

Package ‘spatialsample’

March 4, 2021

Title Spatial Resampling Infrastructure

Version 0.1.0

Description Functions and classes for spatial resampling to use with the 'rsample' package, such as spatial cross-validation (Brenning, 2012) <doi:10.1109/IGARSS.2012.6352393>. The scope of 'rsample' and 'spatialsample' is to provide the basic building blocks for creating and analyzing resamples of a spatial data set, but neither package includes functions for modeling or computing statistics. The resampled spatial data sets created by 'spatialsample' do not contain much overhead in memory.

License MIT + file LICENSE

URL <https://github.com/tidymodels/spatialsample>,
<https://spatialsample.tidymodels.org>

BugReports <https://github.com/tidymodels/spatialsample/issues>

Depends R (>= 3.2)

Imports dplyr (>= 1.0.0), purrr, rlang, rsample (>= 0.0.9), tibble, tidyselect, vctrs (>= 0.3.6)

Suggests ggplot2, knitr, modeldata, rmarkdown, tidyr, testthat (>= 3.0.0), yardstick, covr

Config/testthat/edition 3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1.9001

VignetteBuilder knitr

NeedsCompilation no

Author Julia Silge [aut, cre] (<<https://orcid.org/0000-0002-3671-836X>>), RStudio [cph]

Maintainer Julia Silge <julia.silge@rstudio.com>

Repository CRAN

Date/Publication 2021-03-04 09:30:05 UTC

R topics documented:

spatialsample	2
spatial_clustering_cv	2

Index	4
--------------	----------

spatialsample	<i>spatialsample: Spatial Resampling Infrastructure for R</i>
---------------	---

Description

spatialsample has functions to create resamples of a spatial data set that can be used to evaluate models or to estimate the sampling distribution of some statistic. It is a specialized package designed with the same principles and terminology as [rsample](#).

Terminology

- A **resample** is the result of a split of a data set. For example, in cross-validation, a data set is split into complementary subsets, and different partitions of subsets are used for different purposes. The data structure `rsplit` is used to store a single resample.
- When the data are split in two, the portion that is used to estimate the model or calculate the statistic is called the **analysis** set here. In machine learning this is sometimes called the "training set", but this may be a poor name choice in a resampling context since it might conflict with an initial split of the original data.
- Conversely, the other data in the split are called the **assessment** data. In bootstrapping, these data are often called the "out-of-bag" samples.
- A collection of resamples is contained in an `rset` object.

Basic Functions

The main resampling functions are: [spatial_clustering_cv\(\)](#)

spatial_clustering_cv	<i>Spatial or Cluster Cross-Validation</i>
-----------------------	--

Description

Spatial or cluster cross-validation splits the data into V groups of disjointed sets using k-means clustering of some variables, typically spatial coordinates. A resample of the analysis data consists of $V-1$ of the folds/clusters while the assessment set contains the final fold/cluster. In basic spatial cross-validation (i.e. no repeats), the number of resamples is equal to V .

Usage

```
spatial_clustering_cv(data, coords, v = 10, ...)
```

Arguments

data	A data frame.
coords	A vector of variable names, typically spatial coordinates, to partition the data into disjointed sets via k-means clustering.
v	The number of partitions of the data set.
...	Extra arguments passed on to <code>stats::kmeans()</code> .

Details

The variables in the `coords` argument are used for k-means clustering of the data into disjointed sets, as outlined in Brenning (2012). These clusters are used as the folds for cross-validation. Depending on how the data are distributed spatially, there may not be an equal number of points in each fold.

Value

A tibble with classes `spatial_cv`, `rset`, `tbl_df`, `tbl`, and `data.frame`. The results include a column for the data split objects and an identification variable `id`.

References

A. Brenning, "Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package `sperrorest`," 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, 2012, pp. 5372-5375, doi: 10.1109/IGARSS.2012.6352393.

Examples

```
data(ames, package = "modeldata")
spatial_clustering_cv(ames, coords = c(Latitude, Longitude), v = 5)
```

Index

`rsample`, [2](#)

`spatial_clustering_cv`, [2](#)

`spatial_clustering_cv()`, [2](#)

`spatialsample`, [2](#)

`stats::kmeans()`, [3](#)