GDM Tabular Modeler - Manual

Setting up the GDM tabular modeler in R

The GDM modeler for R comes in two versions, the 32 bit and the 64 bit versions. The version that you decide to use is governed by the version of R that use on your machine.

Both versions come with an R source file and a binary dynamic link library (dll). The R source files are called

GDM_Table_Funcs32_1_0.R and GDM_Table_Funcs64_1_0.R. Use the appropriate file for the version of R that you are using. The source file can be loaded from anywhere via the R “source” command. It will also setup the path to the dll file. This may be located anywhere on the system but the path MUST be edited in the R source files for the “mydllpath” variable at the start of the file to match where you locate the dll.

Fitting a GDM model to tabular data in R with the function gdm.fitfromtable

The GDM fitting function tabular data improves on the former gdm.fit function for R that used a site by species table to generate a Generalised Dissimilarity Model (GDM).

The basic record unit in the input data table is a site pair. Each record contains the response, the weighting, the geographic coordinates and optionally a set of predictor values for a given site pair.

The gdm.fitfromtable function allows users to define the response using any ratio based dissimilarity metric rather than being restricted to a Bray-Curtis dissimilarity measure. The same goes for the weights metrics in the input data table.

Predictor data can be derived from any source as long as there is a valid data value for both sites in a site pair. Categorical data needs to be transformed via a distance table if it is to be used as a predictor as the input data predictors are all expected to be continuous, not categorical.

Geographic distance can be optionally included in a GDM model by setting the ‘geo’ argument to TRUE. All the other predictors in the data table will also be used.

During the model fitting process, all the predictors are converted into I-Splines. The default number of I-Splines per predictor (and the possible minimum) is 3 but can be greater than this by supplying a splines vector as an argument to gdm.fitfromtable having the same length as the number of predictors being used in the model.

Each value in the splines vector contains the number of I-Splines to use for each predictor in the data table (geographic also if included in the ‘geo’ argument is set to TRUE).

Care should be taken not ‘overfit’ the model by using too many I-Splines per predictor.

The number of I-Splines also affects the position of the data quantiles that are used in the model fitting process. The default of 3 splines per predictor will use data quantiles at the 0%, 50% and 100% percentiles.

If a quantiles vector is supplied as an argument to gdm.fitfromtable, and it contains more than 3 quantiles for any predictor, then a splines vector must also be supplied detailing the number of I-Splines per predictor.

A user defined quantiles vector allows one to model with predictors having a greater potential range than suggested by the data.

See the document gdm_fitfromtable.doc for more information about the function arguments.

Data format for the GDM modeler for R tabular version.

The following example details the format for the GDM.fitfromtable input data.

response	weights	S1_X	S1_Y	S2_X	S2_Y	S1_P1	S1_P2	S2_P1	S2_P2
0.72750	1	278.5833	27.2000	280.3836	38.1547	173.3629	78.4853	116.4674	102.6337
0.85762	1	274.7864	41.6237	280.3836	38.1547	71.7503	70.1722	116.4674	102.6337
0.83472	1	280.3836	38.1547	260.0936	44.8361	116.4674	102.6337	72.3174	46.3758
0.61395	1	299.3250	45.6514	270.2744	43.4273	83.3311	87.2174	71.1283	62.1697
0.83438	1	271.6567	40.6758	291.9129	44.6342	79.8585	80.1550	80.2910	72.8016
0.35483	1	292.1265	46.2912	293.9208	44.2639	86.2399	71.0936	86.3301	90.2905
0.39259	1	270.2744	43.4273	293.9208	44.2639	71.1283	62.1697	86.3301	90.2905
0.65483	1	293.9208	44.2639	278.5000	27.5833	86.3301	90.2905	173.3629	78.4853

The first column contains the response data. Where the original GDM for R model function only used the Bray-Curtis distance metric for between-site dissimilarity, the table version will accommodate ANY ratio based distance metric (ratio based distance metrics should range from 0 for no dissimilarity between sites in each site pair to 1.0 for site pairs that share no species in common).

The second column contains the weights data. Where the original GDM for R model function had three modes, “equal” (all weights set to 1.0), standard (sum of presence records) and custom (used defined), this tabular version expects the weights to be included in the second column of the table.

The third and fourth columns have the X and Y coordinates for the FIRST site in the site pair.

The fifth and sixth columns have the X and Y coordinates for the LAST site in the site pair.

The first six columns are mandatory, that is they MUST be included in the data. Even if you don’t intend to use geographic distance as a GDM model predictor or you don’t actually have the coordinates, you still need to include the X and Y columns for both sites as

dummy data. These can be any value if you don't intend to use geographic distance as a GDM model predictor but they cannot be blank records.

The columns following the first six contain all the predictor data for the FIRST site then followed by all the predictor for the LAST site in each site pair. Note the above example has data for two predictors, P1 and P2.

