NAG Library Routine Document

G02KAF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

1 Purpose

G02KAF calculates a ridge regression, optimizing the ridge parameter according to one of four prediction error criteria.

2 Specification

```
SUBROUTINE G02KAF (N, M, X, LDX, ISX, IP, TAU, Y, H, OPT, NITER, TOL, NEP,      &
                   ORIG, B, VIF, RES, RSS, DF, OPTLOO, PERR, IFAIL)
INTEGER            N, M, LDX, ISX(M), IP, OPT, NITER, ORIG, DF, OPTLOO,         &
                   IFAIL
REAL (KIND=nag_wp) X(LDX,M), TAU, Y(N), H, TOL, NEP, B(IP+1), VIF(IP),          &
                   RES(N), RSS, PERR(5)
```

3 Description

A linear model has the form:

$$y = c + X\beta + \epsilon,$$

where

y is an n by 1 matrix of values of a dependent variable;

c is a scalar intercept term;

X is an n by m matrix of values of independent variables;

β is an m by 1 matrix of unknown values of parameters;

ϵ is an n by 1 matrix of unknown random errors such that variance of $\epsilon = \sigma^2 I$.

Let $\tilde{X}$ be the mean-centred X and $\tilde{y}$ the mean-centred y. Furthermore, $\tilde{X}$ is scaled such that the diagonal elements of the cross product matrix $\tilde{X}^{\mathrm{T}}\tilde{X}$ are one. The linear model now takes the form:

$$\tilde{y} = \tilde{X}\tilde{\beta} + \epsilon.$$

Ridge regression estimates the parameters $\tilde{\beta}$ in a penalised least squares sense by finding the $\tilde{b}$ that minimizes

$$\left\|\tilde{X}\tilde{b} - \tilde{y}\right\|^2 + h\left\|\tilde{b}\right\|^2, \quad h > 0,$$

where $\|\cdot\|$ denotes the ℓ_2-norm and h is a scalar regularization or ridge parameter. For a given value of h, the parameter estimates $\tilde{b}$ are found by evaluating

$$\tilde{b} = \left(\tilde{X}^{\mathrm{T}}\tilde{X} + hI\right)^{-1}\tilde{X}^{\mathrm{T}}\tilde{y}.$$

Note that if $h = 0$ the ridge regression solution is equivalent to the ordinary least squares solution.

Rather than calculate the inverse of $(\tilde{X}^{\mathrm{T}}\tilde{X} + hI)$ directly, G02KAF uses the singular value decomposition (SVD) of $\tilde{X}$. After decomposing $\tilde{X}$ into UDV^{T} where U and V are orthogonal matrices and D is a diagonal matrix, the parameter estimates become

$$\tilde{b} = V\left(D^{\mathrm{T}}D + hI\right)^{-1}DU^{\mathrm{T}}\tilde{y}.$$

A consequence of introducing the ridge parameter is that the effective number of parameters, γ, in the model is given by the sum of diagonal elements of

$$D^{\mathrm{T}}D\left(D^{\mathrm{T}}D+hI\right)^{-1},$$

see Moody (1992) for details.

Any multi-collinearity in the design matrix X may be highlighted by calculating the variance inflation factors for the fitted model. The jth variance inflation factor, v_j, is a scaled version of the multiple correlation coefficient between independent variable j and the other independent variables, R_j, and is given by

$$v_j=\frac{1}{1-R_j},\quad j=1,2,\ldots,m.$$

The m variance inflation factors are calculated as the diagonal elements of the matrix:

$$\left(\tilde{X}^{\mathrm{T}}\tilde{X}+hI\right)^{-1}\tilde{X}^{\mathrm{T}}\tilde{X}\left(\tilde{X}^{\mathrm{T}}\tilde{X}+hI\right)^{-1},$$

which, using the SVD of $\tilde{X}$, is equivalent to the diagonal elements of the matrix:

$$V\left(D^{\mathrm{T}}D+hI\right)^{-1}D^{\mathrm{T}}D\left(D^{\mathrm{T}}D+hI\right)^{-1}V^{\mathrm{T}}.$$

Although parameter estimates $\tilde{b}$ are calculated by using $\tilde{X}$, it is usual to report the parameter estimates b associated with X. These are calculated from $\tilde{b}$, and the means and scalings of X. Optionally, either $\tilde{b}$ or b may be calculated.

