

Package ‘EZtune’

October 14, 2018

Type Package

Title Tunes AdaBoost, Support Vector Machines, and Gradient Boosting Machines

Version 1.0.0

Maintainer Jill Lundell <jflundell@gmail.com>

Description Contains two functions that are intended to make tuning supervised learning methods easy. The `eztune` function uses a genetic algorithm or quasi-Newton-Raphson optimizer to find the best set of tuning parameters. The user can choose the optimizer, the learning method, and if optimization will be based on accuracy obtained through resubstitution or cross-validation. The function `eztune.cv` will compute a cross-validated error rate. The purpose of `eztune.cv` is to provide a cross-validated accuracy when resubstitution is used for optimization because resubstitution typically produces an accuracy that is too high.

Depends R (>= 3.1.0)

Imports ada, e1071, GA, gbm, mlbench, parallel, doParallel

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

NeedsCompilation no

Author Jill Lundell [aut, cre]

Repository CRAN

Date/Publication 2018-10-14 16:50:17 UTC

R topics documented:

EZtune-package	2
eztune	2
eztune.cv	4

lichen	5
lichenTest	7
mullein	9
mulleinTest	10

Index	13
--------------	-----------

EZtune-package	<i>Tuning supervised learning models.</i>
----------------	---

Description

The EZtune package contains two functions and four datasets. The functions are `eztune` and `eztune.cv`. The datasets are `lichen`, `lichenTest`, `mullein`, and `mulleinTest`.

Eztune functions

The EZtune functions are `eztune` and `eztune.cv`. The function `eztune` will autotune an adaboost, gradient boosting machine, and support vector machine using a genetic algorithm or a quasi-Newton-Raphson optimizer. The `eztune.cv` function returns the cross-validated error for an object produced by `eztune`.

EZtune datasets

The `lichen` dataset has 840 observations and 40 variables. Seven of the variables are binary response variables. The `lichenTest` dataset is a test dataset for the `lichen` dataset that consists of 300 observations and the same 40 variables.

The `mullein` dataset consists of 12094 observations and 32 variables. One variable is a binary response. The `mulleinTest` dataset is a test dataset for the `mullein` data that consists of 1512 observations and the same 32 variables as `mullein`.

Author(s)

Maintainer: Jill Lundell <jflundell@gmail.com>

eztune	<i>Supervised Learning Function</i>
--------	-------------------------------------

Description

`eztune` is a single function that will tune adaboost, support vector machines, and gradient boosting machines. It currently only tunes models with a binary response. An optimizer is used to find a good set of tuning parameters for the selected model. The function optimizes on either the resubstitution accuracy or the cross-validated accuracy.

Usage

```
eztune(x, y, type = "binary", method = "ada", optimizer = "optim",
       cv = FALSE, fold = 10)
```

Arguments

x	Matrix or data frame containing the dependent variables.
y	Numeric vector of 0s and 1s for the response.
type	Type of response (binary is the only option at this time).
method	Model to be fit. Choices are "ada" for adaboost, "gbm" for gradient boosting machines, and "svm" for support vector machines.
optimizer	Optimization method. Options are "ga" to use a genetic algorithm to optimize and "optim" to use a quasi-Newton-Raphson optimizer.
cv	Indicates if the cross-validation accuracy should be used to fit. FALSE to use resubstitution accuracy to optimize (faster but less accurate) and TRUE to use cross-validation accuracy (slower and more accurate).
fold	The number of folds to use for n-fold cross validation. This is ignored if cv = FALSE.

Value

Function returns a summary of the fitted tuning parameters, the accuracy, and the best model.

summary	Matrix that contains the values of the tuning parameters and the final accuracy.
accuracy	Best accuracy obtained by optimizing algorithm.
best.model	Model using optimized parameters. Adaboost model comes from package ada, gbm model comes from package gbm, svm model comes from package e1071.
iter	Tuning parameter for adaboost.
nu	Tuning parameter for adaboost.
loss	Loss used in adaboost fitting.
interaction.depth	Tuning parameter for gbm.
n.trees	Tuning parameter for gbm.
shrinkage	Tuning parameter for gbm.
cost	Tuning parameter for svm.
epsilon	Tuning parameter for svm.
kernel	Kernel used in svm fitting.

Examples

```
library(mlbench)
data(Glass)

glass <- Glass[as.numeric(as.character(Glass$Type)) < 3, ]
glass <- glass[sample(1:nrow(glass), 80), ]
y <- ifelse(glass$Type == 1, 0, 1)
x <- glass[, 1:9]

eztune(x, y, type = "binary", method = "gbm", optimizer = "optim")
```

eztune.cv

Cross-Validated Accuracy for Supervised Learning Model

Description

eztune.cv is a function that will return the cross-validated accuracy for a model generated by the function eztune. The function eztune can tune a model using resubstitution or cross-validation. If the resubstitution is used to tune the model, the accuracy obtained from the function is inflated. The function eztune.cv will return a cross-validated accuracy for such a model.