The site pairs are represented as S1 and S2.

Note that if you only use the first six columns of data for input and set the 'geo' switch in the gdm.fitfromtable to FALSE (the default), then it will assumed that there are NO predictors and an error message will be returned. If the 'geo' switch was set to TRUE for the same data then only geographic distance would be the only predictor used in the GDM model.

Note also that using this table format, predictor data can be drawn from distance tables in addition to GIS grid data.

R documentation

of '`./gdm_fitfromtable.Rd`' etc.

March 27, 2013

`gdm.fitfromtable` *Fitting Generalized Dissimilarity Models from Tabular Data*

Description

gdm.fitfromtable is used to fit generalized dissimilarity models, specified by giving a data table of site pairs. The table is formatted as follows:

Observed,Weights,X0,Y0,X1,Y1,Pred1SiteA,Pred2SiteA,.,PredNSiteA, Pred1SiteB,Pred2SiteB,.,PredNSiteB

The first column should be a ratio based dissimilarity measure between SitesA and SitesB. The second column defines any weighting to be applied. It should be set to 1.0 as default (no weights) The third and fourth columns, X0 and Y0 represent the coordinates of the first site from a site pair. The fifth and sixth columns, X1 and Y1 represent the coordinates of the second site from a site pair. Note that these columns MUST be included, even if you do not intend to use geographic distance as a predictor. These columns can be loaded with dummy data if the actual coordinates are unknown.

The next columns are for N predictors for SiteA and followed by N predictors for Site B.

Usage

```
gdm.fitfromtable(data, geo=FALSE, splines=NULL, quantiles=NULL)
```

Arguments

`data`	a data frame representing the model data values for a collection of site pairs. The observed response data must be located in the first column. The weights data must be located in the second column. If geo is TRUE, then the X0,Y0 and Y0,Y1 columns will be used for calculating the geographic distance between each site in each site pair for inclusion of the geographic predictor term into the GDM model. If geo is FALSE, the default, then the X0,Y0,X1 and Y1 data columns are ignored. The predictor data for Site A and the predictor data for Site B follows.

geo	set to TRUE if geographic distance between sites is to be included as a model term (refer to the details for the data argument for more details on how to format the input data). Set to FALSE if geographic distance is to be omitted from the model. Default is FALSE.
splines	an optional vector of I-Spline counts to be used in the fitting process. If supplied, it must have the same length as the number of predictors (including geographic distance if in use).
quantiles	an optional vector of quantiles to be used in the fitting process. If quantiles are supplied and splines=NULL, then it must have the same length as the number of predictors * 3. If both quantiles and splines are supplied, then the length of quantiles must be the same as the sum of splines.

Details

A description of the gdm process could be inserted here...

Value

gdm.fitfromtable returns a gdm model object. The function summary (i.e., gdm.summaryfromtable) can be used to obtain or print a summary of the results.

A gdm model object is a list containing at least the following components:

dataname	The name of the table used as the data argument to the model
geo	Whether geographic distance was used in the model
gdmdeviance	The deviance of the gdm model
nulldeviance	the NULL deviance of the gdm model
explained	the percentage of deviance explained by the model
intercept	the intercept value that is added to the overall model
predictors	a list of the names of the predictors that were used to fit
coefficients	a list of the spline values for all the predictors included in the x data. By default only three splines are used for each predictor, so the length of this list will be three times the length of the above predictors list. quantiles a vector of the percentiles derived from the x data (or user defined), for each predictor.
splines	a vector of I-Spline counts for each predictor
creationdate	the date and time of model creation
observed	the observed response for each site pair (from the data column 1)
predicted	the predicted response after applying the GDM link function
ecological	the predicted ecological distance from the GDM model.

Author(s)

The original R implementation of GDM was written by Glenn Manion working for Simon Ferrier at the Department of Environment and Conservation, New South Wales, Australia.

See Also

gdm.summaryfromtable, gdm.plotfromtable, gdm.predictfromtable for GDM tabular methods.

Examples

```
test.mod <- gdm.fitfromtable(test.df, geo=TRUE, splines = c(3,3,4))
```

---

gdm.plotfromtable *Plotting a Generalized Dissimilarity Model*

Description

gdm.plotfromtable is used to plot a generalized dissimilarity model, specified by giving a gdm model object created with the function gdm.fitfromtable.

