Package ‘SAMGEP’

Type Package
Title A Semi-Supervised Method for Prediction of Phenotype Event Times
Version 0.1.0-1
Description A novel semi-supervised machine learning algorithm to predict phenotype event times using Electronic Health Record (EHR) data.
URL https://github.com/celehs/SAMGEP
BugReports https://github.com/celehs/SAMGEP/issues
License GPL-3
Encoding UTF-8
RoxygenNote 7.1.1
Depends R (>= 3.5.0)
Imports stats, mvtnorm, nlme, pROC, abind, nloptr, foreach, doParallel, parallel, Rcpp
LinkingTo Rcpp, RcppArmadillo
Suggests knitr, rmarkdown
VignetteBuilder knitr
LazyData true
NeedsCompilation yes
Author Yuri Ahuja [aut, cre], Tianxi Cai [aut], PARSE LTD [aut]
Maintainer Yuri Ahuja <Yuri_Ahuja@hms.harvard.edu>
Repository CRAN
Date/Publication 2021-01-06 10:00:02 UTC

R topics documented:

SAMGEP-package ............................................................. 2
samgep ........................................................................... 2
simdata ......................................................................... 4
Semi-supervised Adaptive Markov Gaussian Embedding Process (SAMGEP) is a novel semi-supervised machine learning algorithm to predict phenotype event times using Electronic Health Record (EHR) data.

Usage

```r
samgep(
  dat_train = NULL,
  dat_test = NULL,
  Cindices = NULL,
  w = NULL,
  w0 = NULL,
  V = NULL,
  observed = NULL,
  nX = 10,
  covs = NULL,
  survival = FALSE,
  Estep = Estep_partial,
  Xtrain = NULL,
  Xtest = NULL,
  alpha = NULL,
  r = NULL,
  lambda = NULL,
  surrIndex = NULL,
  nCores = 1
)
```
Arguments

dat_train (optional if Xtrain is supplied) Raw training data set, including patient IDs (ID), healthcare utilization feature (H) and censoring time (C)
dat_test (optional) Raw testing data set, including patient IDs (ID), a healthcare utilization feature (H) and censoring time (C)
Cindices (optional if Xtrain is supplied) Column indices of EHR feature counts in dat_train/dat_test
w (optional if Xtrain is supplied) Pre-optimized EHR feature weights
w0 (optional if Xtrain is supplied) Initial (i.e. partially optimized) EHR feature weights
V (optional if Xtrain is supplied) nFeatures x nEmbeddings embeddings matrix
observed (optional if Xtrain is supplied) IDs of patients with observed outcome labels
nX Number of embedding features (defaults to 10)
covs (optional) Baseline covariates to include in model; not yet operational
survival Binary indicator of whether target phenotype is of type survival (i.e. stays positive after incident event) or relapsing-remitting (defaults to FALSE)
Estep E-step function to use (Estep_partial or Estep_full; defaults to Estep_partial)
Xtrain (optional) Embedded training data set, including patient IDs (ID), healthcare utilization feature (H) and censoring time (C)
Xtest (optional) Embedded testing data set, including patient IDs (ID), healthcare utilization feature (H) and censoring time (C)
alpha (optional) Relative weight of semi-supervised to supervised MGP predictors in SAMGEP ensemble
r (optional) Scaling factor of inter-temporal correlation
lambda (optional) L1 regularization hyperparameter for feature weight (w) optimization
surrIndex (optional) Index (within Cindices) of primary surrogate index for outcome event
nCores Number of cores to use for parallelization (defaults to 1)

Value

w_opt Optimized feature weights (w)
r_opt Optimized inter-temporal correlation scaling factor (r)
alpha_opt Optimized semi-supervised:supervised relative weight (alpha)
lambda_opt Optimized L1 regularization hyperparameter (lambda)
margSup Posterior probability predictions of supervised model (MGP Supervised)
margSemisup Posterior probability predictions of semi-supervised model (MGP Semi-supervised)
margMix Posterior probability predictions of SAMGEP
cumSup Cumulative probability predictions of supervised model (MGP Supervised)
cumSemisup Cumulative probability predictions of semi-supervised model (MGP Semi-supervised)
cumMix Cumulative probability predictions of SAMGEP
Description

Click HERE to view details.

Usage

simdata

Format

An object of class list of length 3.

Examples

str(simdata)
Index

* datasets
  simdata, 4

* package
  SAMGEP-package, 2

samgep, 2
SAMGEP-package, 2
simdata, 4