# Package 'iEN'

September 23, 2020

**Title** immunological Elastic-Net model, implementation and optimization

**Version** 0.99.0

**Date** 2019-010-25

**Author** Anthony Culos <tculos@stanford.edu>, Nima Aghaeepour <naghaeep@gmail.com>

**Maintainer** Anthony Culos <tculos@stanford.edu>

**Imports** pROC, Metrics, Matrix, glmnet, knitr, methods, stats, parallel

**Suggests** RUnit, BiocGenerics

**Description** Implements the immunological Elastic-Net model and provides scripts for their performance estimation and optimization via K-fold cross-validation.

**biocViews** Regression, FlowCytometry, CellBiology

**Depends** R (>= 3.5.2)

**License** Artistic-2.0

**Encoding** UTF-8

**RoxygenNote** 6.1.1

**Collate** 'cv_iEN.R' 'iEN.R' 'predict_iEN.R' 'print_iEN.R'

## R topics documented:

---

cv_iEN                          *Optimizes an iEN model via K-fold cross validation gridsearch and*
                                *returns out-of-sample predictions and the associated model meta data.*

---

**Description**

Optimizes an iEN model via K-fold cross validation gridsearch and returns out-of-sample predictions and the associated model meta data.

**Usage**

```
cv_iEN(X, Y, foldid, alphaGrid, phiGrid, nlambda = 100, lambdas = NULL,
    priors, ncores, eval = c("RMSE", "RSS", "wilcox", "ROCAUC", "spearman",
    "pearson"), family = c("binomial", "gaussian"), intercept = TRUE,
    standardize = TRUE, center = TRUE)
```

**Arguments**

| | |
|---|---|
| X | Input matrix of dimensions nobs x nfeat where each row is an observation vector. |
| Y | Response variable. Is continuous vector for family = "gaussian" and categorical (with two levels) for family = "binomial". |
| foldid | Vector that identifies which observations belong to which fold during K-fold Cross-Validation. foldid must consist of at least three folds for optimization and model estimation to occur. |
| alphaGrid | Vector of alpha values for model optimization. |
| phiGrid | Vector of phi values for model optimization. |
| nlambda | Lambda values are generated dynamically during cross-validation to avoid any data leak. nlambda determines the number of lambda values to generate. |
| lambdas | Optional vector of static lambda values. |
| priors | Continuous values which indicates immune features (columns of X) that are consistent with known biology. Values vary between 0 (low consistency) to 1 (highly consistent) for each immune feature which create the column space of X. |
| ncores | Number of cores to use during parallel computing of iEN cross-validation results. For optimal use set ncores = length(alphaGrid) * length(phiGrid). |
| eval | For binomial models evaluations using Wilcoxon P-value and ROCAUC are provided whereas for Gaussian models RMSE, RSS, Pearson P-value, and Spearman P-value are available. |
| family | Type of regression model, currently only "Binomial" and "Gaussian" are supported |
| intercept | Indicator for inclusion of regresstion intercept (default=TRUE). |
| standardize | Indication for X variable standardization prior to model fitting (default=TRUE). |
| center | Indication for X variable centering during scaling (default=TRUE). |

## Value

An object of class "iEN" is returned, which is a class composed of results from the K-fold cross validation and meta data about the analysis. The returned information includes:

Out-of-sample predictions from the K-fold cross validation. Evaluation of the out-of-sample predictions as defined by the eval parameter. Coefficients for each out-of-sample regression model, betas. the optimal parameters (alpha, lambda, phi) calculated for each fold of the analysis.

