CMAT Newsletter: January 2004

Wolfgang M. Hartmann

January 2004

1 General Remarks

2 Modifications of Features

2.1 Fixed Bugs

1. A very serious bug with the transpose of a matrix was fixed. The transpose function reshuffles the matrix inside the same memory and does not need a copy in work space. However, it needs the prime number factorization of the matrix dimension. For very large matrices we had an integer overflow which turned an integer product into a negative number.

2. A large number of bugs where fixed for the `svm()` function. Most of them had to do with the tuning algorithm which was never tested in practical applications.

There are still two known bugs in `svm`:

1. The DQP method does not work properly.

2. Parsing function arguments for nested functions.

2.2 Changes in `svm()` Function

A few new options are now available for `svm()`:

Option Name	`Second Column`	Meaning
`"ppar"`		print vector of parameter estimates α
`"block"`	int	fold number for block cross validation
`"split"`	int	fold number for split cross validation
`"random"`	int	fold number for random cross validation
`"tunitr"`	int	max iterations for tuning
`"tuntim"`	int	max time in seconds for tuning
`"tunftol"`	real	f termination for tuning
`"tunxtol"`	real	x termination for tuning
`"tunsel"`	"test"	test error is tuning selection criterion
	"train"	training error is tuning selection criterion

Note, that the `seed` parameter is now valid also for random cross validation. (It was already needed for the SMO method.)

1. Example for Grid Search Parameter Tuning:

```
data = rspfile("..\\tdata\\heart_scale.dat");
modl = "1 = 2 : 14";
class = 1;

/* [1] tuning by grid search: specify values */
c = [ .001 .1 1. 10. 100. ];
kp1 = [ .2 : .05 : .8 ];
tun_g = c |> kp1 ;
```

   For simplicity we use the fast PSVM method. For nonlinear kernel we could also use the fast methods LSVM or SSVM. For linear kernel we could also use the fast ASVM method.

```
optn = [ "print"          3 ,
         "tun"       "gsrch" ,
         "split"         10 ,
         "kern"       "rbf2" ,
         "meth"        "psv" ];
< alfa,sres,vres,yptr > = svm(data,modl,optn,class,tun_g);
```

   Note, that each line of the following history output requires $fold + 1 = 11$ optimizations. The first one is training with the complete data set, the remaining $fold = 10$ are for the cross validation with about $Nobs/fold$ observations left out from model fitting.

```
          ***************************************
          Results of Grid Search for 2 Parameters
          ***************************************

        N         C        KF1 NMisc  Perct    TrainErr
        1  0.0010000  0.2000000   120  44.44  120.000000
        2             0.2500000   120  44.44  120.000000
        3             0.3000000   120  44.44  120.000000
        4             0.3500000   120  44.44  120.000000
        5             0.4000000   120  44.44  120.000000
        6             0.4500000   120  44.44  120.000000
        7             0.5000000   120  44.44  120.000000
        8             0.5500000   120  44.44  120.000000
```

9		0.6000000	120	44.44	120.000000
10		0.6500000	120	44.44	120.000000
11		0.7000000	120	44.44	120.000000
12		0.7500000	120	44.44	120.000000
13		0.8000000	120	44.44	120.000000
14	0.1000000	0.2000000	45	16.67	38.0000000
15		0.2500000	47	17.41	36.0000000
16		0.3000000	49	18.15	35.0000000
17		0.3500000	46	17.04	35.0000000
18		0.4000000	49	18.15	32.0000000
19		0.4500000	48	17.78	30.0000000
20		0.5000000	52	19.26	30.0000000
21		0.5500000	55	20.37	29.0000000
22		0.6000000	57	21.11	27.0000000
23		0.6500000	63	23.33	25.0000000
24		0.7000000	65	24.07	22.0000000
25		0.7500000	66	24.44	23.0000000
26		0.8000000	69	25.56	23.0000000
27	1.0000000	0.2000000	51	18.89	21.0000000
28		0.2500000	53	19.63	17.0000000
29		0.3000000	52	19.26	17.0000000
30		0.3500000	52	19.26	15.0000000
31		0.4000000	53	19.63	11.0000000
32		0.4500000	53	19.63	11.0000000
33		0.5000000	54	20.00	9.00000000
34		0.5500000	54	20.00	9.00000000
35		0.6000000	56	20.74	8.00000000
36		0.6500000	57	21.11	8.00000000
37		0.7000000	60	22.22	5.00000000
38		0.7500000	61	22.59	5.00000000
39		0.8000000	62	22.96	5.00000000
40	10.000000	0.2000000	57	21.11	4.00000000
41		0.2500000	59	21.85	3.00000000
42		0.3000000	61	22.59	2.00000000
43		0.3500000	62	22.96	1.00000000
44		0.4000000	62	22.96	1.00000000
45		0.4500000	62	22.96	1.00000000
46		0.5000000	62	22.96	1.00000000
47		0.5500000	63	23.33	1.00000000

