

Applicability statement of “K-step Yard sampling method”

Classifying 7000 sample set of Ames test

Challenge for classification and prediction

K-step Yard sampling method KY-method



The most powerful and advanced data analysis method



The most difficult classification problem

6,965 sample of Ames test samples were,

Classified perfectly

Application test of “K-step Yard sampling”

□ Samples

1. Ames test data
 2. Sample population
 - total :6,965
 - Mutagen; 2,932
 - Non-mutagen; 4,033
-

□ Result of KY-method

1. Number of steps : 23 steps ; 22 (2 models) + 1 (1 model)
 2. Classification ratio : 100 %
-

□ Used system

ADMEWORKS / ModelBuilder V 3.0.22

□ Used parameters (Initial condition)

Number of generated parameters : 838

Number of parameters for step 1 : 98

Confidency index (Samples(6965) / Parameters(98)) : 71.1 > 4.0

Application test by various D.A. methods

1. Linear discriminant analysis with linear least-squares method

Classification ratio : total; **73.50**(6965), Mutagen;**73.02**(2932), Non mutagen;**73.84**(4033)
 Number of mis-classified : **(1846)**, **(791)**, **(1055)**
 Prediction ratio (L100 out) 72.58% deviance(0.92%)
 (L500 out) 73.32% deviance(0.18%)

2. SVM (Support Vector Machine with Kernel)

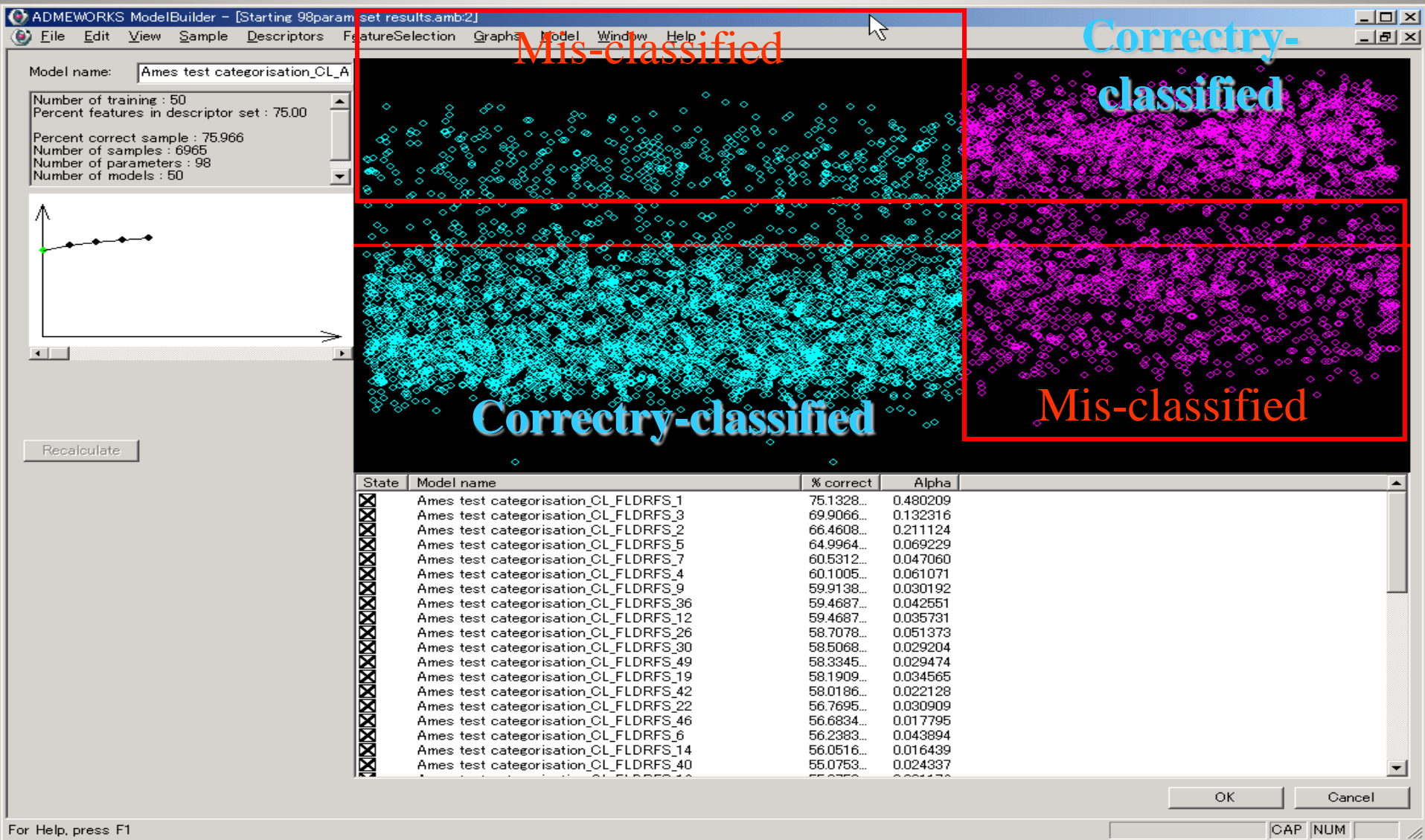
Classification ratio : total; **90.87**(6965), Mutagen;**86.83**(2932) Non mutagen; **93.80**(4033)
 Number of mis-classified : **(636)**, **(386)**, **(250)**
 Prediction ratio (L500 out) 80.99% deviance(9.88%)