Usage

```
eztune.cv(x, y, model, fold = 10)
```

Arguments

x	Matrix or data frame containing the dependent variables used to create the model.
y	Numeric vector of 0s and 1s for the response used to create the model.
model	Object generated with the function eztune.
fold	Number of folds to use for n-fold cross validation.

Value

Function returns a numeric value that represents the cross-validated accuracy of the model.

Examples

```
library(mlbench)
data(Glass)

glass <- Glass[as.numeric(as.character(Glass$Type)) < 3, ]
glass <- glass[sample(1:nrow(glass), 80), ]
y <- ifelse(glass$Type == 1, 0, 1)
x <- glass[, 1:9]

glass_svm <- eztune(x, y, type = "binary", method = "svm")
```

```
eztune.cv(x, y, glass_svm)
```

 lichen

Lichen data from the Current Vegetation Survey

Description

Data were collected between 1993 and 1999 as part of the Lichen Air Quality surveys on public lands in Oregon and southern Washington. Observations were obtained from 1-acre (0.4 ha) plots at Current Vegetation Survey (CVS) sites. Indicator variables denote the presences and absences of 7 lichen species. Data for each sampled plot include the topographic variables elevation, aspect, and slope; bioclimatic predictors including maximum, minimum, daily, and average temperatures, relative humidity precipitation, evapotranspiration, and vapor pressure; and vegetation variables including the average age of the dominant conifer and percent conifer cover. The data in lichenTest were collected from half-acre plots at CVS sites in the same geographical region and contains many of the same variables, including presences and absences for the 7 lichen species. As such, it is a good test dataset for predictive methods applied to the Lichen Air Quality data.

Usage

```
lichen
```

Format

A data frame with 840 observations and 40 variables. One variable is a location identifier, 7 (coded as 0 and 1) identify the presence or absence of a type of lichen species, and 32 are characteristics of the survey site where the data were collected.

In the original data there were 12 monthly values for each of the bioclimatic predictors. Principal components analyses suggested that for each of these predictors 2 principal components explained the vast majority (95.0 indices were created for each set of bioclimatic predictors. The variables with the suffix Ave in the variable name are the average of 12 monthly variables. The variables with the suffix Diff are contrasts between the sum of the April-September monthly values and the sum of the October-December and January-March monthly values, divided by 12. Roughly speaking, these are summer-to-winter contrasts.

The variables are summarized as follows:

PlotNum Identifier of the section of forest from which the data were collected.

LobaOreg *Lobaria oregana* (Absent = 0, Present = 1)

LobaPulm *Lobaria pulmonaria* (Absent = 0, Present = 1)

NephBell *Nephroma bellum* (Absent = 0, Present = 1)

NephHelv *Nephroma helveticum* (Absent = 0, Present = 1)

PseuAnom *Pseudocyphellaria anomala* (Absent = 0, Present = 1)

PseuAnth *Pseudocyphellaria anthraspis* (Absent = 0, Present = 1)

PseuCroc *Pseudocyphellaria crocata* (Absent = 0, Present = 1)

EvapoTransAve Average monthly potential evapotranspiration in mm
EvapoTransDiff Summer-to-winter difference in monthly potential evapotranspiration in mm
MoistIndexAve Average monthly moisture index in cm
MoistIndexDiff Summer-to-winter difference in monthly monthly moisture index in cm
PrecipAve Average monthly precipitation in cm
PrecipDiff Summer-to-winter difference in monthly precipitation in cm
RelHumidAve Average monthly relative humidity in percent
RelHumidDiff Summer-to-winter difference in monthly relative humidity in percent
PotGlobRadAve Average monthly potential global radiation in kJ
PotGlobRadDiff Summer-to-winter difference in monthly potential global radiation in kJ
AveTempAve Average monthly average temperature in degrees Celsius
AveTempDiff Summer-to-winter difference in monthly average temperature in degrees Celsius
MaxTempAve Average monthly maximum temperature in degrees Celsius
MaxTempDiff Summer-to-winter difference in monthly maximum temperature in degrees Celsius
MinTempAve Average monthly minimum temperature in degrees Celsius
MinTempDiff Summer-to-winter difference in monthly minimum temperature in degrees Celsius
DayTempAve Mean average daytime temperature in degrees Celsius
DayTempDiff Summer-to-winter difference in average daytime temperature in degrees Celsius
AmbVapPressAve Average monthly average ambient vapor pressure in Pa
AmbVapPressDiff Summer-to-winter difference in monthly average ambient vapor pressure in Pa
SatVapPressAve Average monthly average saturated vapor pressure in Pa
SatVapPressDiff Summer-to-winter difference in monthly average saturated vapor pressure in Pa
Aspect Aspect in degrees
TransAspect Transformed Aspect: $\text{TransAspect} = (1 - \cos(\text{Aspect})) / 2$
Elevation Elevation in meters
Slope Percent slope
ReserveStatus Reserve Status (Reserve, Matrix)
StandAgeClass Stand Age Class (< 80 years, 80+ years)
ACONIF Average age of the dominant conifer in years
PctVegCov Percent vegetation cover
PctConifCov Percent conifer cover
PctBroadLeafCov Percent broadleaf cover
TreeBiomass Live tree (> 1inch DBH) biomass, above ground, dry weight.

