

DSC 2003 Working Papers (Draft Versions) http://www.ci.tuwien.ac.at/Conferences/DSC-2003/

# Prediction Analysis of Microarrays in Excel

#### Balasubramanian Narasimhan\*

# 1 Introduction

Prediction Analysis of Microarrays (PAM) is a statistical technique for class prediction using gene expression data using shrunken centroids. It is described in [3]. The method of nearest shrunken centroids identifies subsets of genes that best characterize each class. The technique is general and can be used in many other classification problems.

PAM Software for the R [1] has been available for some time now from the http://www-stat.stanford.edu/~tibs/PAM. Using the software, one can train the classifier, perform cross validation to get an idea of the value to use for thresholding, and do predictions. There has been some demand for a GUI version for PAM.

Our experience with SAM (Significance Analysis of Microarrays) [4] software led us to use Excel as a GUI for the package. SAM was written in Visual Basic, using a Java COM library at the core. However, it was clear that this approach would not be an efficient one in the long run for several reasons. Among them:

- Many statistical tools would have to be rewritten in Java, or Visual Basic wasting valuable development time rather than leveraging already available tools.
- The outcome of the market place battle, Microsoft Java versus Sun Java, is hardly clear. Besides, programming in Microsoft Java is like programming with one hand tied behind your back.

<sup>\*</sup>Department of Statistics and Department of Health Research & Policy, Stanford University, Stanford CA 94305. Email: naras@stat.stanford.edu.

• While Microsoft's commitment to DCOM seems quite firm, its commitment to Java COM seems to have stalled.

At DSC 2001, Erich Neuwirth and Thomas Baier [2] demonstrated an R DCOM server and client. We decided to exploit this technology for PAM to produce an environment much like that of the SAM except with R as the computation engine.

Section 2 describes how the PAM software is organized. Section 3 describes some of the capabilities of the software via an example. Section 4 describes some issues that we faced and future work.

## 2 The Structure of PAM

Since the Excel part of PAM is merely a front-end to the R package, one must install PAM for R package first. Then, as is customary for Windows, a point and click **Setup** package is provided to install the Excel addin. Once installed, the user has to activate the addin in the standard fashion.

The Excel GUI is written in Visual Basic using object-oriented VBA throughout. To ensure that a wide audience would be able to use the software, we eschewed newer features and wrote to an Excel 97 target. The software has been tested with R version 1.6.1 on the following platforms: Excel 97 and Windows NT, Excel 2000 and Windows 98, Excel 2000 and Windows ME, Excel 2000 and Windows XP, Excel 2002 and Windows XP.

The software comes with several examples and documentation.

### 3 A Quick Tour of PAM

When PAM is installed, two buttons appear in the Excel toolbar titled **PAM** and **PAM Controller**. PAM uses a data format very similar to SAM (see figure 1). Like SAM, the data is assumed to be normalized.

One highlights an area of the spreadsheet that represents the data. Then clicking on the **PAM** button brings up the dialog shown in figure 2. A brief explanation of the fields shown in the dialog.

- Class Labels Specifies the row number that contains the class labels, 1, 2, or 3.
- Sample Labels Specifies the row number that contains the sample labels. Sample labels are optional and so the field can left blank although they are recommended as a means of identifying samples.
- **Batch Labels** Specifies the row containing the batch labels, if any. Batch labels allow one to combine expression data from different experiments.
- **Imputation Engine** PAM handles missing data (denoted by blank cells) by imputing with K-Nearest Neighbors. One can change the number of neighbors required for the imputation. The default is 10.

