Package 'GowerSom'

May 7, 2026

Type Package

Title Self-Organizing Maps for Mixed-Attribute Data Using Gower Distance

Version 0.1.0

Description Implements a variant of the Self-Organizing Map (SOM) algorithm designed for mixed-attribute datasets. Similarity between observations is computed using the Gower distance, and categorical prototypes are updated via heuristic strategies (weighted mode and multinomial sampling). Provides functions for model fitting, mapping, visualization (U-Matrix and component planes), and evaluation, making SOM applicable to heterogeneous real-world data. For methodological details see Sáez and Salas (2026) <doi:10.1007/s41060-025-00941-6>.

License GPL-2

Encoding UTF-8

Depends R (>= 4.3.0)

Imports StatMatch, dplyr, gower, ggplot2, cluster, reshape2, grid, utils, stats,cli

Suggests knitr, rmarkdown

RoxygenNote 7.3.3

NeedsCompilation yes

Maintainer Patricio Salas <patricioasalas@udec.cl>

Author Patricio Salas [aut, cre] (ORCID: <https://orcid.org/0000-0002-2201-4038>), Patricio Sáez [aut] (ORCID: <https://orcid.org/0000-0002-0113-3644>)

Repository CRAN

Date/Publication 2026-01-27 08:50:02 UTC

Contents

`get_bmu_gower` *Map observations to BMUs (Best Matching Units) using Gower distance*

Description

Computes, for each observation, the index of the best-matching neuron (BMU) in a trained Gower-SOM codebook and the corresponding Gower distance. Also converts BMU indices into grid coordinates (row, col).

Usage

```
get_bmu_gower(data, codebook, n_rows, n_cols)
```

Arguments

`data`	A `data.frame` of observations to map. Must be typed consistently with the training data (numeric, factor, etc.).
`codebook`	A `data.frame` (or coercible matrix) with one row per neuron and the same columns as `data`.
`n_rows, n_cols`	Integers, the SOM grid dimensions.

Value

A `data.frame` with the following columns:

bmu Integer BMU index (1 .. n_rows * n_cols).

distance Numeric, the Gower distance to the BMU.

row Integer, BMU grid row coordinate.

col Integer, BMU grid column coordinate.

Author(s)

Patricio Sáez <patricsaez@udec.cl>; Patricio Salas <patricioasalas@udec.cl>

References

Sáez, P., Salas, P. Gower-SOM: a self-organizing map for mixed data with gower distance and heuristic adaptation for data analytics. Int J Data Sci Anal 22, 26 (2026). https://doi.org/10.1007/s41060-025-00941-6/."

See Also

gsom_predict

Examples

```
set.seed(1)
df <- data.frame(
  x1 = rnorm(10),
  x2 = rnorm(10),
  g  = factor(sample(letters[1:3], 10, TRUE))
)
fit <- gsom_Training(df, grid_rows = 3, grid_cols = 3,
                num_iterations = 5, batch_size = 5)
res <- get_bmu_gower(df, codebook = fit$weights,
                     n_rows = 3, n_cols = 3)
head(res)
```

---

gsom_predict *Predict BMUs for new data using a fitted Gower-SOM*

Description

Maps new observations to their Best Matching Units (BMUs) using the codebook and grid stored in a fitted gowersom object.

Usage

```
gsom_predict(object, newdata, ...)
```

Arguments

object	A gowersom object returned by gsom_Training().
newdata	A data.frame of new observations to map. Must be typed consistently with the training data (numeric, factor, etc.).
...	Additional arguments passed to internal functions (not used).

Details

This function is a convenience wrapper around get_bmu_gower. It automatically extracts the grid dimensions from object\$coords and applies BMU mapping for each observation in newdata.

Value

A data.frame with the following columns:

bmu Integer BMU index (1 .. n_rows * n_cols).

distance Numeric Gower distance to the BMU.

row Integer, BMU grid row coordinate.

col Integer, BMU grid column coordinate.

Author(s)

Patricio Sáez <patricsaez@udec.cl>; Patricio Salas <patricioasalas@udec.cl>

References

Sáez, P., Salas, P. Gower-SOM: a self-organizing map for mixed data with gower distance and heuristic adaptation for data analytics. Int J Data Sci Anal 22, 26 (2026). https://doi.org/10.1007/s41060-025-00941-6/."

See Also

get_bmu_gower

Examples

```
set.seed(1)
df <- data.frame(
  x1 = rnorm(20),
  x2 = rnorm(20),
  g  = factor(sample(letters[1:3], 20, TRUE))
)
fit <- gsom_Training(df, grid_rows = 3, grid_cols = 3,
                num_iterations = 5, batch_size = 4)

# Map observations to BMUs
pred <- gsom_predict(fit, df)
head(pred)
```

---

gsom_Training *Train a Gower-SOM on mixed-attribute data*

Description

Train a Self-Organizing Map (SOM) on datasets with mixed attributes (numeric and categorical) using Gower distance to find the BMU and heuristics to update categorical prototypes.