Source

Cutler, D. Richard., Thomas C. Edwards Jr., Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, and Joshua J. Lawler. 2007. Random Forests for Classification in Ecology. *Ecology* 88(11): 2783-2792.

 lichenTest

Test dataset for lichen data

Description

Data were collected as part of the Northwest Forest Conservation Plan. Data were collected from 300 half-acre (0.2 ha) sites on the Current Vegetation Survey grid in Gifford-Pinchot National Forest, the Umpqua Basin, and the Oregon Coast. Samples were collected between 2002 and 2003. Indicator variables denoted the presence or absence of 7 lichen species. This dataset may be used as a test dataset for the lichen dataset included in this package.

Usage

```
lichenTest
```

Format

A data frame with 300 observations and 40 variables. One variable is a location identifier, 7 identify the presence or absence of the lichen species, and 32 are characteristics of the survey site where the data were collected.

As with the Lichen Air Quality data, the variables with the suffix Ave in the variable name are the average of 12 monthly variables. The variables with the suffix Diff are contrasts between the sum of the April-September monthly values and the sum of the October-December and January-March monthly values, divided by 12. Roughly speaking, these are summer-to-winter contrasts.

The variables are summarized as follows:

PlotNum Identifier of the section of forest from which the data were collected.

LobaOreg Lobaria oregana (Absent = 0, Present = 1)

LobaPulm Lobaria pulmonaria (Absent = 0, Present = 1)

NephBell Nephroma bellum (Absent = 0, Present = 1)

NephHelv Nephroma helveticum (Absent = 0, Present = 1)

PseuAnom Pseudocyphellaria anomala (Absent = 0, Present = 1)

PseuAnth Pseudocyphellaria anthraxis (Absent = 0, Present = 1)

PseuCroc Pseudocyphellaria crocata (Absent = 0, Present = 1)

EvapoTransAve Average monthly potential evapotranspiration in mm

EvapoTransDiff Summer-to-winter difference in monthly potential evapotranspiration in mm

MoistIndexAve Average monthly moisture index in cm

MoistIndexDiff Summer-to-winter difference in monthly monthly moisture index in cm

PrecipAve Average monthly precipitation in cm

PrecipDiff Summer-to-winter difference in monthly precipitation in cm

RelHumidAve Average monthly relative humidity in percent

RelHumidDiff Summer-to-winter difference in monthly relative humidity in percent

PotGlobRadAve Average monthly potential global radiation in kJ

PotGlobRadDiff Summer-to-winter difference in monthly potential global radiation in kJ

AveTempAve Average monthly average temperature in degrees Celsius

AveTempDiff Summer-to-winter difference in monthly average temperature in degrees Celsius

MaxTempAve Average monthly maximum temperature in degrees Celsius

MaxTempDiff Summer-to-winter difference in monthly maximum temperature in degrees Celsius

MinTempAve Average monthly minimum temperature in degrees Celsius

MinTempDiff Summer-to-winter difference in monthly minimum temperature in degrees Celsius

DayTempAve Mean average daytime temperature in degrees Celsius

DayTempDiff Summer-to-winter difference in average daytime temperature in degrees Celsius

AmbVapPressAve Average monthly average ambient vapor pressure in Pa

AmbVapPressDiff Summer-to-winter difference in monthly average ambient vapor pressure in Pa

SatVapPressAve Average monthly average saturated vapor pressure in Pa

SatVapPressDiff Summer-to-winter difference in monthly average saturated vapor pressure in Pa

Aspect Aspect in degrees

TransAspect Transformed Aspect: $\text{TransAspect} = (1 - \cos(\text{Aspect})) / 2$

Elevation Elevation in meters

Slope Percent slope

ReserveStatus Reserve Status (Reserve, Matrix)

StandAgeClass Stand Age Class (< 80 years, 80+ years)

ACONIF Average age of the dominant conifer in years

PctVegCov Percent vegetation cover

PctConifCov Percent conifer cover

PctBroadLeafCov Percent broadleaf cover

TreeBiomass Live tree (> 1inch DBH) biomass, above ground, dry weight.