The method can adopt one of four criteria to minimize while calculating a suitable value for h:

(a) Generalized cross-validation (GCV):

$$\frac{ns}{(n-\gamma)^2};$$

(b) Unbiased estimate of variance (UEV):

$$\frac{s}{n-\gamma};$$

(c) Future prediction error (FPE):

$$\frac{1}{n}\left(s+\frac{2\gamma s}{n-\gamma}\right);$$

(d) Bayesian information criterion (BIC):

$$\frac{1}{n}\left(s+\frac{\log(n)\gamma s}{n-\gamma}\right);$$

where s is the sum of squares of residuals. However, the function returns all four of the above prediction errors regardless of the one selected to minimize the ridge parameter, h. Furthermore, the function will optionally return the leave-one-out cross-validation (LOOCV) prediction error.

4 References

Hastie T, Tibshirani R and Friedman J (2003) *The Elements of Statistical Learning: Data Mining, Inference and Prediction* Springer Series in Statistics

Moody J.E. (1992) The effective number of parameters: An analysis of generalisation and regularisation in nonlinear learning systems *In: Neural Information Processing Systems* (eds J E Moody, S J Hanson, and R P Lippmann) 4 847–854 Morgan Kaufmann San Mateo CA

5 Parameters

1: N – INTEGER *Input*

On entry: n, the number of observations.

Constraint: $\mathrm{N} > 1$.

2: M – INTEGER *Input*

On entry: the number of independent variables available in the data matrix X.

Constraint: $\mathrm{M} \leq \mathrm{N}$.

3: X(LDX,M) – REAL (KIND=nag_wp) array *Input*

On entry: the values of independent variables in the data matrix X.

4: LDX – INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G02KAF is called.

Constraint: $\mathrm{LDX} \geq \mathrm{N}$.

5: ISX(M) – INTEGER array *Input*

On entry: indicates which m independent variables are included in the model.

$\mathrm{ISX}(j) = 1$
The jth variable in X will be included in the model.

$\mathrm{ISX}(j) = 0$
Variable j is excluded.

Constraint: $\mathrm{ISX}(j) = 0$ or 1, for $j = 1, 2, \ldots, \mathrm{M}$.

6: IP – INTEGER *Input*

On entry: m, the number of independent variables in the model.

Constraints:

$1 \leq \mathrm{IP} \leq \mathrm{M}$;
Exactly IP elements of ISX must be equal to 1.

7: TAU – REAL (KIND=nag_wp) *Input*

On entry: singular values less than TAU of the SVD of the data matrix X will be set equal to zero.

Suggested value: $\mathrm{TAU} = 0.0$

Constraint: $\mathrm{TAU} \geq 0.0$.

8: Y(N) – REAL (KIND=nag_wp) array *Input*

On entry: the n values of the dependent variable y.

9: H – REAL (KIND=nag_wp) *Input/Output*

On entry: an initial value for the ridge regression parameter h; used as a starting point for the optimization.

Constraint: $\mathrm{H} > 0.0$.

On exit: H is the optimized value of the ridge regression parameter h.

10: OPT – INTEGER *Input*

On entry: the measure of prediction error used to optimize the ridge regression parameter h. The value of OPT must be set equal to one of:

OPT $= 1$
Generalized cross-validation (GCV);

OPT $= 2$
Unbiased estimate of variance (UEV)

OPT $= 3$
Future prediction error (FPE)

OPT $= 4$
Bayesian information criteron (BIC).

Constraint: OPT $= 1$, 2, 3 or 4.

11: NITER – INTEGER *Input/Output*

On entry: the maximum number of iterations allowed to optimize the ridge regression parameter h.

Constraint: NITER ≥ 1.

On exit: the number of iterations used to optimize the ridge regression parameter h within TOL.

12: TOL – REAL (KIND=nag_wp) *Input*

On entry: iterations of the ridge regression parameter h will halt when consecutive values of h lie within TOL.

Constraint: TOL > 0.0.

13: NEP – REAL (KIND=nag_wp) *Output*

On exit: the number of effective parameters, γ, in the model.

14: ORIG – INTEGER *Input*

On entry: if ORIG $= 1$, the parameter estimates b are calculated for the original data; otherwise ORIG $= 2$ and the parameter estimates $\tilde{b}$ are calculated for the standardized data.

Constraint: ORIG $= 1$ or 2.

