

# Package ‘rcppmlpackexamples’

May 9, 2026

**Type** Package

**Title** Example Use of 'mlpack' from C++ via R

**Version** 0.0.1

**Date** 2025-09-14

**Description** A Minimal Example Package which demonstrates 'mlpack' use via C++ Code from R.

**URL** <https://github.com/eddelbuettel/rcppmlpack-examples>

**BugReports** <https://github.com/eddelbuettel/rcppmlpack-examples/issues>

**License** GPL (>= 2)

**Suggests** tinytest

**Depends** R (>= 3.5.0)

**Imports** Rcpp (>= 1.1.0)

**LinkingTo** Rcpp, RcppArmadillo (>= 15.0.2-1), RcppEnsmallen, mlpack (>= 4.6.3)

**Encoding** UTF-8

**RoxygenNote** 7.3.3

**NeedsCompilation** yes

**Author** Dirk Eddelbuettel [aut, cre],  
Authors of mlpack [aut],  
Constantinos Giachalis [ctb]

**Maintainer** Dirk Eddelbuettel <edd@debian.org>

**Repository** CRAN

**Date/Publication** 2025-09-21 13:30:02 UTC

## Contents

rcppmlpackexamples-package . . . . .	2
coervtype_small . . . . .	2
kMeans . . . . .	3
linearRegression . . . . .	4

loanData . . . . .	5
loanDefaultPrediction . . . . .	5
randomForest . . . . .	6

<b>Index</b>	<b>8</b>
--------------	----------

---

rcppmlpackexamples-package

*Example Use of 'mlpack' from C++ via R*

---

## Description

A Minimal Example Package which demonstrates 'mlpack' use via C++ Code from R.

## Package Content

Index of help topics:

covertime_small	Covertime data subset used for classification
kMeans	Run a k-means clustering analysis
linearRegression	Run a linear regression with optional ridge regression
loanData	Loan data subset used for default prediction
loanDefaultPrediction	loanDefaultPrediction
randomForest	Run a Random Forest classificatio
rcppmlpackexamples-package	Example Use of 'mlpack' from C++ via R

## Maintainer

Dirk Eddelbuettel <edd@debian.org>

## Author(s)

Dirk Eddelbuettel [aut, cre], Authors of mlpack [aut], Constantinos Giachalis [ctb]

---

covertime_small	<i>Covertime data subset used for classification</i>
-----------------	--

---

## Description

A subset of the UCI machine learning data set 'covertime' describing cloud coverage in seven different states of coverage. This smaller subset contains with 100,000 observations and 55 variables. The first 54 variables are explanatory (i.e. "features"), with the last providing the dependent variable ("labels"). The data is in the 'wide' 55 x 100,000 format used by **mlpack**. The dependent variable has been transformed to the range zero to six by subtracting one from the values found in the data file.

**Details**

The original source of the data is the US Forest Service, and the complete file is part of the UC Irvine machine learning data repository.

**Source**

<https://www.mlpack.org/datasets/covertypes-small.csv.gz>

**References**

<https://archive.ics.uci.edu/dataset/31/covertypes>

---

kMeans

*Run a k-means clustering analysis*

---

**Description**

Run a k-means clustering analysis, returning a list of cluster assignments

**Usage**

```
kMeans(data, clusters)
```

**Arguments**

data	A matrix of data values
clusters	An integer specifying the number of clusters

**Details**

This function performs a k-means clustering analysis on the given data set.

**Value**

A list with cluster assignments

**Examples**

```
x <- rbind(matrix(rnorm(100, sd = 0.3), ncol = 2),
            matrix(rnorm(100, mean = 1, sd = 0.3), ncol = 2))
colnames(x) <- c("x", "y")
c1 <- kMeans(x, 2)

data(trees, package="datasets")
c12 <- kMeans(t(trees), 3)
```

---

linearRegression	<i>Run a linear regression with optional ridge regression</i>
------------------	---

---

### Description

Run a linear regression (with optional ridge regression)

### Usage

```
linearRegression(matX, vecY, lambda = 0, intercept = TRUE)
```

### Arguments

matX	A matrix of explanatory variables ('predictors') in standard R format (i.e. 'tall and skinny' to be transposed internally to MLPACK format (i.e. 'short and wide').
vecY	A vector of dependent variables ('responses')
lambda	An optional ridge parameter, defaults to zero
intercept	An optional boolean switch about an intercept, default is true.

