

Package ‘afCEC’

September 13, 2016

Type Package

Title What the package does (short line)

Version 1.0

Date 2015-02-25

Author Your Name

Maintainer Your Name <your@email.com>

Description More about what it does (maybe more than one line)

License GPL (>= 2)

Imports Rcpp (>= 0.11.3), RcppArmadillo

LinkingTo Rcpp, RcppArmadillo

Depends RcppArmadillo

R topics documented:

| | |
|-------------------------|---|
| afCEC-package | 1 |
| afCEC | 2 |
| plot | 4 |

| | |
|--------------|---|
| Index | 7 |
|--------------|---|

| | |
|---------------|---|
| afCEC-package | <i>Active Function Cross-Entropy Clustering</i> |
|---------------|---|

Description

Active Function Cross-Entropy Clustering (afCEC) partitions the n-dimensional data into the clusters by finding the parameters of the mixed generalised multivariate normal distribution, that optimally approximates the scattering of the data in the n-dimensional space, whose density function is of the form:

$$p_1 * N(m_1, \sigma_1, \sigma_1, f_1) + \dots + p_k * N(m_k, \sigma_k, \sigma_k, f_k)$$

Additionally, the Active Function Cross-Entropy Clustering performs the automatic reduction of the unnecessary clusters. It's implemented in the form of a customizable function afCEC.

Details

Package: afCEC
 Type: Package
 Version: 0.9.1
 Date: 2014-10-06
 License: GPL-3

Author(s)

Krzysztof Byrski

See Also

[afCEC](#).

| | |
|-------|--------------|
| afCEC | <i>afCEC</i> |
|-------|--------------|

Description

Performs the Active Function Cross-Entropy Clustering on the data set.

Usage

```

afCEC (
  points,
  maxClusters,
  initialLabels="k-means++",
  cardMin=0.05,
  costThreshold=-0.000001,
  minIterations=0,
  maxIterations=100,
  numberOfStarts=1,
  method="Hartigan",
  values="quadratic",
  interactive=FALSE
)

```

Arguments

| | |
|----------------------------|--|
| <code>points</code> | A (#points x n) matrix of data containing n-dimensional points stored row-by-row. |
| <code>maxClusters</code> | Indicates the maximal number of clusters, that afCEC can partition the data into. |
| <code>initialLabels</code> | Initial labelling determining the data membership to the particular clusters. There are 3 options allowed: <ul style="list-style-type: none"> "random" - causes the labelling to be randomly generated. "k-means++" - causes the labelling to be generated using the k-means++ heuristics. |

| | |
|----------------|---|
| | <ul style="list-style-type: none"> A (#points x numberOfStarts) matrix of type integer containing values in the range: 0...maxClusters - 1. The value x in the row i and column j indicates that in the j-th start, the i-th point will be initially assigned to the (x + 1)-th cluster. <p>The default value is "k-means++".</p> |
| cardMin | Value so that if the number of points in particular cluster relatively to the size of the whole data drops below, then that cluster gets removed and its data is assigned to the other clusters. The default value is 0.05. |
| costTreshold | Negative value so that if the difference between the calculated cost functions in the subsequent iterations is greater than it then the afCEC terminates returning the solution from the terminated iteration as the final solution (in the present start), provided that at least minIterations already passed. The default value is -0.000001; |
| minIterations | Value indicating the minimal number of iterations needed before afCEC can terminate, provided that no error occurred. The default value is 0. |
| maxIterations | Value indicating the maximal number of iterations in one start, that afCEC cannot exceed. The default value is 100. |
| numberOfStarts | Value indicating the number of runs of the algorithm to be performed. The best solution is chosen out of the solutions given in the particular steps, so that it has the lowest value of the cost function among them. The default value is 1. |
| method | <p>Heuristics used to perform the clustering. There are 2 options allowed:</p> <ul style="list-style-type: none"> "Lloyd" - indicates that the Lloyd heuristics will be used to perform the clustering. "Hartigan" - indicates that the Hartigan heuristics will be used to perform the clustering. <p>The default value is "Hartigan".</p> |
| values | <p>Definition of the active function used to perform the clustering. There are 3 options allowed:</p> <ul style="list-style-type: none"> "quadratic" - indicates that the function of the form: $f(x_1, \dots, x_{(n-1)}) = x_1^2 + \dots + x_{(n-1)}^2 + x_1 + \dots + x_{(n-1)} + 1,$ <p>where n is the dimensionality of the data, will be used to perform the clustering.</p> <ul style="list-style-type: none"> "String containing the user-defined formula. See more in User-defined formulas. A ((m*n) x #points) matrix containing the values of the particular components of the active function placed according to the following layout (row by row): <p>First m rows:</p> $\begin{bmatrix} f_1(x_{2_1}, \dots, x_{n_1}), f_1(x_{2_2}, \dots, x_{n_2}), \dots, f_1(x_{2_{\#points}}, \dots, x_{n_{\#points}}) \\ f_2(x_{2_1}, \dots, x_{n_1}), f_2(x_{2_2}, \dots, x_{n_2}), \dots, f_2(x_{2_{\#points}}, \dots, x_{n_{\#points}}) \\ \dots \\ f_m(x_{2_1}, \dots, x_{n_1}), f_m(x_{2_2}, \dots, x_{n_2}), \dots, f_m(x_{2_{\#points}}, \dots, x_{n_{\#points}}) \end{bmatrix}$ <p>Second m rows:</p> $[f_1(x_{1_1}, x_{3_1}, \dots, x_{n_1}), f_1(x_{1_2}, x_{3_2}, \dots, x_{n_2}), \dots]$ |

```
[f_2(x_1_1,x_3_1,...,x_n_1), f_2(x_1_2,x_3_2,...,x_n_2), ...]  
...  
[f_m(x_1_1,x_3_1,...,x_n_1), f_m(x_1_2,x_3_2,...,x_n_2), ...]
```

