追加	0.006	0.220	0.688	-14.050	0	0				
9	X1	追加	0.011	0.180	0.706	-17.129	0					
10	X59	追加	0.012	0.167	0.723	-19.847	0					

step	Var.	F/B	P値	SS	R2	Cp	LDF	QD	Logi	IP	SVM	決定木
11	X3	追加	0.021	0.133	0.736	-21.602	0					
12	X65	追加	0.067	0.082	0.744	-21.909	0					
13	X72	追加	0.071	0.078	0.752	-22.104	0					
14	X99	追加	0.040	0.098	0.762	-22.859	0					
15	X53	追加	0.063	0.078	0.770	-23.059	0					
16	X42	追加	0.066	0.075	0.777	-23.173	0					
17	X43	追加	0.141	0.047	0.782	-22.499	0					
18	X97	追加	0.138	0.047	0.787	-21.833	0					
19	X32	追加	0.180	0.038	0.790	-20.912	0					
20	X31	追加	0.044	0.084	0.799	-21.296	0					
21	X81	追加	0.085	0.060	0.805	-20.994	0					
22	X35	追加	0.100	0.054	0.810	-20.519	0					
23	X21	追加	0.124	0.046	0.815	-19.830	0					
24	X67	追加	0.147	0.041	0.819	-18.981	0					
25	X16	追加	0.120	0.046	0.824	-18.285	0					
26	X20	追加	0.118	0.046	0.828	-17.584	0					
27	X62	追加	0.157	0.037	0.832	-16.631	0					
28	X69	追加	0.175	0.034	0.835	-15.586	0					
29	X26	追加	0.167	0.035	0.839	-14.568	0					
30	X41	追加	0.180	0.032	0.842	-13.482	0					
1	X46	削除	0.974	0.000	0.918	95.001						
2	X48	削除	0.969	0.000	0.918	93.003						
3	X73	削除	0.968	0.000	0.918	91.004						
4	X53	削除	0.961	0.000	0.918	89.006						
5	X57	削除	0.942	0.000	0.918	87.011						
6	X42	削除	0.930	0.000	0.918	85.017						
7	X16	削除	0.930	0.000	0.918	83.023						
8	X70	削除	0.868	0.001	0.918	81.045						
9	X33	削除	0.873	0.001	0.918	79.064						
10	X96	削除	0.820	0.001	0.918	77.102						
11	X30	削除	0.799	0.002	0.918	75.148						
12	X8	削除	0.758	0.002	0.918	73.214						
13	X64	削除	0.720	0.003	0.917	71.301						
14	X63	削除	0.895	0.000	0.917	69.312						
15	X84	削除	0.747	0.002	0.917	67.379						
16	X45	削除	0.656	0.004	0.917	65.503						
17	X47	削除	0.707	0.003	0.916	63.589						
18	X29	削除	0.630	0.005	0.916	61.728						
19	X39	削除	0.634	0.005	0.915	59.861						
20	X3	削除	0.667	0.004	0.915	57.968						
21	X74	削除	0.617	0.005	0.914	56.109						

step	Var.	F/B	P値	SS	R2	Cp	LDF	QD	Logi	IP	SVM	決定木
22	X78	削除	0.634	0.004	0.914	54.235						
23	X86	削除	0.556	0.007	0.913	52.424						
24	X66	削除	0.580	0.006	0.913	50.590						
25	X68	削除	0.595	0.005	0.912	48.740						
26	X49	削除	0.515	0.008	0.911	46.961						
27	X43	削除	0.517	0.008	0.911	45.179						
28	X40	削除	0.501	0.008	0.910	43.410						
29	X35	削除	0.531	0.007	0.909	41.608						
30	X34	削除	0.505	0.008	0.908	39.830						
31	X97	削除	0.459	0.010	0.907	38.101						
32	X100	削除	0.427	0.011	0.906	36.410						
33	X15	削除	0.383	0.013	0.905	34.781						
34	X75	削除	0.370	0.014	0.904	33.171						
35	X95	削除	0.461	0.009	0.903	31.434						
36	X85	削除	0.411	0.012	0.902	29.759						
37	X5	削除	0.393	0.012	0.900	28.108						
38	X6	削除	0.622	0.004	0.900	26.223						
39	X41	削除	0.315	0.017	0.898	24.697						
40	X55	削除	0.325	0.016	0.897	23.153						
41	X54	削除	0.397	0.012	0.895	21.490						
42	X67	削除	0.318	0.017	0.894	19.957						
43	X69	削除	0.265	0.021	0.892	18.541						
44	X93	削除	0.258	0.021	0.890	17.143						
45	X92	削除	0.418	0.011	0.888	15.453						
46	X90	削除	0.372	0.013	0.887	13.826						

(2) ロジスティック回帰

「Logi」列は、ロジスティック回帰の誤分類数 (Number of Misclassifications, NM) である。誤分類数は11個から始まり6変数で0になっている。図2.4は、6変数のNM=0の分析結果である。全変数が収束計算で不安定になっている。推定値に対して標準誤差は大きく、推測統計学上このモデルを採択すべきでない。NM=0のデータでロジスティック回帰に問題があることはすでに指摘されている[1][14]。

名義ロジスティックのあてはめ合否 10

反復の履歴

モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
差	36.765242	6	73.53048	<.0001*
完全	4.5792e-8			
縮小	36.765242			
R2乗(U)			1.0000	
オブザベーション(または重みの合計)			120	

目的関数基準による収束

あてはまりの悪さ(LOF)

要因	自由度	(-1)*対数尤度	カイ2乗	p値(Prob>ChiSq)
あてはまりの悪さ(LOF)	20	4.5792e-8	9.158e-8	
飽和モデル	26	0		
あてはめたモデル	6	4.5792e-8	1.0000	

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切片 不安定	140.155104	323324.48	0.00	0.9997
前3 6 不安定	-37.933455	14429.793	0.00	0.9979
前8 2 不安定	-75.72733	21072.849	0.00	0.9971
前2 5 不安定	-48.468731	336564.5	0.00	0.9999
前9 6 不安定	-38.51828	15508.418	0.00	0.9980
前1 7 不安定	-34.776214	98405.253	0.00	0.9997
前6 1 不安定	-38.195312	12696.629	0.00	0.9976

