# Package ‘Modalclust’

May 7, 2026

**Type** Package

**Title** Hierarchical Modal Clustering

**Version** 0.7

**Date** 2018-11-11

**Author** Surajit Ray and Yansong Cheng

**Maintainer** Surajit Ray <surajit.ray@glasgow.ac.uk>

**Description** Performs Modal Clustering (MAC) including Hierarchical Modal Clustering (HMAC) along with their parallel implementation (PHMAC) over several processors. These model-based non-parametric clustering techniques can extract clusters in very high dimensions with arbitrary density shapes. By default clustering is performed over several resolutions and the results are summarised as a hierarchical tree. Associated plot functions are also provided. There is a package vignette that provides many examples. This version adheres to CRAN policy of not spanning more than two child processes by default.

**Depends** R (>= 2.14.0), mvtnorm, zoo, class

**Suggests** parallel, MASS

**License** GPL-2

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2018-11-14 08:20:03 UTC

## Contents

choose.cluster . . . . . 2
contour.hmac . . . . . 3
cta20 . . . . . 4
disc2d . . . . . 5
findmid . . . . . 6
hard.hmac . . . . . 6
hmac . . . . . 8
khat.inv . . . . . 9
mydmvnorm . . . . . 9

oned . . . . . 10
phmac . . . . . 11
plot.hmac . . . . . 12
soft.hmac . . . . . 13
summary . . . . . 15

**Index** **16**

---

`choose.cluster` *Choosing the cluster which is closest to a specified point*

---

## Description

Choosing the cluster which is closest to a point specified by user. Works only for two dimensional data.

## Usage

```
choose.cluster(hmacobj,x=NULL,level=NULL,n.cluster=NULL)
```

## Arguments

| | |
|---|---|
| `hmacobj` | The output of HMAC analysis. An object of class 'hmac'. |
| `x` | The user-specified location. Deafult value is NULL in which case user chooses a point using the locator function. |
| `level` | The specified level |
| `n.cluster` | The specified number of clusters. Either `level` or `n.cluster` needs to be specified |

## Author(s)

Surajit Ray and Yansong Cheng

## References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

## See Also

`phmac` for front end of using modal clustering and also for parallel implementation of modal clustering.

## Examples

```
data(disc2d.hmac)
#disc2d.hmac is the output of phmac(disc2d,npart=1)

choose.cluster(disc2d.hmac,x=c(0,0),level=3)
choose.cluster(disc2d.hmac,x=c(0,0),n.cluster=2)

# Users can choose anypoint they want by clicking the point
# in the plot after the following command.
# choose.cluster(disc2d.hmac,level=3)
```

---

`contour.hmac` *Plot clusters with different colors for two dimensional data overlayed on the contours of the original data.*

---

## Description

Plot clusters for two dimensional data with contours of the original data

## Usage

```
## S3 method for class 'hmac'
contour(x, n.cluster=NULL,level=NULL,prob=NULL,smoothplot=FALSE,...)
```

## Arguments

| | |
|---|---|
| `x` | The output of HMAC analysis. An object of class 'hmac'. |
| `level` | The specified level |
| `n.cluster` | The specified number of clusters. Either `level` or `n.cluster` needs to be specified |
| `prob` | The specified level of the contour plot. Default value is NULL, plot all levels of the contour plot. Must be between 0 and 1 |
| `smoothplot` | Get the smooth scatter plot of the original data set. Default value is FALSE, which does not provide the smooth scatter plot. |
| `...` | Further arguments passed to or from other methods. |

## Author(s)

Surajit Ray and Yansong Cheng

## References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

## See Also

phmac for front end of using modal clustering and also for parallel implementation of modal clustering. soft.hmac for soft clustering at specified levels. hard.hmac for hard clustering at specified levels. See plot for plotting the whole dendrogram.

