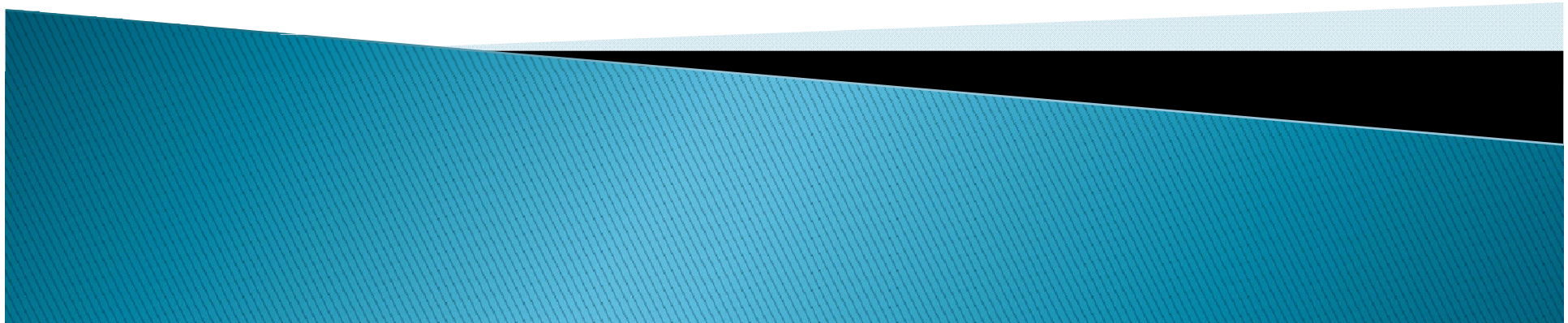


LINGO10の衝撃

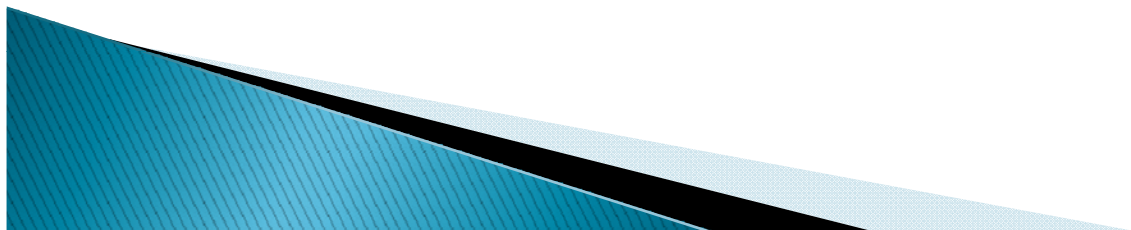
—効率的な数理計画法モデルの開発—

成蹊大学 経済大学
新村秀一



1. 深い悔悟

- ▶ ここ10年, What'sBest! (WB) を用いて, 数理計画法による判別分析の研究
- ▶ 第1期(1998–2004)
 - IP-OLDF: 最小誤分類数 (MMN) 基準による判別
 - 最適凸多様体の発見
 - MMNの単調減少性 ($MMN_p \geq MMN_{(p+1)}$)
 - 一般位置 (Haar条件) を満たさない場合に, 問題
 - Liitschwager & Wang (1978)
 - IP-OLDFの先行研究で無い
 - LP-OLDF: 1970年以降, L1ノルムの研究多い
 - IPLP-OLDF:
 - LP-OLDFで探索域を狭め, IP-OLDFを適用し, 高速化

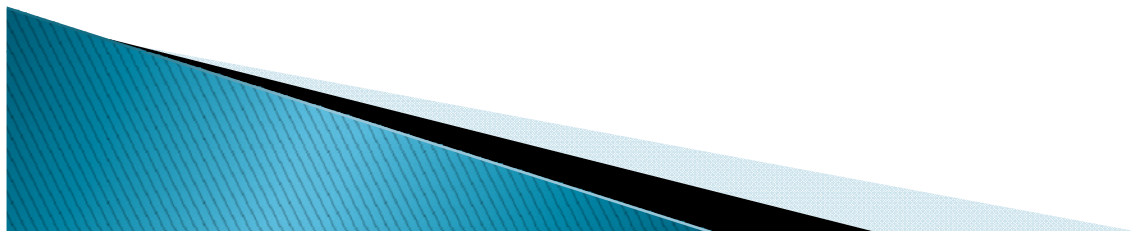


▶ 第2期 (2005–2007)