N	C	KF1	NMisc	Perct	TrainErr
48	10.000000	0.6000000	63	23.33	1.00000000
49		0.6500000	61	22.59	1.00000000

50		0.7000000	64	23.70	1.00000000
51		0.7500000	65	24.07	1.00000000
52		0.8000000	65	24.07	0.00000000
53	100.00000	0.2000000	67	24.81	1.00000000
54		0.2500000	64	23.70	0.00000000
55		0.3000000	63	23.33	0.00000000
56		0.3500000	63	23.33	0.00000000
57		0.4000000	63	23.33	0.00000000
58		0.4500000	64	23.70	0.00000000
59		0.5000000	64	23.70	0.00000000
60		0.5500000	63	23.33	0.00000000
61		0.6000000	63	23.33	0.00000000
62		0.6500000	64	23.70	0.00000000
63		0.7000000	65	24.07	0.00000000
64		0.7500000	66	24.44	0.00000000
65		0.8000000	67	24.81	0.00000000

Best 10 Results of Grid Search

N	C	KF1	NMisc	Perct	TrainErr
14	0.1000000	0.2000000	45	16.67	38.0000000
17	0.1000000	0.3500000	46	17.04	35.0000000
15	0.1000000	0.2500000	47	17.41	36.0000000
19	0.1000000	0.4500000	48	17.78	30.0000000
18	0.1000000	0.4000000	49	18.15	32.0000000
16	0.1000000	0.3000000	49	18.15	35.0000000
27	1.0000000	0.2000000	51	18.89	21.0000000
30	1.0000000	0.3500000	52	19.26	15.0000000
29	1.0000000	0.3000000	52	19.26	17.0000000
20	0.1000000	0.5000000	52	19.26	30.0000000

Best solution with C=0.1 and kernel parameter=0.2 selected.

---Start Training Cycle: Technique= PSVM---

****
Evaluation of Training Data Fit
****

Index	Value	StdErr
Absolute Classification Error	38	.
Classification Accuracy	85.92592593	.

```
Concordant Pairs                72.53333333       .
Discordant Pairs                1.866666667       .
Tied Pairs                      25.60000000       .
Goodman-Kruskal Gamma           0.949820789  0.017697016
Kendall Tau_b                   0.714651854  0.042657886
Stuart Tau_c                    0.697942387  0.044058227
Somers D C|R                    0.706666667  0.043560090


                  Classification Table
                  --------------------

                      |    Predicted
              Observed |       -1       1
              ---------|-----------------
                    -1 |      136      14
                     1 |       24      96


     ---Start CrossValidation: Technique=  PSVM---


               **************************
               K=10 Fold Cross Validation
               **************************

               Fold   Nmiss     +Ch     -Ch
                  1       7       5       2
                  2       5       3       2
                  3       4       4       0
                  4       3       2       1
                  5       6       2       4
                  6       2       1       1
                  7       3       2       1
                  8       8       4       4
                  9       4       3       1
                 10       3       2       1

 Runtime for K=10 fold cross validation in seconds: 2


        ***************************************
        Evaluation of Cross Validation Data Fit
        ***************************************

Index                                 Value       StdErr
```

```
Absolute Classification Error          45          .
Classification Accuracy         83.33333333        .
Concordant Pairs                67.97777778        .
Discordant Pairs                2.644444444        .
Tied Pairs                      29.37777778        .
Goodman-Kruskal Gamma           0.925110132  0.024224006
Kendall Tau_b                   0.661675238  0.045783874
Stuart Tau_c                    0.645267490  0.046741328
Somers D C|R                    0.653333333  0.046482972
```

```
                  Classification Table
                  --------------------

                     |    Predicted
              Observed |       -1       1
              ---------|-----------------
                    -1 |      133      17
                     1 |       28      92
```

```
Regularization Parameter C . . . . . . . . . . . . . . . . . 0.1
Norm of Longest Vector . . . . . . . . . . . . . . . . . 1.41421
Number Misclassifactions (Training Data) . . . . . . . . .  38
Number Misclassifactions (Split Cross Validation). . . . .  45
Total Number of Kernel Calls . . . . . . . . . .  6.80922e+007
Time for Optimization. . . . . . . . . . . . . . . . . . . . . 0
Total Processing Time. . . . . . . . . . . . . . . . . . .  95
Optimization Criterion . . . . . . . . . . . . . . . . . .  .
Geometric Margin . . . . . . . . . .  1.08558 (|w|^2= 3.39419)
Number Support Vectors . . . . . . . . . . . .  270 (100.00 %)
Number Support Vectors on Margin . . . . . . . . . . . . .  38
Bias . . . . . . . . . . . . . . . . . . . . . . . -0.00072126
Radius of Sphere Around SV . . . . . . . . . . . . . . 1.33031
Estimated VCdim of Classifier. . . . . . . . . . . . . 7.00675
KT Threshold . . . . . . . . . . . . . . . . . .  6.99029e-018

Number CV Computations . . . . . . . . . . . . . . . . . .  10
CV Estimate of Error (Nmis=45) . . . . . . . . . . . 16.6667 %
CV Estimate of Recall. . . . . . . . . . . . . . . . 76.6667 %
CV Estimate of Precision . . . . . . . . . . . . . . 84.4037 %
```