## Examples

```
data(disc2d.hmac)
# disc2d.hmac is the output of phmac(disc2d,npart=1)

contour.hmac(disc2d.hmac,level=3,col=gray(0.7))

# Provide contour line at probability density 0.05.
contour(disc2d.hmac,n.cluster=2,prob=0.05)

# Plot using smooth scatter plot.
contour.hmac(disc2d.hmac,n.cluster=2,smoothplot=TRUE)
```

---

cta20 *Two dimensional data in original and log scale*

---

## Description

Two dimensional data in original and log scale and their hierarchical modal clustering. This dataset demonstrates the fact that modal clustering techniques can be used to cluster untransformed data as it does not depend on parametric assumptions. The clustering results, before and after the log transformation both produce nice separation of the three clusters.

## Usage

```
data(cta20)
data(cta20.hmac)
data(logcta20)
data(logcta20.hmac)
```

## Format

cta20 and logcta20 are two dimensional matrices. cta20.hmac and logcta20.hmac are objects of class hmac obtained from applying phmac on cta20 and logcta20 respectively

## Details

The dataset is generated by illumina technology for high throughput genotyping named **GOLDEN GATE**. The data values are actual measurements made by the machine (intensity), after these are normalized (background subtracted etc). The data set is used for making genotype calls by Illumina. The data around X- and Y-axes represents the two homozygous genotypes (e.g. AA and TT), while the cluster along the 45-degree line represents the heterozygous (e.g. AT) genotype. Due to noisy reads, the data points often lie in-between the axes, and cluster detection is used for making automatic genotype calls.

### Author(s)

Surajit Ray and Yansong Cheng

### Examples

```
data(logcta20)
data(logcta20.hmac)
plot(logcta20)
plot(logcta20.hmac)
plot(logcta20.hmac,level=4)
```

---

## disc2d — *Two and three dimensional data representing two half discs*

---

### Description

Two and three dimensional data and their hierarchical modal clustering with 400 observations where the first two dimensions represent the shape of two discs.

### Usage

```
data(disc2d)
data(disc2d.hmac)
data(disc3d)
data(disc3d.hmac)
```

### Format

disc2d and disc3d are two and three dimensional matrices. disc2d.hmac and disc3d.hmac are objects of class hmac obtained from applying `phmac` on disc2d and disc3d respectively

### Details

Two dimensional data with 400 observations representing the shape of two half discs.

### Author(s)

Surajit Ray and Yansong Cheng

### References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

## Examples

```
data(disc2d)
plot(disc2d)
data(disc2d.hmac)
summary.hmac(disc2d.hmac)
hard.hmac(disc2d.hmac,n.clust=2)
```

---

`findmid` *Find the mid point of memberships of each cluster*

---

## Description

Find the mid point of memberships of each cluster. Sub function of `plot.hmac`.

## Usage

```
findmid(x,memb)
```

## Arguments

| | |
|---|---|
| `x` | Input data |
| `memb` | Membership of each observation |

## Author(s)

Surajit Ray and Yansong Cheng

## See Also

`plot` for plotting the dendrogram

---

`hard.hmac` *Plot clusters with different colors.*

---

## Description

Plot clusters with colors obtained from hard density. Plot one dimensional data with density plot. Plot two dimensional data with scatter plot. Pairwise scatter plot will be provided for data with more than two dimensions.

## Usage

```
hard.hmac(hmacobj,level=NULL, n.cluster=NULL,plot=TRUE,colors=1:6,...)
```

### Arguments

| | |
|---|---|
| `hmacobj` | The output of HMAC analysis. An object of class 'hmac'. |
| `level` | The specified level of HMAC output |
| `n.cluster` | The specified number of clusters. If neither `level` nor `n.cluster` is specified, hard clustering output is shown for each level. |
| `plot` | Get the plot of the clusters with different colors. Default value is TRUE, draws a plot on the current graphics device; plot=FALSE indicates do not get the plot and returns the membership of data. |
| `colors` | Colors used to represent different clusters. |
| `...` | Further graphical parameters |

### Value

Returns the membership of each observation of the specified level if plot=FALSE

### Author(s)

Surajit Ray and Yansong Cheng

### References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

### See Also

`phmac` for front end of using modal clustering and also for parallel implementation of modal clustering `soft.hmac` for soft clustering at specified levels. See `plot.hmac`.

