NAG Toolbox for MATLAB

g03eh

1 Purpose

g03eh produces a dendrogram from the results of g03ec.

2 Syntax

```
[c, ifail] = g03eh(orient, dord, dmin, dstep, nsym, lenc, 'n', n)
```

3 Description

Hierarchical cluster analysis, as performed by g03ec, can be represented by a tree that shows at which distance the clusters merge. Such a tree is known as a dendrogram. See Everitt 1974 and Krzanowski 1990 for examples of dendrograms. A simple example is,

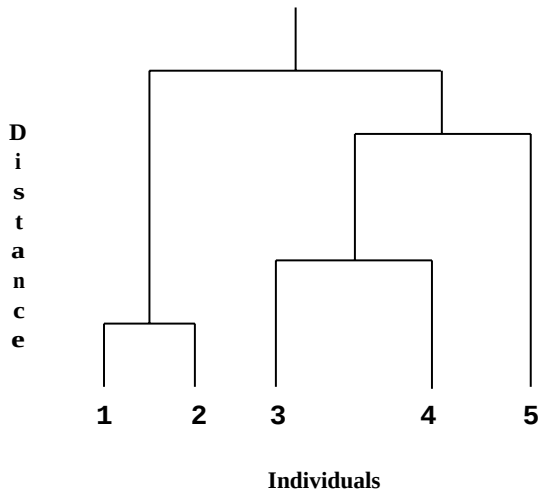

Figure 1

The end points of the dendrogram represent the objects that have been clustered. They should be in a suitable order as given by g03ec. Object 1 is always the first object. In the example above the height represents the distance at which the clusters merge.

The dendrogram is produced in a character array using the ordering and distances provided by g03ec. Suitable characters are used to represent parts of the tree.

There are four possible orientations for the dendrogram. The example above has the end points at the bottom of the diagram which will be referred to as south. If the dendrogram was the other way around with the end points at the top of the diagram then the orientation would be north. If the end points are at the left-hand or right-hand side of the diagram the orientation is west or east. Different symbols are used for east/west and north/south orientations.

4 References

Everitt B S 1974 *Cluster Analysis* Heinemann

Krzanowski W J 1990 *Principles of Multivariate Analysis* Oxford University Press

5 Parameters

5.1 Compulsory Input Parameters

1: **orient – string**

Indicates which orientation the dendrogram is to take.

orient = 'N'

The end points of the dendrogram are to the north.

orient = 'S'

The end points of the dendrogram are to the south.

orient = 'E'

The end points of the dendrogram are to the east.

orient = 'W'

The end points of the dendrogram are to the west.

Constraint: **orient** = 'N', 'S', 'E' or 'W'.

2: **dord**(**n**) – **double array**

The array **dord** as output by g03ec. **dord** contains the distances, in dendrogram order, at which clustering takes place.

Constraint: **dord**(**n**) $\geq$ **dord**(i), for $i = 1, 2, \ldots,$ **n** $- 1$.

3: **dmin – double scalar**

The clustering distance at which the dendrogram begins.

Constraint: **dmin** ≥ 0.0.

4: **dstep – double scalar**

The distance represented by one symbol of the dendrogram.

Constraint: **dstep** > 0.0.

5: **nsym – int32 scalar**

The number of character positions used in the dendrogram. Hence the clustering distance at which the dendrogram terminates is given by **dmin** + **nsym** $\times$ **dstep**.

Constraint: **nsym** ≥ 1.

6: **lenc – int32 scalar**

Constraints:

if **orient** = 'N' or 'S', **lenc** $\geq$ **nsym**;
if **orient** = 'E' or 'W', **lenc** $\geq$ **n**.

5.2 Optional Input Parameters

1: **n – int32 scalar**

Default: The dimension of the array **dord**.

the number of objects in the cluster analysis.

Constraint: **n** ≥ 2.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **c**(**lenc**) – **string array**

Note: the length of each element of **c** must be at least $3 \times$ **n** if **orient** = 'N' or 'S', or at least **nsym** if **orient** = 'E' or 'W'.

The elements of **c** contain consecutive lines of the dendrogram.

2: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail $= 1$

On entry, **n** < 2,
or **nsym** < 1,
or **dmin** < 0.0,
or **dstep** ≤ 0.0,
or **orient** $\neq$ 'N','S','E', or 'W',
or **orient** = 'N' or 'S', **lenc** $<$ **nsym**,
or **orient** = 'E' or 'W', **lenc** $<$ **n**,
or the number of characters that can be stored in each element of array **c** is insufficient for the requested orientation.

ifail $= 2$

On entry, **dord**(**n**) $<$ **dord**(i), for some $i = 1, 2, \ldots, \mathbf{n} - 1$.

7 Accuracy

Not applicable.

8 Further Comments

The scale of the dendrogram is controlled by **dstep**. The smaller the value **dstep** is, the greater the amount of detail that will be given but **nsym** will have to be larger to give the full dendrogram. The range of distances represented by the dendrogram is **dmin** to **nsym** $\times$ **dstep**. The values of **dmin**, **dstep** and **nsym** can thus be set so that only part of the dendrogram is produced.

The dendrogram does not include any labelling of the objects. You can print suitable labels using the ordering given by the array **iord** returned by g03ec.

9 Example

```
orient = 'E';
dord = [1;
     36.25;
     8;
     4;
     36.25];
dmin = 0;
```

```
dstep = 1.1;
nsym = int32(40);
lenc = int32(100);

[c, ifail] = g03eh(orient, dord, dmin, dstep, nsym, lenc)
```

```
c =
     array elided
ifail =
           0
```