# Package 'glmNETpkg'

May 13, 2015

**Type** Package

**Title** Create/predict glmnet models

**Version** 1.0

**Date** 2015-04-23

**Author** Georgia Tsiliki

**Maintainer** Georgia Tsiliki <gtsiliki@central.ntua.gr>

**Depends** RCurl, pmml, glmnet, jsonlite

**Description** This package employs/predicts elastic net models, i.e. lasso, ridge and elastic net regressions. The output is saved in pmml format.

**License** GPL-2

## R topics documented:



---

`glmNETpkg-package` *Reads in json files, create/predict glmnet, save in serialized and pmml format*

---

## Description

This package wrapps up glmnet R package for fitting the lasso or elastic-net regularization path for linear regression, logistic and multinomial regression models, Poisson regression, Cox model, multiple-response Gaussian, and the grouped multinomial. Models are saved in pmml an dserialized forms.

## Details

| | |
|---|---|
| Package: | glmNETpkg |
| Type: | Package |
| Version: | 1.0 |
| Date: | 2015-04-23 |
| License: | GPL-2 |

## Author(s)

Georgia Tsiliki

Maintainer: Georgia Tsiliki <gtsiliki@central.ntua.gr>

## References

The glmnet package help file

## Examples

```
data("dat1p")
data("dat1m")
data("dat1i")

data.file<- read.in.json.for.pred(dat1p, dat1m, dat1i)
```

---

`dat1` *A sample data object*

---

## Description

The dataset for this test is a data frame

## Usage

```
data("dat1")
```

## Format

A list of two objects

**datasetURI** a character vector- ambit data set uri

**dataEntry** a data frame containing two columns: compound and values. Compound is a character vector with all compound anbit uris, and calues is a data frame with all numberic values of the data set (compounds by features)

## Details

There are no more details

## Source

The source of this function is in the

## References

There are no references

## Examples

```
data(dat1)
## maybe str(dat1) ; plot(dat1) ...
```

---

`dat1i` *Information for glmnet.funct*

---

## Description

A list with information for glmNet and prediction

## Usage

```
data("dat1i")
```

## Format

A list with 3 objects:

`indepFeatures.mat` A matrix with two columns showing the independent features names in the function and in reallity (ambit name)

`models.coef` A data.frame as returned by cv.glmnet

`parameters` A list with 3 objects:alpha (The elasticnet mixing parameter, with 0<=alpha<=1), family (as in glmnet one of c("gaussian","binomial","poisson","multinomial","cox","mgaussian")) and s (Value(s) of the penalty parameter lambda at which predictions are required. One in c("lambda.1se","lambda.min"))

## Details

Example dataset to predict with dat1p, dat1m

## Source

The source of this function is in the

## References

There are no references

## Examples

```
data(dat1i)
## maybe str(dat1i) ; plot(dat1i) ...
```

---

# dat1m *Serialized glmnet model file*

---

## Description

A character string for a serialized glmnet model

## Usage

```
data("dat1m")
```

## Format

A character string

## Details

Example glmnet model based on dat1

## Source

The source of this function is in the

## References

There are no references

## Examples

```
data(dat1m)
## maybe str(dat1m) ; plot(dat1m) ...
```

dat1p *A sample data object*

## Description

The dataset for this test is a data frame

## Usage

```
data("dat1p")
```

## Format

A list of two objects

**datasetURI** a character vector- ambit data set uri

**dataEntry** a data frame containing two columns: compound and values. Compound is a character vector with all compound anbit uris, and calues is a data frame with all numberic values of the data set (compounds by features)

## Details

Data set for prediction with dat1m

## Source

The source of this function is in the

## References

There are no references

## Examples

```
data(dat1p)
## maybe str(dat1p) ; plot(dat1p) ...
```

`glmnet.funct` *GLMnet model*

## Description

Does k-fold cross-validation for glmnet, returns a value for lambda using cv.glmnet function. produces a serialized and pmml model.

## Usage

```
glmnet.funct(dataset, predictionFeature, parameters)
```

## Arguments

| | |
|---|---|
| `dataset` | list of 2 objects, datasetURI:= character sring, code name of dataset, dataEntry:= data frame with 2 columns |
| `predictionFeature` | character string specifying which is the prediction feature in dataEntry |
| `parameters` | A list with two objects: alpha (the elasticnet mixing parameter, with 0<=alpha<=1 and alpha=1(lasso), alpha=0.5 (elastic net), alpha=0 (ridge)), family (as in glmnet one of c("gaussian","binomial","poisson","multinomial","cox","mgaussian")) |

## Details

From glmnet package: The function runs glmnet 10+1 times; the first to get the lambda sequence, and then the remainder to compute the fit with each of the folds omitted. The error is accumulated, and the average error and standard deviation over the folds is computed. Note that cv.glmnet does NOT search for values for alpha.