推定値は次の対数オッズに対するものです: 0/1

図2.4 12変数のMNM=0の分析結果

(3) 改定IP OLDFとS-SVM

「IP」列は、改定IP-OLDFで求めたMNMである。1変数の11個から始まって6変数で0になっている。48点以上が合格であることを考えると、僅か6個の設問で合否判定ができたことになる。表2.3は、判別係数と誤分類数を示す。X17の判別係数が0であるから、実際には5変数以下で線形分離可能であることが分かる。これは、逐次変数増加法で選ばれたモデルを順次適用したためである。すなわち ${}_{100}C_5$ 個の5変数モデル、さらに ${}_{100}C_4$ 、 ${}_{100}C_3$ 、 ${}_{100}C_2$ 、 ${}_{100}C_1$ 個の全ての説明変数の組み合わせを検討しないとMNM=0の最小次元は分からない。

表2.3 判別係数と誤分類数

p	X17	X25	X36	X61	X82	X96	定数項	誤分類数
1			2.00000000				-1.00000000	11
2			2.00000000		2.00000000		-1.00000000	7
3		2.00000000	2.00000000		2.00000000		-3.00000000	3
4		1.00000000	1.00000000		1.00000000	1.00000000	-2.00000000	3
5	2.00000000	2.00000000	2.00000000		2.00000000	2.00000000	-5.00000000	2
6	0.00000000	4.00000000	2.00000000	2.00000000	4.00000000	2.00000000	-7.00000000	0

表2.4は、合否と6問の得点の分割表である。列の度数は、「判別得点 = $0 * x_{17} + 4 * x_{25} + 2 * x_{36} + 2 * x_{61} + 4 * x_{82} + 2 * x_{96} - 7$ 」を表している。行の0は不合格であり、1は合格を表す。両群とも $y_i f(x_i) \geq 1$ のSVで完全に判別される。すなわち判別得点が0上にケースはなく誤分類数は0である。

表2.4 合否と6点の得点の分割表

度数	-5	-3	-1	1	-3	5	7	
0	4	1	6	0	0	0	0	11
1	0	0	0	8	19	28	54	109
	4	1	6	8	19	28	54	120

「SVM」列は、S-SVMの誤分類数である。改定IP-OLDFと同じである。表2.5は、S-SVMの判別係数と誤分類数である。判別超平面上にケースはなかった。

表2.5 S-SVMの判別係数と誤分類数

p	X17	X25	X36	X61	X82	X96	定数項	誤分類数
1			0				0	11
2			2		2		-1	7
3		2	2		2		-3	3
4		2	2		2	0	-3	3
5	2	2	2		2	2	-5	2
6	2	2	2	2	4	2	-7	0

(4) パーティション

図2.5は、「合否10」を目的変数として、決定木分析(JMPではパーティション)を行った。分岐は4回行われ5個の葉ノードで停止した。各葉ノードの応答の割合の大きな方に判別することになると、X96で1回目の分岐で誤分類数は11, X25で2回目の分岐で誤分類数は5, X40で3回目の分岐で誤分類数は4, X40で4回目の分岐で誤分類数が4で停止している。結局4変数で誤分類数は4であり、LDFは誤分類数が5とわずかであるが決定木分析より悪かった。

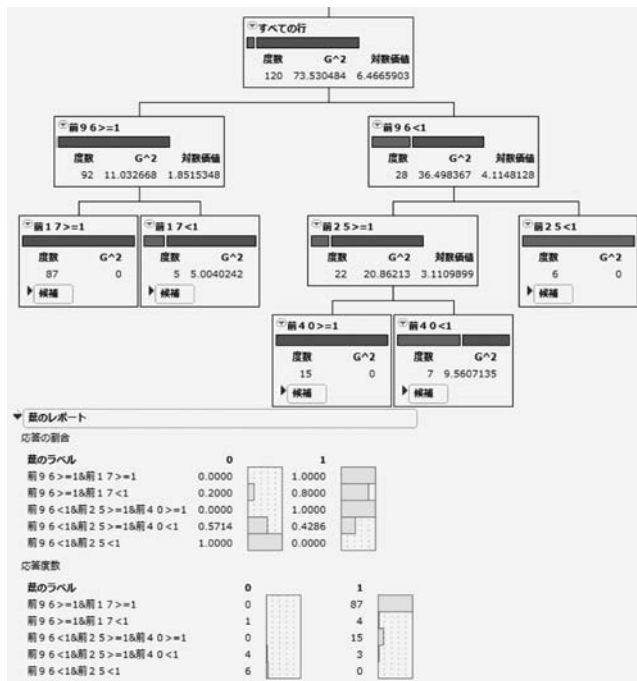


図2.5 決定木分析

2.3 合否50による100項目の分析結果

「合否50」を目的変数とし、100個の設定問を説明変数とした判別を考える。

(1) 逐次変数選択法

表2.6は、「合否50」を1/0の目的変数として、逐次変数増加法と減少法を行った。「Var」と「F/B」列は取り込まれたあるいは掃き出された変数を示す。変数増加法は「X93」から「X53」まで52変数が取り込まれ停止する。その後は、変数減少法で96変数から「X36」が掃き出され、「X32」で停止して32変数のモデルが選ばれる。

その後に、「p値」、「偏差平方和」、「決定係数」、「Cp値」である。Cp値は25変数モデルを選んだ。

「LDF」は、JMPの判別分析で求めた誤分類数である。1変数では28個で、17変数でNM=0になる。「合否10」では7問で合否判定ができた。この違いは、中央値を合否判定の基準に用いると、判別境界上に多くの学生がくる。これらの学生の正答パターンは、「合否10」と比べて多彩なパターンを示すため、判別に有効な説明変数が明確でなくなるためであろう。しかし、LDFの場合はNM=0なことが分かってこの17変数を含むモデルの全てでNM=0であることを保証しない。そこで52変数までの誤分類数を求めると、20、21、23変数で誤分類数が1であることが分かった。

「QD」列は1変数の誤分類数が28から始まり12変数で4にまで減少しているが、13変数で「正則化法」に切り替えて却って悪くなり、合格者の61例の学生が誤判別された。