Image: brance         Image: b		41 900 91 10		2 4 4 3		A STATE A									1
Math         Math <th< th=""><th>Num         Num         Num</th></th<> <th></th> <th>2</th> <th>9</th> <th>1 1</th> <th>-</th> <th>9</th> <th>1</th> <th>-</th> <th>1</th> <th>×</th> <th>- the second</th> <th>-</th> <th></th> <th>0</th>	Num		2	9	1 1	-	9	1	-	1	×	- the second	-		0
Matrix         Matrix<	Image: bit is a bit is bit is a bit bit a bit is a bit a bit is a bit is a bit is a bit	100	Transford.	- 132 mm	Editor.	- anti-	(Miles)	and the	anythin .	in the last	- adapt	in the line	a think a	"Head in	- Lifebur
No.         No. <td>Matrix         Matrix         Matrix&lt;</td> <td>Thank was and the</td> <td>CANS DATE</td> <td>ALCHINE .</td> <td>EAS DEVICE</td> <td>TAVE DATE</td> <td>Sec. State</td> <td>CALL STATE</td> <td>The second</td> <td>THE COLUMN</td> <td>The second secon</td> <td>In summer</td> <td>Will Work with the second</td> <td>EAVE CAVE</td> <td>Ma In</td>	Matrix         Matrix<	Thank was and the	CANS DATE	ALCHINE .	EAS DEVICE	TAVE DATE	Sec. State	CALL STATE	The second	THE COLUMN	The second secon	In summer	Will Work with the second	EAVE CAVE	Ma In
entrop         Control         Control <th< td=""><td></td><td>The state of the state of the</td><td>TIMO STATE</td><td>CLEANE C</td><td>1 54215020</td><td>1 TOTO SALT</td><td>Taxa and</td><td>2 STICKE IT</td><td>TODATE STREET</td><td>1 TWINE CO</td><td>Tanga tanga</td><td>SUCCESSION OF COMPANY</td><td>12.00.000</td><td>C PERSONAL C</td><td></td></th<>		The state of the state of the	TIMO STATE	CLEANE C	1 54215020	1 TOTO SALT	Taxa and	2 STICKE IT	TODATE STREET	1 TWINE CO	Tanga tanga	SUCCESSION OF COMPANY	12.00.000	C PERSONAL C	
		annek/k/c/dence	10 402862070F	12104-1414-00	229-003P( 0)	104602000	15750030810	END OFFICE OF	0 4.895 NUB	2.586.00r	TTHERE II	INTERNES. IF	HELEN THE	1023116201	0.1111-1
Optimize in the state		a pleatantee est	LITERAL CO.	CARDED DATE:		「「「「「「「」」」」	In any second second	IL BOCCEPALITY	11.22/2010/1	ANTERNAL	0.00000000	AL NAROWING AN	「「「「「「「」」」」		THE NAME
Contraction		No. of the District of the Dis	A CONSCURPT IN	STATISTICS OF	01499-02120	TARGER AND	-CEATOGRAFT	Cupili State	Cap Theat I	1.000.0001	Conservation of	NETOSCET 1	THE LIVEOUT	DOCUMPTOR	1.000
Circle         Circle<		TITLE DATE TO A TITLE	TRUNCE I	L'Se Solution	1. 102236024	a Denachrydd	Charles a	1 PSSCHOL	CHANNEL	1.40000-CA	1.001001	L CHANGE OF A	1211123MM	Distanciant P	- Halles
Sector         Control         Control <th< td=""><td>Operation         Operation         <t< td=""><td>Partition at a sur</td><td>0.25500756</td><td>CARDANATER.</td><td>HARTE -II</td><td>THURSDAY D</td><td>10.77144.236=</td><td>IL STATES</td><td>Train Part D</td><td>In THREE AS</td><td>11 WILLIAM D</td><td>41 2012/01/02/02</td><td>D. 1511 million 4900</td><td>IT IS NOT</td><td>11 MAC</td></t<></td></th<>	Operation         Operation <t< td=""><td>Partition at a sur</td><td>0.25500756</td><td>CARDANATER.</td><td>HARTE -II</td><td>THURSDAY D</td><td>10.77144.236=</td><td>IL STATES</td><td>Train Part D</td><td>In THREE AS</td><td>11 WILLIAM D</td><td>41 2012/01/02/02</td><td>D. 1511 million 4900</td><td>IT IS NOT</td><td>11 MAC</td></t<>	Partition at a sur	0.25500756	CARDANATER.	HARTE -II	THURSDAY D	10.77144.236=	IL STATES	Train Part D	In THREE AS	11 WILLIAM D	41 2012/01/02/02	D. 1511 million 4900	IT IS NOT	11 MAC
entrom         entrom<		「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	Dissource of	00100000	0.0010438	0.204630300	10 1: 200 e	0.Californalia	0.000000	C Discertige	0.0000.0000	0.140732746	0000000000	D.205001311	0.04426
Optimize		a transfer of a transferra	IL SUGARANTI	Condition .	ILL'INDUAR II	UNACCOMMENT.	IL WHY Y'TEZA	ALTRACK III	Contraction of the	THREE AND A	MILLING OF	100128142419	SUBSECTION IN	PIRATE AND ADDRESS	In Property lies
worket         formation         f		# yo potenti pote #,	4165CD407	SATISTICS -	1.195721266	01411300	DCS/201	1.159 451-004	1110301300	1,50000161	三日日本 第二二	●(2)第1280 -	110000-301	Decision of	100 100
Operation         Operation <t< td=""><td></td><td>where the set of the set of</td><td>5326.2.3.11</td><td>NUMBER OF</td><td>RUMANPAR', IF</td><td>444.59110</td><td>WALKING W</td><td>STATEMENT C.</td><td>URLAND P</td><td>Un statute .</td><td>ATTACKA P</td><td>HARRING IN</td><td>LINE TRAVERSON</td><td>ALC: ALMERICA</td><td>ULT III</td></t<>		where the set of the set of	5326.2.3.11	NUMBER OF	RUMANPAR', IF	444.59110	WALKING W	STATEMENT C.	URLAND P	Un statute .	ATTACKA P	HARRING IN	LINE TRAVERSON	ALC: ALMERICA	ULT III