### Details

This function performs a linear regression, and serves as a simple test case for accessing an MLPACK function.

### Value

A vector with fitted values

### Examples

```
suppressMessages(library(utils))
data("trees", package="datasets")
X <- with(trees, cbind(log(Girth), log(Height)))
y <- with(trees, log(Volume))
lmfit <- lm(y ~ X)
# summary(fitted(lmfit))
mlfit <- linearRegression(X, y)
# summary(mlfit)
all.equal(unnamed(fitted(lmfit)), as.vector(mlfit))
```

---

loanData	<i>Loan data subset used for default prediction</i>
----------	---

---

**Description**

A four column data set containing a binary variable 'Employed' (with zero denoting unemployment and one employment), a numeric variable 'Bank Balance', a numeric variable 'Annual Salary' and a binary target variable 'Defaulted?' (with zero denoting loan repayment and one denoting default).

**Details**

The original source of the data is not documented by mlpack.

**Source**

<https://datasets.mlpack.org/LoanDefault.csv>

**References**

<https://archive.ics.uci.edu/dataset/31/covertime>

---

loanDefaultPrediction	<i>loanDefaultPrediction</i>
-----------------------	------------------------------

---

**Description**

Predict loan default using a decision tree model

**Usage**

```
loanDefaultPrediction(loanDataFeatures, loanDataTargets, pct = 0.25)
```

**Arguments**

loanDataFeatures	A matrix of dimension 3 by N, i.e. transposed relative to what R uses, with the three explanatory variables
loanDataTargets	A vector of (integer-valued) binary variables loan repayment or default
pct	A numeric variable with the percentage of data to be used for testing, defaults to 25%

**Details**

This functions performs a loan default prediction based on three variables on employment, bank balance and annual salary to predict loan repayment or default

**Value**

A list object with predictions, probabilities, accuracy and a report matrix

**Examples**

```
data(loanData)
res <- loanDefaultPrediction(t(as.matrix(loanData[,-4])), # col 1 to 3, transposed
                             loanData[, 4],           # col 4 is the target
                             0.25)                   # retain 25% for testing

str(res)
res$report
```

---

randomForest

*Run a Random Forest classificatio*

---

**Description**

Run a Random Forest Classifier

**Usage**

```
randomForest(dataset, labels, pct = 0.3, nclasses = 7L, ntrees = 10L)
```

**Arguments**

dataset	A matrix of explanatory variables, i.e. “features”
labels	A vector of the dependent variable as integer values, i.e. “labels”
pct	A numeric value for the percentage of data to be retained for the test set
nclasses	An integer value for the number of a distinct values in labels
ntrees	An integer value for the number of trees

**Details**

This function performs a Random Forest classification on a subset of the standard ‘covertime’ data set

**Value**

A list object

**See Also**

covertime\_small

**Examples**

```
data(covertime_small) # see help(covertime_small)
res <- randomForest(covertime_small[-55,], # features (already transposed)
                   covertime_small[55,], # labels now in [0, 6] range
                   0.3) # percentage used for testing
str(res) # accuracy varies as method is randomized but not seed set here
```

# Index

## \* **data**

covertime\_small, [2](#)

loanData, [5](#)

## \* **package**

rcppmlpackexamples-package, [2](#)

covertime\_small, [2](#)

kMeans, [3](#)

linearRegression, [4](#)

loanData, [5](#)

loanDefaultPrediction, [5](#)

randomForest, [6](#)

rcppmlpackexamples

(rcppmlpackexamples-package), [2](#)

rcppmlpackexamples-package, [2](#)