- 改定IP-OLDF: 最適凸多様体の内点
 - 一般位置になくても対応
 - SVMとの比較が可能
 - 改定LP-OLDF, 改定LPIP-OLDF
- 高速なSoft Margin SVM (制約付きの改定LP-OLDF)の開発

▶ 第3期 (2007年12月末以降～)

- 2007年12月末に, 本発表のため, LINGOによる研究を開始



2. データ

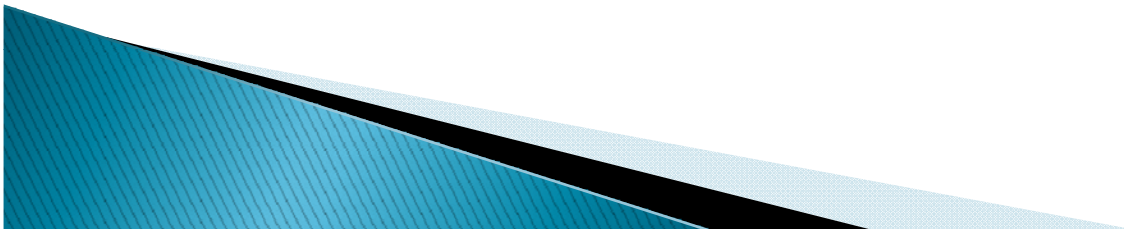
▶ 教師データ

- 学生データ(40人, 5説明変数)
- アイリスデータ(100件, 4変数)
- スイス銀行紙幣データ(200件, 6変数)
- CPDデータ(240人, 19変数, 多重共線性)

▶ 評価データ

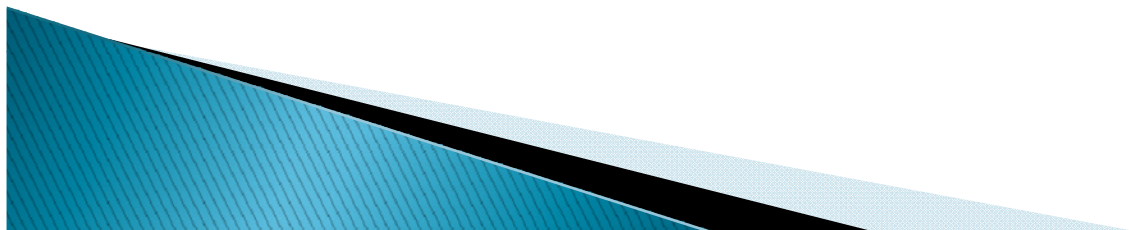
- 教師データと同じ構造を持つ2万件の乱数データ
 - MONACO(数理システム)
- 教師データから, Speakeasyで2万件のリサンプリングデータを作成

▶ 全ての説明変数のモデルで評価



3. LINGOによる研究

- ▶ 2007年12月27日に、本会議に気づく
- ▶ これまでの研究と、
 - 1) やっていないリサンプリングデータによる評価、
 - 2) 改定IPLP-OLDFの評価
- ▶ 2008年1月2日に投稿
- ▶ 2008年1月4日に完成
- ▶ プログラムの説明は最後
 - モデルの汎用化
 - 分析結果のExcelシートが各データ1枚に集約
 - 研究の生産性の向上