15: B$(\mathrm{IP}+1)$ – REAL (KIND=nag_wp) array *Output*

On exit: contains the intercept and parameter estimates for the fitted ridge regression model in the order indicated by ISX. The first element of B contains the estimate for the intercept; B$(j+1)$ contains the parameter estimate for the jth independent variable in the model, for $j = 1, 2, \ldots, \mathrm{IP}$.

16: VIF(IP) – REAL (KIND=nag_wp) array *Output*

On exit: the variance inflation factors in the order indicated by ISX. For the jth independent variable in the model, VIF(j) is the value of v_j, for $j = 1, 2, \ldots, \mathrm{IP}$.

17: RES(N) – REAL (KIND=nag_wp) array *Output*

On exit: RES(i) is the value of the ith residual for the fitted ridge regression model, for $i = 1, 2, \ldots, \mathrm{N}$.

18: RSS – REAL (KIND=nag_wp) *Output*

On exit: the sum of squares of residual values.

19: DF – INTEGER *Output*

On exit: the degrees of freedom for the residual sum of squares RSS.

20: OPTLOO – INTEGER *Input*

On entry: if OPTLOO $= 2$, the leave-one-out cross-validation estimate of prediction error is calculated; otherwise no such estimate is calculated and OPTLOO $= 1$.

Constraint: OPTLOO $= 1$ or 2.

21: PERR(5) – REAL (KIND=nag_wp) array *Output*

On exit: the first four elements contain, in this order, the measures of prediction error: GCV, UEV, FPE and BIC.

If OPTLOO $= 2$, PERR(5) is the LOOCV estimate of prediction error; otherwise PERR(5) is not referenced.

22: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction for details.

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, if you are not familiar with this parameter, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

On exit: IFAIL $= 0$ unless the routine detects an error or a warning has been flagged (see Section 6).

6 Error Indicators and Warnings

If on entry IFAIL $= 0$ or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

IFAIL $= -1$

Maximum number of iterations used.

IFAIL $= 1$

On entry, N ≤ 1;
or TAU < 0.0;
or OPT $\neq 1$, 2, 3 or 4;
or H ≤ 0.0;
or OPTLOO $\neq 1$ or 2;
or TOL ≤ 0.0;
or NITER < 1;
or ORIG $\neq 1$ or 2

IFAIL $= 2$

On entry, M $>$ N;
or LDX $<$ N;
or IP < 1 or IP $>$ M;
or An element of ISX $\neq 0$ or 1;
or IP does not equal the sum of elements in ISX.

IFAIL $= 3$

SVD failed to converge.

IFAIL = 4

Internal error. Check all array sizes and calls to G02KAF. Please contact NAG.

7 Accuracy

Not applicable.

8 Further Comments

G02KAF allocates internally $\max(5 \times (N-1), 2 \times IP \times IP) + (N+3) \times IP + N$ elements of double precision storage.

9 Example

This example reads in data from an experiment to model body fat, and a ridge regression is calculated that optimizes GCV prediction error.

