SAS® Enterprise Miner™ and SAS® Text Miner Procedures Reference for SAS® 9.1.3

The DMNEURL Procedure

(Book Excerpt)

SAS® Documentation

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS® Enterprise Miner™ and SAS® Text Miner Procedures: Reference for SAS® 9.1.3*, Cary, NC: SAS Institute Inc.

SAS® Enterprise Miner™ and SAS® Text Miner Procedures: Reference for SAS 9.1.3

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st electronic book, October 2008

CHAPTER

8

The DMNEURL Procedure

Overview: DMNEURL Procedure

Purpose of PROC DMNEURL

In its current form, PROC DMNEURL tries to establish a nonlinear model for the prediction of a binary or interval scaled response variable (called target in data mining terminology). The algorithm used in DMNEURL was developed to overcome some problems of PROC NEURAL for data mining purposes, especially when the data set contains many highly collinear variables:

1 The nonlinear estimation problem in common Neural Networks is seriously underdetermined yielding to highly rankdeficient Hessian matrices and resulting in extremely slow convergence (close to linear) of nonlinear optimization algorithms.

$\Rightarrow$Full-rank estimation.

2 Each function call in PROC NEURAL corresponds to a single run through the entire (training) data set and normally many function calls are needed for convergent nonlinear optimization with rankdeficient Hessians.

⇒Optimization of discrete problem with all data incore.

3 Because the zero eigenvalues in a Hessian matrix correspond to long and very flat valleys in the shape of the objective function, the traditional Neural Net approach has serious problems to decide when an estimate is close to an appropriate solution and the optimization process can be terminated.

⇒Quadratic convergence.

4 For the same reasons, the common Neural Net algorithms tend toward finding local rather than global optimal solutions and the optimization result often is very sensitive w.r.t. the starting point of the optimization.

⇒Good starting point.

With PROC DMNEURL we deal with specified optimization problems (with full rank Hessian matrices) which have few parameters and for which good starting points can be obtained. The convergence of the nonlinear optimizer is normally very fast, resulting commonly in less than 10 iterations per optimization. The function and derivative calls during the optimization do not need any passes through the data set; however, the search for obtaining good starting points and the final evaluations of the solutions (scoring of all observations) need several passes through the data and needs to perform a number of preliminary tasks. In PROC DMNEURL we fit separately an entire set of activation functions and select the best result. Since the optimization processes for different activation functions do not depend on each other, a possible enhancement is to reduce the processing time through parallel processing. Except for applications where PROC NEURAL would hit a local solution much worse than the global solution, it is not expected that PROC DMNEURL will outperform PROC NEURAL in the precision of the prediction. However, we have found the results of PROC DMNEURL very close to those of PROC NEURAL, and for very large data sets PROC DMNEURL will be faster than PROC NEURAL. For small data sets, PROC NEURAL could be much faster than PROC DMNEURL, especially for an interval target. The most efficient application of PROC DMNEURL is the analysis of a binary target variable without FREQ and WEIGHT statements and without COST variables in the input data set.

Missing Values

Observations with missing values in the target variable (response or dependent variable) are not included in the analysis. Those observations are, however, scored, that is, predicted values are computed. Observations with missing values in the predictor variables (independent variables) are processed depending on the scale type of the variable:

- For numeric variables, missing values are replaced by the (weighted) mean of the variable.
- For class variables, missing values are treated as an additional category.

Syntax: DMNEURL Procedure

PROC DMNEURL <*option(s)*>;
<**DECISION***option(s)*;>
<**FUNCTION** *name(s)*;>

<**FREQ***variable(s)*;>
<**LINK** *name*;>
<**TARGET***variable(s)*;>
<**VAR** *variable(s)*;>
<**WEIGHT***variable(s)*;>

PROC DMNEURL Statement

Invokes the DMNEURL procedure.

PROC DMNEURL <*option(s)*>;

Required Arguments

DATA=*SAS-data-set :*
specifies an input data set. This option must be specified; no default is permitted.

DMDBCAT=*SAS-catalog :*
specifies an input catalog of meta information generated by PROC DMDB. The catalog contains important information (for example, range of variables, number of missing values of each variable, moments of variables) that is used by many other procedures which require a DMDB data set. The DMDBCAT= catalog and the DATA= data set must be insync to obtain proper results! This option must be specified.

Default: No default is permitted.

Options

ABSGCONV, ABSGTOL : $r \geq 0$
specifies an absolute gradient convergence criterion for the default (OPTCRIT=SSE) optimization process. See the document of PROC NLP in SAS/OR for more details.

Default: Default is ABSGCONV=5e-4 in general and ABSCONV=1e-3 for FUNCTION=EXP.

CORRDF :
specifies that the correct number of degrees of freedom is used for the values of RMSE, AIC, and SBC. Without specifying CORRDF the error degrees of freedom are computed as $W - p$, where W is the sum of weights (if the WEIGHT statement is not used, each observation has a weight of 1 assigned, and W is the total number of observations) and p is the number of parameters. When CORRDF is specified the value p is replaced by the rank of the joint Jacobian.

COV, CORR :
specifies that a covariance or correlation matrix is used for computing eigenvalues and eigenvectors compatible with the PRINCOMP procedure. The COV and CORR options are valid only if an OUTSTAT= data set is specified. If neither COV nor CORR are specified, the eigenvalues and eigenvectors of the cross product matrix $X^T X$ are computed and written to the OUTSTAT= data set.

CRITWGT=r: $r > 0$

specifies a positive weight for a weighted least squares fit. Currently this option is valid only for binary target. Values of $r > 1$: will enforce a better fit of the (1,1) entry in the accuracy table which might be useful for fitting rare events. Values of $0 < r < 1$: will enforce a better fit of the (0,0) entry in the accuracy table. Note, that values for r which are far away from $r = 1$ will reduce the fit quality of the remaining entries in the frequency table. At this time values of either $1 < r < 2$ or $.5 < r < 1$ are preferred.

CUTOFF=r : $0 < r < 1$

specifies a cutoff threshold for deciding when a predicted value of a binary response is classified as 0 or 1. If the value of the posterior,$(\widehat{y_i})$, for observation i is smaller the specified cutoff value, the observation is counted in the first column of the accuracy table (that is as 0), otherwise it is counted in the second column (that is as 1). For nonbinary target the cutoff= value is not used.

Default: CUTOFF = .5

FCRIT

specifies that the probability of the F test is being used for the selection of principal components rather than the default R^2 criterion.

Default: R^2 criterion

GCONV, GTOL : $r \geq 0$

specifies a relative gradient convergence criterion for the optimization process. See the document of PROC NLP in SAS/OR for more details.

Default: GCONV=1e-8

MAXCOMP=i : $2 \leq i \leq 8$

specifies an upper bound for the number of components selected for predicting the target in each stage. Good values for MAXCOMP are inbetween 3 and 5. Note, that the computer time and core memory will increase superlinear for larger values than 5. There is one memory allocation which takes n^m long integer values, where n is the value specified with the NPOINT= option and m is the value specified by the MAXCOMP= option. The following table lists values of $4n^m/1000000$ for specific combinations of (n, m). This is the actual memory requirement in Megabytes assuming that a long integer takes 4 bytes storage.

n	m=3	m=4	m=5	m=6	m=7	m=8
5	0	0	0	0	0	2
7	0	0	0	0	3	23*
9	0	0	0	2	19*	172
11	0	0	1	7*	78	857
13	0	0	2*	19	250	3263
15	0	0*	3	46	683	10252
17	0*	0	6	97	1641	27903
19	0	1	10	188	3575	67934

The trailing asterisk indicates the default number of points for a given number of components. Therefore, values larger than 8 for i in

```
MAXCOMP=i
```

are reduced to this upper range. It seems to be better to increase the value i of the MAXSTAGE=i option when higher precision is requested.

MAXFUNC=i : $i \geq 0$
specifies an upper bound for the number of function calls in each optimization. Normally the default number of function calls will be sufficient to reach convergence. Larger values should be used if the iteration history indicates that the optimization process was close to a promising solution but would have needed more than the specified number of function calls. Smaller values should be specified when a faster but suboptimal solution might be sufficient.