4. モデル

- ▶ IP-OLDF (最適凸多様体の頂点)

$$MIN = \sum e_j ;$$

$$y_i \times (x_i' b + 1) \geq -c \times e_j ; \quad (3.1)$$

$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, $y_i = 1$ for $x_i \in class1$, $y_i = -1$ for $x_i \in class2$

b : p 次元判別係数ベクトル, e_i : 各 x_i に対応した0/1決定変数, c : 99999の定数

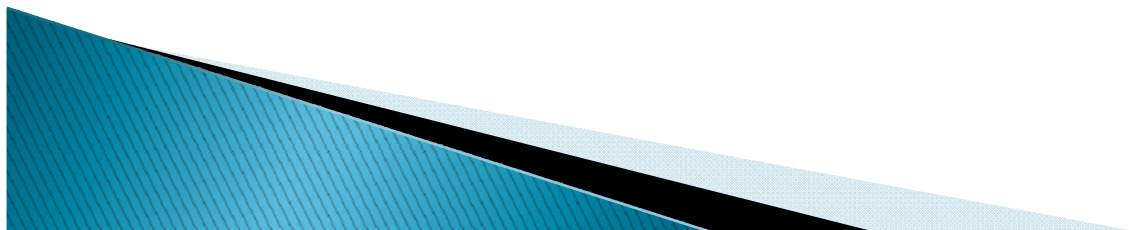
- ▶ 改定IP-OLDF (最適凸多様体の内点)

$$MIN = \sum e_j ;$$

$$y_i \times (x_i' b + b_0) \geq 1 - c \times e_j ; \quad (3.4)$$

- ▶ 改定IPLP-OLDF

- 改定LP-OLDFを適用
- SVで正しく判別されるケースの e_j を0に固定して探索域を狭めた後,
- 改定IP-OLDFを適用し, MMNの近似解を高速で求めることができる



5. 分析結果

5.1 学生データ (5変数, 31モデル)

モデル	IP	IPEC	%	LP	LPEC	%	SV	IPLP	IPLPE C	%
x1,x2,x3,x4,x5	3	2004	-3	4	2391	-2	8	3	2004	-3
x1,x2,x3,x5	3	2004	-3	4	2350	-2	11	3	2004	-3
x1,x2,x3,x4	3	2004	-3	5	2737	-1	11	3	2004	-3
x1,x3,x4,x5	3	2004	-3	6	3464	-2	12	3	2004	-3
x1,x2,x4,x5	4	2099	0	6	3170	-1	13	4	2099	0
x2,x3,x4,x5	3	2004	-3	6	3464	-2	12	3	2004	-3
x1,x2,x3	3	2004	-3	4	2350	-2	11	3	2004	-3
x1,x3,x5	3	2004	-3	4	2350	-2	11	3	2004	-3
x1,x3,x4	5	2486	0	7	3803	-2	13	5	2486	0
x1,x2,x4	5	2486	0	7	3803	-2	13	5	2486	0
x1,x2,x5	3	2004	-3	6	3464	-2	12	3	2004	-3
x2,x3,x4	4	2637	-3	7	4399	-4	10	4	2637	-3
x2,x3,x5	3	2004	-3	4	2350	-2	11	3	2004	-3
x3,x4,x5	3	2004	-3	4	2350	-2	8	3	2004	-3
x1,x4,x5	6	3720	-4	8	4527	-3	15	6	3720	-4
x2,x4,x5	5	2808	-2	7	3333	1	12	5	2808	-2

Elapsed time: 20 秒 对 40秒

x1,x3	5	2831	-2	5	2831	-2	9	5	2831	-2
x1,x2	5	2486	0	9	5145	-3	13	5	2486	0
x2,x3	5	3034	-3	7	3851	-2	12	5	3034	-3
x3,x4	5	2808	-2	7	3333	1	12	5	2808	-2
x3,x5	4	2637	-3	7	4399	-4	10	4	2637	-3
x1,x5	4	2401	-2	6	3464	-2	13	4	2401	-2
x1,x4	7	3587	0	6	2903	0	9	7	3587	0
x2,x4	6	3464	-2	5	2831	-2	9	6	3464	-2
x2,x5	6	3757	-4	3	2004	-3	11	6	3757	-4
x4,x5	13	6290	1	13	6290	1	13	13	6290	1

5.2 あやめのデータ(4変数, 15モデル)