表2.6 逐次変数増加法と減少法

step	Var.	F/B	P値	SS	R2	Cp	LDF	QD	Logi	IP	SVM	決定木
1	X93	追加	0.000	9.245	0.308	246.416	28	28	28	28	61	28
2	X43	追加	0.000	4.513	0.459	169.584	25	28	25	25	61	25
3	X29	追加	0.000	2.663	0.547	125.071	15	15	15	15	15	16
4	X68	追加	0.000	1.570	0.600	99.648	10	11	10	10	10	16
5	X97	追加	0.001	1.021	0.634	83.818	10	13	10	10	10	13
6	X99	追加	0.001	0.948	0.665	69.263	9	12	9	9	10	13
7	X84	追加	0.009	0.596	0.685	60.851	7	7	6	5	7	13
8	X2	追加	0.017	0.471	0.701	54.625	5	6	4	5	5	13
9	X75	追加	0.045	0.324	0.712	50.970	5	8	4	5	5	10
10	X31	追加	0.028	0.376	0.724	46.402	7	6	4	3	5	10
11	X38	追加	0.094	0.213	0.731	44.686	4	5	1	1	1	10
12	X22	追加	0.118	0.183	0.738	43.493	2	4	0	0	0	
13	X67	追加	0.121	0.178	0.743	42.391	1	61				
14	X34	追加	0.098	0.199	0.750	40.916	2	61				
15	X21	追加	0.078	0.221	0.757	39.052	1	61				
16	X58	追加	0.063	0.242	0.766	36.825	1					
17	X91	追加	0.092	0.193	0.772	35.447	0					
18	X33	追加	0.102	0.179	0.778	34.315	0					
19	X76	追加	0.109	0.169	0.784	33.355	0					
20	X17	追加	0.145	0.138	0.788	32.937	1					
21	X13	追加	0.077	0.200	0.795	31.444	1					
22	X45	追加	0.065	0.213	0.802	29.730	0					
23	X47	追加	0.055	0.225	0.810	27.798	1					
24	X27	追加	0.098	0.163	0.815	26.951	0					
25	X24	追加	0.099	0.160	0.820	26.162	0					
26	X77	追加	0.110	0.147	0.825	25.592	0					

step	Var.	F/B	P値	SS	R2	Cp	LDF	QD	Logi	IP	SVM	決定木
27	X41	追加	0.104	0.149	0.830	24.982	0					
28	X50	追加	0.106	0.145	0.835	24.451	0					
29	X14	追加	0.129	0.126	0.839	24.255	0					
30	X73	追加	0.096	0.149	0.844	23.661	0					
31	X46	追加	0.083	0.157	0.849	22.912	0					
32	X60	追加	0.170	0.097	0.853	23.211	0					
33	X56	追加	0.128	0.118	0.857	23.150	0					
34	X96	追加	0.132	0.114	0.860	23.160	0					
35	X88	追加	0.175	0.091	0.863	23.567	0					
36	X61	追加	0.162	0.096	0.867	23.888	0					
37	X30	追加	0.115	0.120	0.871	23.786	0					
38	X98	追加	0.144	0.102	0.874	24.013	0					
39	X79	追加	0.170	0.088	0.877	24.467	0					
40	X26	追加	0.127	0.108	0.881	24.582	0					
41	X4	追加	0.181	0.082	0.883	25.151	0					
42	X86	追加	0.187	0.079	0.886	25.774	0					
43	X42	追加	0.132	0.101	0.889	26.005	0					
44	X48	追加	0.058	0.157	0.895	25.264	0					
45	X57	追加	0.227	0.062	0.897	26.180	0					
46	X54	追加	0.149	0.088	0.900	26.642	0					
47	X62	追加	0.206	0.067	0.902	27.476	0					
48	X63	追加	0.186	0.072	0.904	28.213	0					
49	X32	追加	0.203	0.066	0.906	29.057	0					
50	X52	追加	0.237	0.057	0.908	30.067	0					
51	X1	追加	0.238	0.056	0.910	31.086	0					
52	X53	追加	0.188	0.069	0.912	31.876	0					
1	X36	削除	0.988	0.000	0.956	95.000						
2	X44	削除	0.985	0.000	0.956	93.001						
3	X20	削除	0.971	0.000	0.956	91.002						
4	X38	削除	0.959	0.000	0.956	89.004						
5	X40	削除	0.963	0.000	0.956	87.006						
6	X3	削除	0.926	0.000	0.956	85.013						
7	X5	削除	0.917	0.001	0.956	83.022						
8	X69	削除	0.926	0.000	0.956	81.029						
9	X88	削除	0.893	0.001	0.956	79.042						
10	X97	削除	0.901	0.001	0.956	77.054						
11	X83	削除	0.899	0.001	0.956	75.065						
12	X46	削除	0.899	0.001	0.956	73.076						
13	X99	削除	0.916	0.000	0.956	71.084						
14	X73	削除	0.778	0.003	0.956	69.135						
15	X4	削除	0.836	0.002	0.956	67.163						

step	Var.	F/B	P値	SS	R2	Cp	LDF	QD	Logi	IP	SVM	決定木
16	X87	削除	0.750	0.004	0.956	65.226						
17	X35	削除	0.725	0.004	0.956	63.300						
18	X78	削除	0.680	0.006	0.955	61.401						
19	X100	削除	0.638	0.007	0.955	59.530						
20	X93	削除	0.645	0.007	0.955	57.650						
21	X37	削除	0.568	0.010	0.955	55.832						
22	X51	削除	0.542	0.012	0.954	54.037						
23	X72	削除	0.560	0.011	0.954	52.221						
24	X94	削除	0.638	0.007	0.954	50.339						
25	X79	削除	0.482	0.015	0.953	48.600						
26	X56	削除	0.422	0.019	0.952	46.936						
27	X15	削除	0.407	0.020	0.952	45.292						
28	X16	削除	0.498	0.013	0.951	43.527						
29	X18	削除	0.265	0.036	0.950	42.163						
30	X90	削除	0.250	0.039	0.949	40.845						
31	X89	削除	0.385	0.022	0.948	39.234						
32	X65	削除	0.267	0.036	0.947	37.868						
33	X22	削除	0.275	0.035	0.946	36.484						
34	X23	削除	0.290	0.033	0.945	35.065						
35	X48	削除	0.276	0.035	0.943	33.683						
36	X45	削除	0.263	0.037	0.942	32.337						
37	X49	削除	0.195	0.050	0.940	31.219						
38	X57	削除	0.249	0.040	0.939	29.923						
39	X80	削除	0.237	0.043	0.938	28.669						
40	X19	削除	0.238	0.043	0.936	27.418						
41	X28	削除	0.311	0.032	0.935	25.970						
42	X25	削除	0.336	0.029	0.934	24.469						
43	X30	削除	0.405	0.021	0.933	22.842						
44	X59	削除	0.302	0.033	0.932	21.415						
45	X70	削除	0.255	0.040	0.931	20.111						
46	X61	削除	0.126	0.073	0.929	19.385						
47	X47	削除	0.132	0.072	0.926	18.647						
48	X84	削除	0.178	0.059	0.924	17.672						
49	X62	削除	0.178	0.059	0.922	16.707						
50	X43	削除	0.156	0.067	0.920	15.872						
51	X92	削除	0.122	0.080	0.917	15.273						
52	X60	削除	0.115	0.085	0.915	14.756						
53	X74	削除	0.197	0.058	0.913	13.766						
54	X42	削除	0.201	0.057	0.911	12.768						
55	X55	削除	0.182	0.063	0.909	11.872						
56	X71	削除	0.198	0.059	0.907	10.908						