Current control         Current contro         Current control         Current con	Image: mark and the state of the s	Breespereduction(k)	210125929	NC SECTION	11 1902063	1.012539957	8-10077670	1120512322	1,7612-0019	1.300001248	10000000001 1	1.457 350.0	1-000400001	1.60000000A	COO .
Matrix		「「「「「「「」」」」	ADD DO DO	BUDDECK.	W200808P IF	CHORNELS D	IN ADDITIONAL OF	SHOTSPELL	10122 EARC TO	CONTRACTOR OF	SALES THE REAL PROPERTY OF	DAVAGED DA	ANA CORRECT	Districtions of	PAGE 1
Matrix Mathema         Control		ADATE CANTENES & 24	1000041211	10 11/00 1 MG	0.273273/61	ALITUTADO	O MAGAZORIA	10000000000	10200042212	のたんでありますい	O Preddigter	O MICCINESS	202717192	D. 412 206000	1140004
Production         Control         Contro         Control         Control	Tentom         Control         Control <th< td=""><td>Harnin Mall diaman</td><td>北京に対応す</td><td>0.540.572</td><td>NUMBERSON .</td><td>STREET.U</td><td>INCOMPANY.</td><td>SECONT/ 05</td><td>U.142(01/10)</td><td>DESTINATION OF</td><td>Distances of</td><td>10 57879407</td><td>(Internation)</td><td>U-UBRUSKE</td><td>1067.0</td></th<>	Harnin Mall diaman	北京に対応す	0.540.572	NUMBERSON .	STREET.U	INCOMPANY.	SECONT/ 05	U.142(01/10)	DESTINATION OF	Distances of	10 57879407	(Internation)	U-UBRUSKE	1067.0
Tituto (1)         Control (1)	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	ARPh (Dotter is work)	1202000110	0.211 1022.0	0.0040500.0	Divide Martin	0.127760603	0.03063/776	10230 2110	100000001	見出来のたつ		001 201/PCD	0.712062606	012426
Internation         Control         Contro         Control         Control	(1)         (1)<	The manufactor and	記録を開始ー	6896222799 I	TSTHEED'I	47456237 U	FEBORATE!	80177CHE 8	0.590909706	BUSHBBBBC I	1 1309200661	1 SADUGER	\$ 200306388	100147476	0.90890
Off concrete         Off concre         Off concrete         Off concrete <td></td> <td>INVERTING OCT FREMA</td> <td>112000300</td> <td>EBEEL CANEL P</td> <td>121020307</td> <td>C ERC290/ES</td> <td>1090840P</td> <td>1010230-000</td> <td>ESNIT: BHED</td> <td>1.200345061</td> <td>1 0000033</td> <td>●「「「「「「」」</td> <td>E 212909070</td> <td>113000421</td> <td>0.73040</td>		INVERTING OCT FREMA	112000300	EBEEL CANEL P	121020307	C ERC290/ES	1090840P	1010230-000	ESNIT: BHED	1.200345061	1 0000033	●「「「「「「」」	E 212909070	113000421	0.73040
(1)         (1) <td></td> <td>Deal appointed on</td> <td>何ためのという</td> <td>第一部分が行</td> <td>N 24362004</td> <td>Discosting</td> <td>0.6711662301</td> <td>TO SHORE A</td> <td>T1000-227-1</td> <td>SPLASSING -</td> <td>AL BACKRONALIN</td> <td>DITIERNAY D</td> <td>0.546406093</td> <td>TI 57091400F</td> <td>0.076430.0</td>		Deal appointed on	何ためのという	第一部分が行	N 24362004	Discosting	0.6711662301	TO SHORE A	T1000-227-1	SPLASSING -	AL BACKRONALIN	DITIERNAY D	0.546406093	TI 57091400F	0.076430.0
etype://dim		a varial Catagornia (A No	010830214-0	0.0100000	0.52041614.2	102 112 201	0.000763	0.11 02/2058	C1691 19810	「日本のうない	時日に見つ	0.505-52720	Cirite430510	109/00/001	0.000 00
(1)         (1) <td></td> <td>- it is a to show the short</td> <td>1.1503-16</td> <td>HOLDON .</td> <td>1 57 8001414</td> <td>114071590</td> <td>CLABOULL .</td> <td>Fablance 1</td> <td>110092-000-1</td> <td>140000121</td> <td>1 SCTEPTED</td> <td>4 TRATEDATA</td> <td>0121270918</td> <td>1 2540 20027</td> <td>P384. 1</td>		- it is a to show the short	1.1503-16	HOLDON .	1 57 8001414	114071590	CLABOULL .	Fablance 1	110092-000-1	140000121	1 SCTEPTED	4 TRATEDATA	0121270918	1 2540 20027	P384. 1
0.10.0.1         0.00001         <		A D. W.Y. AALAN M.M.	ADDRESS OF THE OWNER	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「日本」のないで、「	TOPIN BOL	the reaction	C N DOLDER	A PLANTER A	1 Follower a	2110 BUC 2010	ALBRIDGE 1	「「「「「「「「「」」」	THE OWNER WATCHING	1112514
Image: 1			States and	A 11 TSCOL	IL CREATHER I	UNPROVED U	10 SECTORE	BENGER I	PERMITER.	N 1 OPERATION	127735165	STREET OF L	0.2327-16.8	Tech Constraint	A LOT A
Non-Bit (C)         Control         Contro         Control		agricolary, intere	1919304 1919304	1 EAGERTIN		Coxesso:	0.701216-30	1.110846152	0.0804.08746	大学の大学	0.166(E)(0)	11.59657560	#131,000,145.4	1-13466.00	0.00814
		Married Cit 3a Faillin	CONCERNENT	PD6490180 /	-Jacrent-	「「「「「「「「」」」」」	AUTOMOTION.	+ Britistides	1.75C 30000	1 SUBSHIELS	14000001010	CODIECE F. F	38E70930E3	LT STATES	1-30%1
Optimization         Optimization<	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	a total for the stratt	0.0535360	四次の日本部	MIT SALE . IT	の日本の日日日	1987 C 1984	STREET, B	1 180854192	「「二元」の	10.35576.01	11.58F2773F13	ETBELIERE	1.001.5050	0.404.360
		manan wa care branch	1,040,190,0	1 TOOTHER	BOARD	1106-00001 -	140 141 44 Million	0.00070714	144966277 0	- December -	ALL CRATTERS	Distance D	THORYPORT D	IN COMPANY	N 6777N