Even though one training takes less than one second, the complete tuning process requires 93 seconds:

```
  Total Number of Kernel Calls: 6.80922e+007
           Time for Optimization: 0
           Total Processing Time: 95
         Time for Parameter Tuning: 93
```

2. Example for Bounded Optimal Parameter Tuning:

```
/*--- FQP: QPNUSP: with Split Cross Validation ---*/
data = rspfile("..\\tdata\\heart_scale.dat");
modl = "1 = 2 : 14";
class = 1;

/* [2] tuning by automatic NMS: specify bounds */
tun_s = [ 1.e-6   100. ,
          1.e-6     .5 ];


/*--- PSVM: bounded automatic tuning ---*/
optn = [ "print"           3 ,
         "tun"       "bound" ,
         "split"          10 ,
         "kern"       "rbf2" ,
         "meth"        "psv" ];
< alfa,sres,vres,yptr > = svm(data,modl,optn,class,tun_s);


          ---Start Tuning Cycle: Technique=  PSVM---


            ***************************************
            Results of Grid Search for 2 Parameters
            ***************************************

          N          C        KF1 NMisc  Perct    TrainErr
          1  1.00e-006  1.00e-006   120  44.44  120.000000
          2             0.0050010   120  44.44  120.000000
          3             0.0500009   120  44.44  120.000000
          4             0.5000000   120  44.44  120.000000
          5  1.0000010  1.00e-006   120  44.44  120.000000
          6             0.0050010    44  16.30  38.0000000
          7             0.0500009    44  16.30  39.0000000
          8             0.5000000    54  20.00  9.00000000
          9  10.000001  1.00e-006   120  44.44  120.000000
         10             0.0050010    41  15.19  38.0000000
```

```
         11            0.0500009    52  19.26  21.0000000
         12            0.5000000    62  22.96  1.00000000
         13  100.00000  1.00e-006  117  43.33  119.000000
         14            0.0050010    45  16.67  35.0000000
         15            0.0500009    58  21.48  8.00000000
         16            0.5000000    64  23.70  0.00000000
 Best solution with C=10 and kernel parameter=0.00500099 selected.


   Iteration Start:
   N. Variables             2
   N. Bound. Constr.        4       N. Mask  Constr.         0
   Criterion       132.9899993       N. Active Constraints    0

   Iter rest  nfun act  optcrit   nmis  difcrit    std    size
     1    0     5   0  132.856     41  0.49270  0.2080  1.50000
                                         15.0000009  0.00250100
     2    0     6   0  132.856     41  0.13412  6e-002  1.50000
                                         15.0000009  0.00250100
     3    0     8   0  132.856     41  0.03590  1e-002  1.12500
                                         15.0000009  0.00250100
     4    0    10   0  132.848    120  0.02318  1e-002  0.75000
                                         11.8750009  0.00312599
     5    0    12   0  132.825    120  0.03099  1e-002  0.46875
                                         11.7187509  0.00265724
     6    0    13   0  132.825    120  0.02270  1e-002  0.54688
                                         11.7187509  0.00265724
     7    0    15   0  132.825    120   3e-003  1e-003  0.58594
                                         11.7187509  0.00265724
     8    0    17   0  132.821    120   6e-003  2e-003  0.36328
                                         9.98046965  0.00306740
     9    0    19   0  132.821    120   4e-003  1e-003  0.37354
                                         9.98046965  0.00306740
    10    0    21   0  132.819     42   5e-003  2e-003  0.78809
                                         7.93212983  0.00371925
    11    0    23   0  132.819     42   3e-003  1e-003  0.64545
                                         7.93212983  0.00371925
    12    0    25   0  132.816     41   4e-003  2e-003  0.36017
                                         7.89611909  0.00382607
    13    0    27   0  132.816     41   2e-003  1e-003  0.19881
                                         7.89611909  0.00382607
    14    0    28   0  132.816     41   2e-003  8e-004  0.32272
                                         7.89611909  0.00382607
Convergence: Simplex with small differences 0.00179004 in function
values.
```

```
Tuning Result by Optimization F=132.816 (Nmis=41)
*************************************************

          Regularization C     7.8961191
          First Kernel Par     0.0038261


  ---Start Training Cycle: Technique=  PSVM---

         *******************************
         Evaluation of Training Data Fit
         *******************************

Index                                 Value       StdErr
Absolute Classification Error            38         .
Classification Accuracy         85.92592593         .
Concordant Pairs                72.95555556         .
Discordant Pairs                1.955555556         .
Tied Pairs                      25.08888889         .
Goodman-Kruskal Gamma           0.947789973  0.018021781
Kendall Tau_b                   0.714298727  0.042853057
Stuart Tau_c                    0.701234568  0.043906720
Somers D C|R                    0.710000000  0.043392908


               Classification Table
               --------------------

                 |    Predicted
          Observed |       -1       1
          ---------|-----------------
                -1 |      134      16
                 1 |       22      98


  ---Start CrossValidation: Technique=  PSVM---


            *************************
            K=10 Fold Cross Validation
            *************************

            Fold   Nmiss     +Ch     -Ch
               1       5       3       2
```