9.1 Program Text

```
    Program g02kafe

!     G02KAF Example Program Text

!     Mark 24 Release. NAG Copyright 2012.

!     .. Use Statements ..
      Use nag_library, Only: g02kaf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: h, nep, rss, tau, tol
      Integer                          :: df, i, ifail, ip, ldx, m, n, niter,  &
                                          opt, optloo, orig
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: b(:), res(:), vif(:), x(:,:), y(:)
      Real (Kind=nag_wp)               :: perr(5)
      Integer, Allocatable             :: isx(:)
!     .. Intrinsic Procedures ..
      Intrinsic                        :: count
!     .. Executable Statements ..
      Write (nout,*) 'G02KAF Example Program Results'
      Write (nout,*)

!     Skip heading in data file
      Read (nin,*)

!     Read in the problem size
      Read (nin,*) n, m, h, opt, tol, niter, orig, optloo, tau

      ldx = n
      Allocate (x(ldx,m),y(n),isx(m))

!     Read in data
      Read (nin,*)(x(i,1:m),y(i),i=1,n)

!     Read in variable inclusion flags
      Read (nin,*) isx(1:m)

!     Calculate IP
      ip = count(isx(1:m)==1)

      Allocate (b(ip+1),vif(ip),res(n))
```

```
!     Fit ridge regression model
      ifail = -1
      Call g02kaf(n,m,x,ldx,isx,ip,tau,y,h,opt,niter,tol,nep,orig,b,vif,res, &
        rss,df,optloo,perr,ifail)
      If (ifail/=0) Then
        If (ifail/=-1) Then
          Go To 100
        End If
      End If

!     Display results
      Write (nout,99999) 'Value of ridge parameter:', h
      Write (nout,*)
      Write (nout,99998) 'Sum of squares of residuals:', rss
      Write (nout,99997) 'Degrees of freedom: ', df
      Write (nout,99999) 'Number of effective parameters:', nep
      Write (nout,*)
      Write (nout,*) 'Parameter estimates'
      Write (nout,99995)(i,b(i),i=1,ip+1)
      Write (nout,*)
      Write (nout,99996) 'Number of iterations:', niter
      Write (nout,*)
      If (opt==1) Then
        Write (nout,*) 'Ridge parameter minimises GCV'
      Else If (opt==2) Then
        Write (nout,*) 'Ridge parameter minimises UEV'
      Else If (opt==3) Then
        Write (nout,*) 'Ridge parameter minimises FPE'
      Else If (opt==4) Then
        Write (nout,*) 'Ridge parameter minimises BIC'
      End If
      Write (nout,*)
      Write (nout,*) 'Estimated prediction errors:'
      Write (nout,99999) 'GCV    =', perr(1)
      Write (nout,99999) 'UEV    =', perr(2)
      Write (nout,99999) 'FPE    =', perr(3)
      Write (nout,99999) 'BIC    =', perr(4)
      If (optloo==2) Then
        Write (nout,99999) 'LOO CV =', perr(5)
      End If
      Write (nout,*)
      Write (nout,*) 'Residuals'
      Write (nout,99995)(i,res(i),i=1,n)
      Write (nout,*)
      Write (nout,*) 'Variance inflation factors'
      Write (nout,99995)(i,vif(i),i=1,ip)

100   Continue

99999 Format (1X,A,1X,F10.4)
99998 Format (1X,A,E11.4)
99997 Format (1X,A,1X,I5)
99996 Format (1X,A,I16)
99995 Format (1X,I4,1X,F11.4)
    End Program g02kafe
```

9.2 Program Data

```
G02KAF Example Program Data
20 3 0.5 1 1.0e-4 25 2 2 0.0 : N, M, H, OPT, TOL, NITER, ORIG, OPTLOO, TAU
19.5 43.1 29.1 11.9
24.7 49.8 28.2 22.8
30.7 51.9 37.0 18.7
29.8 54.3 31.1 20.1
19.1 42.2 30.9 12.9
25.6 53.9 23.7 21.7
31.4 58.5 27.6 27.1
27.9 52.1 30.6 25.4
22.1 49.9 23.2 21.3
25.5 53.5 24.8 19.3
```

```
31.1 56.6 30.0 25.4
30.4 56.7 28.3 27.2
18.7 46.5 23.0 11.7
19.7 44.2 28.6 17.8
14.6 42.7 21.3 12.8
29.5 54.4 30.1 23.9
27.7 55.3 25.7 22.6
30.2 58.6 24.6 25.4
22.7 48.2 27.1 14.8
25.2 51.0 27.5 21.1 : End of data
1 1 1 : ISX
```

9.3 Program Results

```
 G02KAF Example Program Results

 Value of ridge parameter:     0.0712

 Sum of squares of residuals: 0.1092E+03
 Degrees of freedom:     16
 Number of effective parameters:     2.9059

 Parameter estimates
    1     20.1950
    2      9.7934
    3      9.9576
    4     -2.0125

 Number of iterations:               6

 Ridge parameter minimises GCV

 Estimated prediction errors:
 GCV    =     7.4718
 UEV    =     6.3862
 FPE    =     7.3141
 BIC    =     8.2380
 LOO CV =     7.5495

 Residuals
    1     -1.9894
    2      3.5469
    3     -3.0392
    4     -3.0309
    5     -0.1899
    6     -0.3146
    7      0.9775
    8      4.0157
    9      2.5332
   10     -2.3560
   11      0.5446
   12      2.3989
   13     -4.0876
   14      3.2778
   15      0.2894
   16      0.7330
   17     -0.7116
   18     -0.6092
   19     -2.9995
   20      1.0110

 Variance inflation factors
    1      0.2928
    2      0.4162
    3      0.8089
```