57	X54	削除	0.180	0.065	0.904	10.042
58	X75	削除	0.178	0.066	0.902	9.198
59	X7	削除	0.151	0.076	0.900	8.529
60	X66	削除	0.137	0.083	0.897	7.974
61	X32	削除	0.183	0.067	0.895	7.146

(2) ロジスティック回帰

「Logi」列は、ロジスティック回帰の誤分類数である。誤分類数は28個から始まり改定IP-OLDFと同じく12変数で0になっている。図2.6は、12変数のNM=0の分析結果である。2変数を除いて10変数が収束計算で不安定になっている。すなわち10変数と定数項の推定値はすべて0と推定された。

名義ロジスティックのあてはめ合否 50					
反復の履歴					
モデル全体の検定					
モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)	
差	83.160994	12	166.322	<.0001*	
完全	5.23981e-8				
縮小	83.160994				
R2乗(U)			1.0000		
オブザベーション(または重みの合計)			120		
目的関数基準による収束					
あてはまりの悪さ(LOF)					
要因	自由度	(-1)*対数尤度	カイ2乗	p値(Prob>ChiSq)	
あてはまりの悪さ(LOF)	77	5.23981e-8	1.048e-7		
飽和モデル	89	0			
あてはめたモデル	12	5.23981e-8		1.0000	
パラメータ推定値					
項		推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切片	不安定	416.316379	94783.631	0.00	0.9965
前93	不安定	-80.046296	28438.502	0.00	0.9978
前43	不安定	-78.661165	20486.217	0.00	0.9969
前29	不安定	-78.060387	20435.472	0.00	0.9970
前68	不安定	-78.553386	24868.156	0.00	0.9975
前97	不安定	-40.298023	14677.017	0.00	0.9978
前99	不安定	-39.450297	15343.701	0.00	0.9979
前84	不安定	-38.936139	24113.753	0.00	0.9987
前2	不安定	-38.274251	21894.242	0.00	0.9986
前75		-2.4191638	21326.46	0.00	0.9999
前31		-0.6533539	15786.328	0.00	1.0000
前38	不安定	-40.153893	15226.5	0.00	0.9979
前22	不安定	-80.158728	32550.294	0.00	0.9980

推定値は次の対数オッズに対するものです: 0/1

図2.6 12変数のMNM=0の分析結果

(3) 改定IP OLDFとS-SVM

「IP」列は、改定IP-OLDFで求めたMNMである。1変数の28個から始まって12変数で0になっている。「合否10」の6変数に比べ増えているのは、やはり中央値での合否判定は、正答パターンが多彩になってくるためであろう。表2.7は、1変数から12変数モデルまでの判別係数である。12変数でMNM=0であるが、X31の判別係数が0であり、11変数以下でMNM=0であることが分かる。

表2.7 1変数から12変数までの判別係数

p	X2	X22	X29	X31	X38	X43	X68	X75	X84	X93	X97	X99	定数項	NM
1										2			-1	28
2						2				2			-1	25
3			2			2				2			-3	15
4			2			2	2			2			-5	10
5			2			2	2			2	0		-5	10
6			500000			500002	500000			500000	500000	2	-2E+06	9
7			4			2	4		6	4	4	4	-19	6
8	2		4			4	4		4	4	2	2	-17	4
9	4		4			4	4	0	4	4	2	2	-19	4
10	5		6	2		7	7	1	6	6	4	3	-31	3
11	2		4	0	2	4	4	0	2	4	2	2	-17	1
12	1	6	5	0	3	5	4	1	2	6	3	2	-26	0

「SVM」列は、S-SVMの誤分類数である。改定IP-OLDFと等しいかわずかに大きいものがある。表2.8は、判別係数と誤分類数である。

表2.8 判別係数と誤分類数

p	X2	X22	X29	X31	X38	X43	X68	X75	X84	X93	X97	X99	定数項	NM
1										2.000			-1.000	28
2						2.000				2.000			-1.000	25
3			2.000			2.000				2.000			-3.000	15
4			2.000			2.000	2.000			2.000			-5.000	10
5			2.000			2.000	2.000			2.000	0		-5.000	10
6			2.000			2.000	2.000			2.000	0	0	-5.000	9
7			2.000			2.000	2.000		2.000	2.000	2	2	-9.000	6
8	2.000		2.000			2.000	2.000		2.000	2.000	2	2	-11.000	4
9	2.000		2.000			2.000	2.000	0.000	2.000	2.000	2	2	-11.000	4
10	2.000		2.000	0.000		2.000	2.000	0.000	2.000	2.000	2	2	-11.000	3
11	1.294		3.137	0.157	2	3.137	3.137	0.706	1.686	3.686	2	2	-14.961	1
12	2.000	4	4.000	0.000	2	4.000	4.000	0.000	2.000	4.000	2	2	-21.000	0

(4) パーティション

図2.7は、「合否50」を目的変数として、分類木を行った。分岐はX93 (誤分類数は28), X43(25), X75(16), X29(16), X81(13), X33(13), X42(13), X4(13), X40(10), X45(10), X30(10)の11回で行われて停止した。各葉ノードの応答の割合の大きな方に判別することにする。11変数で誤分類数は10個であり、他の判別手法より多かった。