2	6	3	3
3	5	5	0
4	3	1	2
5	5	1	4
6	0	0	0
7	2	1	1
8	8	4	4
9	4	3	1
10	3	2	1

Runtime for K=10 fold cross validation in seconds: 2

```
***************************************
Evaluation of Cross Validation Data Fit
***************************************
```

Index	Value	StdErr
Absolute Classification Error	41	.
Classification Accuracy	84.81481481	.
Concordant Pairs	71.13333333	.
Discordant Pairs	2.300000000	.
Tied Pairs	26.56666667	.
Goodman-Kruskal Gamma	0.937358148	0.020748504
Kendall Tau_b	0.691703996	0.044255926
Stuart Tau_c	0.679835391	0.045074591
Somers D C|R	0.688333333	0.044666408

```
Classification Table
--------------------

          |    Predicted
 Observed |       -1       1
----------|-----------------
       -1 |      132      18
        1 |       23      97
```

```
Regularization Parameter C . . . . . . . . . . . . . . . 7.89612
Norm of Longest Vector . . . . . . . . . . . . . . . . . 1.41421
Number Misclassifactions (Training Data) . . . . . . . . .  38
Number Misclassifactions (Split Cross Validation). . . . .  41
Total Number of Kernel Calls . . . . . . . . . .  4.91617e+007
Time for Optimization. . . . . . . . . . . . . . . . . . . . 0
Total Processing Time. . . . . . . . . . . . . . . . . . .  76
```

```
Optimization Criterion . . . . . . . . . . . . . . . . . . .   .
Geometric Margin . . . . . . . . . . . 0.245851 (|w|^2= 66.1786)
Number Support Vectors . . . . . . . . . . . .  270 (100.00 %)
Number Support Vectors on Margin . . . . . . . . . . . . . .  38
Bias . . . . . . . . . . . . . . . . . . . . . . . . -5.67557e-005
Radius of Sphere Around SV . . . . . . . . . . . . .  0.284634
Estimated VCdim of Classifier. . . . . . . . . . . . . . 6.36157
KT Threshold . . . . . . . . . . . . . . . . . . . . . 1.2747e-017

Number CV Computations . . . . . . . . . . . . . . . . . . . .  10
CV Estimate of Error (Nmis=41) . . . . . . . . . . . . 15.1852 %
CV Estimate of Recall. . . . . . . . . . . . . . . . . 80.8333 %
CV Estimate of Precision . . . . . . . . . . . . . . . 84.3478 %
```

The full QP method takes much more time than the PSVM method:

```
/*--- FQP: bounded automatic tuning ---*/
optn = [ "print"          3 ,
         "tun"      "bound" ,
         "split"         10 ,
         "kern"      "rbf2" ,
         "meth"       "fqp" ];
< alfa,sres,vres,yptr > = svm(data,modl,optn,class,tun_s);
```

```
           ---Start Tuning Cycle: Technique=  FQP---


            ***************************************
            Results of Grid Search for 2 Parameters
            ***************************************
```