モデル	P	IP	IPEC	%	LP	LPEC	%	SV	IPLP	IPLP	%
X1,X2,X3,X4	4	1	204	0.0	2	411	-0.1	3	1	204	0.0
X2,X3,X4	3	2	411	-0.1	2	411	-0.1	5	2	411	-0.1
X1,X3,X4	3	2	414	-0.1	2	414	-0.1	6	2	414	-0.1
X1,X2,X4	3	4	799	0.0	7	1413	-0.1	11	4	799	0.0
X1,X2,X3	3	2	402	0.0	3	616	-0.1	10	2	402	0.0
X2,X4	2	5	1020	-0.1	6	1232	-0.2	11	5	1024	-0.1
X3,X4	2	3	622	-0.1	6	1252	-0.3	6	3	622	-0.1
X1,X3	2	4	817	-0.1	5	1031	-0.2	10	4	823	-0.1
X1,X4	2	5	1024	-0.1	6	1232	-0.2	10	5	1024	-0.1
X2,X3	2	6	1209	0.0	6	1209	0.0	14	6	1213	-0.1
X1,X2	2	25	4924	0.4	27	5391	0.0	61	25	4975	0.1
X4	1	6	1232	-0.2	6	1232	-0.2	10	6	1232	-0.2
X3	1	7	1413	-0.1	7	1408	0.0	12	7	1408	0.0
X1	1	27	5362	0.2	25	4954	0.2	57	27	5362	0.2
X2	1	37	7351	0.2	34	6841	-0.2	78	37	7351	0.2

CPU of IP : 7:26

CPU of IPLP : 30

5.3 スイス銀行紙幣データ(6変数, 63モデル)

モデルの数	p	CPU	IP	IPEC	%	LP	$\frac{LPE}{C}$	%	SV	$\frac{IPL}{P}$	$\frac{IPLPE}{C}$	%
X1,X2,X3,X4,X5,X6	6	0	0	0	0.0	0	0	0.0	0 0	0	0	0.0
X2,X3,X4,X5,X6	5	0	0	0	0.0	0	0	0.0	0 0	0	0	0.0
X1,X3,X4,X5,X6	5	0	0	<u>95</u>	-0.5	0	0	0.0	0 0	0	<u>0</u>	0.0
X1,X2,X4,X5,X6	5	0	0	<u>799</u>	-4.0	0	0	0.0	0 0	0	<u>0</u>	0.0
X1,X2,X3,X4,X6	5	0	0	<u>807</u>	-4.0	0	531	-2.7	0 0	0	<u>531</u>	-2.7
X1,X2,X3,X4,X5	5	1	1	<u>371</u>	-1.4	2	496	-1.5	3 0	1	<u>389</u>	-1.4
X1,X2,X3,X5,X6	5	0	1	115	-0.1	1	115	-0.1	1 0	1	115	-0.1

X4,X5,X6	3	0	0	0	0.0	0	0	0.0	0	0	0	0	0.0
X3,X4,X6	3	0	0	0	0.0	0	0	0.0	0	0	0	0	0.0
X1,X4,X6	3	0	0	<u>1303</u>	-6.5	0	531	-2.7	0	0	0	<u>531</u>	-2.7
X2,X4,X6	3	0	0	0	0.0	0	0	0.0	0	0	0	0	0.0
X3,X4,X5	3	0	2	198	0.0	3	282	0.1	5	0	2	198	0.0
X2,X4,X5	3	0	2	198	0.0	2	198	0.0	7	0	2	198	0.0
X1,X4,X5	3	0	2	<u>480</u>	-1.4	3	480	-0.9	5	0	2	<u>455</u>	-1.3
X3,X5,X6	3	0	1	115	-0.1	1	115	-0.1	2	0	1	115	-0.1
X1,X3,X6	3	0	1	2200	-10.5	1	115	-0.1	2	0	1	2200	-10.5
X2,X3,X6	3	0	1	115	-0.1	1	115	-0.1	3	0	1	115	-0.1
X2,X5,X6	3	0	1	115	-0.1	1	115	-0.1	3	0	1	115	-0.1
X1,X5,X6	3	0	1	<u>115</u>	-0.1	1	115	-0.1	3	0	1	<u>2110</u>	-10.1
X1,X2,X6	3	0	1	115	-0.1	1	115	-0.1	1	0	1	115	-0.1
X1,X3,X4	3	6:22	11	<u>1900</u>	-4.0	18	2002	-1.0	42	0	11	<u>1793</u>	-3.5
X2,X3,X4	3	0	13	1292	0.0	17	1703	0.0	43	0	13	1292	0.0
X1,X2,X4	3	3:41	13	<u>2628</u>	-6.6	17	2086	-1.9	45	0	13	<u>1789</u>	-2.4
X1,X3,X5	3	10:38	22	2124	0.4	27	2522	0.9	73	0	22	2124	0.4
X2,X3,X5	3	9:10: 15	24	2402	0.0	30	3056	-0.3	72	0	24	2402	0.0
X1,X2,X5	3	10:38				32	2882	1.6	80	0	29	2886	0.1
X1,X2,X3	3	1:02: 38				42	3818	1.9	10 0	0	38	3758	0.2