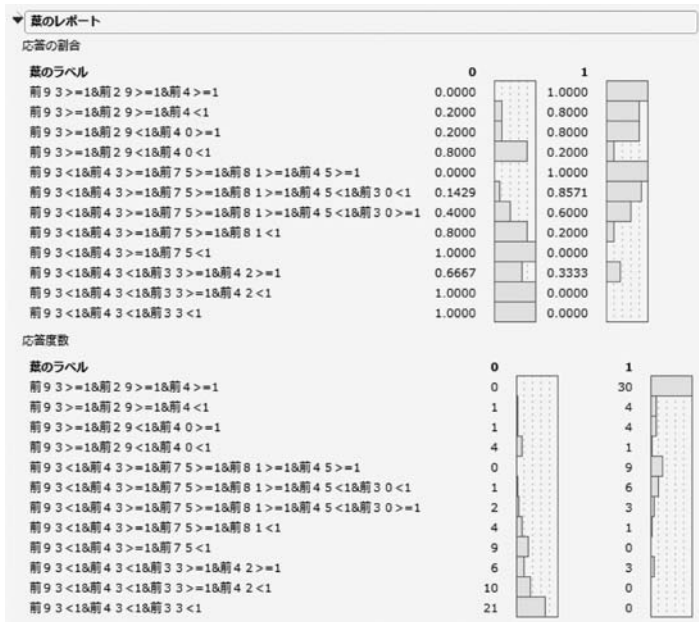


図2.7 合否50を目的変数とした分類木

3 中間試験9項目の分析

中間試験の100項目を、試験内容で9個の設問群に分けて集計した得点を説明変数にして分析する。

3.1 概略

図3.1は、主成分分析の固有値と寄与率である。第2主成分までが固有値1以上で、累積寄与率は53%である。

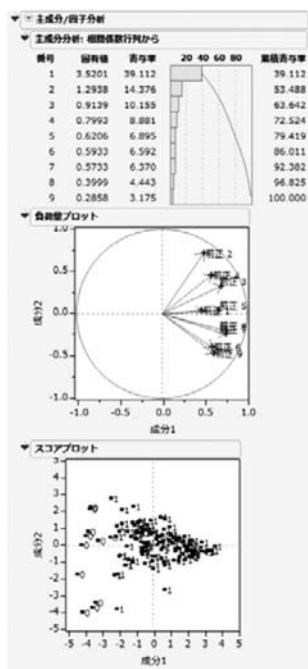


図3.1 主成分分析の固有値，寄与率，因子負荷プロット，スコアプロット

因子負荷量プロットは，第4象限と第1象限にあり，第1主成分が合否判定を表している。スコアプロットでは，不合格の学生が第1主成分のほぼ-2以下に布置していることから，MNM=0の可能性が主成分分析でもうかがわれる。

3.2 「合否10」の9変数の分析

(1) 逐次変数選択法

表3.1は，「合否10」の9変数の逐次変数選択法の出力である。「Var」と「F/B」列は取り込まれたあるいは掃き出された変数を示す。変数増加法と減少法は9変数モデルを選んだ。その後，P値，偏差平方和，決定係数，Cp統計量を示す。Cp統計量は3変数モデルを選んだ。

LDFは線形判別関数の誤分類数であるがNM=0であることは発見できない。2次判別関数は9変数のフルモデルでNM=0である。Logiはロジスティック回帰の誤分類数で，4変数でNM=0である。LDFと2次判別関数とロジスティック回帰は，100項目でNM=0であることを発見できたが，設問項目でまとめたためLDFは形分離可能であることは発見できなかったと考えられる。IPはMNMであり4変数で，SVMは5変数で線形分離可能である。

表3.1 「合否10」の9変数の全ての回帰モデル

step	Var.	F/B	P値	SS	R2	Cp	LDF	QD	Logi	IP	SVM	決定木
1	X7	追加	0.000	3.519	0.352	42.466	14	14	8	7	7	11
2	X4	追加	0.000	1.279	0.480	13.158	8	7	1	1	2	3
3	X2	追加	0.001	0.450	0.525	4.137	7	5	1	1	1	3
4	X6	追加	0.054	0.151	0.540	2.439	6	4	0	0	1	3
5	X9	追加	0.153	0.082	0.549	2.434	5	5		0	0	
6	X3	追加	0.590	0.012	0.550	4.150	7	3		0	0	
7	X8	追加	0.768	0.004	0.550	6.064	7	1		0	0	
8	X1	追加	0.807	0.002	0.550	8.004	7	2		0	0	
9	X5	追加	0.951	0.000	0.550	10.000	7	0		0	0	
10	X5	削除	0.951	0.000	0.550	8.004						
11	X1	削除	0.807	0.002	0.550	6.064						
12	X8	削除	0.768	0.004	0.550	4.150						
13	X3	削除	0.590	0.012	0.549	2.434						
14	X9	削除	0.153	0.082	0.540	2.439						
15	X6	削除	0.054	0.151	0.525	4.137						
16	X2	削除	0.001	0.450	0.480	13.158						
17	X4	削除	0.000	1.279	0.352	42.466						
18	X7	削除	0.000	3.519	0.000	126.602						

表3.2は、改定IP-OLDFの判別係数である。4変数でMNM=0で、5変数から9変数のモデルのX1, X3, X5, X8とX9の判別係数は0になっている。結局、4変数以下でMNM=0であることが分かる。

表3.2 改定IP-OLDFの判別係数

p	X1	X2	X3	X4	X5	X6	X7	X8 X9	定数項	誤分類数
1							0.666667		-5.000000	7
2			866665.800000				333333.000000		-6399992.600000	1
3	1.666667		4.666667				2.333333		-54.000000	1
4	30.000000		91.000000	-5.000000			44.000000		-995.000000	0
5	30.000000		91.000000	-5.000000			44.000000	0	-995.000000	0
6	30.000000	0	91.000000	-5.000000			44.000000	0	-995.000000	0
7	30.000000	0	91.000000	-5.000000			44.000000	0 0	-995.000000	0
8	0 30.000000	0	91.000000	-5.000000			44.000000	0 0	-995.000000	0
9	0 30.000000	0	91.000000	0 -5.000000			44.000000	0 0	-995.000000	0