Default: MAXFUNC=500

MAXITER=i : $i \geq 0$
specifies an upper bound for the number of iterations in each optimization. Normally the default number of iterations will be sufficient to reach convergence. Larger values should be used if the iteration history indicates that the optimization process was close to a promising solution but would have needed more than the specified number of iterations. Smaller values should be specified when a faster but suboptimal solution might be sufficient.

Default: MAXITER=200

MAXROWS=i : $i \geq 1$
specifies an upper bound for the number of independent variables selected for the model. More specific, this is an upper bound for the rows and columns of the$X'X$ matrix of the regression problem. Note, that the $X^T X$ matrix used for the stepwise regression takes $n_{rows}(n_{rows}+1)/2$ double precision values storage in RAM. For the default maximum size of $n_{rows} = 3000$ you will need more than $3000 * 1500 * 8$ bytes RAM, which is slightly more than 36 megabytes.

Default: MAXROWS=3000

MAXSTAGE=i : $i \geq 1$
specifies an upper bound for the number of stages of estimation. When a missing value is specified, the multistage estimation process is terminated

- if the sum-of-squares residual in the component selection process changes by less than 1%
- or when an upper range of 100 stages are processed.

That means, not specifying MAXSTAGE= or specifying a missing value are treated differently. Large values for MAXSTAGE= might result in numerical problems: the discretization error might be too large and the fit criterion does no longer improve and can actually become worse. In such a case the stagewise process is terminated with the last good stage.

Default: MAXSTAGE=5

MAXSTPT=i : $i \geq 1$
specifies the number of values of the objective function inspected for the start of the optimization process. Larger values than the default value can improve the result of the optimization especially when more than three components are used.

Default: MAXSTPT=250

MAXVECT=i : $i \geq 2$
specifies an upper bound for the number of eigenvectors made available for selection. The default is MAXVECT=400. Smaller values should be used only if there are memory problems for storing the eigenvectors when too many variables are included in the analysis. The specified value for MAXVECT= cannot be smaller than that for MINCOMP=. If the specified value of MAXVECT= is larger than the value for MAXROWS= it is reduced to the value of MAXROWS=.

Default: MAXVECT=400

MEMSIZ=i : $i \geq 1$

For interval targets and in a multiple stage process some memory consuming operations are being performed. For very large data sets the computations can significantly depend on the size of the available RAM memory for those computations. By default MEMSIZ=8 specifies the availability of 8 mb of RAM for such operations. Since other operations need additional memory not more than 25 percent of the total amount of memory should be specified here. If you are running out of memory during the DMNEURL run, you can actually specify a smaller amount than the default 8 mb.

Default: MEMSIZ=8

MINCOMP=i : $2 \leq i \leq 8$

specifies a lower bound for the number of components selected for predicting the target in each stage. The default is MINCOMP=2. The specified value for MINCOMP= cannot be larger than that for MAXCOMP=. The MINCOMP= specification can permit the selection of components which otherwise would be rejected by the STOPR2= option. PROC DMNEURL can override the specified value when the rank of the $X'X$ matrix is less than the specified value.

Default: MINCOMP=2

NOMONITOR

suppresses the output of the status monitor indicating the progress made in the computations.

NOPRINT

suppresses all output printed in the output window.

NPOINT=i : $5 \leq i \leq 19$

number of discretization points (should be even inbetween 5 and 19). By default NPOINT= is selected depending on the number of components selected in the model using the MINCOMP= and MAXCOMP= options.

OPTCRIT=*SSE*|*ACC*|*WSSE*

specifies the criterion for the optimization:

OPTCRIT=SSE	the sum-of-squares error is minimized.
OPTCRIT=ACC	a measure of the accuracy rate is maximized. (For interval target the Goodman-Kruskalγ is applied on a frequency table defined by deciles of the actual target value.)
OPTCRIT=WSSE	a weighted sum-of-squares criterion is minimized. When this option is specified the weight must be specified using the CRITWGT= option. Currently this option is valid only for binary target.

OUT=*SASdataset*

specifies an output data set generated by PROC DMNEURL which contains the predicted values (posteriors) and residuals for all observations in the DATA= input data set.

Variables of the output data set:

$idvarnam_i$	values of all ID variables
TARGET	(character) name of the target
STAGE	number of stage
_P_	predicted value ($\hat{y}$)
_R_	residual ($y - \hat{y}$)

The following variables are added if a DECISION statement is used:

BSTDEC

CONSEQ

EVALUE expected profit or cost value

$decvar_i$ expected values for all decision variables
The number of observations in the OUT= data set agrees with that of the DATA= input data set.

OUTCLASS=*SASdataset*
specifies an output data set generated by PROC DMNEURL that contains the mapping inbetween compound variable names and the names of variables and categories of CLASS variables used in the model. The compound variable names are used to denote dummy variables that are created for each category of a CLASS variable with more than two categories. Since the compound names of dummy variables are used for variable names in other data sets, the user must know to which category each compound name corresponds. The OUTCLASS= data set has only three character variables

NAME
contains compound name used as variable names in other output data sets

VAR
contains variable name used in DATA= input data set

LEVEL
contains level name of variable as used in DATA= input data set.
Note, if the DATA= input data set does not contain any CLASS variables the OUTCLASS= data set is not written.

OUTEST=*SASdataset*
specifies an output data set generated by PROC DMNEURL that contains all the model information necessary for scoring additional cases or data sets. Variables of the output data set:

TARGET
(character) name of the target

TYPE
(character) type of observation

NAME
(character) name of observation

STAGE
number of stage

MEAN
contains different numeric information based on different _TYPE_ observations (See below for more details.)

STDEV
contains different numeric information based on different _TYPE_ observations (See below for more details.)

$varname_i$

variables in the model variables. The first variables correspond to CLASS (categorical); the remaining variables are continuously (interval or ratio) scaled. Note that for nonbinary CLASS (nominal or ordinal categorical) variables, a set of

binary dummy variables is created. In those cases the prefix of variable names $varname_i$ used for a group of variables in the data set can be the same for a successive group of variables which differs only by a numeric suffix.
This data set contains all the model information necessary to compute the predicted model values (scores).

1 The _TYPE_=_V_MAP_ and _TYPE_=_C_MAP_ observations contain the mapping indices between the variables used in the model and the number of the variable in the data set.
   - The _MEAN_ variable contains the number of index mappings.
   - The _STDEV_ variable contains the index of the target (response) variable in the data set for the _TYPE_=_V_MAP_ observation. For _TYPE_=_C_MAP_ it contains the level (category) number of a categorical target variable that corresponds to missing values.

2 The _TYPE_=_EIGVAL_ observation contains the sorted eigenvalues of the $X'X$ matrix. Here, the _MEAN_ variable contains the number of model variables (rows/columns of the model$X'X$ matrix) and the _STDEV_ variable contains the number c of model components.

3 For each stage of the estimation process two groups of observations are written to the OUTEST= data set:

   a The _TYPE_=_EIGVEC_ observations contain a set of c principal components which are used as predictor variables for the estimation of the original target value y (in stage 0) or for the prediction of the stage i residual. Here, the _MEAN_ variable contains the value for the criterion used to include the component into the model which is normally the R^2 value. The _STDEV_ variable contains the eigenvalue number to which the eigenvector corresponds.

Function	**Formula**
TANH	$a * \tanh\left(b * x\right)$
ARCTAN	$a * a\tan\left(b * x\right)$
LOGIST	$\exp\left(a * x\right) / \left(1 + \exp\left(b * x\right)\right)$
GAUSS	$a * \exp\left(-\left(b * x\right)^2\right)$
SIN	$a * \sin\left(b * x\right)$
COS	$a * cos\left(b * x\right)$
EXP	$a * \exp\left(b * x\right)$

   The _NAME_ variable reports the corresponding name of the best activation function found.

   b The _TYPE_=_PARMS_ observations contain for each activation function the$p = 2c + 1$ parameter estimates. Here, the _MEAN_ variable contains the value for the optimization criterion and the _STDEV_ variable contains the accuracy value of the prediction.

OUTFIT=*SASdataset*

specifies an output data set generated by PROC DMNEURL which contains a number of fit indices for each stage and for the final model estimates. For a binary target (response variable) it also contains the frequencies of the 2 2 accuracy table of the best fit at the final stage. The same information is additionally provided if a TESTDATA= input data set is specified.