X4,X6	2	1	0	0	0.0	0	0	0.0	0	0	0	0	0.0
X4,X5	2	0	3	<u>277</u>	0.1	4	382	0.1	8	0	3	<u>282</u>	0.1
X3,X6	2	0	1	115	-0.1	1	115	-0.1	3	0	1	115	-0.1
X5,X6	2	0	1	115	-0.1	1	115	-0.1	2	0	1	115	-0.1
X2,X6	2	0	1	115	-0.1	1	115	-0.1	4	0	1	115	-0.1
X1,X6	2	0	1	115	-0.1	1	427	-1.6	3	0	1	115	-0.1
X3,X4	2	0:54	14	1392	0.0	17	1703	0.0	43	0	14	1392	0.0
X2,X4	2	6:18	14	1392	0.0	19	1902	0.0	46	0	14	<u>1394</u>	0.0
X1,X4	2	18:26	13	2536	-6.2	15	2081	-2.9	47	0	13	<u>2440</u>	-5.7
X3,X5	2	18:15	26	2613	-0.1	29	2902	0.0	68	0	26	2613	-0.1
X2,X5	2	4:23	31	3049	0.3	31	3081	0.1	83	0	31	3049	0.3
X1,X3	2	23:45	38	4420	-3.1	43	3640	3.3	92	0	38	4420	-3.1
X1,X5	2	1:08: 49				49	4875	0.1	98	0	41	5028	-4.6
X2,X3	2	1:02:3 8				42	4240	-0.2	10 2	0	39	3928	-0.1
X1,X2	2	1:37:4 8				48	4327	2.4	11 8	0	45	5257	-3.8