表3.3はS SVMの判別係数と誤分類数である。5変数でNM=0であるが、判別係数は0のものはない。

表3.3 S-SVMの判別係数と誤分類数

p	X1	X2	X3	X4	X5	X6	X7	X8	X9	定数項	NM
1							0.182			-1.182	7
2				1.625			0.625			-11.000	2
3		0.444		2.711			1.111			-24.422	1
4		0.414		0.851		0.221	0.492			-12.409	1
5		0.324		2.876		0.162	0.949		2.686	-25.427	0
6		0.253	0.904	1.326		-0.022	0.897		0.608	-23.769	0
7		0.023	0.433	0.699		0.057	0.303	0.532	0.653	-10.612	0
8	0.083	0.023	0.440	0.675		0.037	0.286	0.506	0.689	-10.761	0
9	0.049	0.114	0.144	0.405	0.356	0.059	0.301	0.215	0.219	-8.677	0

(2) ロジスティック回帰

ロジスティック回帰は、4変数でNM=0になるが、4変数から9変数の全てのモデルで計算が不安定になる。

名義ロジスティックのあてはめ合否 10

反復の履歴

モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
差	36.765242	4	73.53048	<.0001*
完全	2.56619e-8			
縮小	36.765242			

R2乗(U) 1.0000
 オプザベーション(または重みの合計) 120
 目的関数基準による収束

あてはまりの悪さ(LOF)

要因	自由度	(-1)*対数尤度	カイ2乗
あてはまりの悪さ(LOF)	96	2.56619e-8	5.132e-8
総和モデル	100	0	p値(Prob>ChiSq)
あてはめたモデル	4	2.56619e-8	1.0000

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)	
切片	不安定	18497.6626	6259171.1	0.00	0.9976
前正 7	不安定	-817.98194	276765.38	0.00	0.9976
前正 4	不安定	-1691.6549	572380.18	0.00	0.9976
前正 2	不安定	-557.46074	188973.43	0.00	0.9976
前正 6	不安定	92.4281915	32324.105	0.00	0.9977

推定値は次の対数オッズに対するものです: 0/1

効果の尤度比検定

要因	パラメータ数	自由度	尤度比カイ2乗	p値(Prob>ChiSq)
前正 7	1	1	27.3327912	<.0001*
前正 4	1	1	18.2765413	<.0001*
前正 2	1	1	5.84431111	0.0156*
前正 6	1	1	3.46504296	0.0627

図3.2 1変数のロジスティック回帰の結果

(3) 決定木分析

図3.3は決定木分析の結果である。X9 (誤分類数は11), X4(3), X5(3), X1(3)の順に分岐し, 4変数で誤分類数は3になって停止した。表3.1の4変数モデルのLDF, 2次判別関数, ロジスティック回帰, 改定IP-OLDF, S-SVMの誤分類数は6, 4, 0, 0, 1であり, LDF, 2次判別関数の誤分類数は決定木分析より悪い。

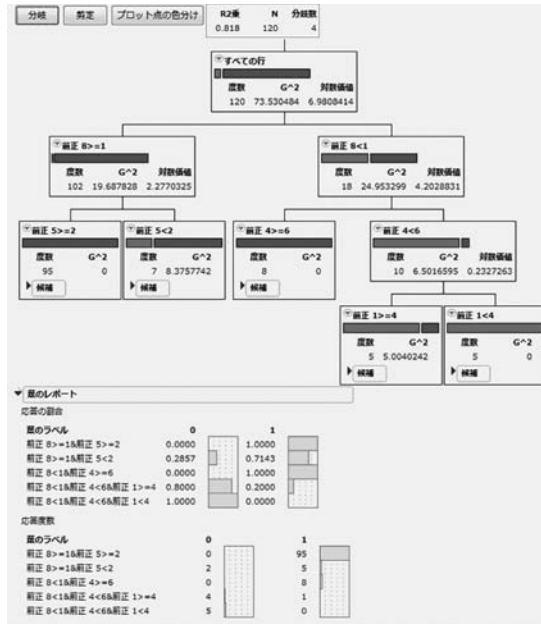


図3.3 決定木分析の結果

3.3 「合否50」による分析

(1) 逐次変数選択法での検討

表3.4は, 「合否50」の9変数の逐次変数選択法の出力である。「Var」と「F/B」列は取り込まれたあるいは掃き出された変数を示す。逐次変数選択法は, 9変数モデルを選んだ。その後に, P値, 偏差平方和, 決定係数, Cp統計量を示す。Cp統計量は7変数モデルを選んだ。

その後のLDFとQDは線形分離可能であることは発見できない。LDFとロジスティック回帰は, 100項目で線形分離可能であることを発見できたが, 設問項目でまとめたためと考えられる。ロジスティック回帰とIPとS-SVMは9変数でNM=0になった。S-SVMは, 1変数では判別超平面上に10人の学生が拘束された。

表3.4 「合否50」の9変数の全ての回帰モデル

step	Var.	F/B	P値	SS	R2	Cp	LDF	QD	Logi	IP	SVM	決定木
1	X8	追加	0.000	11.426	0.381	88.806	28	28	28	28	23+10	20
2	X3	追加	0.000	4.635	0.536	39.676	16	17	15	15	16	19
3	X5	追加	0.000	2.040	0.604	19.176	11	13	9	9	12	19
4	X7	追加	0.007	0.730	0.628	13.120	9	9	7	5	9	16
5	X6	追加	0.023	0.497	0.644	9.635	7	5	5	3	6	10
6	X1	追加	0.079	0.288	0.654	8.454	7	5	4	2	4	10
7	X9	追加	0.093	0.260	0.663	7.586	6	5	5	2	4	10
8	X4	追加	0.273	0.109	0.666	8.381	7	3	1	1	2	9
9	X2	追加	0.539	0.035	0.668	10.000	7	6	0	0	0	9
10	X2	削除	0.539	0.035	0.666	8.381						
11	X4	削除	0.273	0.109	0.663	7.586						
12	X9	削除	0.093	0.260	0.654	8.454						
13	X1	削除	0.079	0.288	0.644	9.635						
14	X6	削除	0.023	0.497	0.628	13.120						
15	X7	削除	0.007	0.730	0.604	19.176						
16	X5	削除	0.000	2.040	0.536	39.676						
17	X3	削除	0.000	4.635	0.381	88.806						
18	X8	削除	0.000	11.426	0.000	212.846						