Optimum         Optimum <t< td=""><td></td><td>The drawer all the second</td><td>CONTRACTOR OF</td><td>1001001010101</td><td>-Distantion:</td><td>LTUNESS (ULL</td><td>12.108023510</td><td>7-2288//2 F</td><td>CLARGE BETCH</td><td>- 1988 CT12</td><td>111001027</td><td>T DALANCE</td><td>120000000000000000000000000000000000000</td><td>TRUE NUMBER</td><td>0.95614</td></t<>		The drawer all the second	CONTRACTOR OF	1001001010101	-Distantion:	LTUNESS (ULL	12.108023510	7-2288//2 F	CLARGE BETCH	- 1988 CT12	111001027	T DALANCE	120000000000000000000000000000000000000	TRUE NUMBER	0.95614
Control         Control <t< td=""><td></td><td>A Severative contraction (1)</td><td>ADDRESS OF</td><td>CONCRETE OF</td><td>11 7646 1105 2</td><td>TUBOCCUT D</td><td>10 2710 51</td><td>0.TTMCACA</td><td>1 0 4420 10001</td><td>なないに通っ</td><td>Dian arona</td><td>L TTRADUND</td><td>Stratters 1</td><td>21 200 (342 / C</td><td>D'Alpha</td></t<>		A Severative contraction (1)	ADDRESS OF	CONCRETE OF	11 7646 1105 2	TUBOCCUT D	10 2710 51	0.TTMCACA	1 0 4420 10001	なないに通っ	Dian arona	L TTRADUND	Stratters 1	21 200 (342 / C	D'Alpha
Operation of circle		and the second se	O SCHOOL ST	A THEFT AND	CRAME I	C 1 1 2000 - 000	in oprocession	1 201020020	Butter +	- Generation	HULS IN ALL	DECEMBER OF	CHOICE AND INC. M.	N TY SUDDAG	L'SOLU
Matrix (a)         Matrix (a) <thmatrix (a)<="" th="">         Matrix (a)         Matrix (</thmatrix>		HANNER MUSIC C. M.	D Design The	CERTICITY OF	111 8 MIL	TABIDICEST	CLUCLIPERI D	Rection Tra	0.678.6028	11/10/2006	Distance of the	N MATTER O	D TELEVISION	TOTAL COC 1	o arrest
Operation         Operation <t< td=""><td></td><td>and the same of the same</td><td>Batra o</td><td>Contraction of</td><td>1 LISTON</td><td>CARL AND A D</td><td>WELSUPE E</td><td>WOLLBOATT ST</td><td>Participation a</td><td>COMPLETENCE</td><td>「「「「「「「</td><td>NUMBER OF</td><td>Lincols Vice V</td><td>"I STUTTER"</td><td>O DOWN</td></t<>		and the same of the same	Batra o	Contraction of	1 LISTON	CARL AND A D	WELSUPE E	WOLLBOATT ST	Participation a	COMPLETENCE	「「「「「「「	NUMBER OF	Lincols Vice V	"I STUTTER"	O DOWN
environ         1000000000000000000000000000000000000		A MARTY AT MANALINE	D C 10 40 2000	011762001	0.74TTDOCI	10.000000	「日本にない」の	CTLAUSE -	IT - E & TOMOD E	C. D. IESU 2000	Contraction of	010000-00-01	1 762 7 9000		CIETTER O
Normality         Observation		a later at a state	0172-17-10	LUR PESSEL 9	市に行きたちの	Photo Pauline L	LOCATION IN	Stat Fuller	1 200817157	FROMULT	- TURNET	BURNET -	5.000000000	TRUNCTION V	1-1205.4
		THE PARTY COLVER AS	東京・アイト	10112854	1.045405122	100000000	100,000	1.2286278	L'Strender	1 Martine a	184 81916	1 45022364	1011-100107#	D. 64000000	U.6.94
Control         Control <t< td=""><td></td><td>「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」</td><td>01/26/FF/00 UF</td><td>92.73.2.0</td><td>1 294203467 U</td><td>0.3200626.0</td><td>Contrasto.</td><td>CONTRACTOR OF</td><td>182221222.0</td><td>PERSONAL PROPERTY.</td><td>Manager of</td><td>6110595 · UT</td><td>1112202011</td><td>TISKEDH VIDEL</td><td>CORE L</td></t<>		「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	01/26/FF/00 UF	92.73.2.0	1 294203467 U	0.3200626.0	Contrasto.	CONTRACTOR OF	182221222.0	PERSONAL PROPERTY.	Manager of	6110595 · UT	1112202011	TISKEDH VIDEL	CORE L
		a scored a born dotte	1 206460303	THE RELEASE	11000000001	C Lessenbert L	0.806.78022	0.5305(6-5	1.1807-1966.4	E. HEREELE	まであるの	● 四八日二四十一	11114-244	200-00-00C	U CTRACK
Contraction		a fit wanter bear fait	STRATTON IL	たとしたのでしてい	N241287-217	Since and a	なんないとなる	Andreading to	0 67710161616	CALINADADA -	いたいにないない	In the Thirt Party Party	n molectiona	2010111000	Social U
Contraction         Contract		a for an above property to	15554000	Concernent	TRADER I	日本市のあり日	10 (0 School)	11-3472-5CB	COMPAREMENT IN	C. Z. C. Martel Stor	0.65774 Nat	REAL PROPERTY	書になるのであるので	1292.8.202	1 Teers
Answer volta         Collection         Colle		to per series and a series in the series	12 13 19 100 11 IL	ARCHINE B	Fallender U	Without Bt T	WALLFLOID	- I METERMAN		CARACTERIA IN	THUR HADA	PERSONAL PROPERTY.	BARBAN PATE	122219-0015	AT 401776
Optimizer wire         CBGRFII         C.3/1000         C/0000         C/0000 <thc 000<="" th=""> <t< td=""><td></td><td>a hardward of the law</td><td>No. C. Land</td><td>Constant and the</td><td>TOWN NEWS</td><td>Participant 1</td><td>ATTENTION -</td><td>1 Later to the</td><td>· Discharte</td><td>1.06136034</td><td>1 STATED</td><td>1154 246 2711</td><td>1 Notesting</td><td>A Canadian</td><td>(1. Second</td></t<></thc>		a hardward of the law	No. C. Land	Constant and the	TOWN NEWS	Participant 1	ATTENTION -	1 Later to the	· Discharte	1.06136034	1 STATED	1154 246 2711	1 Notesting	A Canadian	(1. Second