Last m rows:

```
[f_1(x_1_1,...,x_(n-1)_1), f_1(x_1_2,...,x_(n-1)_2), ...]  
[f_2(x_1_1,...,x_(n-1)_1), f_2(x_1_2,...,x_(n-1)_2), ...]  
...  
[f_m(x_1_1,...,x_(n-1)_1), f_m(x_1_2,...,x_(n-1)_2), ...],
```

where: n - dimensionality of the data, x_{i_j} - j-th coordinate of the i-th point of data, m - number of components of the active function.

The default value is "quadratic".

Remarks:

In the case of the third way of defining the function, the returned object doesn't contain the information about the means coordinates corresponding to the active direction. Moreover, the plotting capabilities of the afCEC package are in that case severely impaired. See plot.

interactive Indicates if the algorithm runs in the "interactive" mode. The "interactive" mode allows to track the intermediate steps of the afCEC. Instead of one object of the afCEC class, the whole list of such objects is returned, where each item of the list corresponds to the intermediate step of the algorithm. See value section for more details. The default value is FALSE.

Value

- Empty list - if clustering failed.
- Object of class afCEC containing the parameters of the best fitted mixed generalised multivariate normal distribution - if argument interactive=FALSE and clustering succeeded.
- List of k objects of class afCEC containing the parameters of the best fitted mixed generalised multivariate normal distribution in the k intermediate steps of running of the algorithm - if argument interactive=TRUE and clustering succeeded.

See Also

[User-defined formulas](#)
[plot](#)

| | |
|------|------|
| plot | plot |
|------|------|

Description

Plots the chart showing the clustering of the data. Depending on the dimensionality of the data passed to the afCEC function, it draws either 2D or 3D chart. The function doesn't work with the other dimensions.

Usage

```
plot (
  x,
  draw_points=TRUE,
  draw_means=TRUE,
  draw_ellipsoids=TRUE,
  draw_surfaces=TRUE,
  confidence=0.95,
  grid_resolution=32
)
```

Arguments

| | |
|-------------|--|
| x | Object of class afCEC returned from the afCEC function. |
| draw_points | If the value is TRUE, function draws the points belonging to the particular clusters, so that the points belonging to the different clusters have the different colors. The default value is TRUE. |
| draw_means | If the value is TRUE, function draws the means of the clusters as the big black dots / spheres. The default value is TRUE. |

Remarks:

This functionality doesn't work if object x was given after clustering the data with the active function defined by the matrix of it's components values.

| | |
|-----------------|--|
| draw_ellipsoids | If the value is TRUE, function draws the curved ellipses / ellipsoids of confidence of the particular clusters. The default value is TRUE. |
|-----------------|--|

Remarks:

This functionality doesn't work if object x was given after clustering the data with the active function defined by the matrix of it's components values.

| | |
|---------------|--|
| draw_surfaces | If the value is TRUE, function draws the line / curved surface spanned on the inactive principal axes of the curved ellipses / ellipsoids of confidence of the particular clusters. The default value is TRUE. |
|---------------|--|

Remarks:

This functionality doesn't work if object x was given after clustering the data with the active function defined by the matrix of it's components values.

| | |
|------------|--|
| confidence | Determines the percentile of data belonging to the particular clusters, the corresponding ellipses / ellipsoids of confidence should contain. For example, if the value 0.5 is specified, then, for each cluster, it's ellipse / ellipsoid of confidence will contain $0.5 * 100$ [%] of assigned points. The default value is 0.95. |
|------------|--|

| | |
|-----------------|--|
| grid_resolution | Determines the grid resolution using to approximate the curves / surfaces drawn in the plot function. The default value is 32. |
|-----------------|--|

Value

None.

See Also[afCEC](#)

Index

*Topic **afCEC**

afCEC, [2](#)

*Topic **clustering, entropy, gaussian,
kmeans**

afCEC-package, [1](#)

afCEC, [2](#), [2](#), [6](#)

afCEC-package, [1](#)

plot, [4](#), [4](#)