## Examples

```
data(test_data)

alphaGrid <- seq(0,1, length.out=2)
phiGrid <- exp(seq(log(1),log(10), length.out=2))
nlambda <- 3
ncores <- 2
eval <- "RSS"
family <- "gaussian"
intercept <- TRUE
standardize <- TRUE
center <- TRUE

model <- cv_iEN(X, Y, foldid, alphaGrid, phiGrid, nlambda, NULL, priors, ncores, eval, family, intercept, standardi
```

---

iEN-class                    *An S4 class representative of the immunological Elastic Net model*

---

## Description

An S4 class representative of the immunological Elastic Net model

## Slots

cv.preds  out-of-sample predictions made during k-Fold cross validation.

cv.eval  out-of-sample evaluation of predictive analysis, as defined by cv.kEN parameter eval, this estimates the models ability at predicting unseen data.

betas  matrix of model coefficients for each optimized out-of-sample model of the K-fold cross validation.

dfs  vector of values corresponding to the number of non-zero elements in each out-of-sample model

nobs  number of observations the provided dataset contained.

nfeat  number of features the provided dataset contained.

opt.params  optimal parameters (alpha, phi, lambda) for each out-of-sample model generated from the initial search parameters provided.

call  function call as string from the cv.kEN call which generated this particular object.

---

predict,iEN-method        *predicts with a iEN object, defualt of the model is to predict the new*
                          *data with the mean coefficient model over each out of sample instance.*
                          *With the default phi applied to to newx being the mean phi from opti-*
                          *mal models found during CV.*

---

### Description

predicts with a iEN object, defualt of the model is to predict the new data with the mean coefficient
model over each out of sample instance. With the default phi applied to to newx being the mean phi
from optimal models found during CV.

### Usage

```
## S4 method for signature 'iEN'
predict(object, newx, priors, phi)
```

### Arguments

object            object of type iEN

newx              new data to predict

priors            priors indicating how phi will be applied, should be identical to the priors used
                  during cross-validation.

phi               The amount of prioritization to use on newx, this should be the same phi as is
                  used to construct the iEN model during prediction. Default for this method is to
                  apply the mean phi from out of sample models.

### Value

vector of predicted values using object to predict newx

### Examples

```
data(test_data)

alphaGrid <- seq(0,1, length.out=2)
phiGrid <- exp(seq(log(1),log(10), length.out=2))
nlambda <- 3
ncores <- 2
eval <- "RSS"
family <- "gaussian"
intercept <- TRUE
standardize <- TRUE
center <- TRUE

model <- cv_iEN(X, Y, foldid, alphaGrid, phiGrid, nlambda, NULL, priors, ncores, eval, family, intercept, standardi
predict(model, X, priors)
```

---

| | |
|---|---|
| `print,iEN-method` | *Prints iEN object in a formatted way using the knitr package, displays optimized hyper parameters for each out of sample instance during cross-validated gridsearch.* |

---

### Description

Prints iEN object in a formatted way using the knitr package, displays optimized hyper parameters for each out of sample instance during cross-validated gridsearch.

### Usage

```
## S4 method for signature 'iEN'
print(x, ...)
```

### Arguments

| | |
|---|---|
| x | object of type iEN |
| ... | additional parameters fed to print |

### Value

formatted print

### Examples

```
data(test_data)

alphaGrid <- seq(0,1, length.out=2)
phiGrid <- exp(seq(log(1),log(10), length.out=2))
nlambda <- 3
ncores <- 2
eval <- "RSS"
family <- "gaussian"
intercept <- TRUE
standardize <- TRUE
center <- TRUE

model <- cv_iEN(X, Y, foldid, alphaGrid, phiGrid, nlambda, NULL, priors, ncores, eval, family, intercept, standardi
print(model)
```

---

test_data                              *test_data*

---

#### Description

A dataset containing CyTOF data relating to a study on the immunological progression of pregnancy. Includes immune features X and corresponding gestational age Y, as well as corresponding biological priors and fold identification per individual to maintain patient separation during cross-validation.

#### Usage

```
data(test_data)
```

#### Details

This dataset consists of a single data matrix and three vectors X CyTOF data Y Response vector of gestational age at time of sampling `priors` Priors indicating prioritization of immune features `foldid` identification of individuals for Cross-Validation

#### Author(s)

Anthony Culos <<tculos@stanford.edu>> and Nima Aghaeepour <<naghaeep@gmail.com>>.

#### References

https://nalab.stanford.edu/

#### Examples

```
data(test_data)
```

# Index