N	C	KF1	NMisc	Perct	TrainErr
1	1.00e-006	1.00e-006	105	38.89	42.0000000
2		0.0050010	118	43.70	42.0000000
3		0.0500009	112	41.48	120.000000
4		0.5000000	120	44.44	120.000000
5	1.0000010	1.00e-006	118	43.70	120.000000
6		0.0050010	45	16.67	42.0000000
7		0.0500009	45	16.67	41.0000000
8		0.5000000	55	20.37	19.0000000
9	10.000001	1.00e-006	93	34.44	72.0000000
10		0.0050010	46	17.04	41.0000000
11		0.0500009	51	18.89	25.0000000
12		0.5000000	64	23.70	1.00000000

```
      13  100.00000  1.00e-006    44  16.30  44.0000000
      14  100.00000  0.0050010    48  17.78  38.0000000
      15             0.0500009    57  21.11  8.00000000
      16             0.5000000    59  21.85  0.00000000
  Best solution with C=100 and kernel parameter=1e-006 selected.


    Iteration Start:
    N. Variables              2
    N. Bound. Constr.         4        N. Mask  Constr.          0
    Criterion       752.9234286        N. Active Constraints     0

    Iter rest  nfun act  optcrit   nmis  difcrit     std     size
      1    0     5   1  255.076      44  516.397  239.18  3.00000
                                           97.5000000  1.000e-006
      2    0     6   1  255.076      44  497.848  234.51  2.25000
                                           97.5000000  1.000e-006
      3    0     8   2  254.708      44  1.10504  0.4595  1.00000
                                           100.000000  1.000e-006
      4    1    12   2  254.708      44  0.00000  0.0000  0.00000
                                           100.000000  1.000e-006
  Convergence: Simplex with small differences 0 in function values.


         Tuning Result by Optimization F=254.708 (Nmis=44)
         *************************************************

                   Regularization C     100.00000
                   First Kernel Par     1.00e-006


            ---Start Training Cycle: Technique=  FQP---


                  *******************************
                  Evaluation of Training Data Fit
                  *******************************
```

Index	Value	StdErr
Absolute Classification Error	44	.
Classification Accuracy	83.70370370	.
Concordant Pairs	68.48888889	.
Discordant Pairs	2.488888889	.
Tied Pairs	29.02222222	.
Goodman-Kruskal Gamma	0.929868503	0.023102582
Kendall Tau_b	0.669438681	0.045290515

```
Stuart Tau_c                    0.651851852  0.046395230
Somers D C|R                    0.660000000  0.046108568


                    Classification Table
                    --------------------

                          |    Predicted
                 Observed |       -1       1
                 ---------|-----------------
                       -1 |      134      16
                        1 |       28      92


      ---Start CrossValidation: Technique=  FQP---

                 **************************
                 K=10 Fold Cross Validation
                 **************************

                 Fold   Nmiss     +Ch     -Ch
                    1       4       4       0
                    2       5       3       2
                    3       6       6       0
                    4       3       2       1
                    5       5       1       4
                    6       1       1       0
                    7       4       3       1
                    8       8       4       4
                    9       4       3       1
                   10       4       2       2

 Runtime for K=10 fold cross validation in seconds: 8


        ***************************************
        Evaluation of Cross Validation Data Fit
        ***************************************

Index                                   Value       StdErr
Absolute Classification Error              44      .
Classification Accuracy         83.70370370      .
Concordant Pairs                68.25000000      .
Discordant Pairs                2.416666667      .
Tied Pairs                      29.33333333      .
Goodman-Kruskal Gamma           0.931603774  0.022839544
```

```
Kendall Tau_b                  0.669895779  0.045131914
Stuart Tau_c                   0.650205761  0.046403557
Somers D C|R                   0.658333333  0.046121568
```

```
                  Classification Table
                  --------------------

              |    Predicted
     Observed |       -1       1
     ---------|-----------------
           -1 |      135      15
            1 |       29      91
```

```
Regularization Parameter C . . . . . . . . . . . . . . . . . . 100
Norm of Longest Vector . . . . . . . . . . . . . . . . . . . . . 1
Number Misclassifactions (Training Data) . . . . . . . . .  44
Number Misclassifactions (Split Cross Validation). . . . .  44
Total Number of Kernel Calls . . . . . . . . . . .  3.37982e+007
Time for Optimization. . . . . . . . . . . . . . . . . . . . . 1
Total Processing Time. . . . . . . . . . . . . . . . . . . . 413
Optimization Criterion . . . . . . . . . . . . . . .  -23658.2
Infinity Norm of Gradient. . . . . . . . . . . .  2.09388e-013
Geometric Margin . . . . . . . . .  0.0764978 (|w|^2= 683.536)
Number Support Vectors . . . . . . . . . . . .  241 ( 89.26 %)
Number Support Vectors on Margin . . . . . . . . . . . . . . 238
Bias . . . . . . . . . . . . . . . . . . . . . . . . -0.00178453
Radius of Sphere Around SV . . . . . . . . . . . .  0.00461052
Estimated VCdim of Classifier. . . . . . . . . . . . . . 1.01453
KT Threshold . . . . . . . . . . . . . . . . . . . . . . . . . 0

Number CV Computations . . . . . . . . . . . . . . . . . . .  10
CV Estimate of Error (Nmis=44) . . . . . . . . . . . 16.2963 %
CV Estimate of Recall. . . . . . . . . . . . . . . . . 75.8333 %
CV Estimate of Precision . . . . . . . . . . . . . . 85.8491 %
```

2.3 New Functions

New functions are implemented:

sir Sliced Inverse Regression (S. Weisberg, 2002)

generead not finished yet.

