Package 'aum'

October 12, 2022

Type Package

Title Area Under Minimum of False Positives and Negatives

Version 2022.2.7

Author Toby Dylan Hocking

Maintainer Toby Dylan Hocking <toby.hocking@r-project.org>

Description Standard template library sort is
used to implement an efficient
algorithm <arXiv:2107.01285> for computing Area Under Minimum and
directional derivatives.

License GPL-3

LinkingTo Rcpp

URL https://github.com/tdhock/aum

BugReports https://github.com/tdhock/aum/issues

Imports Rcpp, data.table

Suggests testthat, ggplot2, WeightedROC, penaltyLearning, knitr,
markdown, mlbench, directlabels, microbenchmark, covr

VignetteBuilder knitr

NeedsCompilation yes

Repository CRAN

Date/Publication 2022-02-08 16:30:05 UTC

R topics documented:

aum *aum*

Description

Compute the Area Under Minimum of False Positives and False Negatives, and its directional derivatives.

Usage

```
aum(error.diff.df, pred.vec)
```

Arguments

error.diff.df	data frame of error differences, typically computed via aum_diffs_binary or aum_diffs_penalty. There should be one row for each change in error functions. "example" column indicates example ID (int from 1 to N), "pred" column indicates predicted value where there is a change in the error function(s), "fp_diff" and "fn_diff" columns indicate differences in false positives and false negatives at that predicted value. Note that this representation assumes that each error function has fp=0 at pred=-Inf and fn=0 at pred=Inf.
pred.vec	numeric vector of N predicted values.

Value

Named list of two items: aum is numeric scalar loss value, derivative_mat is N x 2 matrix of directional derivatives (first column is derivative from left, second column is derivative from right). If

Author(s)

Toby Dylan Hocking

Examples

```
(bin.diffs <- aum::aum_diffs_binary(c(0,1)))
aum::aum(bin.diffs, c(-10,10))
aum::aum(bin.diffs, c(0,0))
aum::aum(bin.diffs, c(10,-10))
```

aum_diffs *aum diffs*

Description

Create error differences data table which can be used as input to aum function. Typical users should not use this function directly, and instead use aum_diffs_binary for binary classification, and aum_diffs_penalty for error defined as a function of non-negative penalty.

Usage

```
aum_diffs(example, pred,
    fp_diff, fn_diff,
    pred.name.vec)
```

Arguments

example	Integer or character vector identifying different examples.
pred	Numeric vector of predicted values at which the error changes.
fp_diff	Numeric vector of difference in fp at pred.
fn_diff	Numeric vector of difference in fn at pred.
pred.name.vec	Character vector of example names for predictions.

Value

data table of class "aum_diffs" in which each rows represents a breakpoint in an error function. Columns are interpreted as follows: there is a change of "fp_diff","fn_diff" at predicted value "pred" for example/observation "example". This can be used for computing Area Under Minimum via aum function, and plotted via plot.aum_diffs.

Author(s)

Toby Dylan Hocking

Examples

```
aum::aum_diffs_binary(c(0,1))
aum::aum_diffs(c("positive", "negative"), 0, c(0,1), c(-1,1), c("negative", "positive"))
rbind(aum::aum_diffs(0L, 0, 1, 0), aum_diffs(1L, 0, 0, -1))
```

`aum_diffs_binary` *aum diffs binary*

Description

Convert binary labels to error differences.

Usage

```
aum_diffs_binary(label.vec,
    pred.name.vec, denominator = "count")
```

Arguments

`label.vec`	Numeric vector representing binary labels (either all 0,1 or all -1,1). If named, names are used to identify each example.
`pred.name.vec`	Character vector of prediction example names, used to convert names of `label.vec` to integers.
`denominator`	Type of diffs, either "count" or "rate".

Value

data table of class "aum_diffs" in which each rows represents a breakpoint in an error function. Columns are interpreted as follows: there is a change of "fp_diff","fn_diff" at predicted value "pred" for example/observation "example". This can be used for computing Area Under Minimum via `aum` function, and plotted via `plot.aum_diffs`.

Author(s)

Toby Dylan Hocking

Examples

```
aum_diffs_binary(c(0,1))
aum_diffs_binary(c(-1,1))
aum_diffs_binary(c(a=0,b=1,c=0), pred.name.vec=c("c","b"))
aum_diffs_binary(c(0,0,1,1,1), denominator="rate")
```

aum_diffs_penalty *aum diffs penalty*

Description

Convert penalized errors to error differences. A typical use case is for penalized optimal change-point models, for which small penalty values result in large fp/fn, and large penalty values result in small fp/fn.