Control (model)         Contro (model)         Contro (mod		A state of section in the line of	語になるのになり、	Distriction of	REPORTED T	LATTRIAL L	A number	Distriction in	I I I I I I I I I I I I I I I I I I I	Paul Paul Paul Paul Paul Paul Paul Paul	D.00771-94 100	ATTENT D	United a child	and a laboration of	
Virtual Internation         Original Internation         Operation         Oper		a statute of a second second	Constantion of the		Diameter and	Statute and the	Nov Sectors	Concerne of	Contraction of the	Contraction of the second	Thursday of the	A TANKING	1100.0000	Contraction of	
Anternation and an extension of the contrast of the contrast		a populate an ere.	South States		ALLONG TALL O	CALOR PROPERTY AND	CONTRACTOR OF	STATE AND A DESCRIPTION	COLOR LONG	Services -	The state of the s	O NATE OF ALL	1 STORES	T DO TOTAL	Annual I
Answerster wirden     The first of the first		The second secon	Trans and the	A NUMBER OF TAXABLE	Contraction of the second	And the second second	The second se	The second secon	COLUMN TO THE	Tart and the				Town of the second second	
		The second second second	CITIC COMPANY	STOCK	Contraction of the	NUCLEAR AND ADDRESS OF	LATING AND	The statement of the st	TALE OF TALE OF TALE	Concentration of the	Contraction of the local division of the loc	I MANUAL IN I	THE PARTY OF THE P	A TRATA BANK	and a second sec
Anywartoware Condect Condect Lange Condect		BARNESS BOOK STORES	A PERSONAL PROPERTY AND INCOME.	Technol -		COLUMN THE PARTY	GI T MAN	Contraction of the local division of the loc	C.C. TRA	D.T. 1302-11	1 412 120029	the charge in	101/200 010	A DOUGLE	
<pre>Activation (12.000 (12.000) (12.001) (12.00</pre>	<pre>Accompany accompany a</pre>	a screptic tablear	In all office in the second	Z-F-VOH27110	at all the second	THE REAL PROPERTY OF	10101110102	IL MERCHAND	In the second second		THE REPORT OF	10111111111111111111111111111111111111	NAMES OF TAXABLE PARTY.		-
The second secon	$ \left  \left( \frac{1}{2} \right) \left( \frac{1}{2}$	INCOME OF TAXABLE PARTY	COLORO IND	Cier mart	Cite Cite	bill branching	ISANGUIC. O	13410000	10140000	C ANDINOS	NUL BERG	ETS MUSIC D	State of the local division of the local div	Distances of	THE A
	AND THE OFFICE A CONTRACT OF A	+ Wernwoor S.M	COMPANY OF	Control Control	Turning and	-c.Th. Andall	10000000	Unicial to a	F 181 CALL AND INC.	-1 152 47 158971	WATER D.	A LINE ALL AND	TANK T	- LINE	New York
Adversion of the second structure advector fragments and the second structure advector structure adve Advector structure advector structure advect	$\left  \left\{ 1 \right\} \left$	arteal much wat rite au	CLORED L	A DESCRIPTION	0.290, P.296	1011623402	I Deby Lat	1040003008	0.100304627	1 AGUSTEEP	5. ST ST ST 1	1 DESCRIPTION	PER MANUTE	HILLING I	0 10 100
the first product of the second structure of the second structure operation of the second structure of	<pre>add the interview of the interview</pre>	WARAN PRIMA IN	TUNES SECON	41.44 whether 24	PERSONAL PROPERTY AND INC.	AL DIAL IN COLUMN	401224300	This house the state	WEITERS ??	- THEN AND AND AND AND AND AND AND AND AND AN	VILLIABILITY	ALTER-DREAM	PLANE DATE	-1.2453275-	11/1/11
states 1, total a constant a transfer a constant a material and the structure a second a person a terrary a second a constant a second a constant a second	a a na strategia de la constante d la constante de la consta	Adding of prospection.	510802822.0	0.184662-12	0.584200526	0.200303010	00.50040530	DAX NOT	0.12425304	1 040009679	D TRUNCIES	た日本の日本の日本	CREAKING	11220001	D RIVER
Constants (reserve data) & the Testant /	Constant (reverse data) ( the fact and /) = (	and and all the dealers	NACESCO DE COMPLEX	- I TUTOLOGIA	1. 42 STREET	10527251 1	MINCHOOD	T. BUNKACH	1 THOULE	COLL 1990 1	A SAMPAGE IN	1126212251	A 14170 047	d > Put million	-
	6 0.000 mode >> > 10 000 million >> 20 - 20 million == 10 000 million >> 10 0000 million >> 10 000 million >> 10 0000 million >> 10 00000 million >> 10 00000 million >> 10 00000 million >> 10 00000000 million >> 10 00000000000000000000000000000000	Shar Data (reissing data	a) / Khan Tetri Dal	1 1						- in the second					-
	5、100000000×1、×10.0100米回水量×鱼×目前用用用用用用用	Contraction of the second seco	Contraction of the local distance of the loc	and the second se	Contraction of the local division of the loc	Consecution of the				1					
												Contraction of the local division of the loc	11 - 120X		1000

