

# Package ‘kpodclustr’

October 13, 2022

**Title** Method for Clustering Partially Observed Data

**Version** 1.1

**Description** Software for k-means clustering of partially observed data from Chi, Chi, and Baraniuk (2016) <[doi:10.1080/00031305.2015.1086685](https://doi.org/10.1080/00031305.2015.1086685)>.

**URL** <http://jocelynchi.com/kpodclustr>

**Depends** R (>= 3.1.0)

**License** MIT + file LICENSE

**LazyData** true

**RoxygenNote** 7.1.0

**Encoding** UTF-8

**NeedsCompilation** no

**Author** Jocelyn T. Chi [aut, cre],  
Eric C. Chi [aut, ctb],  
Richard G. Baraniuk [aut]

**Maintainer** Jocelyn T. Chi <[jtchi@ncsu.edu](mailto:jtchi@ncsu.edu)>

**Repository** CRAN

**Date/Publication** 2020-06-24 09:10:06 UTC

## R topics documented:

assign_clustpp . . . . .	2
findMissing . . . . .	2
initialImpute . . . . .	3
kmpp . . . . .	4
kpod . . . . .	5
makeData . . . . .	6
<b>Index</b>	<b>7</b>

---

assign\_clustpp      *Function for assigning clusters to rows in a matrix*

---

**Description**

assign\_clustpp Function for assigning clusters to rows in a matrix

**Usage**

```
assign_clustpp(X, init_centers, kmpp_flag = TRUE, max_iter = 20)
```

**Arguments**

X	Data matrix containing missing entries whose rows are observations and columns are features
init_centers	Centers for initializing k-means
kmpp_flag	(Optional) Indicator for whether or not to initialize with k-means++
max_iter	(Optional) Maximum number of iterations

**Author(s)**

Jocelyn T. Chi

**Examples**

```
p <- 2
n <- 100
k <- 3
sigma <- 0.25
missing <- 0.05
Data <- makeData(p,n,k,sigma,missing)
X <- Data$Missing
Orig <- Data$Orig

clusts <- assign_clustpp(Orig, k)
```

---

findMissing      *Function for finding indices of missing data in a matrix*

---

**Description**

findMissing Function for finding indices of missing data in a matrix

**Usage**

```
findMissing(X)
```

**Arguments**

X                    Data matrix containing missing entries whose rows are observations and columns are features

**Value**

A numeric vector containing indices of the missing entries in X

**Author(s)**

Jocelyn T. Chi

**Examples**

```
p <- 2
n <- 100
k <- 3
sigma <- 0.25
missing <- 0.05
Data <- makeData(p,n,k,sigma,missing)
X <- Data$Missing
missing <- findMissing(X)
```

---

*initialImpute*

*Function for initial imputation for k-means*

---

**Description**

*initialImpute* Initial imputation for k-means

**Usage**

```
initialImpute(X)
```

**Arguments**

X                    Data matrix containing missing entries whose rows are observations and columns are features

**Value**

A data matrix containing no missing entries

**Author(s)**

Jocelyn T. Chi

**Examples**

```
p <- 2
n <- 100
k <- 3
sigma <- 0.25
missing <- 0.05
Data <- makeData(p,n,k,sigma,missing)
X <- Data$Missing
X_copy <- initialImpute(X)
```

---

kmpp

*k-means++*

---

**Description**

kmpp Computes initial centroids via kmeans++

**Usage**

```
kmpp(X, k)
```

**Arguments**

X                    Data matrix whose rows are observations and columns are features  
k                    Number of clusters.

**Value**

A data matrix whose rows contain initial centroids for the k clusters

**Examples**

```
n <- 10
p <- 2
X <- matrix(rnorm(n*p),n,p)
k <- 3
kmpp(X,k)
```

---

kpod *Function for performing k-POD*

---

**Description**

kpod Function for performing k-POD, a method for k-means clustering on partially observed data

**Usage**

```
kpod(X, k, kmpp_flag = TRUE, maxiter = 100)
```

**Arguments**

X	Data matrix containing missing entries whose rows are observations and columns are features
k	Number of clusters
kmpp_flag	(Optional) Indicator for whether or not to initialize with k-means++
maxiter	(Optional) Maximum number of iterations

**Value**

cluster: Clustering assignment obtained with k-POD  
cluster\_list: List containing clustering assignments obtained in each iteration  
obj\_vals: List containing the k-means objective function in each iteration  
fit: Fit of clustering assignment obtained with k-POD (calculated as 1-(total withinss/totss))  
fit\_list: List containing fit of clustering assignment obtained in each iteration

**Author(s)**

Jocelyn T. Chi

**Examples**

```
p <- 5
n <- 200
k <- 3
sigma <- 0.15
missing <- 0.20
Data <- makeData(p,n,k,sigma,missing)
X <- Data$Missing
Orig <- Data$Orig
truth <- Data$truth

kpod_result <- kpod(X,k)
kpodclusters <- kpod_result$cluster
```

---

`makeData`*Make test data*

---

**Description**

`makeData` Function for making test data

**Usage**

```
makeData(p, n, k, sigma, missing, seed = 12345)
```

**Arguments**

<code>p</code>	Number of features (or variables)
<code>n</code>	Number of observations
<code>k</code>	Number of clusters
<code>sigma</code>	Variance
<code>missing</code>	Desired missingness percentage
<code>seed</code>	(Optional) Seed (default seed is 12345)

**Author(s)**

Jocelyn T. Chi

**Examples**

```
p <- 2
n <- 100
k <- 3
sigma <- 0.25
missing <- 0.05

X <- makeData(p,n,k,sigma,missing)$Orig
```

# Index

`assign_clustpp`, 2

`findMissing`, 2

`initialImpute`, 3

`kmpp`, 4

`kpod`, 5

`makeData`, 6