Package 'CRPClustering'

May 2, 2018

Type Package

Title Bayesian Nonparametric Chinese Restaurant Process Clustering
with Entropy

Version 1.2

Date 2018-05-02

Author Masashi Okada [aut, cre]

Maintainer Masashi Okada <okadaalgorithm@gmail.com>

Description Chinese restaurant process [Pitman (1995) <doi:10.1007/BF01213386>] is used in or-
der to compose Dirichlet process [Ferguson (1973) <doi:10.1214/aos/1176342360>]. The clus-
tering which uses Chinese restaurant process does not need to decide the number of clus-
ters in advance. This algorithm automatically adjusts it. And this package calculates the ambigu-
ity of clusters as entropy [Yngvason (1999) <doi:10.1016/S0370-1573(98)00082-9>].

Depends R (>= 3.2.3)

License GPL (>= 2)

Imports MASS, mvtnorm, randomcoloR, lucid, dplyr, graphics, stats,
utils, png

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

Suggests R.rsp

VignetteBuilder R.rsp

NeedsCompilation no

Repository CRAN

Date/Publication 2018-05-01 22:05:51 UTC

R topics documented:

---

crp_gibbs   *Markov chain Monte Carlo methods for CRP clustering*

Description

Markov chain Monte Carlo methods for CRP clustering

Usage

```
crp_gibbs(data, mu = c(0, 0), sigma_table = 14, alpha = 0.3, ro_0 = 0.1,
  burn_in = 40, iteration = 200)
```

Arguments

data	: a matrix of data for clustering. Row is each data_i and column is dimensions of each data_i.
mu	: a vector of center points of data. If data is 3 dimensions, a vector of 3 elements like c(2,4,2).
sigma_table	: a numeric of CRP variance.
alpha	: a numeric of a CRP concentrate rate.
ro_0	: a numeric of a CRP mu change rate.
burn_in	: an iteration integer of burn in.
iteration	: an iteration integer.

Value

z_result : an array expresses cluster numbers for each data_i.

Examples

```
data <- matrix(c(1.8,1.9,2.1,2.5,5.6,5.2,6,6.1), 4, 2)
z_result <- crp_gibbs(
                    data,
                    mu=c(0,0),
                    sigma_table=14,
                    alpha=0.3,
                    ro_0=0.1,
                    burn_in=10,
                    iteration=100
                    )
```

---

crp_graph_2d   *CRP clustering visualization*

Description

CRP clustering visualization

Usage

```
crp_graph_2d(data = data, z_result = z_result)
```

Arguments

data   : a matrix of data for clustering. Row is each data_i and column is dimensions
                of each data_i.

z_result   : an array denotes the number of a cluster for each data_i and it is the output of
                the method "crp_gibbs".

---

data   *Generated from three normal distributions.*

Description

Generated from three normal distributions.

Usage

```
data
```

Format

A data frame with 1000 rows and 2 dimensions

Index