37時間
14分

5.4 CPD (19変数, 40モデル)

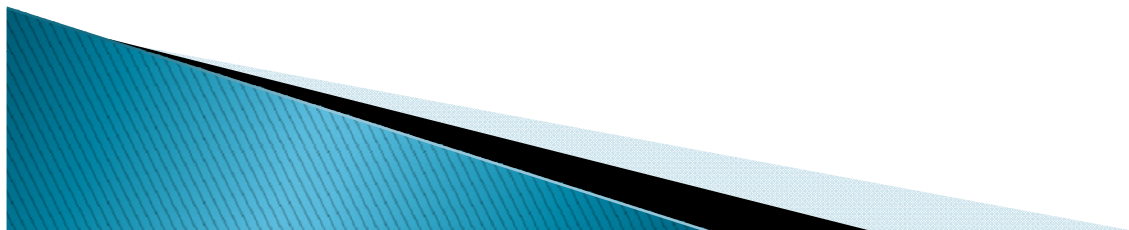
P	Ran k	Type	IP	IPEC	%	LP	LPEC	%	SV	IPLP	IPLPEC	%	Time1
1	1	FBfb	20	2142	-2.4	20	2142	-2.4	50	20	2142	-2.4	1:17
2	2	FBfb	13	1815	-3.7	17	1931	-2.6	38	13	1815	-3.7	56
3	3	FBfb	12	1647	-3.2	18	1991	-2.5	37	12	1524	-2.6	3:06
4	1	Ffb	10	1285	-2.3	13	1378	-1.5	32	10	1285	-2.3	3:27
4	3	B	11	1468	-2.8	19	2159	-2.9	36	11	1468	-2.8	7:02
5	1	Ff	11	1468	-2.8	19	2159	-2.9	35	11	1468	-2.8	6:45:56
5	2	b	7	1043	-2.3	13	1477	-2.0	26	7	1043	-2.3	1:03
5	3	B	11	1468	-2.8	18	2094	-3.0	33	11	1468	-2.8	11:19
6	1	B	9	1136	-1.9	13	1469	-1.9	30	9	1136	-1.9	5:09
6	2	b	7	1043	-2.3	14	1626	-2.3	24	7	1043	-2.3	1:27
6	3	Ff	7	1043	-2.3	14	1523	-1.8	24	7	1043	-2.3	3:26
6	*	DOC1	12	1533	-2.7	18	2097	-3.0	35	12	1533	-2.7	6:34:27
6	*	DOC2	11	1361	-2.2	17	1927	-2.6	36	11	1361	-2.2	1:17:35
7	1	B	9	1136	-1.9	13	1469	-1.9	29	9	1136	-1.9	36:43:00
7	2	Ffb	6	887	-1.9	14	1523	-1.8	24	6	887	-1.9	36:19:00
8	1	F	6	887	-1.9	12	1289	-1.4	23	6	887	-1.9	
8	2	B	8	980	-1.6	15	1577	-1.6	29	8	980	-1.6	
8	5	fb	6	887	-1.9	12	1401	-2.0	24	6	887	-1.9	
9	1	B	6	887	-1.9	13	1343	-1.3	23	6	887	-1.9	
9	2	F	4	408	-0.4	8	960	-1.5	19	4	408	-0.4	

10時間36分10秒 対 6分20秒

9	3	fb	4	539	-1.0	9	1016	-1.3	17	4	539	-1.0	6:58
10	1	B	6	887	-1.9	14	1485	-1.6	22	6	887	-1.9	
10	6	F	4	539	-1.0	7	799	-1.1	18	4	<u>408</u>	-0.4	
10	*	fb	3	370	-0.6	8	873	-1.0	16	3	370	-0.6	
11	1	B	4	408	-0.4	9	1012	-1.3	16	4	408	-0.4	
11	*	F	4	539	-1.0	9	1012	-1.3	15	4	<u>408</u>	-0.4	
11	*	fb	3	370	-0.6	8	855	-0.9	14	3	370	-0.6	
12	1	FB	4	539	-1.0	9	1012	-1.3	15	4	<u>408</u>	-0.4	
12	*	fb	3	370	-0.6	8	855	-0.9	13	3	370	-0.6	
13	1	FB	3	240	0.1	8	946	-1.4	14	3	<u>370</u>	-0.6	
13	*	fb	3	390	-0.7	9	1020	-1.4	12	3	390	-0.7	
14	1	FB	3	370	-0.6	7	777	-1.0	15	3	370	-0.6	
14	*	fb	2	214	-0.2	7	807	-1.1	14	2	214	-0.2	
15	1	FB	3	370	-0.6	8	855	-0.9	13	3	370	-0.6	
15	*	fb	2	202	-0.2	5	591	-0.9	10	2	202	-0.2	
16	1	FB	2	202	-0.2	5	482	-0.3	14	2	202	-0.2	
16	*	fb	2	214	-0.2	5	481	-0.3	8	2	<u>202</u>	-0.2	
17	1	FB	2	334	-0.8	8	744	-0.4	8	2	<u>214</u>	-0.2	
18	1	FB	2	334	-0.8	5	591	-0.9	7	2	<u>214</u>	-0.2	
19	1	FB	2	221	-0.3	6	542	-0.2	7	2	<u>102</u>	0.3	