Variables of the output data set:

TARGET (character) name of the target

DATA (character) specifies the data set to which the fit criteria correspond: =TRAINING: fit criteria belong to DATA= input data set =TESTDATA: fit criteria belong to TESTDATA= input data set

TYPE=_FITIND_
for fit indices;

TYPE=_ACCTAB_
for frequencies of accuracy table (only for binary target)

STAGE number of stages in the estimation process

SSE sum-of-squared error of solution

RMSE root mean squared error of solution

ACCU percentage of accuracy of prediction (only for categorical target)

AIC Akaike information criterion

SBC Schwarz' information criterion

The following variables are added if a DECISION statement is used:
PROF
APROF
LOSS
ALOSS
_IC_
ROI

OUTSTAT=*SASdataset*
specifies an output data set generated by PROC DMNEURL which contains all eigenvalues and eigenvectors of the $X'X$ matrix. When this option is specified, no other computations are performed and the procedure terminates after writing this data set.
Variables of the OUTSTAT= output data set:

TYPE (character) type of observation

EIGVAL contains different numeric information

varname$_i$ variables in the model; the first variables correspond to CLASS (categorical) the remaining variables are continuously (interval or ratio) scaled. Note, that for nonbinary CLASS (nominal or ordinal categorical) variables a set of binary dummy variables is created. In those cases the prefix of variable names *varname*$_i$ used for a group of variables in the data set can be the same for a successive group of variables which differs only by a numeric suffix.

Observations of the OUTSTAT= output data set:

1 The first three observations, _TYPE_=_V_MAP_ and _TYPE_=_C_MAP_, contain the mapping indices between the variables used in the model and the

number of the variables in the data set. The _EIGVAL_ variable contains the number of index mappings. This is the same information as in the first observation of the OUTEST= data set, except that here the _TYPE_=_EIGVAL_ variables replaces the _TYPE_=_MEAN_ variable in the OUTEST= data set.

2 The _TYPE_=_EIGVAL_ observation contains the sorted eigenvalues of the $X'X$ matrix.

3 The _TYPE_=_EIGVEC_ observations contain a set of n eigenvectors of the $X'X$ matrix. Here, the _EIGVAL_ variable contains the eigenvalue to which the eigenvector corresponds.

PALL

- If an OUTSTAT= data set is specified, that is only principal components are being computed, the following table illustrates the output options:

Output	**PSHORT**	**Default**	**PALL**
Simple Stat	x	x	x
Eigenvalues	x	x	x

- If no OUTSTAT= data set is specified, that is a nonlinear model based on activation and link functions is being optimized. The following is a list of the output options:

  Output
  NOPRINT
  PSHORT
  default
  PALL

PMATRIX
This option is valid only if an OUTSTAT= data set is specified, that is when DMNEURL is used only for computing eigenvalues and eigenvectors of the $X'X$, covariance, or correlation matrix. If PMATRIX is specified, this matrix is being printed. Since this matrix might be very large its printout is not included by that of the PALL option.

POPTHIS
print the detailed histories of all optimization processes. The PALL option includes only the summarized forms of the history output (header and result).

PSHORT
see the PALL option for the amount of output being printed.

PTABLE
specifies the output of accuracy tables. This option is invoked automatically if the PALL option is specified.

SELCRIT=*SSE | ACC | WSSE*
specifies the criterion for selecting the best result among all of the activation functions:

SELCRIT=SSE	select solution with smallest sum-of-squares error.
SELCRIT=ACC	select solution with largest accuracy rate. (For interval target the Goodman-Kruskal γ is applied on a frequency table defined by deciles of the actual target value.)

SELCRIT=WSSE select solution with smallest weighted sum-of-squares error. This option is valid only for binary target. When this option is specified the weight must be specified using the CRITWGT= option.

SINGULAR=r
specifies a criterion for the singularity test.

Default: The default is SINGULAR=$1.e-8$ and should not be changed if there are no significant reasons to do so.

STOPR2=r
specifies a lower value for the incremental model R^2 value at which the variable selection process is stopped. The STOPR2= criterion is used only for the R2 values of the components selected in the range specified by the MINCOMP= and MAXCOMP= values.

Default: STOPR2=$5e-5$.

TESTDATA=*SASdataset*
specifies a second input data set, and which must contain all variables of the DATA= input data set which are used in the model. The variables not used in the model can be different. The order of variables is not relevant. If TESTDATA= is specified, you can specify a TESTOUT= output data set (containing predicted values and residuals) which relates to the TESTDATA= input data set the same as the OUT= data set relates to the DATA= input training data set. When specifying the TESTDMDB option you can use a data set generated by PROC DMDB as the TESTDATA= input data set.

TESTDMDB
permits the use of a data set generated by PROC DMDB to be specified as a TESTDATA= input data set. If this option is not specified, the data set specified with TESTDATA= must be a normal SAS data set.

TESTOUT=*SASdataset*
specifies an output data set which is in structure identical to the OUT= output data set but relates to the information given in the TESTDATA= input data set rather than that of the DATA= input data set used in the OUT= output data set. The number of observations in the TESTOUT= data set agrees with that of the TESTDATA= input data set.

DECISION Statement

DECISION *option(s)*;

For the syntax of the DECISION statement see the document of PROC DECIDE.

FUNCTION and LINK Statement

FUNCTION *name(s)*;

LINK *name*;

An activation function f and a link function g can be specified for the mapping inbetween the component scores s_{ij} and the values y_i of the response variable (stage=0) (or the residuals in stage > 0),

$$\hat{y}_i = g\left(f^{(k)}\left(s_{ij}, \theta_j\right)\right),\ i = 1, ..., N\ j = 1, ..., p$$

for each activation function $f^{(k)}, k = 1, ..., K$. The FUNCTION and LINK statement can be used to specify the functions $f^{(k)}$ and g:

FREQ or FREQUENCY Statement

FREQ *onevar;*

FREQUENCY *onevar;*

One numeric (interval scaled) variable can be specified as a FREQ variable. Note, that a rational value is truncated to the next integer. It is recommended to specify the FREQ variable already in the PROC DMDB run. Then the information is saved in the catalog and that variable is used automatically as a FREQ variable in PROC DMNEURL. This also ensures that the FREQ variable is being used automatically by all other PROCs in the EM project.

FUNCTION Statement

FUNCTION *name(s)*;

One or more of the following activation functions f can be specified

Function	**Formula**
SQUARE	$(a + b * x) * x$
TANH	$a * \tanh(b * x)$
ARCTAN	$a * a\tan(b * x)$
LOGIST	$\exp(a * x) / (1 + \exp(b * x))$
GAUSS	$a * \exp\left(-(b * x)^2\right)$
SIN	$a * \sin(b * x)$
COS	$a * \cos(b * x)$
EXP	$a * \exp(b * x)$

If more than one function $f^{(k)}$ is specified, each of the specified functions is evaluated during the estimation process and the best result w.r.t. to the sum-of-squares residual or accuracy (see SELCRIT= option) is selected. By default all available activation functions are used.

LINK Statement

LINK *name*;

Currently only one of the following link functions can be used for the outer function g:

Function	Formula
IDENT	x
LOGIST	$\exp(x) / (1 + \exp(x))$
RECIPR	$1/x$

By default, the LOGIST function is used for a binary target and the IDENT(ity) function is used for interval target. In a parallelized version of PROC DMNEURL, multiple functions g could be feasible.

TARGET Statement

TARGET *onevar*;

One variable name can be specified identifying the target (response) variable for the two regressions. Note, that one or more target variables might be specified already with the PROC DMDB run.

VAR or VARIABLES Statement

VAR *varlist;*

VARIABLES *varlist;*

All variables, numeric (interval) and categorical (CLASS) variables which can be used for independent variables are specified with the VAR statement.

WEIGHT or WEIGHTS Statement

WEIGHT *onevar;*

WEIGHTS *onevar;*

One numeric (interval scaled) variable can be specified as a WEIGHT variable. It is recommended to specify the WEIGHT variable already in the PROC DMDB invocation.

Then the information is saved in the catalog and that variable is used automatically as a FREQ variable in PROC DMNEURL.

Details: DMNEURL Procedure

Scoring the Model Using the OUTEST= Data Set

The score value $\hat{y}_i$ is computed for each observation $i = 1, ..., N_{obs}$ with nonmissing value of the target (response) variable y of the input data set. All information needed for scoring an observation of the DMDB data set is contained in the output of the OUTEST= data set. First an observation from the input data set is mapped into a vector v of n new values in which

1 CLASS predictor variables with K categories are replaced by $K + 1$ or K dummy (binary) variables, depending on the fact whether the variable has missing values or not.

