Package 'treatppmx'

September 29, 2022

Type Package

Title Personalized Treatment Selection

Version 1.0

Author Matteo Pedone

Maintainer Matteo Pedone <matteo.pedone@unifi.it>

Description Bayesian treatment selection rule at a single decision point.

License GPL-3

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.1

LinkingTo Rcpp, RcppArmadillo

Imports Rcpp, coda, mcclust, mcclust.ext, mclust, mvtnorm, ggplot2, reshape2, vweights, cluster, sn

URL https://github.com/mattpedone/treatppmx

BugReports https://github.com/mattpedone/treatppmx/issues

Depends R (>= 2.10)

R topics documented:

genmech *genmech*

Description

Generates the data for the simulations scenarios 1a and 1b reported in the paper. It follows the strategy designed by Ma et al. (2019). In particular, the outcome is a categorical variable representing $K = 3$ benefit-increasing levels. Patients ($n = 152$) are assigned to $T = 2$ competing treatments.

Usage

```
genmech(
  npred = 10,
  progscen = 1,
  predscen = 1,
  nnoise = 15,
  nset = 30,
  save = FALSE,
  filename = "myscenario"
)
```

Arguments

npred	number of Q predictive covariates used to generate the outcome
progscen	Prognostic covariates option: 0 - No prognostic biomarkers; 1 - Prognostic biomarkers are considered in the original scale (default); 2 - Prognostic biomarkers are transformed
predscen	Predictive covariates option: 1 - npred Predictive biomarkers are considered to generate outcomes (default); 2 - nnoise noisy std normals are added to the design matrix of predictive covariates in addition to npred predictive biomarkers considered to generate outcomes
nnoise	number of noisy covariates added to predictive biomarkers
nset	number of replicated scenarios generated
save	logical. if TRUE the function save the results in a .rda file. Default is FALSE
filename	Name given to the file is results are saved in .rda file Default is myscenario.rda

Value

a list of 5 elements.

- Y: $n \times K \times$nset array. It contains nset matrices that store the outcome in the form of a multinomial experiment
- Yord: $n \times$nset matrix. It contains the ordinal outcome for each replicated dataset
- treatment: $n-$dimensional vector. It contains the treatment assigned
- cov: $n \times (Q + 2)$ matrix. It contains all the biomarkers. The last two columns are the two predictive biomarkers
- prob: List of T. Each element is a $n \times K$ matrix containing the response probabilities

References

Ma, J., Stingo, F. C., & Hobbs, B. P. (2016). Bayesian predictive modeling for genomic based personalized treatment selection. *Biometrics*, **72**(2), 575-583. https://onlinelibrary.wiley.com/doi/full/10.1111/biom.12448

Ma, J., Stingo, F. C., & Hobbs, B. P. (2019). Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants. *Biometrical Journal*, **61**(4), 902-917. https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.201700323

`genmech_clu` *genmech clustering*

Description

Generates the data for the simulations scenarios S1 reported in the Supplementary Material. It follows the strategy designed by Ma et al. (2019). In particular, the outcome is a categorical variable representing $K = 3$ benefit-increasing levels. Patients (n) are assigned to $T = 2$ competing treatments. Moreover, the clustering is known and fixed. We generate predictive biomarkers as in Argiento et al. (2022).

Usage

```
genmech_clu(npred = 4, n = 200, nnoise = 7, nset = 50)
```

Arguments

`npred`	number of Q predictive covariates used to generate the outcome
`n`	number of observations
`nnoise`	number of noisy variables
`nset`	number of replicated scenarios generated

Value

a list of 8 elements.