Figure 1: Data Format and Highlighting data



Figure 2: The PAM Dialog Box



Figure 3: The PAM Controller









- Web Link Option PAM can hyperlink worksheet cells containing gene ids to the SOURCE database at Stanford so that one can easily search the web database for other information about the gene.
- Additional Sheets A single Excel spreadsheet can have a maximum of 256 columns. When one has more than 256 samples, additional sheets can be specified to overcome this limitation.

Clicking **OK** sends the data over to R. Any missing data are imputed in a new sheet, several other sheets are added and a **PAM Controller** similar to the one shown in figure 3 pops up.

The following is a brief description of the controls.

- **Current Threshold** is a textbox where one would enter a threshold that specifies the degree of shrinkage used by the classified. This field becomes visible after training has been done. The choice of the threshold is typically made after a judicious examination of training errors and the cross-validation results. Until a threshold is chosen, many controls remain inactive.
- **Train** Trains the classifier. This is required and is always the first step. Until this is done, several other fields and the buttons remain disabled.
- Plot Training Error will plot the training error and place the plot in the PAM Plots sheet.
- **Confusion Matrix** button will output a training confusion matrix for a given threshold. If a threshold has not been entered already, then you are asked to enter one.
- **Cross Validate** will do a 10-fold cross validation to help one to choose a threshold that minimizes classification errors.
- **Plot CV Curves** will plot the misclassification errors obtained by cross-validation for various values of the threshold.
- **Plot CV Probabilities** will plot classification probabilities for a specified threshold.
- Plot Centroids will plot the shrunken centroids for a specified threshold.
- List Genes will list the significant genes with the associated score for each class as shown in figure 6.
- **Predict Test Set** can be used to predict a test set.
- **Plot Test Error** can be used to plot the prediction errors for a number of values of the threshold.
- **Plot Test Probabilities** will plot the class probabilities for each sample in the test set for a specified threshold.