2 Missing values in interval predictor variables are replaced by the mean value of this variable in the DMDB data set. This mean value is taken from the catalog of the DMDB data set.

3 The values of a WEIGHT or FREQ variable are multiplied into the observation.

4 For an interval target variable y its value is transformed into the interval [0,1] by the relationship

$$y_i^{new} = \frac{y_i - y_{min}}{y_{max} - y_{min}}$$

5 Allpredictor variables are transformed into values with zero mean and unit standard deviation by

$$x_{ij}^{new} = \frac{x_{ij} - Mean(x_j)}{StDev(x_j)}$$

The values for $Mean\left(x_j\right)$ and $StDev\left(x_j\right)$ are listed in the OUTEST= data set.

This means, that in the presence of CLASS variables the n-vector v has more entries than the observation in the data set. The scoring is additive across the stages. The following information is available for scoring each stage

- c components (eigenvectors) z_l each of dimension n
- the best activation function f and a specified link function g
- the $p = 2c + 1$ optimal parameter estimates θ_j

For each component z_l we compute the component score u_l,

$$u_l = \sum_{j=1}^{n} z_l v_j$$

similar to principal component analysis. With those values u_l the model can be expressed as

$$\hat{y} = \sum_{i\,stage}^{n\,stage} g\left(f\left(u, \theta\right)\right)$$

where$f$ is the best activation function and g is the specified link function. In other words, this means, that given the u_l the value w is computed from

$$w = \theta_o + \sum_l (u_l, a_l, b_l)$$

where a_l and b_l are two of the $p = 2c + 1$ optimal parameters and f is defined as

Function	**Formula**
SQUARE	$(a + b * u) * u$
TANH	$a * \tanh(b * u)$
ARCTAN	$a * a \tan(b * u)$
LOGIST	$\exp(a * u) / (1 + \exp(b * u))$
GAUSS	$a * \exp\left(-(b * u)^2\right)$
SIN	$a * \sin(b * u)$
COS	$a * \cos(b * u)$
EXP	$a * \exp(b * u)$

For the first component $a_1 = \theta_1$ and $b_1 = \theta_2$, for the second component $a_2 = \theta_3$ and $b_2 = \theta_4$, and for the last component $a_c = \theta_{p-1}$ and$b_c = \theta_p$ are used. The link function g is applied on w and yields to h

Function	**Formula**
IDENT	$h = w$
LOGIST	$h = \exp(w) / (1 + \exp(w))$
RECIPR	$1/w$

Across all stages the values of h are added to the predicted value (posterior) $\hat{y}$.

Examples: DMNEURL Procedure

Example 1: Application: HMEQ Data Set: Binary Target BAD

To illustrate the use of PROC DMNEURL we choose the HMEQ data set:

```
libname sampsio '/sas/a612/dmine/sampsio';
proc dmdb batch data=sampsio.hmeq out=dmdbout dmdbcat=outcat;
    var LOAN MORTDUE VALUE YOJ DELINQ CLAGE NINQ CLNO DEBTINC;
    class BAD(DESC) REASON(ASC) JOB(ASC) DEROG(ASC);
    target BAD;
```

```
run;
```

When selecting the binary target variable BAD a typical run of PROC DMNEURL would be the following:

```
proc dmneurl data=sampsio.hmeq dmdbcat=outcat
         outclass=oclass outest=estout out=dsout outfit=ofit
         ptable maxcomp=3 maxstage=5;
    var LOAN MORTDUE VALUE REASON JOB YOJ DEROG DELINQ
          CLAGE NINQ CLNO DEBTINC;
    target BAD;
run;
```

The number of parameters p estimated in each stage of the optimization is $p = 2 * c + 1$, where c is the number of components that is selected at the stage. Since here $c = 3$ is specified with the MAXCOMP= option each optimization process estimates only $p = 7$ parameters. First some general information are printed and the four moments of the numeric data set variables involved in the analysis.

```
                         The DMNEURL Procedure

                  Binary Target                      BAD
                  Number Observations               5960
                  NOBS w/o Missing Target           5960
                  Link Function                   LOGIST
                  Selection Criterion                SSE
                  Optimization Criterion             SSE
                  Estimation Stages                    5
                  Max. Number Components               3
                  Minimum R2 Value              0.000050
                  Number Grid Points                  17


                   Response Profile for Target: BAD

           Level         Nobs        Frequency           Weight

           1             1189             1189      1189.000000
           0             4771             4771      4771.000000


    Variable            Mean         Std Dev       Skewness       Kurtosis

    LOAN               18608           11207        2.02378        6.93259
    MORTDUE            67350           44458        1.81448        6.48187
    VALUE              99863           57386        3.05334       24.36280
    YOJ              8.15130         7.57398        0.98846        0.37207
    DELINQ           0.40570         1.12727        4.02315       23.56545
    CLAGE          170.47634        85.81009        1.34341        7.59955
    NINQ             1.08456         1.72867        2.62198        9.78651
    CLNO            20.50285        10.13893        0.77505        1.15767
    DEBTINC         26.59885         8.60175        2.85235       50.50404
```

For the first stage we select three eigenvectors corresponding to the 4th, 11th, and 2nd largest eigenvalues. There is no relationship between

- the R^2 value which measures the prediction of the response (target) variable by each eigenvector
- and the eigenvalue corresponding to each eigenvector that measures the variance explained in the $X^T X$ data matrix.

The eigenvalues are not used in the analysis of PROC DMNEURL and are printed only for curiosity.

```
             Component Selection: SS(y) and R2 (SS_total=4771)

 Comp        Eigval       R-Square        F Value   p-Value            SSE

    4   9397.769045       0.017419     105.640645    <.0001    4687.893424
   11   6327.041282       0.006317      38.550835    <.0001    4657.755732
    2         13164       0.005931      36.408247    <.0001    4629.461194
```

The optimization history indicates a maximum of 11 iterations for the activation function LOGIST:

```
     ---------- Optimization Cycle (Stage=0) ----------
-------------------- Activation= SQUARE (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=5 Crit=0.06782364: SSE=1616.91564 Acc= 81.6443
-------------------- Activation= TANH (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=4 Crit=0.06802595: SSE=1621.73865 Acc= 81.6275
-------------------- Activation= ARCTAN (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=5 Crit=0.06795346: SSE=1620.01041 Acc= 81.6611
-------------------- Activation= LOGIST (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=11 Crit=0.06802943: SSE= 1621.8217 Acc= 81.6107
-------------------- Activation= GAUSS (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
GAUSS: Iter=12 Crit=0.07716305: SSE=1839.56717 Acc= 80.2517
-------------------- Activation= SIN (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=5 Crit=0.06811774: SSE= 1623.9269 Acc= 81.6611
-------------------- Activation= COS (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
COS: Iter=9 Crit=0.07419096: SSE=1768.71252 Acc= 81.1913
-------------------- Activation= EXP (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=9 Crit=0.06798656: SSE=1620.79948 Acc= 81.5436
```

The following approximate accuracy rates for each activation function are based on the discrete values of the predictor (x) variables:

```
              Approximate Goodness-of-Fit Criteria (Stage 0)

     Run    Activation       Criterion             SSE        Accuracy

       1    SQUARE            0.067824     1616.915639       81.644295
       3    ARCTAN            0.067953     1620.010407       81.661074
       8    EXP               0.067987     1620.799480       81.543624
```

2	TANH	0.068026	1621.738647	81.627517
4	LOGIST	0.068029	1621.821701	81.610738
6	SIN	0.068118	1623.926900	81.661074
7	COS	0.074191	1768.712522	81.191275
5	GAUSS	0.077163	1839.567167	80.251678

After running through the data set we obtain the correct accuracy tables:

Classification Table for CUTOFF = 0.5000

Activation	Accuracy	Observed	Predicted 1	0
SQUARE	81.610738	1	229.0	960.0
	1610.380526	0	136.0	4635.0
ARCTAN	81.778523	1	232.0	957.0
	1611.782124	0	129.0	4642.0
EXP	81.593960	1	231.0	958.0
	1613.330656	0	139.0	4632.0
LOGIST	81.778523	1	238.0	951.0
	1614.606262	0	135.0	4636.0
TANH	81.778523	1	227.0	962.0
	1615.441761	0	124.0	4647.0
SIN	81.291946	1	224.0	965.0
	1618.630662	0	150.0	4621.0
COS	81.359060	1	101.0	1088.0
	1763.371571	0	23.0000	4748.0
GAUSS	80.201342	1	9.0000	1180.0
	1891.947600	0	0	4771.0