- `Y`: $n \times K \times$`nset` array. It contains `nset` matrices that store the outcome in the form of a multinomial experiment
- `Yord`: $n \times$`nset` matrix. It contains the ordinal outcome for each replicated dataset
- `treatment`: $n-$dimensional vector. It contains the treatment assigned
- `pred`: $n \times$(`npred`+`nnoise`) matrix. It contains the predictive biomarkers (only the first `npred` are effectively used to generate the response)
- `pred`: $n \times 2$ matrix. It contains the prognostic biomarkers
- `clu1`: contains the cluster label for patients assigned to Treatment 1
- `clu2`: contains the cluster label for patients assigned to Treatment 2
- `prob`: List of T. Each element is a $n \times K$ matrix containing the response probabilities

References

Argiento, R., Corradin, R., and Guglielmi, A. (2022). A Bayesian nonparametric model for covariate driven clustering: improved insights of blood donors data. *Unpublished Manuscript*.

Ma, J., Stingo, F. C., & Hobbs, B. P. (2019). Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants. *Biometrical Journal*, **61**(4), 902-917. https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.201700323

`genmech_clu2` *genmech clustering 2*

Description

Generates the data for the simulations scenarios S2 reported in the Supplementary Material. It follows the strategy designed by Ma et al. (2019). In particular, the outcome is a categorical variable representing $K = 3$ benefit-increasing levels. n patients are assigned to $T = 2$ competing treatments. Moreover, the clustering is known and fixed. We generate predictive biomarkers as Scenarion 4 in Page and Quintana (2018)

Usage

```
genmech_clu2(nset = 50)
```

Arguments

`nset`	number of replicated scenarios generated

Value

a list of 8 elements.

- `Y`: $n \times K \times$ `nset` array. It contains `nset` matrices that store the outcome in the form of a multinomial experiment
- `Yord`: $n \times$ `nset` matrix. It contains the ordinal outcome for each replicated dataset
- `treatment`: $n-$dimensional vector. It contains the treatment assigned
- `pred`: $n \times$ (`npred`$+$`nnoise`) matrix. It contains the predictive biomarkers (only the first `npred` are effectively used to generate the response)
- `pred`: $n \times 2$ matrix. It contains the prognostic biomarkers
- `clu1`: contains the cluster label for patients assigned to Treatment 1
- `clu2`: contains the cluster label for patients assigned to Treatment 2
- `prob`: List of T. Each element is a $n \times K$ matrix containing the response probabilities

References

Ma, J., Stingo, F. C., & Hobbs, B. P. (2019). Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants. *Biometrical Journal*, **61**(4), 902-917. https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.201700323

Page, G. L. and Quintana, F. A. (2018). Calibrating covariate informed product partition models. *Statistics and Computing*, **28**(5), 1009–1031.

`genmech_clu3` *genmech clustering 3*

Description

Generates the data for the simulations scenarios S3 reported in the Supplementary Material. It follows the strategy designed by Ma et al. (2019). In particular, the outcome is a categorical variable representing $K = 3$ benefit-increasing levels. Patients (n) are assigned to $T = 2$ competing treatments. Moreover, the clustering is known and fixed. We generate predictive biomarkers as Scenarion 4 in Page and Quintana (2018)

Usage

```
genmech_clu3(npred = 10, nset = 1)
```

Arguments

`npred`	number of predictive covariates used to generate the outcome
`nset`	number of replicated scenarios generated

Value

a list of 8 elements.

- `Y`: $n \times K \times$`nset` array. It contains `nset` matrices that store the outcome in the form of a multinomial experiment
- `Yord`: $n \times$`nset` matrix. It contains the ordinal outcome for each replicated dataset
- `treatment`: $n-$dimensional vector. It contains the treatment assigned
- `pred`: $n \times$(`npred`+`nnoise`) matrix. It contains the predictive biomarkers (only the first `npred` are effectively used to generate the response)
- `pred`: $n \times 2$ matrix. It contains the prognostic biomarkers
- `clu1`: contains the cluster label for patients assigned to Treatment 1
- `clu2`: contains the cluster label for patients assigned to Treatment 2
- `prob`: List of T. Each element is a $n \times K$ matrix containing the response probabilities

References

Ma, J., Stingo, F. C., & Hobbs, B. P. (2019). Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants. *Biometrical Journal*, **61**(4), 902-917. https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.201700323

Page, G. L. and Quintana, F. A. (2018). Calibrating covariate informed product partition models. *Statistics and Computing*, **28**(5), 1009–1031.