3 New Developments

3.1 Function `sir`

< gof,parm,atst,ptst,res > = sir(a,model<,"sopt"<,optn<,slic<,class>>>>)

Purpose: The `sir` function implements a number of methods for finding a valid subspace of predictors for given uni- or multivariate response **Y**. It implements asymptotic tests as well as a permutation test for determining the correct dimension.

Input: a the $N \times M$ input data set containing the $N \times p$ predictor matrix **X** and the $N \times k$ response vector or matrix **Y** in its columns. Which columns are selected must be specified with the model string.

model : The analysis model is specified in form of a string, e.g. `model= "3=1 2"`, containing column numbers for variables. The model string specifies which variables (columns) are independend (predictors, covariates) and which are dependend (response variables) and is of the form

$$Y_1 \dots Y_r = X_{effect_1} \dots X_{effect_n}$$

where X_{effect_i} may be one of the following:

- a single variable x_i is an effect
- effects that are interactions among multiple x variables, seprated by the $*$ operator, $x_i * \dots * x_l$
- nested effects: that are single variables or interactions followed by a list of variables inside parentheses, $(x_i \dots x_l)$. Note that only categorical variables without interactions can be listed inside parentheses.

The parameters of the effects are estimated for each response variable Y_i separately. Additionally for the `glim` function you may specify so-called "effects/trials" response variables in the form

$$Y_{effects}/Y_{trials} = X_{effect_1} \dots X_{effect_n}$$

i.e. by listing two data column numbers separated by a forward slash. There are a number of additional features which can make it easier to use such a kind of model specification:

- You may use the | (bar) operator, for the shorter notation of `model = "3 = 1 2 1*2"` by `model="3 = 1 | 2"`.
- You may use the colon operator to express a successive number single variable effects, e.g. `"4= 1 :  3 * 5"` is the same as `"4= 1 2 3*5"`.

- Instead of specifying the entire model in one long string variable you may use a vector of strings. However, each effect must be entirely contained in one vector entry and cannot be continued with the next entry. E.g. the above example could be split into the vector `v` as `v[1]="4="; v[2]="1 2"; v[3]="3*5";` . Note, that the order of the effect listing is important for stepwise regression and Type I estimates.

The syntax of the `model` string argument is the same as for the `glmod()` function except for the additional *events / trial* response specification.

sopt This must be a string specifying the method:

"sir" sliced inverse regression (Li, 1991; Cook, 1998)

"save" sliced average variance estimation (Cook & Weisberg, 1991)

"rphd" principal Hessian direction (residuals) (Li, 1992; Cook, 1998)

"yphd" principal Hessian direction (response y)

"qphd" quadratic principal Hessian direction (Li, 1992)

optn : The option argument is specified in form of a two column matrix where the first column defines the option as string value (in quotes) and the second column can be used for a numeric or string specification of the option. Some of the options are similar to those of the `glmod()` and `glmixd()` functions. See table below for content.

slic This can be a scalar (for one response) or a vector of k values, where k is the number of response variables. It specifies the number of slices for each response y_j, $j = 1, \ldots, k$.

class : This optional argument should be an integer scalar or vector of integer scalars naming the number of columns which are considered categorical (nominal scaled) variables.

Options Matrix Argument: The option argument is specified in form of a two column matrix:

Option Name	Second Column	Meaning
`"alpha"`	real	the significance level for p values; default is $\alpha = 0.05$
`"freq"`	int	column number of FREQ variable
`"nperm"`	int	replications for permutation test
`"numdir"`	int	number directions to be testd, default is $rank(X)$
`"offset"`	int	column number of offset variable
`"print"`	int	amount of printed output
`"seed"`	real	seed value for random generator
`"slic"`	int	number of slices (only for univariate response)
`"vardef"`	string	variance divisor
`"weight"`	int	column number of weight variable

Output: gof goodness-of-fit coefficients

parm eigenvalues and eigenvectors

atst asymptotic test table

ptst permutation test table

res residuals

Restrictions: 1. The input data set should not contain missing values, string or complex data.