Usage

```
gdm.plotfromtable(model, plot.layout = c(2,2), plot.color = rgb(0.0,0.0,1.0),
plot.linewidth = 2.0)
```

Arguments

model	a gdm model object returned from gdm.fitfromtable.
plot.layout	This sets the row and column layout for the predictor plots. Default is 2 rows by 2 columns. To produce one predictor plot per page set plot.layout to c(1,1). The overall model plots are always produced on a single page each i.e. a single page plot of the raw prediction against the observed response data and a single page plot of the predicted data after applying the link function of $1 - e^{-y}$. This layout parameter only affects the layout of the respective plot for those predictors that had featured in the model fitting process.
plot.colour	an RGB vector describing the colour of the data points that are plotted for the overall plots. The default is blue rgb(0.0,0.0,1.0).
plot.linewidth	A value describing the desired linewidth for the regression over-plotted in the two overall plots to optimise the display of the regression line over the data points.

Details

A gdm.plotfromtable produces a series of plots detailing the results of a gdm model. The overall fit is plotted with the raw data in the first window. The overall fit after applying the link function $1 - e^{-y}$ is plotted in the second window. Individual plots of the predictors that were included in the gdm model are plotted in successive windows determined by the plot.layout parameter.

Value

gdm.plotfromtable

Author(s)

The original R implementation of GDM was written by Glenn Manion working for Simon Ferrier at the Department of Environment and Conservation, New South Wales, Australia.

References

Ferrier, S., Manion, G., Elith, J. and Richardson, K. (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, **13**, 252–264

See Also

gdm.fitfromtable, gdm.summaryfromtable, gdm.predictfromtable for GDM tabular methods.

Examples

```
gdm.plotfromtable(test.mod,test.df,plot.layout=c(3,3),+ plot.color=rgb(0.5,0.25,0.25))
```

---

gdm.predictfromtable *Predicting GDM Response Distances Using a Generalized Dissimilarity Model*

Description

This function predicts the response values data using the results of a model object returned from gdm.fitfromtable and a data frame of environmental data for a set of sites formatted as follows

Observed,Weights,X0,Y0,X1,Y1,Pred1SiteA,Pred2SiteA,.,PredNSiteA, Pred1SiteB,Pred2SiteB,.,PredNSiteB

The first column should be a ratio based dissimilarity measure between SitesA and SitesB. The second column defines any weighting to be applied. It should be set to 1.0 as default (no weights) The third and fourth columns, X0 and Y0 represent the coordinates of the first site from a site pair. The fifth and sixth columns, X1 and Y1 represent the coordinates of the second site from a site pair. Note that these columns MUST be included, even if you do not intend to use geographic distance as a predictor. These columns can be loaded with dummy data if the actual coordinates are unknown.

The next columns are for N predictors for SiteA and followed by N predictors for Site B.

In the case of gdm.predictfromtable, the first column could be set with dummy data as it is only the predictor data (and the XY columns if geo=TRUE) that will be used for the predictions.

Usage

```
gdm.predictfromtable(model, data)
```

Arguments

model	a gdm model object resulting from a call to gdm.fitfromtable.
data	a data frame representing the model data values for a collection of site pairs. The observed response data must be located in the first column. The weights data must be located in the second column. If geo is TRUE, then the X0,Y0 and Y0,Y1 columns will be used for calculating the geographic distance between each site in each site pair for inclusion of the geographic predictor term into the GDM model. If geo is FALSE, the default, then the X0,Y0,X1 and Y1 data columns are ignored. The predictor data for Site A and the predictor data for Site B follows.

Details

gdm.predictfromtable uses the model object to predict the response values from a set of site pairs supplied for the data argument.

Value

gdm.predictfromtable returns a response vector with the same length as the number of rows in each of the input data frame.

Author(s)

The original R implementation of GDM was written by Glenn Manion working for Simon Ferrier at the Department of Environment and Conservation, New South Wales, Australia.

See Also

gdm.fitfromtable, gdm.summaryfromtable, gdm.plotfromtable for GDM tabular methods.

Examples

```
myResponse <- gdm.predictfromtable(test.mod,data2predict)
```

---

gdm.summaryfromtable *Summarize a Fitted Generalized Dissimilarity Model*

Description

This function summarizes the model object returned by gdm.fitfromtable.

Usage

```
gdm.summaryfromtable(model_object)
```

Arguments

`object`	a gdm model object resulting from a call to gdm.fitfromtable.

Details

gdm.summaryfromtable displays the contents of a gdm model object to the R console window. It prints the names of the list items in the model object and formats the description of each predictor giving the name, the 0%, 50% and 100% quantiles and the respective coefficients derived for each predictor.

Value

gdm.summaryfromtable prints its output to the R Console windows and returns no value.

Author(s)

The original R implementation of GDM was written by Glenn Manion working for Simon Ferrier at the Department of Environment and Conservation, New South Wales, Australia.

References

Ferrier, S., Manion, G., Elith, J. and Richardson, K. (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, **13**, 252–264

See Also

`gdm.fitfromtable`, `gdm.plotfromtable`, `gdm.predictfromtable` for GDM tabular methods.

Examples

```
## Not run: gdm.summaryfromtable(my.gdm)
```

Index