The activation function SQUARE seems to be most appropriate for the first stage (stage=0) of estimation based on SSE. However, the TANH, ARCTAN, and LOGIST activation functions yield an even higher accuracy rate:

Goodness-of-Fit Criteria (Ordered by SSE, Stage 0)

Run	Activation	SSE	ASE	Accuracy
1	SQUARE	1610.4	0.270198	81.610738
3	ARCTAN	1611.8	0.270433	81.778523
8	EXP	1613.3	0.270693	81.593960
4	LOGIST	1614.6	0.270907	81.778523
2	TANH	1615.4	0.271047	81.778523
6	SIN	1618.6	0.271582	81.291946
7	COS	1763.4	0.295868	81.359060
5	GAUSS	1891.9	0.317441	80.201342

The following is the start of the second stage of estimation (stage=1). It starts with selecting three eigenvectors which might predict the residuals best:

Component Selection: SS(y) and R2 (Stage=1)

Comp	Eigval	R-Square	F Value	p-Value

```
23     4763.193233        0.023292      142.109442     <.0001
21     5192.070258        0.018366      114.178467     <.0001
24     4514.317020        0.017493      110.756118     <.0001
```

When fitting the first order residuals the average value of the objective function dropped from 0.068 to 0.063. For time reasons the approximate accuracy rates are not computed after the first stage:

```
     ---------- Optimization Cycle (Stage=1) ----------
-------------------- Activation= SQUARE (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.06280042: SSE=1483.03474 Acc= 83.1376
-------------------- Activation= TANH (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=4 Crit=0.06299539: SSE=1496.24113 Acc= 83.0705
-------------------- Activation= ARCTAN (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=5 Crit=0.06300879: SSE=1496.51115 Acc= 83.0705
-------------------- Activation= LOGIST (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=7 Crit=0.06302038: SSE=1498.11974 Acc= 83.1208
-------------------- Activation= GAUSS (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
GAUSS: Iter=37 Crit=0.06583983: SSE=1566.67082 Acc= 82.2987
-------------------- Activation= SIN (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=5 Crit=0.06294376: SSE=1495.72652 Acc= 83.0201
-------------------- Activation= COS (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
COS: Iter=6 Crit=0.06575231: SSE= 1564.5619 Acc= 82.2819
-------------------- Activation= EXP (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=5 Crit=0.06318155: SSE=1493.64595 Acc= 83.1208
```

The best accuracy increased from 81.61 to 83.28. The number of correctly predicted observations, with target=1, increased from 229 to 319 (which is still less than that of TANH in the first stage):

```
                 Classification Table for CUTOFF = 0.5000

                                                      Predicted
Activation        Accuracy    Observed             1            0

SQUARE           83.288591    1                319.0        870.0
               1483.955212    0                126.0       4645.0
EXP              83.187919    1                304.0        885.0
               1494.498743    0                117.0       4654.0
SIN              83.120805    1                306.0        883.0
               1496.568504    0                123.0       4648.0
TANH             83.221477    1                309.0        880.0
```

```
                 1496.873853    0              120.0       4651.0
ARCTAN             83.255034    1              310.0        879.0
                 1497.229795    0              119.0       4652.0
LOGIST             83.171141    1              305.0        884.0
                 1498.619513    0              119.0       4652.0
COS                82.583893    1              309.0        880.0
                 1562.685434    0              158.0       4613.0
GAUSS              82.567114    1              309.0        880.0
                 1564.972115    0              159.0       4612.0
```

```
      Goodness-of-Fit Criteria (Ordered by SSE, Stage 1)

Run    Activation             SSE              ASE        Accuracy

  1    SQUARE              1484.0         0.248986       83.288591
  8    EXP                 1494.5         0.250755       83.187919
  6    SIN                 1496.6         0.251102       83.120805
  2    TANH                1496.9         0.251153       83.221477
  3    ARCTAN              1497.2         0.251213       83.255034
  4    LOGIST              1498.6         0.251446       83.171141
  7    COS                 1562.7         0.262196       82.583893
  5    GAUSS               1565.0         0.262579       82.567114
```

Here starts the third stage (stage=2):

```
         Component Selection: SS(y) and R2 (Stage=2)

Comp           Eigval        R-Square         F Value    p-Value

   1            15337        0.006514       39.068994     <.0001
   5      8117.555354        0.005566       33.564983     <.0001
   7      7371.205837        0.005429       32.918782     <.0001
```

When fitting the second order residuals the average value of the objective function (that is, the value of crit) dropped from 0.063 to 0.061:

```
     ---------- Optimization Cycle (Stage=2) ----------
-------------------- Activation= SQUARE (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.06128736: SSE=1441.99306 Acc= 83.4228
-------------------- Activation= TANH (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=5 Crit=0.06164516: SSE=1452.25582 Acc= 83.6242
-------------------- Activation= ARCTAN (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=5 Crit=0.06167716: SSE=1452.93789 Acc= 83.6074
-------------------- Activation= LOGIST (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=4 Crit=0.06162208: SSE=1452.53306 Acc= 83.6745
-------------------- Activation= GAUSS (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
GAUSS: Iter=2 Crit=0.06272854: SSE=1481.49104 Acc= 83.3054
```

```
-------------------- Activation= SIN (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=6 Crit=0.06164909: SSE= 1452.3708 Acc= 83.6074
-------------------- Activation= COS (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
COS: Iter=3 Crit=0.06277476: SSE=1482.69978 Acc= 83.2886
-------------------- Activation= EXP (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=4 Crit=0.06131052: SSE=1443.35818 Acc= 83.4396
```

The best accuracy increased from 83.29 to 83.47 and the number of correctly predicted observations, with target=1, increased from 319 to 343:

Classification Table for CUTOFF = 0.5000

Activation	Accuracy	Observed	Predicted 1	0
SQUARE	83.473154	1	343.0	846.0
	1442.237048	0	139.0	4632.0
EXP	83.422819	1	337.0	852.0
	1442.603820	0	136.0	4635.0
TANH	83.607383	1	340.0	849.0
	1451.107616	0	128.0	4643.0
ARCTAN	83.607383	1	341.0	848.0
	1451.449336	0	129.0	4642.0
SIN	83.607383	1	340.0	849.0
	1451.779559	0	128.0	4643.0
LOGIST	83.640940	1	335.0	854.0
	1452.211780	0	121.0	4650.0
GAUSS	83.372483	1	317.0	872.0
	1482.942814	0	119.0	4652.0
COS	83.322148	1	316.0	873.0
	1483.336312	0	121.0	4650.0

Even though SQUARE shows the best SSE, the accuracy rates for some other functions (for example LOGIST) are slightly better:

Goodness-of-Fit Criteria (Ordered by SSE, Stage 2)

Run	Activation	SSE	ASE	Accuracy
1	SQUARE	1442.2	0.241986	83.473154
8	EXP	1442.6	0.242048	83.422819
2	TANH	1451.1	0.243474	83.607383
3	ARCTAN	1451.4	0.243532	83.607383
6	SIN	1451.8	0.243587	83.607383
4	LOGIST	1452.2	0.243660	83.640940
5	GAUSS	1482.9	0.248816	83.372483
7	COS	1483.3	0.248882	83.322148

Component selection w.r.t. the residuals of the stage 2 starts the estimation of stage 3. Note, that the R^2 values decrease.

```
                Component Selection: SS(y) and R2 (Stage=3)

     Comp          Eigval        R-Square         F Value    p-Value

        8     6938.083228        0.005571       33.383374     <.0001
       20     5345.603436        0.004223       25.409312     <.0001
       12     6136.575271        0.004059       24.517995     <.0001
         Component Selection: SS(y) and R2 (Stage=3)
```

Note also that the size of the objective function at the optimization results decreases:

```
     ---------- Optimization Cycle (Stage=3) ----------
-------------------- Activation= SQUARE (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.06049339: SSE=1421.03255 Acc= 83.7081
-------------------- Activation= TANH (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=4 Crit=0.06052425: SSE=1420.79227 Acc= 83.7752
-------------------- Activation= ARCTAN (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=3 Crit=0.06052607: SSE=1420.97943 Acc= 83.7081
-------------------- Activation= LOGIST (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=6 Crit=0.06055936: SSE=1422.10914 Acc= 83.6577
-------------------- Activation= GAUSS (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
GAUSS: Iter=6 Crit=0.06111674: SSE=1438.83388 Acc= 83.3725
-------------------- Activation= SIN (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=3 Crit=0.06051959: SSE=1420.61742 Acc= 83.8087
-------------------- Activation= COS (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=2.
COS: Iter=6 Crit=0.06117044: SSE=1438.52462 Acc= 83.3725
-------------------- Activation= EXP (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=2 Crit=0.06051088: SSE=1421.62112 Acc= 83.7081
```

The accuracy of the best fit improves slightly from 83.47 to 83.79 and the number of correctly predicted observations, with target=1, increases from 343 to 364.