### Examples

```
data(disc2d.hmac)
#disc2d.hmac is the output of phmac(disc2d,npart=1)

hard.hmac(disc2d.hmac,level=3)

#returns the membership of each observation
disc2d.2clus=hard.hmac(hmacobj=disc2d.hmac,n.cluster=2,plot=FALSE)
table(disc2d.2clus)

#hard.hmac(disc2d.hmac)

iris.hmac=phmac(iris[,-5])
# For more than two dimensions it produces the pairs plot
hard.hmac(iris.hmac,n.cluster=2)
```

hmac *Perform Modal Clustering in serial mode only*

## Description

Performs Modal Cluster with specified smoothing paramters. Used as a sub function of phmac.

## Usage

```
hmac(dat,Sigmas,G=NULL,member=NULL)
```

## Arguments

| | |
|---|---|
| dat | Matrix of data points |
| Sigmas | Specified smoothing levels |
| G | Specified values of modes. A matrix with number of rows equal to the number of modes and number of columns equal to the dimension of the data. Defualt value is NULL |
| member | Membership of the observations to the modes given in G. Default value is NULL |

## Value

| | |
|---|---|
| data | Same as the input dat. |
| n.cluster | Number of clusters at each level. |
| level | Levels corresponding to each smoothing parameter. |
| Sigmas | Same as input sigmas. |
| mode | List of modes at each distinct levels. |
| membership | List of memmbership to modes at each distinct levels. |

## Author(s)

Surajit Ray and Yansong Cheng

## References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

## See Also

phmac for front end of using modal clustering and also for parallel implementation of modal clustering.

khat.inv *Calculate the smoothing paramters for implementation of Modal Clustering.*

## Description

these set of functions are based on the concept of pseudo degrees of freedom (Lindsay et al 2008) and are used to calculate the Sigmas that are used for the 'hmac' function

## Usage

```
khat.inv(p,len=10)
sdofnorm(h,p)
khat(dof,p)
```

## Arguments

| | |
|---|---|
| len | Number of smoothing parameters. |
| h | Smoothing parameter |
| p | Number of column of data |
| dof | Degrees of freedom |

## Author(s)

Surajit Ray

## References

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

## See Also

phmac for front end of using HMAC and also for parallel implementation of modal clustering.

mydmvnorm *Calculate Density of Multivariate Normal for diagonal covariance*

## Description

Faster calculation of density of multivariate normal with diagonal covariance matrix

## Usage

```
mydmvnorm(x, mean, sigmasq)
```

### Arguments

| | |
|---|---|
| x | The input data |
| mean | The vector of mean values |
| sigmasq | The variance of each dimension. Assume the variance are the same for all dimensions. |

### Author(s)

Surajit Ray and Yansong Cheng

---

## oned — *One dimensional data with two main clusters*

---

### Description

A one dimensional data and its hierarchical modal clustering with 2 main clusters

### Usage

```
data(oned)
data(oned.hmac)
```

### Format

oned is a one dimensional data with 2 main clusters and several subclusters. oned.hmac is an object of class 'hmac' obtained from applying phmac on disc2d and disc3d respectively

### Author(s)

Surajit Ray and Yansong Cheng

### References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

### Examples

```
data(oned)
hist(oned,col="lavender",n=15)
data(oned.hmac)
plot(oned.hmac)
plot(oned.hmac,n.clust=2)
```

phmac *Main function for performing Modal Clusters either parallel or serial mode.*