Relationships: factor(), sem(), pls()

Examples: 1. Australian Institute of Sport (ais) data set (Weisberg, 2002):

```
options NOECHO;
ais = [
%inc "..\\tdata\\ais.dat";
];
options ECHO;
cname = [ "Sex" "Ht" "Wt" "LBM" "RCC" "WCC" "Hc" "Hg" ];
ais = shape(ais,.,8);
nobs = nrow(ais); nvar = ncol(ais);
ais = ais -> log(ais[,5]);
```

Using the "sir" method:

```
modl = "4 = 2 3 9 6";
optn = [ "slic"        8 ,
         "dec"      "svd" ,
         "numdir"      4 ,
         "print"       3 ,
         "nperm"     499 ,
         "seed"      123 ];
< gof,parm,atst,ptst > = sir(ais,modl,"sir",optn);
```

```
                       *****************
                       Model Information
                       *****************

                 Number Valid Observations  202
                 Response Variable         Y[4]
                 N Independend Variables      4
                 Reduction Method           SIR
                 Number of Sclices            8
                 N Test Directions            4
                 Significance Level:  0.0500000
                 Random Seed                123
```

Permutation Test 499

****
Model Effects
****

X2 + X3 + X9 + X6

****
Simple Statistics
****

Column	Nobs	Mean	Std Dev	Skewness	Kurtosis
Y[4]	202	64.873713	13.070197	0.3611965	-0.2158696
X[2]	202	180.10396	9.7344945	-0.2007970	0.5717271
X[3]	202	75.008168	13.925574	0.2424318	0.4254397
X[9]	202	1.5468840	0.0963023	0.1185203	-0.1197264
X[6]	202	7.1086634	1.8005490	0.8413109	1.5149690

****
Parameter Information
****

Parameter	Meaning
1	X2
2	X3
3	X9
4	X6

Slice Sizes
----

1 :	26	26	25	25	25
6 :	27	30	18		

Eigenvectors

	Dir1	Dir2	Dir3	Dir4
X1	0.01054752	1.569e-004	0.10750718	0.00919795
X2	0.02374812	-0.00409125	-0.06248766	-0.01928675
X3	0.99960915	0.99996148	-0.99211547	0.75708759
X4	-0.01031144	-0.00776400	0.01563290	0.65296386

```
                 Dir1        Dir2        Dir3        Dir4
Eigenvalues  0.87789585  0.15017504  0.03972711  0.01737281
R^2(OLS|dr)  0.70079174  0.95966284  0.99681662  1.00000000


          Asymptotic Chi-Square Tests for Dimension
          *****************************************

                            ChiSqu      df  p-value
           0D vs >= 1D  219.204503      28   0.0000
           1D vs >= 2D  41.8695414      18   0.0012
           2D vs >= 3D  11.5341836      10   0.3174
           3D vs >= 4D  3.50930731       4   0.4765


             Permutation Test Results (Nperm=499)
             ************************************

                            ChiSqu  p-value
               0D vs >= 1D  219.204503   0.0000
               1D vs >= 2D  41.8695414   0.0020
               2D vs >= 3D  11.5341836   0.3080
               3D vs >= 4D  3.50930731   0.4020
```

```
print "GOF=",gof;
print "Parm=",parm;
print "Atst=",atst;
print "Ptst=",ptst;
```

```
             |        X1        X2        X3        X4
-------------------------------------------------------
  Eigenvalues |   0.87790   0.15018   0.03973   0.01737
  R^2(OLS|dr) |   0.70079   0.95966   0.99682   1.00000
           X1 |   0.01055   0.00016   0.10751   0.00920
           X2 |   0.02375  -0.00409  -0.06249  -0.01929
           X3 |   0.99961   0.99996  -0.99212   0.75709
           X4 |  -0.01031  -0.00776   0.01563   0.65296


Atst=
             |   ChiSquare          DF     p-value
---------------------------------------------------
  0D vs >= 1D |     219.205      28.000      2e-031
  1D vs >= 2D |      41.870      18.000       0.001
```

```
  2D vs >= 3D |      11.534      10.000       0.317
  3D vs >= 4D |       3.509       4.000       0.476


Ptst=
          LOW |   ChiSquare     p-value
---------------------------------------
  0D vs >= 1D |     219.205           0
  1D vs >= 2D |      41.870       0.002
  2D vs >= 3D |      11.534       0.308
  3D vs >= 4D |       3.509       0.402
```

