Package ‘malani’

September 26, 2016

Type Package
Title Machine Learning Assisted Network Inference
Version 1.0
Author Mehrab Ghanat Bari
Maintainer Mehrab Ghanat Bari <m.ghanatbari@gmail.com>
Description Find dark genes. These genes are often disregarded due to no detected mutation or differential expression, but are important in coordinating the functionality in cancer networks.
License GPL-3
LazyData TRUE
Depends e1071, stats
RoxygenNote 5.0.1
NeedsCompilation no
Repository CRAN
Date/Publication 2016-09-26 23:44:53

R topics documented:

dat ................................................................. 2
grp ............................................................... 2
Gsvmod .......................................................... 2
intGenes ....................................................... 3
malanidata ..................................................... 4
pairmod ....................................................... 4

Index 5
dat  
*A matrix of expression values.*

**Description**

A numeric matrix 100*20.

**Usage**

dat

**Format**

matrix.

grp  
*A vector of class labels for dat.*

**Description**

Vector length of 20.

**Usage**

grp

**Format**

vector

---

**Gsvmod**  
*G SVM models.*

**Description**

Returns accuracy performance of all genes. G support vector machine (SVM) classifiers trained using G different data matrixes, are used to predict labels in test data. Models are ranked based on prediction performances.

**Usage**

Gsvmod(dat.train, lab.train, dat.test, lab.test)
**Arguments**

- `dat.train`: Train data with G features and \((k-1)\times S/k\) samples. Parameter \(k\) comes from cross-validation scheme and is specified by user (default is 2).
- `lab.train`: Class labels for train data.
- `dat.test`: Test data with G features and \(S/k\) samples.
- `lab.test`: Class labels for test data.

**Value**

Accuracy scores for models. Each model represents one gene.

---

**intGenes**

*Select initial gene list from original data matrix.*

**Description**

Train G-1 SVM models in \(k\)-fold cross validation scheme to select initial genes list.

**Usage**

```r
intGenes(dat, grp, nfolds.out = 2, top.per = 0.05)
```

**Arguments**

- `dat`: Original gene expression data matrix with G rows (number of genes) and S column (number of samples).
- `grp`: Class labels.
- `nfolds.out`: Outer cross validation number (default is 2).
- `top.per`: All genes are ranked based on their models performance and `top.per`% of them are selected as initial genes.

**Value**

Selected initial genes.

**Examples**

```r
data(malanidata)
int <- intGenes(dat,grp)
print(int$top.genes)
```
malanidata  

*Dataset for malani package*

**Description**

A numeric matrix $G*S$ contains gene expressions data. $G$ are the genes (rows) and $S$ are the samples (columns).

**Usage**

malanidata

**Format**

A matrix of numeric values, 100 genes, 20 samples and class labels.

**Examples**

data(malanidata)

---

pairmod  

*Find best performing pairs*

**Description**

Combine each gene in initial set with all genes in the original set. Top $npair$ pairs are selected to construct the $Q$ matrix.

**Usage**

pairmod(X, L, theta, npair = 10)

**Arguments**

- **X**: Original gene expression data matrix. With $G$ rows (number of genes) and $S$ column (number of samples).
- **L**: Class labels.
- **theta**: Initial gene set.
- **npair**: Given a gene in initial set, top $npair$ best performing pairs correspond to that gene are selected (Default is 10).

**Value**

Best ($npair*G/20$) performing pairs.
Index

*Topic datasets
  dat, 2
  grp, 2
  malanidata, 4

  dat, 2, 2

  grp, 2
  Gsvmmod, 2

  intGenes, 3
  malanidata, 4
  pairmod, 4