## Description

Performing Modal Clustering

## Usage

```
phmac(dat, length = 10, npart = 1, parallel = TRUE, sigmaselect = NULL,
G= NULL)
modalclust(dat, length = 10, npart = 1, parallel = TRUE, sigmaselect = NULL,
G= NULL)
```

## Arguments

| | |
|---|---|
| `dat` | Matrix of data points |
| `length` | number of smoothing levels. Default is 10 |
| `sigmaselect` | Specified Smoothing levels. Default NULL will calculate the Sigma levels using concept of spectral degrees of freedom given in Lindsay et al (2008) |
| `npart` | Number of random partitions when using parallel computing. If using several processors of a machine one option is to choose the number of partitions equal to the number of processors |
| `parallel` | If TRUE uses parallel comptation using `npart` processors. Requires the package `multicore` to perform parallel computing |
| `G` | Specified values of modes. A matrix with number or rows equal to the number of modes and number of columns equal to the dimension of the data. Defualt value is NULL |

## Value

| | |
|---|---|
| `data` | Same as the input Data |
| `n.cluster` | Number of clusters at each level. |
| `level` | Levels corresponding to each smoothing parameter. |
| `sigmas` | Same as input sigmaselect if provided or dynamically calculated smoothing levels based on Spectral Degrees of Freedom criterion. Uses the function khat.inv |
| `mode` | List of modes at each distinct levels. |
| `membership` | List of memmbership to modes at each distinct levels. |

## Author(s)

Surajit Ray and Yansong Cheng

## References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

## See Also

`soft.hmac` for soft clustering at specified levels. `hard.hmac` for hard clustering at specified levels. See `plot.hmac`.

## Examples

```
data(disc2d)
## Not run: disc2d.hmac=phmac(disc2d,npart=1)

plot.hmac(disc2d.hmac,level=2)


## For parallel implementation
## Not run: disc2d.hmac.parallel=phmac(disc2d,npart=2,parallel=TRUE)

soft.hmac(disc2d.hmac,level=2)
soft.hmac(disc2d.hmac,n.cluster=3)

hard.hmac(disc2d.hmac,n.cluster=3)
```

---

`plot.hmac` *Plots of heierarchical tree for a 'hmac' object*

---

## Description

Plots the dendrogram of the entire heierarchical tree for a 'hmac' object starting from any specified smoothing level.

## Usage

```
## S3 method for class 'hmac'
plot(x,mycol=1:6,level=1,n.cluster=NULL,userclus=NULL,sep=.1,...)
```

## Arguments

| | |
|---|---|
| `x` | The output of HMAC analysis. An object of class 'hmac'. |
| `mycol` | Colors used to represent different clusters. |
| `level` | The specified level that dendrogram starts. Default value is 1. |
| `n.cluster` | The specified number of clusters. If neither `level` nor `n.cluster` is specified, the full tree is plotted. |

| | |
|---|---|
| `userclus` | If user provides membership, the tree colors the node according to this membership and the tree can be used for validation. |
| `sep` | It provides the distance between the lowest layer of nodes of the clusters. |
| `...` | further arguments passed to or from other methods. |

## Author(s)

Surajit Ray and Yansong Cheng

## References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

## See Also

`phmac` for front end of using modal clustering and also for parallel implementation of modal clustering. `hard.hmac` for hard clustering at specified levels. `soft.hmac` for soft clustering at specified levels.