## Value

A List

| | |
|---|---|
| `rawModel` | A serialized cv.glmnet object (class raw) |
| `pmmlModel` | A pmml cv.glmnet object |
| `predictedFeatures` | A character string specifying the name of the predicted feature |
| `independentFeatures` | A list with all independent features included in the glmnet model |
| `additionalInfo` | A list with 3 objects: indepFeatures.mat(A matrix with two columns showing the independent features names in the function and in reallity (ambit name)), models.coef(A data.frame as returned by cv.glmnet with coefficients), parameters (A list with 2 objects: alpha (The elasticnet mixing parameter, with 0<=alpha<=1), family (as in glmnet one of c("gaussian","binomial","poisson","multinomial","cox","mgaussian"))) |

## Note

No notes for this function

## Author(s)

Georgia Tsiliki

## References

The help file glmnet

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

data("dat1")

x1<- "https://apps.ideaconsult.net/ambit2/feature/22137"


x2<- list(alpha=1,family="gaussian")

lasso.mod<- glmnet.funct(dat1,x1,x2)
```

---

`pred.funct` *cv.glmnet prediction*

---

## Description

predict.glmnet function expecting list from fromJSON

## Usage

```
pred.funct(dataset, rawModel, additionalInfo)
```

## Arguments

| | |
|---|---|
| `dataset` | Data for prediction. A list of two objects: datasetURI (a character string ), dataEntry (a data frame). |
| `rawModel` | R model serialized. |
| `additionalInfo` | A list with 3 objects: indepFeatures.mat (A matrix with two columns showing the independent features names in the function and in reallity (ambit name)), models.coef (A data.frame as returned by cv.glmnet), parameters (A list with 3 objects:alpha (The elasticnet mixing parameter, with 0<=alpha<=1), family (as in glmnet one of c("gaussian","binomial","poisson","multinomial","cox","mgaussian")) |

and s (Value(s) of the penalty parameter lambda at which predictions are required. One in c("lambda.1se","lambda.min")))

## Details

The function returns an error if the supplied dataset does not include all the independent features in the raw model.

## Value

A data.frame with prediction values

## Note

No notes for this function

## Author(s)

Georgia Tsiliki

## References

The glmnet package help file

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

data("dat1p")
data("dat1m")
data("dat1i")

pred.res<- pred.funct(dat1p, dat1m, dat1i)
```

---

# read.in.json *Read in function for json files*

---

## Description

This function reads in a json data file and produces a list with y and x variables

## Usage

```
read.in.json(dataset, predictionFeature, parameters)
```

## Arguments

| | |
|---|---|
| dataset | list of 2 objects, datasetURI:= character sring, code name of dataset, dataEntry:= data frame with 2 columns |
| predictionFeature | character string specifying which is the prediction feature in dataEntry |
| parameters | A list with two objects: alpha (the elasticnet mixing parameter, with 0<=alpha<=1), family (as in glmnet one of c("gaussian","binomial","poisson","multinomial","cox","mgaussian" |

## Details

json file includes uris, data, indicator of y variable, model's parameters list

## Value

A list is returned

| | |
|---|---|
| x.mat | data frame with independent variables |
| y.mat | singular data frame with dependent variable |
| data.names | names of independent features |
| par.list | list of parameters |

## Note

No notes for this function

## Author(s)

Georgia Tsiliki

## References

No references for this function

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

data("dat1")

x1<- "https://apps.ideaconsult.net/ambit2/feature/22137"

x2<- list(alpha=1,family="gaussian")

r.dat1<- read.in.json(dat1,x1,x2)
```

`read.in.json.for.pred` *Read in function for json files for prediction*

## Description

This function reads in a json data file and produces a list with y and x variables

## Usage

```
read.in.json.for.pred(dataset, rawModel, additionalInfo)
```

## Arguments

| | |
|---|---|
| `dataset` | Data for prediction. A list of two objects: datasetURI (a character string ), dataEntry (a data frame). |
| `rawModel` | R model serialized. |
| `additionalInfo` | A list with 3 objects: indepFeatures.mat (A matrix with two columns showing the independent features names in the function and in reallity (ambit name)), models.coef (A data.frame as returned by cv.glmnet), parameters (A list with 3 objects:alpha (The elasticnet mixing parameter, with 0<=alpha<=1), family (as in glmnet one of c("gaussian","binomial","poisson","multinomial","cox","mgaussian")) and s (Value(s) of the penalty parameter lambda at which predictions are required. One in c("lambda.1se","lambda.min"))) |

## Details

The function returns an error if the supplied dataset does not include all the independent features in the raw model.

## Value

A List including:

| | |
|---|---|
| `x.mat` | data frame with independent variables |
| `model` | R cv.glmnet model |
| `additionalInfo` | A list with 4 objects: depFeature (a character string for the dependent feature in the data set (ambit name)), indepFeatures.mat(A matrix with two columns showing the independent features names in the function and in reallity (ambit name)), models.coef(A data.frame as returned by cv.glmnet with coefficients), parameters (A list with 3 objects:alpha (The elasticnet mixing parameter, with 0<=alpha<=1), family (as in glmnet one of c("gaussian","binomial","poisson","multinomial","cox","mgaussian and s (Value(s) of the penalty parameter lambda at which predictions are required. One in c("lambda.1se","lambda.min"))) |

## Note

No notes for this function

## Author(s)

Georgia Tsiliki

## References

No reference for this function

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

data("dat1p")
data("dat1m")
data("dat1i")

data.file<- read.in.json.for.pred(dat1p, dat1m, dat1i)
```

# Index