表3.5は、改定IP-OLDFの判別係数である。MNMは28から0まで単調に減少し、判別係数で0になるものはないので、おそらく8変数以下でMNM=0にはならないと考えられる。

表3.5 改定IP-OLDFの判別係数

p	X1	X2	X3	X4	X5	X6	X7	X8	X9	定数項	NM
1								1.249		-4.746	28
2			374999.125					416665.917		-5499987.500	15
3			6.000		4.000			10.000		-111.000	9
4			279220.682		214285.500		58441.864	266234.045		-5714291.000	5
5			17.200		8.400	0.800	4.800	11.600		-341.000	3
6	5.961		9.588		4.451	3.980	7.647	9.353		-341.196	2
7	5.961		9.588		4.451	3.980	7.647	9.353	0.000	-341.196	2
8	2.257		2.720	1.403	2.104	2.253	1.868	2.773	0.795	-115.537	1
9	0.778	1.365	1.282	0.848	0.809	1.061	0.726	1.196	0.234	-65.090	0

表3.6はS-SVMの判別係数と誤分類数である。4変数で誤分類数は10であり、6変数まで同じである。このためX97とX99の判別係数が0になっている。このことは9変数と10変数モデルでもX31とX75が0であり、誤分類された5と変わらないことに対応している。12変数モデルで判別係数が2個0である。

表3.6 S-SVMの判別係数と誤分類数

p	X2	X22	X29	X31	X38	X43	X68	X75	X84	X93	X97	X99	定数項	NM
1										2.000			-1.000	28
2						2.000				2.000			-1.000	25
3			2.000			2.000				2.000			-3.000	15
4			2.000			2.000	2.000			2.000			-5.000	10
5			2.000			2.000	2.000			2.000	0		-5.000	10
6			2.000			2.000	2.000			2.000	0	0	-5.000	10
7			2.000			2.000	2.000		2.000	2.000	2	2	-9.000	7
8	2.000		2.000			2.000	2.000		2.000	2.000	2	2	-11.000	5
9	2.000		2.000			2.000	2.000	0.000	2.000	2.000	2	2	-11.000	5
10	2.000		2.000	0.000		2.000	2.000	0.000	2.000	2.000	2	2	-11.000	5
11	1.294		3.137	0.157	2	3.137	3.137	0.706	1.686	3.686	2	2	-14.961	1
12	2.000	4	4.000	0.000	2	4.000	4.000	0.000	2.000	4.000	2	2	-21.000	0

(2) ロジスティック回帰

図3.4は9変数のフルモデルのロジスティック回帰である。9変数でNM=0であるが、収束計算上で1変数を除いて不安定になっている。

▼ 名義ロジスティックのあてはめ合否 50

▶ 反復の履歴

▼ モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
差	83.160994	9	166.322	<.0001*
完全	5.37974e-8			
縮小	83.160994			
R2乗(U)			1.0000	
オブザベーション(または重みの合計)			120	

目的関数基準による収束

▼ あてはまりの重さ(LOF)

要因	自由度	(-1)*対数尤度	カイ2乗
あてはまりの重さ(LOF)	110	5.37974e-8	1.076e-7
飽和モデル	119	0	p値(Prob>ChiSq)
あてはめたモデル	9	5.37974e-8	1.0000

▼ パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切片 不安定	1162.25183	275333.16	0.00	0.9966
前正 8 不安定	-24.148668	7716.8102	0.00	0.9975
前正 3 不安定	-23.990066	6636.4578	0.00	0.9971
前正 5 不安定	-15.06776	4185.95	0.00	0.9971
前正 7 不安定	-12.995892	4880.7714	0.00	0.9979
前正 6 不安定	-18.191015	6443.9938	0.00	0.9977
前正 1 不安定	-13.171178	10725.362	0.00	0.9990
前正 9	-0.7982985	10449.049	0.00	0.9999
前正 4 不安定	-15.244734	5244.7062	0.00	0.9977
前正 2 不安定	-23.373996	13125.872	0.00	0.9986

推定値は次の対数オッズに対するものです: 0/1

図3.4 9変数のフルモデルのロジスティック回帰

(3) 決定木分析

図3.5は決定木分析の結果である。X7(誤分類数は20), X5(19), X3(19), X2(16), X6(10), X9(10), X3(10), X9(9), X8(9)の順に9回分岐した。ただし, 7変数しか使われていない。7変数の誤分類数を比較すると, 他の判別手法のそれは6以下である。