	M Administration Bakanavatido		CONTRACTOR CONTRACTOR DATA				
4LuesLuesLuesLuesNata AttentMaterialLues10001000100010001000Lues1000100010001000Lues1000100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues100010001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues10001000Lues1000	A antoninate	-	Load Street	Corr Loss Mr.	Instead - 1	F 6 8 1	4
	Radian and the	ALLEN	BL BEDY -	Ma anona Ma	SCUIN R	SAL HOUR	
	A A PARTY A DECK	CULUM:		The second	-	11.7720 A 4000	
	Production sisters	10 March		4110		0	
	Lark crock	COLUMN -	-500	1	* *	2.0	-
	three assessmentable a	2016-2010		ee		1940	
	100	Selection of	New Y		i e	0	
Microsconstant         Cut/0	I Lawrence And	Summer of	200	0.00	• •	0 0	ł
	Herst stores and a	20461219				91926	
	bit to burner and	The last	2 are	10		2.0	
MANNAL         Seller         D         D           MANNAL         SELE         D         D         D <td>CONTRACTOR STREET</td> <td>and the</td> <td></td> <td>Lave I</td> <td>114</td> <td>0.0</td> <td></td>	CONTRACTOR STREET	and the		Lave I	114	0.0	
	1 BACKING THE REAL	SHELVES		1 T		0.312	-
Matter (Matter (Mattet (Mattet (Mattet (Matter (Matter (Matter (Matter (Matter (Matter	I DESCRIPTION OF	121 Mar	La y	C. Same	-		
Material Matrix         Material Matrix         Material Matrix         Matr	Toole Produced	ALCHER	1	(PTC D			Ŧ
Matrix metrix         Matrix m	- obertrathered	NAME.	AML .	Real Property in	-0	0	
	******	2016000		0.2600		0.0	
Intra         Control         Control <thcontrol< th=""> <thcontrol< th=""> <thcon< td=""><td>N INCOMENDATION</td><td>and the second</td><td>1</td><td>9.29</td><td></td><td></td><td></td></thcon<></thcontrol<></thcontrol<>	N INCOMENDATION	and the second	1	9.29			
Instruction         Construction         Construction </td <td></td> <td>CHI MAN</td> <td></td> <td></td> <td>15.0</td> <td>(=)</td> <td></td>		CHI MAN			15.0	(=)	
Statutistication         Section         Section         Section         Section           Section         Section         Section         Section         Sectio	A PARTICULA ACCURA	And and	a and	é e		0.405	
Instruction         Series         Construction         Series         Construction           Address         Series         Series         Construction         Series         Construction           Address         Series         Construction         Series         Construction         Series           Address         Series         Construction         Series         Construction         Series           Address         Series         Construction         Series         Construction         Series           Series         Series         Construction         Series         Construction         Series           Series         Series         Construction         Series         Construction         Series           Series         Series         Construction         Series         Series         Series           Series         Series         Construction         Series         Series         Series           Series         Series         Series         Series         Series         Series           Series         Series         Series         Series         Series         Series           Series         Series         Series         Series         Series         Series <tr< td=""><td>P. Dokt toward the</td><td>30010010</td><td></td><td></td><td>302.0</td><td>0</td><td></td></tr<>	P. Dokt toward the	30010010			302.0	0	
Interaction         Control         Contro         Control         Control	A REPAIR A RECORDING	(USA)		ð t	a street	0.052	-
Mathematical According (1990)         Control (1990)         Control (1990)         Control (1990) <thcontrol (1990)<="" th="">         Control (1990)</thcontrol>	A LOUGH AND	Source	2.696	19		a	
	and the second se	ALWAY A	144.2	e	-	0 4 4	
Stream (Note)         Cold (Note) <thcold (note)<="" th=""> <thcold (note)<="" th=""></thcold></thcold>	A TANDERGE ALS	NULL NO.	****			0 100	
International and constrained and const	resource spaces in	11141		30	17.0	0	
International Anticipation (1) Anticipation (1) Ant	The second	ACTURINAL VIEW	3.4	äe		0.00	ł
Mathematicant     Section     1112     0       Mathematicant     Section     Labor     Mathematicant	I thur to see the	100000	21%	0	*	e	
Instruction     Control     Control     Control     Control       Instruction     Set (Pr)     Control     1     0       Instruction     Set (Pr)     Control     0     0	No. 60 Stores of Second	2010100		20192	•	0.5	Ī
Mathematical All Sectors     Name (all Sectors     Name (all Sectors     Name (all Sectors       All Sectors     Sectors     Class     0       All Sectors     Sectors     0     0       Sectors     Sectors     1     0       Sectors     Sectors     1     0       Sectors     1     0     0       Sectors     1     0     0       Sectors     1     0     0       Sectors     1     0     0       Sectors     1	P.S. kutson hanh.	New York	1.000	* 17.1	- 11	BIRCK	
All Matrix Lands     States     States     States       All Matres     States <td>author sub starts</td> <td>COMP.</td> <td>2,000</td> <td>ė.</td> <td></td> <td>AMITH A</td> <td>+</td>	author sub starts	COMP.	2,000	ė.		AMITH A	+
Matrix form         Attach         At	MADE AND A	10000	1	0		ognis	
Mathematican Alternational (Mathematican	DESCRIPTION OF THE PARTY	Sec. Sec.	1000			19460	1
S. Markenson, Software     Software     Software       S. Markenson     Software     Software       Software     Software    <	The second second second	TAL MI		i.	1020		
S. M. and Arrinn         2000         2.200         0.000           Boost (Browth         2000         0.000         0.000           Boost (Browth<	Chicadoletication of	NOREI DI		10111	÷	00:00	
a source (transition     2000.11     1.201     0     0       a source (transition     2001.12     1.201     0     0       a source (transition     2001.12     0     0     0       a to the source (transit (	Statement in the second second second	TOTEXOS.		0007		0.00	-
Mill         Control Test         Control         Contro         Control         Control         <	a merce Dependent	の事業の	NT:	0.354		0	
0     E.C.     Second     1 million       1     1     Million     Million	a cost Hi TRACT ra-	Contractory of the		eie	-	10074	-
el des construes 2000 and 2000 and 2000 al 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Per-	Contraction of	Mary -				
FILE IN Reaction & Reaction (200)	Hard good sector sheep 2	20162007	200212	0		0	*
	I. F. M. M. M. M. October & March 1	-0.0×2.0+	Indian Tudan In	V PWM WO	A dest IDONC	PERIO PERIOD PURCHARDED Text	