Classification Table for CUTOFF = 0.5000

Activation	Accuracy	Observed	Predicted 1	Predicted 0
SIN	83.791946	1	364.0	825.0
	1419.557263	0	141.0	4630.0
TANH	83.758389	1	363.0	826.0
	1419.778512	0	142.0	4629.0
ARCTAN	83.708054	1	361.0	828.0
	1420.072781	0	143.0	4628.0
SQUARE	83.724832	1	355.0	834.0
	1420.150396	0	136.0	4635.0

```
EXP               83.741611    1              356.0       833.0
                1420.424317    0              136.0      4635.0
LOGIST            83.691275    1              357.0       832.0
                1421.645293    0              140.0      4631.0
COS               83.355705    1              340.0       849.0
                1437.890025    0              143.0      4628.0
GAUSS             83.288591    1              328.0       861.0
                1438.539930    0              135.0      4636.0
```

```
      Goodness-of-Fit Criteria (Ordered by SSE, Stage 3)

Run    Activation           SSE            ASE         Accuracy

  6    SIN               1419.6       0.238181        83.791946
  2    TANH              1419.8       0.238218        83.758389
  3    ARCTAN            1420.1       0.238267        83.708054
  1    SQUARE            1420.2       0.238280        83.724832
  8    EXP               1420.4       0.238326        83.741611
  4    LOGIST            1421.6       0.238531        83.691275
  7    COS               1437.9       0.241257        83.355705
  5    GAUSS             1438.5       0.241366        83.288591
```

The residuals are computed and components are selected for the last estimation stage:

```
         Component Selection: SS(y) and R2 (Stage=4)

Comp          Eigval         R-Square          F Value    p-Value

  28     1195.710958         0.003997        23.916548     <.0001
  27     3456.490592         0.001822        10.919693     0.0010
  25     3935.018952         0.001803        10.824185     0.0010
```

The optimization processes complete without error:

```
     ---------- Optimization Cycle (Stage=4) ----------
-------------------- Activation= SQUARE (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.05983921: SSE=1407.33854 Acc= 83.6913
-------------------- Activation= TANH (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=5 Crit=0.06015823: SSE=1412.95394 Acc= 83.6074
-------------------- Activation= ARCTAN (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=3 Crit=0.06013359: SSE=1412.42466 Acc= 83.7081
-------------------- Activation= LOGIST (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=3 Crit=0.06017552: SSE=1413.70283 Acc= 83.7919
-------------------- Activation= GAUSS (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
GAUSS: Iter=4 Crit=0.06032127: SSE=1417.14339 Acc= 83.8255
-------------------- Activation= SIN (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
```

```
NOTE: Singular Hessian at result, nullity=1.
SIN: Iter=3 Crit=0.06014411: SSE=1412.80581 Acc= 83.6745
-------------------- Activation= COS (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
COS: Iter=4 Crit=0.06007575: SSE=1414.03361 Acc= 83.8087
-------------------- Activation= EXP (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=3 Crit=0.05983526: SSE=1407.86753 Acc= 83.6074
```

The accuracy drops from 83.79 to 83.72, and also the number of correctly predicted observations, with target=1, decreased from 365 to 363. This can happen only when the discretization error becomes too large in relation to the goodness of fit of the nonlinear model. Perhaps the specification of larger values for MAXCOMP= and NPOINT= could improve the solution, but in most applications we would see this behavior as a sign that no further improvement of the model fit is possible.

Classification Table for CUTOFF = 0.5000

			Predicted	
Activation	Accuracy	Observed	1	0
SQUARE	83.724832	1	363.0	826.0
	1405.799588	0	144.0	4627.0
EXP	83.691275	1	361.0	828.0
	1406.591128	0	144.0	4627.0
ARCTAN	83.775168	1	364.0	825.0
	1410.486170	0	142.0	4629.0
SIN	83.691275	1	363.0	826.0
	1411.016319	0	146.0	4625.0
TANH	83.708054	1	362.0	827.0
	1411.269012	0	144.0	4627.0
LOGIST	83.708054	1	360.0	829.0
	1411.465189	0	142.0	4629.0
COS	83.842282	1	364.0	825.0
	1414.584867	0	138.0	4633.0
GAUSS	83.791946	1	362.0	827.0
	1417.319507	0	139.0	4632.0

Even though accuracy did not improve, the SSE value still dropped from 1419.6 to 1405.8 during the last stage.

Goodness-of-Fit Criteria (Ordered by SSE, Stage 4)

Run	Activation	SSE	ASE	Accuracy
1	SQUARE	1405.8	0.235872	83.724832
8	EXP	1406.6	0.236005	83.691275
3	ARCTAN	1410.5	0.236659	83.775168
6	SIN	1411.0	0.236748	83.691275
2	TANH	1411.3	0.236790	83.708054
4	LOGIST	1411.5	0.236823	83.708054
7	COS	1414.6	0.237346	83.842282
5	GAUSS	1417.3	0.237805	83.791946

The following summary table shows the improvements in SSE and Accuracy rates across the five stages:

```
                    Summary Table Across Stages

Stage Activation Link          SSE        RMSE    Accuracy         AIC

    0 SQUARE     LOGIST     1610.4    0.520111   81.610738 -7785.255870
    1 SQUARE     IDENT      1484.0    0.499572   83.288591 -8258.542723
    2 SQUARE     IDENT      1442.2    0.492790   83.473154 -8414.495404
    3 SIN        IDENT      1419.6    0.489188   83.791946 -8494.963649
    4 SQUARE     IDENT      1405.8    0.487099   83.724832 -8539.006860

                   Summary Table Across Stages

                         Stage         SBC

                             0 -7738.406090
                             1 -8164.843163
                             2 -8273.946063
                             3 -8307.564528
                             4 -8304.757959
```

All 40 optimizations were very efficient with about 5 iterations per optimization and less than 10 function calls per optimization:

```
*** Total Number of Runs through Data :       27
*** Total Number of NL Optimizations  :       40
*** Total Number of Iterations in NLP :      217
*** Total Number Function Calls in NLP:      431
```

In this application those solutions were selected which had the smallest Sum-of-Squares Error. By specifying the *selcrit=acc* option we can instead select the solutions with the largest accuracy rate:

```
proc dmneurl data=sampsio.hmeq dmdbcat=outcat
    outclass=oclass outest=estout out=dsout outfit=ofit
    ptable maxcomp=3 maxstage=5 selcrit=acc;
    var LOAN MORTDUE VALUE REASON JOB YOJ DEROG DELINQ
          CLAGE NINQ CLNO DEBTINC;
    target BAD;
run;
```

The following output only shows the summary table. For this example, the total accuracy was slightly increased in all stages except the second, but this behavior might not occur for other examples.

```
                    Summary Table Across Stages

Stage Activation Link          SSE        RMSE    Accuracy         AIC

    0 ARCTAN     LOGIST     1611.8    0.520338   81.778523 -7780.070830
    1 SQUARE     IDENT      1493.2    0.501133   83.271812 -8221.350308
    2 LOGIST     IDENT      1461.5    0.496062   83.859060 -8335.596710
```

```
    3 EXP        IDENT        1428.2     0.490668    84.026846 -8458.969424
    4 COS        IDENT        1423.2     0.490098    83.976510 -8465.850656

                    Summary Table Across Stages

                        Stage          SBC

                            0 -7733.221050
                            1 -8127.650748
                            2 -8195.047369
                            3 -8271.570302
                            4 -8231.601754
```