Usage

```
gsom_Training(data, grid_rows = 5, grid_cols = 5,
         learning_rate = 0.1, num_iterations = 100,
         radius = NULL, batch_size = 10,
         sampling = TRUE, set_seed = 123)
```

Arguments

`data`	`data.frame` with correctly typed columns (numeric, factor, etc.).
`grid_rows`, `grid_cols`	SOM grid dimensions (rows x cols).
`learning_rate`	Initial learning rate (decays exponentially).
`num_iterations`	Number of iterations.
`radius`	Initial neighborhood radius; default `max(grid_rows, grid_cols)/2`.
`batch_size`	Mini-batch size per iteration.
`sampling`	Logical; if `TRUE`, multinomial sampling for categorical updates, else weighted mode.
`set_seed`	Integer random seed for reproducibility.

Details

Learning rate and neighborhood radius decay exponentially per iteration:

$$\alpha_t = \alpha_0 \exp(-t/T), \quad r_t = r_0 \exp(-t/(T/\log r_0))$$

where T is `num_iterations` and r_0 is `radius` (default `max(grid_rows, grid_cols)/2`). For categorical variables, the prototype combines current and input values weighted by α_t and the neighborhood kernel; if `sampling = TRUE`, a weighted draw is used; otherwise a weighted mode is applied.

Value

An object of class `gowersom` with:

- `weights`: data.frame of trained neuron prototypes.
- `coords`: data.frame of grid coordinates per neuron.

Author(s)

Patricio Sáez <patricsaez@udec.cl>; Patricio Salas <patricioasalas@udec.cl>

References

Sáez, P., Salas, P. Gower-SOM: a self-organizing map for mixed data with gower distance and heuristic adaptation for data analytics. Int J Data Sci Anal 22, 26 (2026). https://doi.org/10.1007/s41060-025-00941-6/."

Examples

```
set.seed(1)
df <- data.frame(
  x1 = rnorm(50),
  x2 = rnorm(50),
  g  = factor(sample(letters[1:3], 50, TRUE))
)
fit <- gsom_Training(df, grid_rows = 3, grid_cols = 3,
```

```
                learning_rate = 0.1, num_iterations = 10,
                batch_size = 8, sampling = TRUE, set_seed = 123)
str(fit)
```

---

gsom_Umatrix *Compute the U-Matrix for a trained Gower-SOM*

Description

Calculates the U-Matrix (unified distance matrix) to visualize the topology and cluster structure of a Self-Organizing Map trained on mixed-attribute data. Each entry contains the average Gower distance between a neuron and its immediate neighbors in the rectangular grid.

Usage

```
gsom_Umatrix(codebook, n_rows, n_cols)
```

Arguments

codebook	A data.frame or matrix containing the SOM prototypes (weights), with one row per neuron.
n_rows	Integer, number of rows in the SOM grid.
n_cols	Integer, number of columns in the SOM grid.

Details

The function assumes a rectangular topology where each neuron has up to four direct neighbors (up, down, left, right). For each neuron, the mean Gower distance to its valid neighbors is computed using daisy with metric = "gower".

Value

A numeric matrix of size n_rows x n_cols, where each cell contains the average distance between the corresponding neuron and its neighbors.

Author(s)

Patricio Sáez <patricsaez@udec.cl>; Patricio Salas <patricioasalas@udec.cl>

References

Sáez, P., Salas, P. Gower-SOM: a self-organizing map for mixed data with gower distance and heuristic adaptation for data analytics. Int J Data Sci Anal 22, 26 (2026). https://doi.org/10.1007/s41060-025-00941-6/."

See Also

daisy

Examples

```
set.seed(1)
df <- data.frame(
  x1 = rnorm(20),
  x2 = rnorm(20),
  g  = factor(sample(letters[1:3], 20, TRUE))
)
fit <- gsom_Training(df, grid_rows = 3, grid_cols = 3,
                num_iterations = 5, batch_size = 4)
U <- gsom_Umatrix(fit$weights, n_rows = 3, n_cols = 3)
plot_Umatrix(U)
```

`plot_Umatrix` *Plot the U-Matrix of a Gower-SOM*

Description

Visualizes the U-Matrix of a trained Gower-SOM using **ggplot2**. The U-Matrix reveals cluster boundaries and topological structures in the map.

Usage

```
plot_Umatrix(u_matrix, fill_palette = "C")
```

Arguments

`u_matrix`	Numeric matrix as returned by `gsom_Umatrix` (`n_rows` x `n_cols`).
`fill_palette`	Character string, viridis option for the fill scale (default `"C"`).

Details

The function reshapes the U-Matrix into long format and draws a raster heatmap with `geom_raster`. By default, it uses perceptually uniform viridis palettes for improved interpretability, but the palette can be changed through `fill_palette`.

Value

A `ggplot` object displaying the U-Matrix as a heatmap.

See Also

`gsom_Umatrix`

Examples

```
set.seed(1)
df <- data.frame(
  x1 = rnorm(20),
  x2 = rnorm(20),
  g  = factor(sample(letters[1:3], 20, TRUE))
)
fit <- gsom_Training(df, grid_rows = 3, grid_cols = 3,
                num_iterations = 5, batch_size = 4)
U <- gsom_Umatrix(fit$weights, n_rows = 3, n_cols = 3)
plot_Umatrix(U)
```

Index