**Show Prediction** will create a worksheet with the prediction confusion matrix, if computable, and a list of actual and predicted class labels along with the prediction probabilities for each class.

Figures 4 and 5 show some of the plots produced by PAM.

#### 3.1 Prediction

Class labels for test data can be predicted by once again highlighting an area and clicking on the **Predict Test Set** button. A prediction dialog (see 7) allows the user to specify the characteristics of the test set. A new worksheet with the predicted class labels along with the posterior probabilities is output. If the test data has all class labels specified, then a confusion matrix is also provided. See figure 8 where some class labels are missing and highlighted in red.

#### 4 Discussion

The current version of PAM uses Thomas Baier's DCOM server version 0.99. This version has some limitations in handling mixed data types and therefore such data have to be parsed and handled in Visual Basic before sending them over to R. Baier has since released a newer version of the DCOM server and Duncan Temple Lang also has an R DCOM server. Newer versions of PAM will be modified to use these servers as they become available.

Excel's facilities for graphics use a very different model compared to R. Reproducing some R plots in Excel is troublesome and tedious, if not impossible, although many tricks are available. For example, we were unable to reproduce the centroid plot and resorted to embedding the plot in a worksheet. On the other hand, Excel plots are interactive and many users seem comfortable in dealing with them, so that yields an advantage.

With the sizeable number of plots that PAM produces, there are presentation issues that come to the fore. Plots tend to overlap one another and the user has to physically relocate them. Some generated plots usually are more comprehensible when resized. Embedding the plots in worksheets, as PAM does, alleviates the problem to some extent.

Slapping a GUI in front of a package can sometimes obscure the details for more sophisticated users. Such users might wish to see the actual results of the computation rather than just the final results. We have addressed this by creating the special worksheet named **PAM Worksheet (Do Not Edit!)**. This sheet contains intermediate results of computations done by PAM. Every column in this worksheet is given a heading and an elaborate comment is provided so that the user can discern what it represents. If desired, these results can be used for further computations.

In the two months the software has been available, we have found that the installation process needs improvement. The current version demands that the user install an R package and then follow it up with an installation of the Excel Addin.

e first field. The default values for othe	r fields are thos	e that you specifie
er the training data.		
Do not include gene id and	l gene name c	olumns!
Test Set Range		_
Class labels in selection ro	w 2	(Optional)
Sample labels in selection row	1	
Expression data starts in selection row	<b>4</b> 3	

Figure 7: The PAM Controller





Despite elaborate instructions, this is one step too many for the common Windows user and it would be ideal to make this a one-step process. In addition, error handling needs to be improved.

The ideas used in developing PAM should enable us to build a similar interface for Bioconductor.

PAM is freely available from the PAM Website at http://www-stat.stanford.edu/~tibs/PAM.

#### References

- Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5(3):299–314, 1996.
- [2] Erich Neuwirth and Thomas Baier. Embedding R in standard software, and the other way round. In Kurt Hornik and Friedrich Leisch, editors, Proceedings of the 2nd International Workshop on Distributed Statistical Computing, March 15-17, 2001, Technische Universität Wien, Vienna, Austria, 2001. ISSN 1609-395X.
- [3] Robert J. Tibshirani, Trevor J. Hastie, Balasubramanian Narasimhan, and Gilbert Chu. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences*, 99(10):6567–6572, May 2002.
- [4] V. Tusher, R. Tibshirani, and C Chu. Significance analysis of microarrays applied to ionizing radiation response. *Proceedings of the National Academy of Sciences*, 98(9):5116–5121, April 2001.