`genmech_het` *genmech_het*

Description

Generates the data for the simulations scenarios reported in the paper. It follows the strategy designed by Ma et al. (2019). In particular, the outcome is a categorical variable representing $K = 3$ benefit-increasing levels. Patients ($n = 152$) are assigned to $T = 2$ competing treatments.

Usage

```
genmech_het(npred = 10, nset = 30, overlap = 0.8)
```

Arguments

`npred`	number of Q predictive covariates used to generate the outcome
`nset`	number of replicated scenarios generated
`overlap`	proportion of predictors used to generate the response in both the train and the validation set

Value

a list of 6 elements.

- `yord`: List of `nset`. Each element is a `n`-dimensional vector of the ordinal outcome
- `ymat`: List of `nset`. Each element is a `n\times K` matrix containing the ordinal outcome in the form of a Multinomial experiment
- `pred`: List of `nset`. Each element is a `n\times Q` matrix containing the predictive biomarkers
- `pred`: List of `nset`. Each element is a `n\times 2` matrix containing the prognostic biomarkers
- `trtsgn`: List of `nset`. Each element is a `n`-dimensional vector of the treatment assigned
- `prob`: List of `nset`. Each element of the list is a list of T $n \times K$ matrices containing the response probabilities

References

Ma, J., Stingo, F. C., & Hobbs, B. P. (2016). Bayesian predictive modeling for genomic based personalized treatment selection. *Biometrics*, **72**(2), 575-583. https://onlinelibrary.wiley.com/doi/full/10.1111/biom.12448

Ma, J., Stingo, F. C., & Hobbs, B. P. (2019). Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants. *Biometrical Journal*, **61**(4), 902-917. https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.201700323

---

npc *npc*

Description

Function to calculate the npc

Usage

```
npc(output, trtsgn, myoutot)
```

Arguments

output	Must be an array storing the probabilities attributed to nobs patients to the 3 different benefit levels of the two competing treatments for nset replicas of the simulation study. Its dimension are nobs x 6 x nset, where the first three columns are the probabilities for treatment 1 and the latter 3 columns store the probabilities for treatment 2.
trtsgn	treatment assigned to the patients by design.
myoutot	List of nset outcome variables in ordinal notation

Details

See Ma et al. (2019) for all the details.

Value

a nobs vector storing the npc for each patient,

References

Ma, J., Stingo, F. C., & Hobbs, B. P. (2019). Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants. *Biometrical Journal*, **61**(4), 902-917. https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.201700323

ppmxct *ppmxct*

Description

Function to predict personalized treatment for n_{test} new untreated patients, given their biomarkers. It leverages response to treatment, P prognostic, Q predictive biomarkers of n_{train} historical patients. It accounts for K ordinal response level and T competing treatments.

Usage

```
ppmxct(
  y,
  X = NULL,
  Xpred = NULL,
  Z = NULL,
  Zpred = NULL,
  asstreat = NULL,
  PPMx = 1,
  cohesion = 2,
  kappa = c(1, 30, 10, 1),
  sigma = c(0.005, 1, 10),
  similarity = 1,
  consim = 1,
  similparam,
  calibration = 0,
  coardegree = 1,
  modelpriors,
  update_hierarchy = 1,
  hsp = 1,
  iter = 1100,
  burn = 100,
  thin = 1,
  mhtunepar = c(0.05, 0.05),
  CC = 3,
  reuse = 1,
  nclu_init = 5
)
```