5.5 まとめ

- ▶ 改定IPLP-OLDFの誤分類数は、全てMMNに等しい
- ▶ 教師と評価データの、誤分類確率の差は、多くは1%以内
- ▶ 計算時間
 - 学生データ : 20秒 : 40秒
 - アイリスデータ : 7分26秒 : 30秒 (15倍)
 - 銀行紙幣 : 37時間3分19秒 : 44分48秒 (49倍)
 - CPDデータ : 10時間36分10秒 : 6分20秒(100倍)



6. LINGOの集合モデル

▶ 集合節

MODEL:

SETS: !SETS節の始まり;

P/X1..X6/: VARK;

N/1..40/: E,SCORE;

D(N,P): IS;

MS/1..31/: NDROPK,LPIC,LPEC,LPIC2,IPLP,IPLPEC;

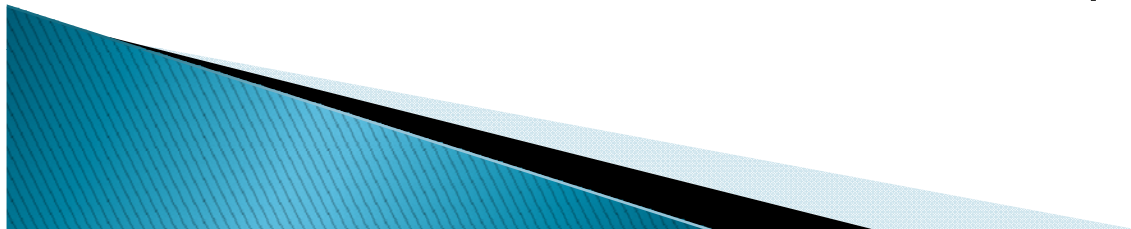
C(N,MS): CONSTANT;

MB(MS,P): CHOICE,VAR,VAR2;

N2/1..20000/: E2;

D2(E2,P): ES;

ENDSETS ! SETS節の終わり;



6.2 DATA節

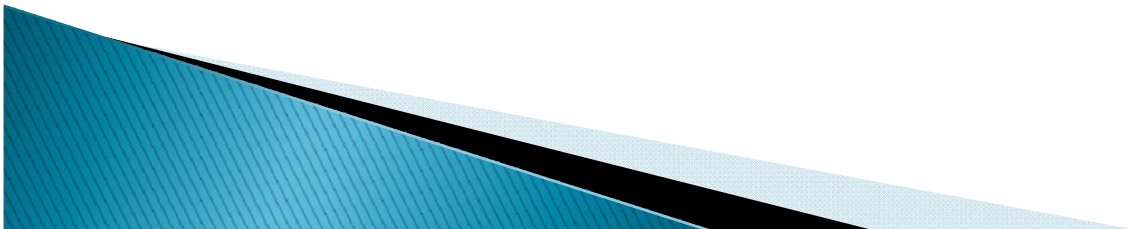
DATA: ! Data節の開始;

IS=@OLE(); ! 教師データを入力;

CHOICE=@OLE(); ! 判別モデルパターン入力;

ES=@OLE(); ! 2万件の評価データを入力;

ENDDATA



6.3 SUBMODEL節

```
SUBMODEL sub1:    ! Calc節で使用するモデルを定義;  
MIN= NDROPK(k);  NDROPK(k) =@SUM(N(i):E(i));  
@FOR(N(i):  
    @SUM(P(j):IS(i,j)*VARK(j)*CHOICE(k,j)) > 1 -  
    9999*E(i));  
@FOR(P(j): @FREE(VARK(j)));  
!@FOR(N(i):@BIN(E(i)));  
ENDSUBMODEL
```

```
SUBMODEL sub2:    !SUB2で改定IPLP-OLDFの定式化;  
MIN= NDROPK(k);  NDROPK(k) =@SUM(N(i):E(i));  
@FOR(N(i):  
    @SUM(P(j):IS(i,j)*VARK(j)*CHOICE(k,j)) > 1 -9999*E(i));  
@FOR(P(j):@FREE(VARK(j)));  
@FOR(N(i) | CONSTANT(i,k) #NE# 0: @BIN(E(i)));  
@FOR(N(i) | CONSTANT(i,k) #EQ# 0: E(i)=0);  
ENDSUBMODEL
```