Source

Cutler, D. Richard., Thomas C. Edwards Jr., Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, and Joshua J. Lawler. 2007. Random Forests for Classification in Ecology. *Ecology* 88(11): 2783-2792.

mullein

*Mullein data from Lava Beds National Monument***Description**

This dataset contains information about the presence and absence of common mullein (*verbascum thapsus*) at Lava Beds National Monument. The park was digitally divided into 30m by 30m pixels. Park personnel provided data on 6,047 sites at which mullein was detected and treated between 2000 and 2005, and these data were augmented by 6,047 randomly selected pseudo-absences. For each 30m by 30m site there are data on elevation, aspect, slope, proximity to roads and trails, and interpolated bioclimatic variables such as minimum, maximum, and average temperature, precipitation, relative humidity, and evapotranspiration. The dataset called *mulleinTest* is a test dataset collected in Lava Beds National Monument in 2006 that can be used to verify evaluate predictive statistical procedures done using applied to the mullein dataset.

Usage

mullein

Format

A data frame with 12,094 observations and 32 variables. One variable identifies the presence or absence of mullein in a 30m by 30m site and 31 variables are characteristics of the site where the data were collected.

In the original data there were 12 monthly values for each of the bioclimatic predictors. Principal components analyses suggested that for each of these predictors 2 principal components explained the vast majority (95.0 indices were created for each set of bioclimatic predictors. The variables with the suffix *Ave* in the variable name are the average of 12 monthly variables. The variables with the suffix *Diff* are contrasts between the sum of the April-September monthly values and the sum of the October-December and January-March monthly values, divided by 12. Roughly speaking, these are summer-to-winter contrasts. The variables are summarized as follows:

VerbThap Presence or absence of *Verbascum thapsus*, common mullein, (Absent = 0, Present = 1)

DegreeDays Degree days in degrees Celsius

EvapoTransAve Average monthly potential evapotranspiration in mm

EvapoTransDiff Summer-to-winter difference in monthly potential evapotranspiration in mm

MoistIndAve Average monthly moisture index in cm

MoistIndDiff Summer-to-winter difference in monthly moisture index in cm

PrecipAve Average monthly precipitation in cm

PrecipDiff Summer-to-winter difference in monthly precipitation in cm

RelHumidAve Average monthly relative humidity in percent

RelHumidDiff Summer-to-winter difference in monthly relative humidity in percent

PotGlobRadAve Average monthly potential global radiation in kJ

PotGlobRadDiff Summer-to-winter difference in monthly potential global radiation in kJ
AveTempAve Average monthly average temperature in degrees Celsius
AveTempDiff Summer-to-winter difference in monthly average temperature in degrees Celsius
MinTempAve Average monthly minimum temperature in degrees Celsius
MinTempDiff Summer-to-winter difference in monthly minimum temperature in degrees Celsius
MaxTempAve Average monthly maximum temperature in degrees Celsius
MaxTempDiff Summer-to-winter difference in monthly maximum temperature in degrees Celsius
DayTempAve Mean average daytime temperature in degrees Celsius
DayTempDiff Summer-to-winter difference in average daytime temperature in degrees Celsius
AmbVapPressAve Average monthly average ambient vapor pressure in Pa
AmbVapPressDiff Summer-to-winter difference in monthly average ambient vapor pressure in Pa
SatVapPressAve Average monthly average saturated vapor pressure in Pa
SatVapPressDiff Summer-to-winter difference in monthly average saturated vapor pressure in Pa
VapPressDefAve Average monthly average vapor pressure deficit in Pa
VapPressDefDiff Summer-to-winter difference in monthly average vapor pressure deficit in Pa
Elevation Elevation in meters
Slope Percent slope
TransAspect Transformed Aspect: $\text{TransAspect} = (1 - \cos(\text{Aspect})) / 2$
DistRoad Distance to the nearest road in meters
DistTrail Distance to the nearest trail in meters
DistRoadTrail Distance to the nearest road or trail in meters

Source

Cutler, D. Richard., Thomas C. Edwards Jr., Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, and Joshua J. Lawler. 2007. Random Forests for Classification in Ecology. *Ecology* 88(11): 2783-2792.

mulleinTest

Mullein data from Lava Beds National Monument - test dataset

Description

This dataset contains information about the presence and absence of common mullein (*Verbascum thapsus*) at 1,512 randomly selected sites in Lava Beds National Monument. The data were collected in summer 2006. This dataset may be used to evaluate predictive statistical procedures that have been fit on the mullein dataset.