Using the "save" method:

```
  modl = "4 = 2 3 9 6";
  optn = [ "slic"         8 ,
           "dec"      "svd" ,
           "numdir"       4 ,
           "print"        3 ,
           "nperm"      499 ,
           "seed"       123 ];
  < gof,parm,atst,ptst > = sir(ais,modl,"save",optn);
```

```
 Slice Sizes
 -----------

   1 :        26       26       25       25       25
   6 :        27       30       18


                               Eigenvectors

                   Dir1         Dir2         Dir3         Dir4
       X1  2.550e-004  0.25690992 -0.00542506  0.00365180
       X2  0.02785912 -0.14344669 -0.00152016 -0.00561149
       X3  0.99935272 -0.88860377  0.99899078  0.99882543
       X4 -0.02275826  0.35185747  0.04456093 -0.04798888


                  Dir1         Dir2         Dir3         Dir4
   Eigenvalues  0.80277774  0.38219979  0.15647959  0.09020895
   R^2(OLS|dr)  0.70079174  0.95966284  0.99681662  1.00000000


                Permutation Test Results (Nperm=499)
```

```
                  **********************************

                                    ChiSqu  p-value
                    0D vs >= 1D  289.196546   0.0020
                    1D vs >= 2D  127.035442   0.1700
                    2D vs >= 3D  49.8310850   0.8240
                    3D vs >= 4D  18.2222083   0.7800
```

Using the "rphd" method:

```
modl = "4 = 2 3 9 6";
optn = [ "slic"        8 ,
         "dec"      "svd" ,
         "numdir"      4 ,
         "print"       3 ,
         "nperm"     499 ,
         "seed"      123 ];
< gof,parm,atst,ptst > = sir(ais,modl,"rphd",optn);

print "GOF=",gof;
print "Parm=",parm;
print "Atst=",atst;
print "Ptst=",ptst;
```

```
                                Eigenvectors

                         Dir1         Dir2         Dir3         Dir4
              X1  0.12764099 -3.378e-004  0.00555008 -0.02548627
              X2 -0.02163115  0.03261381 -0.00734152  0.01343129
              X3 -0.74347735  0.98164633  0.99993021  0.99908616
              X4  0.65611074 -0.18790078 -0.00740800  0.03157356


                   Dir1         Dir2         Dir3         Dir4
 Eigenvalues  1.43025337  1.17504352  1.12441046  0.39992542
 R^2(OLS|dr)  0.70079174  0.95966284  0.99681662  1.00000000


             Asymptotic Chi-Square Tests for Dimension
             *****************************************

                    ChiSqu      df  p-value   IndepT  General
   0D vs >= 1D  35.0150809      10   0.0001   0.0054   0.0181
   1D vs >= 2D  20.2482800       6   0.0025    .       0.0320
   2D vs >= 3D  10.2811903       3   0.0163    .       0.0553
   3D vs >= 4D  1.15456526       1   0.2826    .       0.2663
```

```
          Permutation Test Results (Nperm=499)
          ************************************

                           ChiSqu  p-value
            0D vs >= 1D  35.0150809   0.0080
            1D vs >= 2D  20.2482800   0.0140
            2D vs >= 3D  10.2811903   0.0520
            3D vs >= 4D  1.15456526   0.3240
```

Using the "yphd" method:

```
modl = "4 = 2 3 9 6";
optn = [ "slic"        8 ,
         "dec"     "svd" ,
         "numdir"      4 ,
         "print"       3 ,
         "nperm"     499 ,
         "seed"      123 ];
< gof,parm,atst,ptst > = sir(ais,modl,"yphd",optn);
```

```
                           Eigenvectors

                  Dir1        Dir2        Dir3        Dir4
        X1  0.07148043 -0.01776378 -0.00294621  0.03686273
        X2 -0.06585991  0.01185685 -0.00301142 -0.00523703
        X3 -0.98952448 -0.96941019  0.99998202  0.99890219
        X4  0.10674416  0.24451532  0.00426841  0.02842752
```

```
                     Dir1        Dir2        Dir3        Dir4
 Eigenvalues   10.4455178  1.75295405  1.73428918  1.56858401
 R^2(OLS|dr)   0.70079174  0.95966284  0.99681662  1.00000000
```

```
        Asymptotic Chi-Square Tests for Dimension
        *****************************************

                        ChiSqu      df  p-value
         0D vs >= 1D  69.5582534      10   0.0000
         1D vs >= 2D  5.04973979       6   0.5374
         2D vs >= 3D  3.23297730       3   0.3571
         3D vs >= 4D  1.45469739       1   0.2278
```

```
Permutation Test Results (Nperm=499)
************************************

                      ChiSqu  p-value
  0D vs >= 1D  69.5582534   0.0000
  1D vs >= 2D  5.04973979   0.5440
  2D vs >= 3D  3.23297730   0.2440
  3D vs >= 4D  1.45469739   0.0980
```

3.2 Function `generead`