3. AdaBoost

Classification ratio : total; **77.24**(6965), Mutagen;**66.13**(2932) Non-mutagen; **85.32**(4033)
 Number of mis-classified : **(1585)**, **(993)**, **(592)**
 Prediction ratio (L500 out) 75.16% deviance(2.08%)

Classification results by AdaBoost

Sample distribution of 6,965 of 77.24%



“K-step Yard sampling (KY) Method”

Total steps : 23 steps (2 models) + 1 step (1 model)

ステップID(KY法)	Starting samples(Total)	Mutagen (Initial)	Non-mutagen (Initial)	Grey sample (Initial)
	Final samples	Mutagen (Final)	Non-mutagen (Final)	Grey sample (Final)
	Determined samples(Total)	Determined samples(Mut.)	Determined samples(Non-mu)	Grey ratio(%) (Grey/Total)
1	6965	2932	4033	0
	5864	2413	3451	5864
	1101	519	582	84.19
2	5864	2413	3451	5864
	5108	2142	2966	5108
	756	271	485	87.11
3	5108	2142	2966	5108
	4486	1919	2567	4486
	622	223	399	87.82
4	4486	1919	2567	4486
	4133	1779	2354	4133
	353	140	213	92.13
5	4133	1779	2354	4133
	3794	1651	2143	3794
	339	128	211	91.8
6	3794	1651	2143	3794
	3462	1485	1977	3462
	332	166	166	91.25
7	3462	1485	1977	3462
	3090	1345	1745	3090
	372	140	232	89.25
8	3090	1345	1745	3090
	2826	1220	1606	2826
	264	125	139	91.46
9	2826	1220	1606	2826
	2592	1139	1453	2592
	234	81	153	90.63
10	2592	1139	1453	2592
	2384	1047	1337	2384
	208	92	116	91.98

“K-step Yard sampling (KY) Method”

12	2095	931	1164	2095
	1848	829	1019	1848
	247	102	145	88.21
13	1848	829	1019	1848
	1607	733	874	1607
	241	96	145	86.96
14	1607	733	874	1607
	1380	623	757	1380
	227	110	117	85.87
15	1380	623	757	1380
	1028	466	562	1028
	352	157	195	74.49
16	1028	466	562	1028
	787	358	429	787
	241	108	133	76.56
17	787	358	429	787
	529	234	295	529
	258	124	134	67.22
18	529	234	295	529
	392	201	191	392
	137	33	104	74.1
19	392	201	191	392
	279	141	138	279
	113	60	53	71.17
20	279	141	138	279
	184	105	79	184
	95	36	59	65.95
21	184	105	79	184
	112	66	46	112
	72	39	33	60.87
22	112	66	46	112
	66	39	27	66
	46	27	19	58.93
23(1 model)	66	39	27	66
	0	0	0	0
	66	39	27	0

“K-step Yard sampling (KY) Method”

Classification results by 3 steps

Microsoft Excel - MB_summery.xls

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H) Adobe PDF(B)

MS Pゴシック 11

	A	B	C	D	E	F	G	H	I	J
1	Sample ID	ステップ1		ステップ2		ステップ3				
2		AP	AN	AP	AN	AP	AN	ステップ1	ステップ2	ステップ3
3	1	nonmutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	ネガ		
4	2	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
5	3	mutagen	nonmutagen	mutagen	nonmutagen	nonmutagen	nonmutagen	グレー	グレー	ネガ
6	4	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
7	5	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	mutagen	グレー	グレー	ポジ
8	6	nonmutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	ネガ		
9	7	nonmutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	ネガ		
10	8	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
11	9	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
12	10	mutagen	mutagen	mutagen	mutagen	mutagen	nonmutagen	ポジ		
13	11	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
14	12	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
15	13	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
16	14	mutagen	nonmutagen	mutagen	mutagen	mutagen	nonmutagen	グレー	ポジ	

3 STEP TEST / コマンド

CAPS NUM

Spatial features of “K-step Yard sampling”

□ Summary

■ Features

1. KY-method is a meta-method
Used with various DA and Fitting methods.
2. Perfect classification is achieved in any condition

■ Disadvantages

1. Relatively complex operation to generate discriminant functions
2. Need powerful computer power

Spatial features of “K-step Yard sampling”

□ Summary

■ Advantages

1. Sample number free approach
2. Sample distribution free approach
3. Perfect classification is achieved in any condition

■ Disadvantages

1. Relatively complex operation to generate discriminant functions
2. Need powerful computer power

In Silico Data

Miracles by the KY-methods

Kohtaro Yuta

In Silico Data, Ltd.