## Examples

```
data(disc2d.hmac)
# disc2d.hmac is the output of phmac(disc2d,npart=1)
plot(disc2d.hmac)

set.seed(20)
mix4=data.frame(rbind(rmvnorm(20,rep(0,4)), rmvnorm(20,rep(2,4)),
               rmvnorm(20,rep(10,4)),rmvnorm(20,rep(13,4))))
mix4.hmac=phmac(mix4,npart=1)
plot(mix4.hmac,col=1:6)

# Verifying with user provided groups
plot(mix4.hmac,userclus=rep(c(1,2,3,4),each=20),col=1:6)
```

---

`soft.hmac` *Plot soft clusters from Modal Clustering output*

---

## Description

Plot clusters for two dimensional data with colors representing the posterior probability of belonging to clusters. Additionally boundary points between the clusters, with specified thresholds are also

## Usage

```
soft.hmac(hmacobj,n.cluster=NULL,level=NULL,boundlevel=0.4,plot=TRUE)
```

### Arguments

| | |
|---|---|
| `hmacobj` | The output of HMAC analysis. An object of class 'hmac'. |
| `level` | The specified level of HMAC output |
| `n.cluster` | The specified number of clusters. If neither `level` nor `n.cluster` is specified, soft clustering output is shown for each level. |
| `boundlevel` | Posterior probability threshold. Points having posterior probability below `boundlevel` are assigned as boundary points and colored in gray. Default value is 0.4. |
| `plot` | Get the two dimensional plot of the clusters with different colors. Default value is TRUE, which returns the two dimensional plot on the current graphics device; plot=FALSE returns the posterior probability of each observation. |

### Value

Returns the list that contains the posterior probability of each observation and boundary points at specified level if plot=FALSE

### Author(s)

Surajit Ray and Yansong Cheng

### References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

### See Also

`phmac` for front end of using modal clustering and also for parallel implementation of modal clustering `hard.hmac` for hard clustering at specified levels.

### Examples

```
data(logcta20.hmac)
#logcta20.hmac is the output of phmac(logcta20,npart=1)

soft.hmac(logcta20.hmac,n.cluster=3)

#return the posterior probability of each observation and boundary points.
postprob=soft.hmac(hmacobj=logcta20.hmac,n.cluster=3,plot=FALSE)
```

summary *Summary of HMAC output*

## Description

Gives the summary of output of a 'hmac' object.

## Usage

```
## S3 method for class 'hmac'
summary(object,...)
```

## Arguments

| | |
|---|---|
| `object` | The output of HMAC analysis. An object of class 'hmac'. |
| `...` | further arguments passed to or from other methods. |

## Author(s)

Surajit Ray and Yansong Cheng

## References

Li. J, Ray. S, Lindsay. B. G, "A nonparametric statistical approach to clustering via mode identification," Journal of Machine Learning Research , 8(8):1687-1723, 2007.

Lindsay, B.G., Markatou M., Ray, S., Yang, K., Chen, S.C. "Quadratic distances on probabilities: the foundations," The Annals of Statistics Vol. 36, No. 2, page 983–1006, 2008.

## See Also

`phmac` for front end of using modal clustering and also for parallel implementation of modal clustering.

## Examples

```
data(disc2d.hmac)
summary(disc2d.hmac)
```

# Index


∗ **cluster, hierarchical, nested, modal**
- choose.cluster, 2
- contour.hmac, 3
- hard.hmac, 6
- hmac, 8
- phmac, 11
- plot.hmac, 12
- soft.hmac, 13
- summary, 15

∗ **data**
- cta20, 4
- disc2d, 5
- oned, 10

choose.cluster, 2
contour *(*contour.hmac*)*, 3
contour.hmac, 3
cta20, 4

disc2d, 5
disc3d *(*disc2d*)*, 5
dmvnorm *(*mydmvnorm*)*, 9

findmid, 6

hard.hmac, *4*, 6, *12–14*
hmac, 8

khat *(*khat.inv*)*, 9
khat.inv, 9

logcta20 *(*cta20*)*, 4

modalclust *(*phmac*)*, 11
mydmvnorm, 9

oned, 10

phmac, *2*, *4*, *7–9*, 11, *13–15*
plot, *4*, *6*
plot *(*plot.hmac*)*, 12
plot.hmac, *6*, *7*, *12*, 12

sdofnorm *(*khat.inv*)*, 9
soft.hmac, *4*, *7*, *12*, *13*, 13
summary, 15