Arguments

y	$n_{train} \times K$ matrix of ordinal-valued response variable
X	$n_{train} \times Q$ dataframe of predictive biomarkers
Xpred	$n_test \times Q$ dataframe of predictive covariates of new untreated patient
Z	$n_{train} \times P$ dataframe of prognostic covariates
Zpred	$n_test \times P$ dataframe of prognostic covariates of new untreated patient

asstreat	n_test vector of integers encoding treatment received by historical patients
PPMx	logical. option for the use of product partition model with covariates. (default is yes)
cohesion	type of cohesion function that is employed for the PPMx prior on partitions. Options are 1 - DirichletProcess-like cohesion (DP) cohesion 2 - Normalized Generalized Gamma Process (NGG) cohesion
kappa	vector of possible values for κ for cohesion function (concentration parameter in DP and NGG)
sigma	vector of possible value for σ parameter in the cohesion function (reinforcement parameter in NGG)
similarity	type of similarity function that is employed for the PPMx prior on partitions. Options are 1 - Auxiliary similarity 2 - Double dipper similarity
consim	integer 1 or 2. 1 implies sim for con var is NN. 2 implies sim is NNIG
similparam	vector containing similarity functions paramaters
calibration	If the similarity function is Auxiliary or Double Dipper, the similarity can be calibrated. Options are 0 - no calibration 1 - standardize similarity value for each covariate 2 - coarsening is applied so that each similarity is raised to the 1/p power
coardegree	If the similarity is coarsened, it is possible to temper the coarsening 1 - $g(x^*)^{1/p}$ 2 - $g(x^*)^{1/p^{1/2}}$
modelpriors	vector containing prior values for model
update_hierarchy	should hyperparameter for BNP intercept be updated? if 1 yes (default)
hsp	parameter for employ horseshoe prior for coefficients for prognostic markers
iter	MCMC number of iteration
burn	MCMC iteration discarded due to burnin
thin	thinning for MCMC
mhtunepar	vector containing tuning parameters for MCMC updates
CC	number of auxiliary parameters for Algorithm 8 by Neal (2000)
reuse	option for the reuse algorithm by Favaro and Teh (2013). integer 0 or 1. 0 - reuse algorithm is not adopted 1 - reuse algorithm is adopted
nclu_init	number of cluster used for partial correlation initialization for prognostic covariates coefficient (default)

Value

a list of 16 elements.

- label: List of T matrices. Each element is a $n^a_{test}\times$nout matrix. nout is the number of MCMC iterations after burnin and thinning The row of each matrix contains the vector of cluster labels of the historical patient assigned to given treatment
- asstreat: Vector of n_{test} treatment assignment

- num_treat: Vector of dimension T. Contains the number of patients assigned to each treatment n^a, for $a = 1, \ldots, T$.
- nclu: $T \times$nout matrix of total number of cluster at each MCMC iteration for each treatment
- eta: nout$\times T$ List of matrices. Each matrix has dimension $K \times n^a$. Each element of the list is the linear predictor matrix for treatment a at each MCMC iteration.
- beta: Array of dimensions $P \times K \times$nout. It contains the matrix of prognostic factors at each MCMC iteration
- acc_beta: $P \times K$ matrix containing the counts each prognostic coefficient has been accepted in the MH step. Note that these counts should be divided by iter Each element of the list is a list of 2 matrices (one for each treatment).
- pi_out: $n \times K \times$nout array. It contains the Multinomial parameters for each patients at each MCMC iteration
- isypred: $n \times K \times$nout array. It contains the outcome in-sample-prediction for each patients at each MCMC iteration
- WAIC: scalar. It is the average WAIC
- lpml: scalar. It is the average lpml
- sigmangg: $T \times$nout matrix. It contains the σ parameter of NGGP for each treatment at each MCMC iteration
- kappangg: $T \times$nout matrix. It contains the κ parameter of NGGP for each treatment at each MCMC iteration
- ypred: $n_{pred} \times K \times T \times$nout array. It contains the predicted outcome for each untreated patients, under all the competing for each treatment at each MCMC iteration
- ypred: $n_{pred} \times K \times T \times$nout array. It contains the Multinomial parameters for each untreated patients, under all the competing for each treatment at each MCMC iteration

References

Page, G. L. and Quintana, F. A. (2016). Calibrating covariate informed product partition models. *Statistics and Computing*, **28**(5), 1009–1031.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**(2): 249–265.