Usage

mulleinTest

Format

A data frame with 1512 observations and 32 variables. One variable identifies the presence or absence of mullein in a 30m by 30m site and 31 variables are characteristics of the site where the data were collected.

In the original data there were 12 monthly values for each of the bioclimatic predictors. Principal components analyses suggested that for each of these predictors 2 principal components explained the vast majority (95.0 these analyses, indices were created for each set of bioclimatic predictors. The variables with the suffix Ave in the variable name are the average of 12 monthly variables. The variables with the suffix Diff are contrasts between the sum of the April-September monthly values and the sum of the October-December and January-March monthly values, divided by 12. Roughly speaking, these are summer-to-winter contrasts.

The variables are summarized as follows:

VerbThap Presence or absence of *Verbascum thapsus*, common mullein, (Absent = 0, Present = 1)

DegreeDays Degree days in degrees Celsius

EvapoTransAve Average monthly potential evapotranspiration in mm

EvapoTransDiff Summer-to-winter difference in monthly potential evapotranspiration in mm

MoistIndAve Average monthly moisture index in cm

MoistIndDiff Summer-to-winter difference in monthly moisture index in cm

PrecipAve Average monthly precipitation in cm

PrecipDiff Summer-to-winter difference in monthly precipitation in cm

RelHumidAve Average monthly relative humidity in percent

RelHumidDiff Summer-to-winter difference in monthly relative humidity in percent

PotGlobRadAve Average monthly potential global radiation in kJ

PotGlobRadDiff Summer-to-winter difference in monthly potential global radiation in kJ

AveTempAve Average monthly average temperature in degrees Celsius

AveTempDiff Summer-to-winter difference in monthly average temperature in degrees Celsius

MinTempAve Average monthly minimum temperature in degrees Celsius

MinTempDiff Summer-to-winter difference in monthly minimum temperature in degrees Celsius

MaxTempAve Average monthly maximum temperature in degrees Celsius

MaxTempDiff Summer-to-winter difference in monthly maximum temperature in degrees Celsius

DayTempAve Mean average daytime temperature in degrees Celsius

DayTempDiff Summer-to-winter difference in average daytime temperature in degrees Celsius

AmbVapPressAve Average monthly average ambient vapor pressure in Pa

AmbVapPressDiff Summer-to-winter difference in monthly average ambient vapor pressure in Pa

SatVapPressAve Average monthly average saturated vapor pressure in Pa

SatVapPressDiff Summer-to-winter difference in monthly average saturated vapor pressure in Pa

VapPressDefAve Average monthly average vapor pressure deficit in Pa

VapPressDefDiff Summer-to-winter difference in monthly average vapor pressure deficit in Pa

Elevation Elevation in meters

Slope Percent slope

TransAspect Transformed Aspect: $\text{TransAspect} = (1 - \cos(\text{Aspect})) / 2$

DistRoad Distance to the nearest road in meters

DistTrail Distance to the nearest trail in meters

DistRoadTrail Distance to the nearest road or trail in meters

Source

Cutler, D. Richard., Thomas C. Edwards Jr., Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, and Joshua J. Lawler. 2007. Random Forests for Classification in Ecology. *Ecology* 88(11): 2783-2792.

Index

- *Topic **adaboost**,
 - [eztune](#), [2](#)
 - [eztune.cv](#), [4](#)
 - *Topic **cross-validation**
 - [eztune.cv](#), [4](#)
 - *Topic **datasets**
 - [lichen](#), [5](#)
 - [lichenTest](#), [7](#)
 - [mullein](#), [9](#)
 - [mulleinTest](#), [10](#)
 - *Topic **gbm**,
 - [eztune](#), [2](#)
 - [eztune.cv](#), [4](#)
 - *Topic **svm**,
 - [eztune](#), [2](#)
 - [eztune.cv](#), [4](#)
 - *Topic **tuning**,
 - [eztune.cv](#), [4](#)
 - *Topic **tuning**
 - [eztune](#), [2](#)
- [EZtune \(EZtune-package\)](#), [2](#)
[eztune](#), [2](#)
[EZtune-package](#), [2](#)
[eztune.cv](#), [4](#)
- [lichen](#), [5](#)
[lichenTest](#), [7](#)
- [mullein](#), [9](#)
[mulleinTest](#), [10](#)