6.4 CALC節

```
CALC:    !CALC節の始まり;
@SET('DEFAULT'); @SET('TERSEO',2);
K=1; Lend=@SIZE(MS);
MNI=0; MNE=0; MNI2=0;
@WHILE(k #LE# Lend:
    @SOLVE(sub1);
    @FOR(N(i): SCORE(i)=@SUM(P(j):IS(i,j)*VARK(j)*CHOICE(k,j)));
    @FOR(N(i): @IFC(SCORE(i)#LT#0:MNI=MNI+1));
    @FOR(N(i): @IFC(E(i) #EQ# 0: CONSTANT(i,k)=0;
                @ELSE CONSTANT(i,k)=1;));
    @FOR(N(i): @IFC(E(i) #GT# 0: MNI2=MNI2+1));;
    @FOR(N2(l): E2(l)=@SUM(P(j):ES(l,j)*VARK(j)*CHOICE(k,j)));
    @FOR(N2(l): @IFC(E2(l) #LT# 0:MNE=MNE+1));
    @FOR( P(j):VAR(k,j) = CHOICE(k,j)*VARK(j););
    LPIC(K)=MNI; LPEC(K)=MNE;LPIC2(k)=MNI2;
MNI=0;MNE=0;MNI2=0;
    K=K+1);
ENDCALC
```

6.5 DATA節

DATA:

@OLE()=VAR;

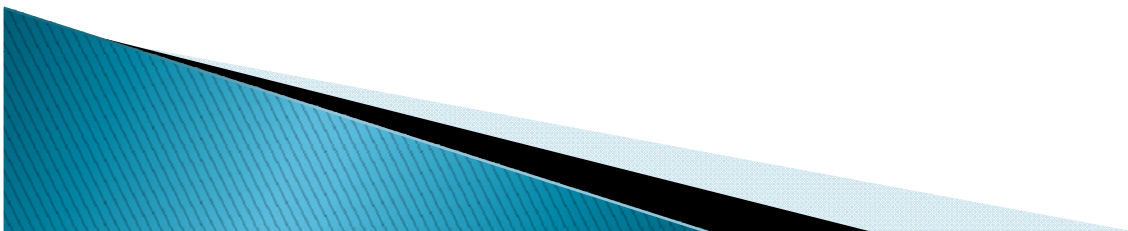
@OLE()=CONSTANT;

@OLE()=LPIC;

@OLE()=LPEC;

@OLE()=LPIC2;

ENDDATA



セル名の指定

	choice					& 1			
	M	N	O	P	Q	R	S	T	
1	X1	X2	X3	X4	X5	LPIC	LPEC	LPIC2	
2	1	1	1	1	1	34	411	53	
3	0	1	1	1	1	34	411	81	
4	1	0	1	1	1	53	622	91	
5	1	1	0	1	1	115	1408	180	
6	1	1	1	0	1	66	817	172	
7	0	1	0	1	1	126	1232	235	
8	0	0	1	1	1	119	1252	156	
9	1	0	1	0	1	136	1416	194	
10	1	0	0	1	1	126	1232	202	
11	0	1	1	0	1	131	1417	266	
12	1	1	0	0	1	526	5391	1189	
13	0	0	0	1	1	126	0	202	
14	0	0	1	0	1	129	0	241	
15	1	0	0	0	1	523	0	1199	
16	0	1	0	0	1	803	0	1477	

7. まとめ

- ▶ 数理計画法ソフトは、制約条件付きの関数の最大/最小値を求める
- ▶ しかし、これまでソフトの能力が低く、とてもこのことを強調できなかった
 - 漸く、数理計画法ソフトはこの桎梏から解き放たれた。
- ▶ これから、数式で記述できる全ての現象、問題の解決ができる
 - OR, 統計, 理学, 工学, 農学, 経済.
- ▶ 道具の選択で、知的生産性が著しく向上