Favaro, S., Teh, Y. W., et al. (2013). MCMC for normalized random measure mixture models. *Statistical Science*, **28**(3): 335–359.

---

prior_ppmx *ppmx prior*

Description

ppmx prior

Usage

```
prior_ppmx(
  X = NULL,
  PPMx = 1,
  cohesion = 2,
  alpha = 1,
  sigma = 0.2,
  similarity = 1,
  consim = 1,
  similparam,
  calibration = 0,
  coardegree = 1,
  iter = 1100,
  burn = 100,
  thin = 1,
  nclu_init = 5
)
```

Arguments

X	covariates
PPMx	option for the use of product partition model with covariates. (default is yes)
cohesion	type of cohesion function that is employed for the PPMx prior on partitions. Options are 1 - DirichletProcess-like cohesion (DP) cohesion 2 - Normalized Generalized Gamma Process (NGG) cohesion
alpha	value of α for cohesion function (concentration parameter in DP)
sigma	value of σ for cohesion function (reinforcement parameter in NGG)
similarity	type of similarity function that is employed for the PPMx prior on partitions. Options are 1 - Auxiliary similarity 2 - Double dipper similarity
consim	integer 1 or 2. 1 implies sim for con var is NN. 2 implies sim is NNIG
similparam	vector containing similarity functions paramaters
calibration	If the similarity function is Auxiliary or Double Dipper, the similarity can be calibrated. Options are 0 - no calibration 1 - standardize similarity value for each covariate 2 - coarsening is applied so that each similarity is raised to the 1/p power
coardegree	If the similarity is coarsened, it is possible to temper the coarsening 1 - $g(x^*)^{1/p}$ 2 - $g(x^*)^{1/p^{1/2}}$
iter	MCMC number of iteration
burn	MCMC iteration discarded due to burnin
thin	thinning for MCMC
nclu_init	number of cluster used for partial correlation initialization for prognostic covariates coefficient (default)

Value

Number of cluster at each MCMC iteration

simupats *152 simulated patients data.*

Description

A data frame 92 genes coming from 152 patients. This data were obtained to emulate the dependence structure arising in sequencing data. The simulated scenario starts from a well-known dataset of leukemia. The process to obtained these data from the original Leukemia dataset is extensively detailed in Ma et al. (2019).

Usage

```
simupats
```

Format

A data frame containing 152 observation of 92 variables:

1-90 predictive covariates

91-92 prognostic covariates

Source

https://www.pnas.org/content/101/12/4164.full?tab=ds

References

Ma, J., Stingo, F. C., & Hobbs, B. P. (2019). Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants. *Biometrical Journal*, **61**(4), 902-917. https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.201700323

The Leukemia dataset containing gene expression levels for a total of 5,000 genes across 38 patients.

treatppmx *Treatment PPMx*

Description

This package implements the treatment selection rule at a single decision point proposed by Pedone M., Argiento R., and Stingo F. It is a Bayesian predictive model for personalized treatment selection for new untreated patients, which leverages known predictive and prognostic biomarkers. The method assumes that the data come from historical patients that received two or more treatments. After a clinically relevant period, a response to treatment is measured. The clinical outcome must be a categorical variable. Biomarkers are assumed to be measured at baseline. Our approach accounts for continuous and categorical variables. The user must specify which biomarkers are prognostic and which are predictive. In particular, predictive biomarkers are exploited to inform a product partition model with covariates (PPMx) to obtain homogeneous clusters. The implementation has been done in C++ through the use of Rcpp and RcppArmadillo.

Author(s)

Matteo Pedone

Maintainer: Matteo Pedone <matteo.pedone@unifi.it>

References

TBA

Index