Usage

```
aum_diffs_penalty(errors.df,
    pred.name.vec, denominator = "count")
```

Arguments

errors.df	data.frame which describes error as a function of penalty/lambda, with at least columns example, min.lambda, fp, fn. Interpreted as follows: fp/fn occur from all penalties from min.lambda to the next value of min.lambda within the current value of example.
pred.name.vec	Character vector of prediction example names, used to convert names of label.vec to integers.
denominator	Type of diffs, either "count" or "rate".

Value

data table of class "aum_diffs" in which each rows represents a breakpoint in an error function. Columns are interpreted as follows: there is a change of "fp_diff","fn_diff" at predicted value "pred" for example/observation "example". This can be used for computing Area Under Minimum via aum function, and plotted via plot.aum_diffs.

Author(s)

Toby Dylan Hocking

Examples

```
## Simple synthetic example with two changes in error function.
simple.df <- data.frame(
  example=1L,
  min.lambda=c(0, exp(1), exp(2), exp(3)),
  fp=c(6,2,2,0),
  fn=c(0,1,1,5))
(simple.diffs <- aum::aum_diffs_penalty(simple.df))
if(requireNamespace("ggplot2"))plot(simple.diffs)
(simple.rates <- aum::aum_diffs_penalty(simple.df, denominator="rate"))
if(requireNamespace("ggplot2"))plot(simple.rates)
```

```
## Simple real data with four example, one has non-monotonic fn.
if(requireNamespace("penaltyLearning")){
  data(neuroblastomaProcessed, package="penaltyLearning", envir=environment())
  ## assume min.lambda, max.lambda columns only? use names?
  nb.err <- with(neuroblastomaProcessed$errors, data.frame(
    example=paste0(profile.id, ".", chromosome),
    min.lambda,
    max.lambda,
    fp, fn))
  (nb.diffs <- aum::aum_diffs_penalty(nb.err, c("1.2", "1.1", "4.1", "4.2")))
  if(requireNamespace("ggplot2"))plot(nb.diffs)
}

## More complex real data example
data(fn.not.zero, package="aum", envir=environment())
pred.names <- unique(fn.not.zero$example)
(fn.not.zero.diffs <- aum::aum_diffs_penalty(fn.not.zero, pred.names))
if(requireNamespace("ggplot2"))plot(fn.not.zero.diffs)

if(require("ggplot2")){
  name2id <- structure(seq(0, length(pred.names)-1L), names=pred.names)
 fn.not.zero.wide <- fn.not.zero[, .(example=name2id[example], min.lambda, max.lambda, fp, fn)]
  fn.not.zero.tall <- data.table::melt(fn.not.zero.wide, measure=c("fp", "fn"))
  ggplot()+
    geom_segment(aes(
      -log(min.lambda), value,
      xend=-log(max.lambda), yend=value,
      color=variable, size=variable),
      data=fn.not.zero.tall)+
    geom_point(aes(
      -log(min.lambda), value,
      fill=variable),
      color="black",
      shape=21,
      data=fn.not.zero.tall)+
    geom_vline(aes(
      xintercept=pred),
      data=fn.not.zero.diffs)+
    scale_size_manual(values=c(fp=2, fn=1))+
    facet_grid(example ~ ., labeller=label_both)
}
```

---

aum_errors *aum errors*

Description

Convert diffs to canonical errors, used internally in plot.aum_diffs.

Usage

```
aum_errors(diffs.df)
```

Arguments

diffs.df	data.table of diffs from aum_diffs.

Value

data.table suitable for plotting piecewise constant error functions, with columns example, min.pred, max.pred, fp, fn.

Author(s)

Toby Dylan Hocking

Examples

```
(bin.diffs <- aum::aum_diffs_binary(c(0,1)))
if(requireNamespace("ggplot2"))plot(bin.diffs)
aum::aum_errors(bin.diffs)
```

---

fn.not.zero *Penalized models with non-zero fn at penalty=0*

Description

Usually we assume that fn must be zero at penalty=0, but this is not always the case in real data/labels. For example in the PeakSegDisk model with penalty=0, there are peaks almost everywhere but if a positive label is too small or misplaced with respect to the detected peaks, then there can be false negatives.

Usage

```
data("fn.not.zero")
```

Format

A data frame with 156 observations on the following 5 variables.

`example` a character vector

`min.lambda` a numeric vector

`max.lambda` a numeric vector

`fp` a numeric vector

`fn` a numeric vector

Source

https://github.com/tdhock/feature-learning-benchmark

`plot.aum_diffs` *plot aum diffs*

Description

Plot method for `aum_diffs` which shows piecewise constant error functions. Uses `aum_errors` internally to compute error functions which are plotted. Not recommended for large number of examples (>20).

Usage

```
## S3 method for class 'aum_diffs'
plot(x, ...)
```

Arguments

`x`	data table with class "aum_diffs".
`...`	ignored.

Value

ggplot of error functions, each example in a different panel.

Author(s)

Toby Dylan Hocking

Index