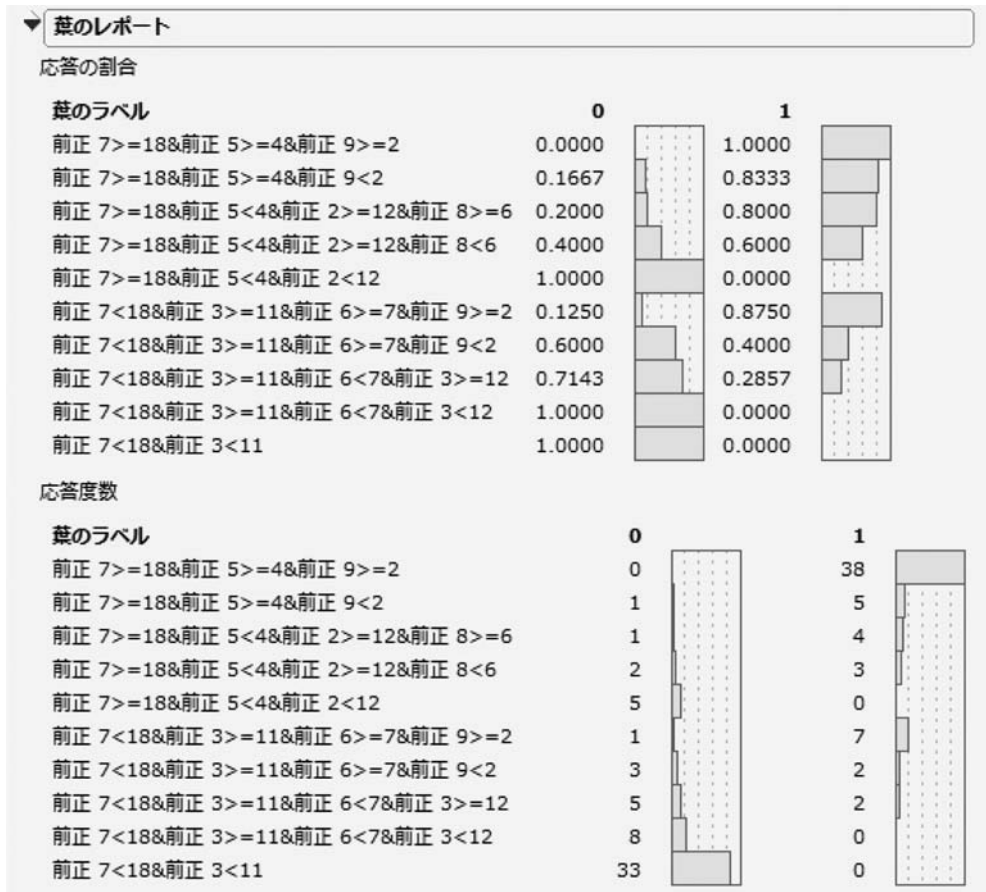


図3.5 決定木分析の結果

4. まとめ

本研究では、2010年度の中間に行った成蹊大学経済学部1年生の「統計入門の中間試験の合否判定」の判別の比較検討を行った。すでに期末試験の分析も終わっているが、中間試験は「1変数の基本統計量」が試験範囲であり、「相関，単回帰，分割表」という広範囲な期末に比べ比較的容易な試験内容である。このため，期末試験では改定IP-OLDFは，他の判別手法より判別結果が良いという特徴があるが，中間試験では大差はなかった。

一方，「合否10」は全体の10%を不合格とし，「合否50」はほぼ合否で2分割しているので，「合否10」の方が「Fisherの仮説」を満たさないと考えられる。どの判別手法でも逐次変数選択法や C_p 統計量で選ぶモデル，そして各判別手法が線形分離可能になる説明変数の個数が少なかった。

これに対して、100個の設問を9個にまとめた変数を用いると、LDFだけが線形分離可能にならなかった。

判別手法別にみると、逐次変数選択法はより高次のモデルを選ぶのに対して、100変数ではLDFでもより説明変数の少ないモデルで線形分離可能になる。

2次判別関数は、線形分離可能になったり、変数を増やしていくと誤分類数が増えたり、分析結果が不安定である。

ロジスティック回帰は、線形分離可能なデータでは問題であることが指摘されている[1][14]。今回の全ての分析で、実際に推定結果が不安定であり、回帰係数の標準誤差は非常に大きな値になり、95%信頼区間は0を含んだ。

改定IP-OLDFは、全ての分析結果で線形分離可能なことを示し、選ばれた説明変数も一番少なかった。

S-SVMは、これまでの分析に用いてきた4種の実データ（アヤメのデータ、CPDデータ[2][11]、学生データ[3]、スイス銀行紙幣データ）では、判別成績や判別境界上にケースがくるといふ問題点が見られたが、今回のデータではほぼ改定IP-OLDFと似た結果になった。

決定木分析で、分岐に用いた変数の個数と誤分類数を調べた。「合否10」の100個と9個の説明変数の分析ではLDFと2次判別関数の誤分類数が多かった。「各判別手法の誤分類数は決定木分析の誤分類数より少なくなるべき」という判別分析の評価に用いるという新しい試みは有効であることが分かった。

（成蹊大学経済学部教授）

文献

- [1] 大倉征幸，鎌倉稔成（2007）「精確ロジスティック回帰の近似推定」『応用統計学』36(2&3),87-98頁
- [2] 新村秀一（1996）「重回帰分析と判別分析のモデル決定（2） - 19変数を持つC.P.D.データのモデル決定 - 」『成蹊大学経済学部論集』第27巻第1号,180-203頁
- [3] ——（2004）『JMP活用統計学とおき勉強法』講談社
- [4] ——（2007）『ExcelとLINGOで学ぶ数理計画法』丸善
- [5] ——（2007）『JMPによる統計レポート作成法』丸善
- [6] ——（2007）「数理計画法による判別分析の10年」『計算機統計学』, 20(1/2), 59-94頁
- [7] ——（2010）『最適線形判別関数』日科技連出版社
- [8] ——（2010a）「Fisherの判別分析を越えて」, 『成蹊大学経済論集』, 41-1,63-101頁
- [9] ——（2010b）「マークシート試験によるFDの一提案」, 『成蹊大学一般研究報告』, 第44巻第4分冊, 1-26頁

- [10] —— (2010c) 「試験の合否判定データの最適線形判別関数による分析」, 『成蹊大学一般研究報告』, 第44巻第5分冊, 1-44頁
- [11] 新村秀一・三宅章彦 (1983) 「C.P.Dデータの多重共線性の解消」 『医療情報学』, 3-3, 107-124頁
- [12] A.Edgar (1935) “The irises of the Gaspé Peninsula”, *Bulletin of the American Iris Society*, 59, 2-5.
- [13] B.Flury & H.Rieduyll(1988).*Multivariate statistics : A Practical Approach*. Cambridge University Press. [田端吉雄 (1900) 『多変量解析とその応用.現代数学社』
- [14] Hirji, Karim F., Mehta, Cyrus R., and Patel, Nitin R. (1987) “Computing Distributions for Exact Logistic Regression,” *JASA*, 82, 1110-1117.
- [15] J.P.Sall, L.Creighton & A.Lehman (2004) 『JMPを用いた統計およびデータ分析入門 (第3版)』 SAS Institute Japan (株). [新村秀一監修]
- [16] L. Schrage (1981), *LINDO - An Optimization Modeling System* -, The Scientific Press. [新村秀一・高森寛 (1992). 実践数理計画法.朝倉書店]
- [17] R.A. Fisher(1936) “The Use of Multiple Measurements in Taxonomic Problems”, *Annals of Eugenics*, 7, 179-188.
- [18] V.Vapnik(1995), *The Nature of Statistical Learning Theory*, Springer-Verlag,1995.