Example 2: Application: HMEQ Data Set: Interval Target LOAN

Now we show the specification and results of PROC DMNEURL for the interval target LOAN. First we have to obtain the DMDB data set and catalog from the raw data set:

```
libname sampsio '/sas/a612/dmine/sampsio';
proc dmdb batch data=sampsio.hmeq out=dmdbout dmdbcat=outcat;
    var LOAN MORTDUE VALUE YOJ DELINQ CLAGE NINQ CLNO DEBTINC;
    class BAD(ASC) REASON(ASC) JOB(ASC) DEROG(ASC);
    target LOAN;
run;
```

The PROC DMNEURL call is very similar, but here five stages each with three components and seven parameters are specified:

```
proc dmneurl data=sampsio.hmeq dmdbcat=outcat
    outclass=oclass outest=estout out=dsout outfit=ofit
    ptable maxcomp=3 maxstage=6;
    var BAD MORTDUE VALUE REASON JOB YOJ DEROG DELINQ
          CLAGE NINQ CLNO DEBTINC;
    target LOAN;
run;
```

The link function for interval target is by default specified as the identity:

```
                    The DMNEURL Procedure

          Interval Target                       LOAN
          Number Observations                   5960
          NOBS w/o Missing Target               5960
          Target Range                   [ 1100, 89900]
          Link Function                        IDENT
          Selection Criterion                    SSE
          Optimization Criterion                 SSE
          Estimation Stages                        6
          Max. Number Components                   3
          Minimum R2 Value                  0.000050
```

```
Number Grid Points                    17
```

Variable	Mean	Std Dev	Skewness	Kurtosis
LOAN	18608	11207	2.02378	6.93259
MORTDUE	67350	44458	1.81448	6.48187
VALUE	99863	57386	3.05334	24.36280
YOJ	8.15130	7.57398	0.98846	0.37207
DELINQ	0.40570	1.12727	4.02315	23.56545
CLAGE	170.47634	85.81009	1.34341	7.59955
NINQ	1.08456	1.72867	2.62198	9.78651
CLNO	20.50285	10.13893	0.77505	1.15767
DEBTINC	26.59885	8.60175	2.85235	50.50404

For an interval target the percentiles of the response (target) variable are computed as an aside of the preliminary runs through the data.

Percentiles of Target LOAN in [1100 : 89900]

	Nobs	Y Value	Label
1	596	7600.000000	0.073198198
2	1192	10000	0.100225225
3	1788	12100	0.123873874
4	2384	14400	0.149774775
5	2980	16300	0.171171171
6	3576	18800	0.199324324
7	4172	21700	0.231981982
8	4768	25000	0.269144144
9	5364	30500	0.331081081
10	5960	89900	1

The first estimation stage starts with the selection of the best predictor components (eigenvectors):

Component Selection: SS(y) and R2 (SS_total=326.60303927)

Comp	Eigval	R-Square	F Value	p-Value	SSE
2	14414	0.015964	96.672480	<.0001	321.389163
28	1232.230727	0.005739	34.949673	<.0001	319.514886
11	6335.576701	0.005490	33.620923	<.0001	317.721686

Component Selection: SS(y) and R2 (SS_total=326.60303927)

A maximum of eight iterations is needed for convergence:

```
     ---------- Optimization Cycle (Stage=0) ----------
-------------------- Activation= SQUARE (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.00719589: SSE=6.76374E11 Acc= 32.7484
-------------------- Activation= TANH (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=7 Crit=0.00729363: SSE=6.85561E11 Acc= 29.2423
```

```
-------------------- Activation= ARCTAN (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=4 Crit=0.00730523: SSE=6.86651E11 Acc= 29.2427
-------------------- Activation= LOGIST (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=5 Crit=0.00730061: SSE=6.86217E11 Acc= 29.0029
-------------------- Activation= GAUSS (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
GAUSS: Iter=8 Crit=0.00753383: SSE=7.08138E11 Acc= 15.3436
-------------------- Activation= SIN (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=2 Crit=0.00732518: SSE=6.88526E11 Acc= 29.0399
-------------------- Activation= COS (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
COS: Iter=5 Crit=0.00753876: SSE=7.08602E11 Acc= 22.5534
-------------------- Activation= EXP (Stage=0) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=2 Crit=0.00724534: SSE=6.81022E11 Acc= 29.9011
```

For interval target y the accuracy is computed as the Goodman-Kruskal γ coefficient for an observed-predicted frequency table using the percentiles of y for row and column definitions. (Note, that the Goodman-Kruskal γ can have negative values for an extremely bad fit.)

Approximate Goodness-of-Fit Criteria (Stage 0)

Run	Activation	Criterion	SSE	Accuracy
1	SQUARE	0.007196	676373814905	32.748384
8	EXP	0.007245	681021814295	29.901149
2	TANH	0.007294	685560807525	29.242251
4	LOGIST	0.007301	686216550016	29.002936
3	ARCTAN	0.007305	686651267193	29.242724
6	SIN	0.007325	688526431720	29.039929
5	GAUSS	0.007534	708138467963	15.343567
7	COS	0.007539	708601735523	22.553358

The Sum Of Squares Error SSE for the first stage is 6.68237E11:

Goodness-of-Fit Criteria (Ordered by SSE, Stage 0)

Run	Activation	SSE	ASE	Accuracy
1	SQUARE	6.68237E11	112120251	33.925841
4	LOGIST	6.73986E11	113084875	30.721154
2	TANH	6.74156E11	113113490	30.776431
8	EXP	6.77111E11	113609246	30.503925
3	ARCTAN	6.78503E11	113842802	29.832932
6	SIN	6.81758E11	114388913	29.402115
5	GAUSS	7.0371E11	118072079	20.824223
7	COS	7.04934E11	118277529	20.022492

The second stage (stage=1) starts with selecting the best principal components for predicting the residual:

```
            Component Selection: SS(y) and R2 (Stage=1)

Comp        Eigval       R-Square        F Value    p-Value

   1         16197       0.023135     141.126005     <.0001
   3         12170       0.017130     106.340108     <.0001
  20   5623.081574       0.012121      76.193667     <.0001
```

Now a maximum of five iterations is needed for convergence:

```
     ---------- Optimization Cycle (Stage=1) ----------
-------------------- Activation= SQUARE (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.00675824: SSE=6.35237E11 Acc= 39.9782
-------------------- Activation= TANH (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=4 Crit=0.00677155: SSE=6.36489E11 Acc= 40.1296
-------------------- Activation= ARCTAN (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=4 Crit=0.00675928: SSE=6.35335E11 Acc= 41.0832
-------------------- Activation= LOGIST (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=2 Crit=0.00677728: SSE=6.37026E11 Acc= 40.5476
-------------------- Activation= GAUSS (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
GAUSS: Iter=5 Crit=0.00701114: SSE=6.59009E11 Acc= 37.5715
-------------------- Activation= SIN (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=4 Crit=0.00676857: SSE=6.36208E11 Acc= 40.1257
-------------------- Activation= COS (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
COS: Iter=5 Crit=0.00699351: SSE=6.57351E11 Acc= 36.1925
-------------------- Activation= EXP (Stage=1) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=1 Crit=0.00676817: SSE= 6.3617E11 Acc= 40.3710
```

The SSE dropped from 6.68237E11 to 6.31521E11:

```
         Goodness-of-Fit Criteria (Ordered by SSE, Stage 1)

Run    Activation          SSE            ASE        Accuracy

  1    SQUARE       6.31521E11      105959851       41.275893
  3    ARCTAN       6.32669E11      106152521       40.218692
  8    EXP          6.32795E11      106173614       40.947613
  6    SIN          6.32908E11      106192666       40.499603
  2    TANH          6.3331E11      106260141       40.464591
  4    LOGIST       6.33926E11      106363485       41.328405
  7    COS          6.56699E11      110184474       36.381117
```

```
   5    GAUSS         6.58666E11       110514467       37.785522
Goodness-of-Fit Criteria (Ordered by SSE, Stage 1)
```

The third stage starts with selecting the best eigenvectors for prediction of the residuals of the last stage:

```
                 Component Selection: SS(y) and R2 (Stage=2)

         Comp          Eigval        R-Square         F Value    p-Value

           23     4811.081772        0.011805       71.186100     <.0001
           28     1232.230727        0.007479       45.434565     <.0001
           18     5865.674624        0.006724       41.123895     <.0001
```

