Package 'MVpower'

February 14, 2012

Type Package

Title Give power for a given effect size using multivariate classification methods

Version 2.0

Date 2010-07-30

Author Yu Guo and Raji Balasubramanian

Maintainer Yu Guo <yguo@bg-medicine.com>

Depends R (>= 2.7.1), randomForest, pamr, kernlab, class

Description Calculate power for a given effect size and sample size using multivariate classification methods: Random forest (RF),Prediction analysis of microarrays (PAM), K-nearest neighbor (KNN), or Support vector machine (SVM)

License GPL-2

lazyload yes

Repository CRAN

Date/Publication 2010-07-31 04:57:04

R topics documented:

MVpower *Calculate power for a given effect size and sample size using multivariate classification methods*

Description

Calculate power for a given effect size and sample size using multivariate classification methods: Random forest (RF), Prediction analysis of microarrays (PAM), K-nearest neighbor (KNN), or Support vector machine (SVM)

Usage

```
MVpower(method, prior = c(0.5, 0.5), nfold = 4, delta = 1, NumFts = 100, NumImp = 5, n1 = 100, n2 = 100, N.
```

Arguments

method	a character string of multivariate classification method used, "RF", , "PAM", "KNN", "SVM"
prior	numeric vector of length 2, prior distribution for the two groups, default is c(0.5, 0.5)
nfold	number of cross validation folds to be run, default is 4
delta	numeric value, effect size. Used only if user provided alternative data (alt.data) is not provided
NumFts	number of total features used for prediction. Used only if user provided alternative data (alt.data) is not provided
NumImp	number of total markers. Used only if user provided alternative data (alt.data) is not provided
n1	number of samples in group 1. Used only if user provided alternative data (alt.data) is not provided
n2	number of samples in group 2. Used only if user provided alternative data (alt.data) is not provided
N.null	number of simulations to get null distribution, default is 100. Used only if user provided null data (null.data) is not provided
N.pow	number of simulations to get alternative distribution, default is 100. Used only if user provided alternative data (alt.data) is not provided
alt.data	User generated alternative distribution. Default is NULL, if NULL, run N.pow simulation to generate alternative distribution; or a list specifying alternative distribution the user generated, each element is a data matrix with rows as samples and columns as features.
null.data	User generated null distribution. Default is NULL, if NULL, run N.null simulation to generate null distribution; or a list specifying null distribution the user generated, each element is a data matrix with rows as samples and columns as features.
my.groups	User provided group indicator. Default is NULL, if NULL, it is set to be a factor with 2 levels, and n1 and n2 samples in each level, individually.

Details

The classifiers are evaluated in the setting in which the number of features measured exceeds the number of subjects in the study.

Power for the four listed classifiers is calculated by the following simulation that incorporates a recursive feature elimination procedure to discover an optimal subset of features that resulted in maximum success of the estimated classifier in predicting each subject's class membership.

For each simulated dataset, a p value associated with the classification metric (classification accuracy) was calculated based on comparison of the observed value of the statistic to its distribution under the null, obtained when all features were non-discriminating between cases and controls. Statistical power was estimated as the proportion of simulated datasets resulting in classifiers with associated p values less than or equal to 0.05. The results reported in the paper are averages based on `N.pow` simulated datasets.

Value

a list of the following components

`effect.size`	effect size
`n1`	number of samples in group 1
`n2`	number of samples in group 2
`NumFts`	number of total features used for prediction
`NumImp`	number of total markers
`method`	multivariate classification method used
`my.prop.rejected`	power
`null.dist`	Null distribution
`alt.dist`	Alternative distribution

Author(s)

Yu Guo and Raji Balasubramanian

Examples

```
#an exmaple using MVpower function to generate null and alternative distribution
MVpower(method="PAM", #"PAM", "KNN", "RF", "SVM",
        prior=c(0.5, 0.5), ## prior dist
        nfold=4, ## number of CV-folds
        delta=.5, #effect size mu/sigma
        NumFts=100, ## total number of features
        NumImp=5, ## number of markers
        n1=100, #number of samples
        n2=100,
        N.null=10, #Number of simulations for null distribution
        N.pow=10,   #Number of simulations for power calculation
        alt.data=NULL, null.data=NULL, my.groups=NULL)
```

```
# an example using user generated null and alternative distribution
# simulate normally distributed data with 10 samples in each group,
# 30 analytes, 5 markers, effect.size=1
null.data <- alt.data <- NULL
for(i in 1:10)
    {
        my.data <- matrix(rnorm(20*30,sd=1),ncol=30)
        my.data[11:20,1:5] <- my.data[11:20,1:5]
        null.data[[i]] <- my.data
    }
for(i in 1:10)
    {
        my.data <- matrix(rnorm(20*30,sd=1),ncol=30)
        my.data[11:20,1:5] <- my.data[11:20,1:5] + 1
        alt.data[[i]] <- my.data
    }
my.groups <- factor(rep(1:2,each=10))

MVpower(method="PAM", #"PAM", "KNN", "RF", "SVM",
        prior=c(0.5, 0.5), ## prior dist
        nfold=4, ## number of CV-folds
        delta=NULL, NumFts=NULL, NumImp=NULL, n1=NULL, n2=NULL, N.null=NULL, N.pow=NULL,
        alt.data=alt.data,
        null.data=null.data,
        my.groups=my.groups)
```

Index