Now, the maximum of iterations is four!

```
     ---------- Optimization Cycle (Stage=2) ----------
-------------------- Activation= SQUARE (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.00652258: SSE=6.13086E11 Acc= 42.4729
-------------------- Activation= TANH (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=3 Crit=0.00650972: SSE=6.11878E11 Acc= 41.5654
-------------------- Activation= ARCTAN (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=3 Crit=0.00650403: SSE=6.11342E11 Acc= 42.2974
-------------------- Activation= LOGIST (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=4 Crit=0.00641908: SSE=6.03358E11 Acc= 43.3779
-------------------- Activation= GAUSS (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=2.
GAUSS: Iter=2 Crit=0.00671802: SSE=6.31456E11 Acc= 41.2660
-------------------- Activation= SIN (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=4 Crit=0.00651654: SSE=6.12519E11 Acc= 41.9809
-------------------- Activation= COS (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=2.
COS: Iter=4 Crit=0.00671738: SSE=6.31396E11 Acc= 41.2783
-------------------- Activation= EXP (Stage=2) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=0 Crit=0.00656615: SSE=6.17182E11 Acc= 41.6251
```

The RMSE dropped from 6.31521E11 to 6.00171E11:

```
             Goodness-of-Fit Criteria (Ordered by SSE, Stage 2)

         Run    Activation            SSE             ASE        Accuracy

           4    LOGIST        6.00171E11       100699874       44.076324
           3    ARCTAN        6.11611E11       102619248       42.722853
           2    TANH          6.12574E11       102780849       42.454925
           6    SIN           6.12902E11       102835885       42.618536
```

```
  1    SQUARE       6.14545E11      103111590       43.452922
  8    EXP          6.17964E11      103685248       42.588823
  7    COS          6.31415E11      105942050       41.153349
  5    GAUSS        6.31533E11      105961918       41.028769
```

In stage 3, components are selected w.r.t. the residuals from stage 2:

```
                Component Selection: SS(y) and R2 (Stage=3)

       Comp          Eigval        R-Square         F Value    p-Value

          5     8108.233368        0.008115       48.751302     <.0001
          7     7678.598513        0.004569       27.574638     <.0001
         27     3496.302840        0.003929       23.802006     <.0001
```

```
     ---------- Optimization Cycle (Stage=3) ----------
-------------------- Activation= SQUARE (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.00627437: SSE=5.89756E11 Acc= 46.5664
-------------------- Activation= TANH (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=2 Crit=0.00628152: SSE=5.90428E11 Acc= 46.2812
-------------------- Activation= ARCTAN (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=2 Crit=0.00628195: SSE=5.90469E11 Acc= 46.2125
-------------------- Activation= LOGIST (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=5 Crit=0.00628127: SSE=5.90404E11 Acc= 45.8040
-------------------- Activation= GAUSS (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
GAUSS: Iter=2 Crit=0.00637032: SSE=5.98775E11 Acc= 45.3250
-------------------- Activation= SIN (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=6 Crit=0.00627884: SSE=5.90176E11 Acc= 46.0972
-------------------- Activation= COS (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=3.
COS: Iter=5 Crit=0.00637738: SSE=5.99438E11 Acc= 44.6439
-------------------- Activation= EXP (Stage=3) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=1 Crit=0.00628498: SSE=5.90753E11 Acc= 46.4650
```

The SSE dropped from 6.00171E11 to 5.88794E11:

```
             Goodness-of-Fit Criteria (Ordered by SSE, Stage 3)

       Run   Activation          SSE              ASE          Accuracy

         1   SQUARE       5.88794E11         98790914         46.874168
         6   SIN          5.89308E11         98877106         46.045231
```

4	LOGIST	5.89532E11	98914767	45.936035
2	TANH	5.89658E11	98935963	46.521683
3	ARCTAN	5.89714E11	98945236	46.462470
8	EXP	5.89994E11	98992341	46.590155
5	GAUSS	5.98777E11	100465890	45.338954
7	COS	5.99432E11	100575773	44.563112

Again, the new stage 4 starts with component selection w.r.t. the residuals of the last stage3:

Component Selection: SS(y) and R2 (Stage=4)

Comp	Eigval	R-Square	F Value	p-Value
14	5977.581155	0.004044	24.196310	<.0001
24	4589.938565	0.002803	16.817296	<.0001
8	7098.575517	0.002425	14.583721	0.0001

A maximum of seven iterations is needed for convergence:

```
     ---------- Optimization Cycle (Stage=4) ----------
-------------------- Activation= SQUARE (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SQUARE: Iter=1 Crit=0.00618628: SSE=5.81476E11 Acc= 46.9112
-------------------- Activation= TANH (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
TANH: Iter=5 Crit= 0.0061812: SSE=5.80998E11 Acc= 45.5545
-------------------- Activation= ARCTAN (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
ARCTAN: Iter=3 Crit=0.00618984: SSE= 5.8181E11 Acc= 45.7876
-------------------- Activation= LOGIST (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
LOGIST: Iter=2 Crit=0.00618964: SSE=5.81791E11 Acc= 46.2233
-------------------- Activation= GAUSS (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
GAUSS: Iter=3 Crit=0.00617962: SSE= 5.8085E11 Acc= 46.3454
-------------------- Activation= SIN (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
SIN: Iter=7 Crit=0.00614634: SSE=5.77722E11 Acc= 45.3470
-------------------- Activation= COS (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
NOTE: Singular Hessian at result, nullity=1.
COS: Iter=2 Crit=0.00619158: SSE=5.81974E11 Acc= 47.5278
-------------------- Activation= EXP (Stage=4) --------------------
NOTE: ABSGCONV convergence criterion satisfied.
EXP: Iter=2 Crit=0.00620074: SSE=5.82835E11 Acc= 47.1760
```

The RMSE dropped from 5.88794E11 to 5.82844E11:

Goodness-of-Fit Criteria (Ordered by SSE, Stage 4)

Run	Activation	SSE	ASE	Accuracy

5	GAUSS	5.82844E11	97792594	46.529424
1	SQUARE	5.82906E11	97803038	47.202054
7	COS	5.83336E11	97875104	47.340247
8	EXP	5.83553E11	97911654	47.907199
4	LOGIST	5.84275E11	98032716	47.245956
2	TANH	5.8489E11	98135835	46.238301
3	ARCTAN	5.88716E11	98777913	46.013181
6	SIN	6.12035E11	102690480	45.185468

For space reasons we are skipping the results of stage 5 except the following table which shows that the RMSE dropped again.

Goodness-of-Fit Criteria (Ordered by SSE, Stage 5)

Run	Activation	SSE	ASE	Accuracy
4	LOGIST	5.77698E11	96929208	47.464184
1	SQUARE	5.78114E11	96998924	47.337721
8	EXP	5.78394E11	97045985	47.342324
3	ARCTAN	5.79057E11	97157141	46.609720
2	TANH	5.8133E11	97538668	46.144529
5	GAUSS	5.82144E11	97675226	46.792103
7	COS	5.82405E11	97718917	46.540469
6	SIN	6.10096E11	102365171	45.303003

This is a summary table for the first six estimation stages:

Summary Table Across Stages

Stage	Activation	Link	SSE	RMSE	Accuracy	AIC
0	SQUARE	IDENT	6.68237E11	10595	33.925841	110483
1	SQUARE	IDENT	6.31521E11	10306	41.275893	110160
2	LOGIST	IDENT	6.00171E11	10053	44.076324	109871
3	SQUARE	IDENT	5.88794E11	9962.791914	46.874168	109771
4	GAUSS	IDENT	5.82844E11	9918.178853	46.529424	109724
5	LOGIST	IDENT	5.77698E11	9880.137343	47.464184	109685

Summary Table Across Stages

Stage	SBC
0	110530
1	110254
2	110011
3	109958
4	109958
5	109966

The six stages took 48 objective function optimizations (each with 7 parameters) and 33 runs through the data. On average, less than four iterations and about seven function calls are needed for each optimization:

```
*** Total Number of Runs through Data :       33
*** Total Number of NL Optimizations  :       48
*** Total Number of Iterations in NLP :      157
*** Total Number Function Calls in NLP:      395
```